Evolution of cooperation among individuals with limited payoff memory

Christian Hilbe, Nikoleta E. Glynatsi, Alex McAvoy

Abstract

1 Introduction

Evolutionary game theory is a mathematical framework for modeling evolution in biological, social and economical systems. Compared to classical game theory where it is assumed that players are rational, evolutionary game theory does not require individuals to be rational. Instead individuals adapt their strategies based on exploration and mutation. In this context, strategies spread based on fitness either because the individuals who adopt them have more offspring (fitness-based processes), or they are imitated more often (pairwise comparison processes) [1]. Fitness is not constant but it depends on the composition of the population. Individuals interact with members of the population according to pre-define games, and their payoffs are then translated into fitness.

Evolutionary models often work with idealized assumptions when estimating fitness. It is commonly assumed that fitness is given by the mean payoff an individual achieved over the different types in the population following multiple intercalations. Referred to as the *expected payoffs*. Expected payoffs imply that individuals have a perfect a memory, as they are able to recall several of their interactions with each player in the population. However most evolutionary models, 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] when modeling how individuals make decisions in each round. This creates a curious inconsistency. Individuals make decisions in each round whist having very limited memory, however, when modeling how individuals update their strategies over time, individuals are assumed to have perfect memory.

One of the most important applications of evolutionary game theory is the evolution of cooperation. Though cooperation can been seen to be at odd, several mechanisms can allow for cooperation to emerge. One of these mechanisms is direct reciprocity. However, given these idealized assumptions it feels like our understanding of reciprocity is thus still incomplete. In this work we will explore how robust our understanding of cooperation is as models deviate from the perfect memory assumption.

This work is not the first to relax assumption made at the fitness level. 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 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. In Section 3, we explore the effect of the payoffs in the cooperation rate given the donation game. We explore the difference between the ratios as well as the effect of different model parameters. In Section 4 we explore the cooperation rate for all the important case of symmetric 2×2 games. For the later part of the section we also allow individuals to use more memory.

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.

1. In the game stage, each player is randomly matched with some other player in the population to interact in a number of instances of the game $\begin{pmatrix} R & S \\ T & P \end{pmatrix}$. The number of interactions is not fixed, on the contrast we consider that another interaction has a probability δ to continue following each turn. 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.

2. 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 with probability

$$\rho(u_L, u_{RL}) = \frac{1}{1+\exp\left[-\beta(u_E - u_{RL})\right]}. \tag{1}$$

where $\beta \geq 0$ corresponds to 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.

We iterate this basic evolutionary step 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 $S_2' = (y_2', p_2', q_2')$ (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.xw The fixation probability of the mutant strategy then takes the standard form [7],

$$\varphi = \frac{1}{1 + \sum_{i=1}^{N-1} \prod_k^i \frac{\lambda_k^-}{\lambda_k^+}}. \tag{2}$$

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

We compare this process for stochastic payoff evaluation with the analogous process where players update their strategies with respect to their *expected* payoffs,

$$\begin{aligned} \pi_1 &= \frac{N-k-1}{N-1} \cdot \langle \mathbf{v}(S_1, S_1), \mathbf{u} \rangle + \frac{k}{N-1} \cdot \langle \mathbf{v}(S_1, S_2), \mathbf{u} \rangle, \\ \pi_2 &= \frac{N-k}{N-1} \cdot \langle \mathbf{v}(S_2, S_1), \mathbf{u} \rangle + \frac{k-1}{N-1} \cdot \langle \mathbf{v}(S_2, S_2), \mathbf{u} \rangle. \end{aligned} \tag{3}$$

In the limit of no discounting, $\delta \to 1$, this process based on expected payoffs has been considered in [8].

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. Figure 1 shows simulation results for the described process of Section A.

In Section 3, we will specifically consider the donation game where 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 (**??**) is give by:

$$\begin{pmatrix} b-c & -c \\ b & 0 \end{pmatrix} \tag{4}$$

Figure 1 depicts the evolving conditional cooperation probabilities p and q (assuming that the discount factor δ and the benefit b are comparably high). The left panel 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 figure suggests that when updating is based on expected payoffs, players tend to be more generous (their q-values are higher on average). In addition, players are generally more cooperative.

There is a statistically significant different in the cooperation distribution of the population over time. The cooperation rate falls significantly yo under 50%. There results are true for a strong selection strength ($\beta = 10$). The follow up question that arises is what effect does the selection strength have on the cooperation rate?

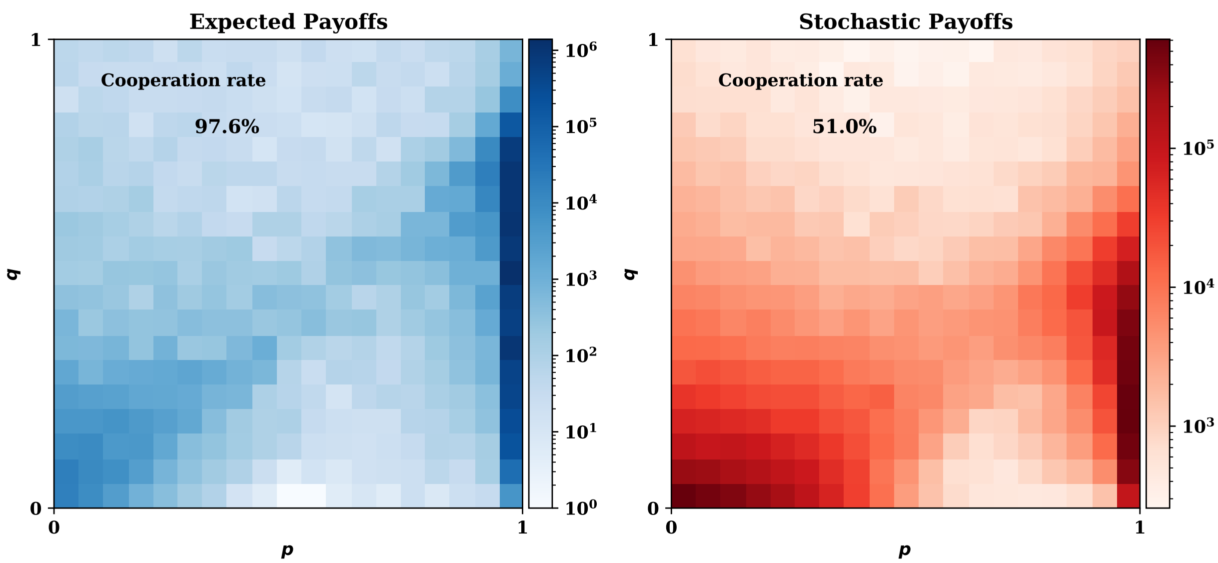


Figure 1: Evolutionary dynamics under expected payoffs and stochastic payoffs. We have run two simulations of the evolutionary process described in Section A for $T = 10^7$ time steps. For each time step, we have recorded the current resident population (y, p, q). Since simulations are run for a relatively 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 illustrates the results for different runs of the evolutionary process where $\beta \in \{10^-2, 10^-1, 10^0, 10^1, 10^2\}$. In the case where $\beta = 10^-2$ the process is almost random, as the effect of the payoffs is very small. The cooperation rate for this case and the case of $\beta = 10^-1$ between the stochastic and expected payoffs are the same. However, it is clear that this does not apply given $\beta \geq 1$. The difference in the cooperation rate increases, in the case where for the stochastic payoffs it can drop to 10%.

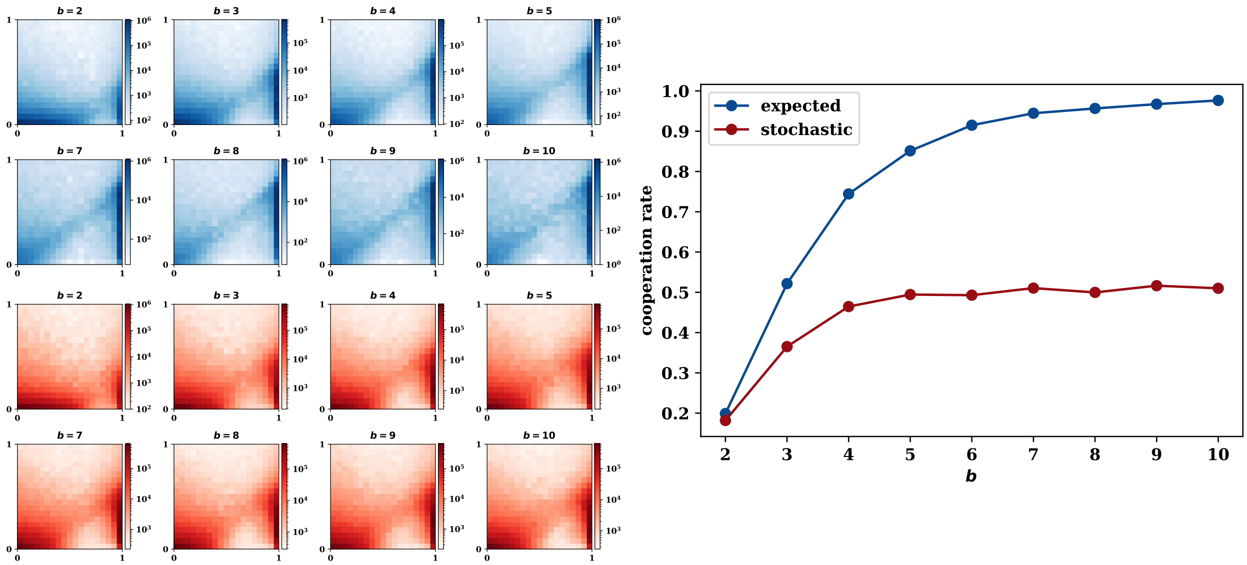


Figure 2: The evolution of cooperation for different parameter values. While the previous figures depict the evolutionary outcome for fixed parameter values, here we vary the benefit of cooperation b and the strength of selection β. In all cases, stochastic payoff evaluation tends to reduce the evolving 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.

Figure 2 illustrates the results for different runs of the evolutionary process where $\beta \in \{10^{-}2, 10^{-}1, 10^0, 10^1, 10^2\}$. In the case where $\beta = 10^{-}2$ the process is almost random, as the effect of the payoffs is very small. The cooperation rate for this case and the case of $\beta = 10^{-}1$ between the stochastic and expected payoffs are the same. However, it is clear that this does not apply given $\beta \geq 1$. The difference in the cooperation rate increases, in the case where for the stochastic payoffs it can drop to 10%.

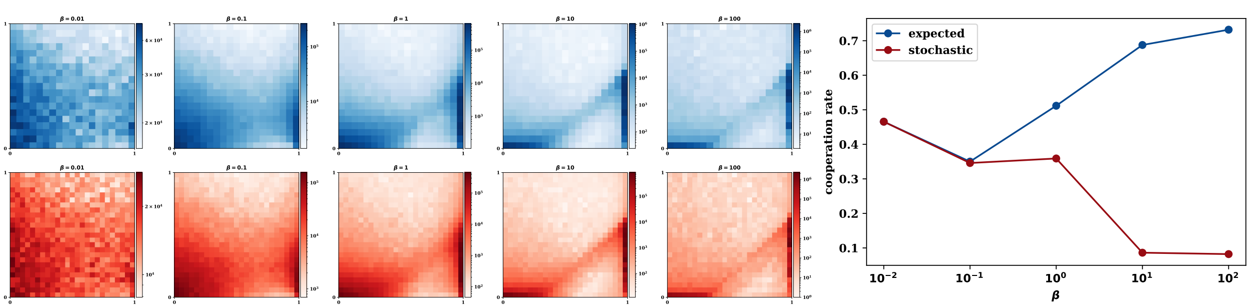


Figure 3: The evolution of cooperation for different parameter values. While the previous figures depict the evolutionary outcome for fixed parameter values, here we vary the benefit of cooperation b and the strength of selection β. In all cases, stochastic payoff evaluation tends to reduce the evolving 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

In evolution context 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. Suppose there are $N-k$ players who use the strategy $s_1 = (y_1, p_1, q_1)$, whereas k players use the strategy $s_2 = (y_2, p_2, q_2)$. We refer to these two player types as 'residents' and 'mutants', respectively.

Each step of the evolutionary process consists of two stages, a game stage and an updating stage.

1. In the game stage, each player is randomly matched with some other player in the population to interact in one instance of the IPD.

2. In the updating stage, two players are randomly drawn from the population, a 'learner' and a 'exemplar'. Given that the learner's payoff in the last round is $u_L \in \mathcal{U}$ and that the exemplar's last round's payoff $u_E \in \mathcal{U}$, we assume the learner adopts the role model's strategy with probability

$$\rho(u_L, u_E) = \frac{1}{1+\exp\left[-\beta(u_E - u_L)\right]}. \tag{5}$$

where $\beta \geq 0$ corresponds to the strength of selection.

We iterate this basic evolutionary step 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 $s_2' = (y_2', p_2', q_2')$ (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.

B Expected Payoffs

We compare this process for what we defined as **stochastic payoff evaluation** with the analogous process where players update their strategies with respect to their **expected** payoffs,

$$\begin{aligned}
\pi_1 &= \frac{N-k-1}{N-1} \cdot \langle \mathbf{v}(s_1, s_1), \mathbf{U} \rangle + \frac{k}{N-1} \cdot \langle \mathbf{v}(s_1, s_2), \mathbf{U} \rangle, \\
\pi_2 &= \frac{N-k}{N-1} \cdot \langle \mathbf{v}(s_2, s_1), \mathbf{U} \rangle + \frac{k-1}{N-1} \cdot \langle \mathbf{v}(s_2, s_2), \mathbf{U} \rangle.
\end{aligned} \tag{6}$$

In the limit of no discounting, $\delta \to 1$, this process based on expected payoffs has been considered in [8].

C Stochastic Payoffs

We define **stochastic payoff** as the average payoff $u \in \mathcal{U}$ a player receives in the last n rounds of the game given that they interact with m players.

Case $n = m = 1$.

Initially, consider the situation where $n = m = 1$. The player's stochastic payoff is what they receive in the last round against a single opponent. There only four possible outcomes for the last round, those are CC, CD, DC, DD. Consider two players with reactive strategies $S_1 = (y_1, p_1, q_1)$ and $S_2 = (y_2, p_2, q_2)$ who interact in a repeated prisoner's dilemma with continuation probability δ, the probability that are in each of the four possible states in the last round is given by:

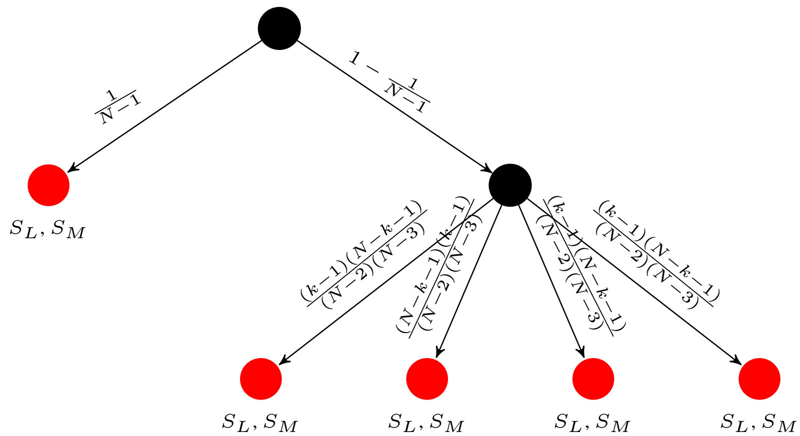
$$\mathbf{v}(s_1, s_2) = \Big(\mathbf{v}_R(s_1, s_2), \mathbf{v}_S(s_1, s_2), \mathbf{v}_T(s_1, s_2), \mathbf{v}_P(s_1, s_2)\Big). \tag{7}$$

where,

$$\begin{aligned}
\mathbf{v}_R(S_1, S_2) &= (1-\delta)\frac{y_1 y_2}{1-\delta^2 r_1 r_2} + \delta\frac{\Big(q_1 + r_1\big((1-\delta)y_2 + \delta q_2\big)\Big)\Big(q_2 + r_2\big((1-\delta)y_1 + \delta q_1\big)\Big)}{(1-\delta r_1 r_2)(1-\delta^2 r_1 r_2)}, \\
\mathbf{v}_S(S_1, S_2) &= (1-\delta)\frac{y_1 \bar{y}_2}{1-\delta^2 r_1 r_2} + \delta\frac{\Big(q_1 + r_1\big((1-\delta)y_2 + \delta q_2\big)\Big)\Big(\bar{q}_2 + \bar{r}_2\big((1-\delta)y_1 + \delta p_1\big)\Big)}{(1-\delta r_1 r_2)(1-\delta^2 r_1 r_2)}, \\
\mathbf{v}_T(S_1, S_2) &= (1-\delta)\frac{\bar{y}_1 y_2}{1-\delta^2 r_1 r_2} + \delta\frac{\Big(\bar{q}_1 + \bar{r}_1\big((1-\delta)y_2 + \delta p_2\big)\Big)\Big(q_2 + r_2\big((1-\delta)y_1 + \delta q_1\big)\Big)}{(1-\delta r_1 r_2)(1-\delta^2 r_1 r_2)}, \\
\mathbf{v}_P(S_1, S_2) &= (1-\delta)\frac{\bar{y}_1 \bar{y}_2}{1-\delta^2 r_1 r_2} + \delta\frac{\Big(\bar{q}_1 + \bar{r}_1\big((1-\delta)y_2 + \delta p_2\big)\Big)\Big(\bar{q}_2 + \bar{r}_2\big((1-\delta)y_1 + \delta p_1\big)\Big)}{(1-\delta r_1 r_2)(1-\delta^2 r_1 r_2)}.
\end{aligned} \tag{8}$$

Proof. Assume a repeated prisoner's dilemma between two reactive strategies. Given the continuation probability δ, probability that the game ends in the after the first round $(1-\delta)$ and the expected distribution of the four outcomes in the very first round is $\mathbf{v_0}$ defined as. Following the first round the, the outcome of the next rounds with a probability δ is M such that,

$$\dots$$



It can shown that, $(1-\delta)\mathbf{v_0}(I_4-\delta M)^{-1}$ and with some algebraic manipulation we derive to Equation 8.

□

Equation 8 is the probability vector that players S_1, S_2 are in each of the possible states of the last round. Given the population N with k mutants in the last round there only possible combinations of interactions are:

We compare this process for stochastic payoff evaluation with the analogous process where players update their strategies with respect to their *expected* payoffs,

$$\begin{aligned} \pi_1 &= \frac{N-k-1}{N-1} \cdot \langle \mathbf{v}(S_1, S_1), \mathbf{u} \rangle + \frac{k}{N-1} \cdot \langle \mathbf{v}(S_1, S_2), \mathbf{u} \rangle, \\ \pi_2 &= \frac{N-k}{N-1} \cdot \langle \mathbf{v}(S_2, S_1), \mathbf{u} \rangle + \frac{k-1}{N-1} \cdot \langle \mathbf{v}(S_2, S_2), \mathbf{u} \rangle. \end{aligned} \tag{9}$$

In the limit of no discounting, $\delta \to 1$, this process based on expected payoffs has been considered in [8].

C.1 Fixation probabilities under stochastic payoff evaluation

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

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

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. Given that the payoffs are u_1 and u_2, the imitation probability is then given by $\rho(u_1,u_2)$, as specified by Eq. (5). The probability that the respective payoffs of the players are given by u_1 and u_2 can be calculated as

$$\begin{aligned} x(u_1,u_2) = \quad & \frac{1}{N-1} \cdot v_{u_1}(S_1,S_2) \cdot 1_{(u_1,u_2)\in\mathcal{U}_F^2} \\ & + \left(1-\frac{1}{N-1}\right) \Big[\tfrac{k-1}{N-2}\tfrac{k-2}{N-3} v_{u_1}(S_1,S_2) v_{u_2}(S_2,S_2) + \tfrac{k-1}{N-2}\tfrac{N-k-1}{N-3} v_{u_1}(S_1,S_2) v_{u_2}(S_2,S_1) \\ & \qquad + \tfrac{N-k-1}{N-2}\tfrac{k-1}{N-3} v_{u_1}(S_1,S_1) v_{u_2}(S_2,S_2) + \tfrac{N-k-1}{N-2}\tfrac{N-k-2}{N-3} v_{u_1}(S_1,S_1) v_{u_2}(S_2,S_1) \Big]. \end{aligned} \tag{11}$$

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, $(u_1,u_2) \in \mathcal{U}_F^2 := \big\{(R,R),(S,T),(T,S),(P,P)\big\}$, 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.

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.