Supplementary Information: Evolution of cooperation among individuals with limited updating payoff memory

Nikoleta E. Glynatsi, Christian Hilbe, Alex McAvoy

Section 1 gives a brief overview of the pairwise comparison process. The pairwise comparison process consists of three phases; (1) the mutation phase (2) the game phase and (3) the update phase. In the update phase an individual adopts the strategy of another individual based on their "updating payoffs". In Section 2 we present/describe the conventional approach for calculating updating payoffs, as well as, our newly introduced approach.

1 Pairwise comparison process

Pairwise comparison process is a stochastic process for modelling the evolution of a finite population. The process starts with assigning all individuals of the population the same strategy. A strategy is a set of rules of how an individual should behave in an interaction with another individual. Each elementary time step of the process consists of three phases; (1) the **mutation phase** (2) the **game phase** and (3) the **update phase**. These are summarised in Figure 1.

In the **mutation phase** one individual is chosen to switch to a new mutant strategy with a probability μ. In the **game phase** individuals are randomly matched with other individuals in the population, and they engage in a repeated game where each subsequent turn occurs with a fixed probability δ. At each turn the individuals decide on an action based on their strategies. In repeated games there are infinite many strategies, however, it is commonly assumed that individuals can only choose strategies from a restricted set. One such set is that of reactive strategies. A reactive strategy considers only the previous action of the other player, and thus, a reactive strategy s can be written as a three-dimensional vector $s = (y, p, q)$. The parameter 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 **update stage** where two individuals are randomly selected. From the two individuals, one serves as the 'learner' and the other as the 'role model'. The learner adopts the role model's strategy with a probability ρ given by,

$$\rho(\pi_L, \pi_{RM}) = \frac{1}{1+e^{-\beta(\pi_{RM}-\pi_L)}}. \tag{1}$$

π_L and π_{RM} are the updating payoffs of the learner and the role model respectively. The updating payoffs are a measure of how successful individuals are in the current standing of the population. The parameter β is known as the selection strength, namely, it shows how important the payoff difference is when the learner is considering adopting the strategy of the role model.

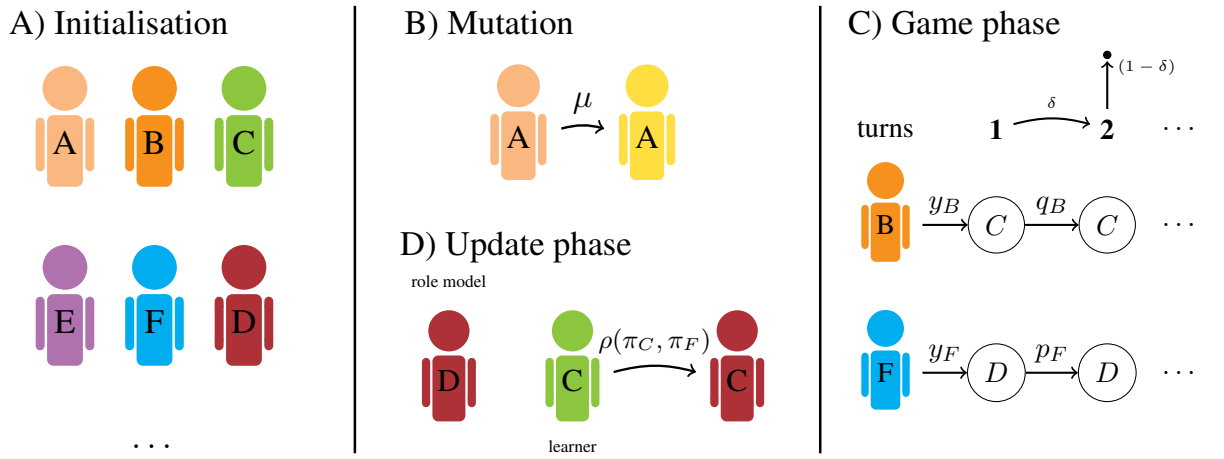

Figure 1: Pairwise comparison process phases. A) Initialisation The process begins with a finite population where each member is assigned a given strategy. Each color represents a different strategy, and the members are labelled by letters. **B) Mutation phase.** An individual is selected (in the example individual A) and with a given probability μ that individual adopts a new strategy. **C) Game phase.** Individuals are selected to interact in a repeated social dilemma with other individuals. We demonstrate the case where individuals B and F have been selected to interact. They use the reactive strategies $s_B = (y_B, p_B, q_B)$ and $s_F = (y_F, p_F, q_F)$ respectively. The opening moves depend on their y_i probability. In turn 1, individual B cooperated, thus, F cooperates with a probability p_F in turn 2. On the opposite, individual F defected in turn 1, and so B cooperates in the next turn with a probability q_B. At each turn there is a probability δ that a subsequent turn will occur, and with a probability $(1-\delta)$ the interaction ends. **D) Update phase.** At the updating phase two individuals are chosen; one serves as the role of the learner and the other one as the role model. In our example C adopts D's strategy with a probability $\rho(\pi_C, \pi_D)$ where π_C, π_D denote the updating payoffs of the individuals.

This elementary step of the process (mutation, game and update phases) is repeated for a large number of time steps, and at each time step we record the state of the population.

1.1 Low mutation $\mu \to 0$

In the case of low mutation ($\mu \to 0$) we assume that mutations are rare. In fact, so rare that only two different strategies can be present in the population at any given time. The case of low mutation is vastly adopted because it allows us to explicitly calculate the fixation probability of a newly introduced mutant.

More specifically, the process again starts with a population where all members are of the same strategy. At each step one individual adopts a mutant strategy randomly selected from the set of feasible strategies. The fixation probability ϕ_M of the mutant strategy can be calculated explicitly,

$$\varphi = \frac{1}{1 + \sum_{i=1}^{N-1} \prod_{k}^{i} \frac{\lambda_k^-}{\lambda_k^+}}, \tag{2}$$

where λ_k^-, λ_k^+ are the probabilities that the number of mutants decreases and increases respectively, N is the size of the population, and k is the number of mutants. The probabilities λ_k^- and λ_k^+ depend on the updating payoffs of the mutant and the resident strategies. 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. The process is summarised by Algorithm 1.

Algorithm 1: Evolutionary process

```
N ← population size;
k ← 1;
resident ← starting resident;
while t < maximum number of steps do
    mutant ← random: {∅} → R^n;
    fixation probability ← φ;
    if φ > random: i → [0, 1] then
        resident ← mutant;
    end
end
```

Most of the results we present in this work consider the case of low mutation, however, we have also verified that the main result holds in the case of high mutation rates.

2 Updating Payoffs

The updating payoffs depend on the interactions of the individuals at the game phase. In this work we assume that in the game phase individuals are matched in pairs and that they participate in a repeated 2 persons donation game. In the donation game there are two actions: cooperation (C) and defection (D). By cooperating a player provides a benefit b to the other player at their cost c, with $0 < c < b$. Thus the payoffs for a player in each turn are,

$$\begin{array}{c} \\ \text{cooperate} \\ \text{defect} \end{array} \begin{array}{c} \begin{array}{cc} \text{cooperate} & \text{defect} \end{array} \\ \begin{pmatrix} b-c & -c \\ b & 0 \end{pmatrix} \end{array}. \tag{3}$$

Let $\mathbf{u} = (b - c, -c, b, 0)$ be payoffs in a vector format, and let $\mathcal{U} = \{r, s, t, p\}$ denote the set of feasible payoffs, where r denotes the payoff of mutual cooperation, s the sucker's payoff, t the temptation to defect payoff, and p the punishment payoff.

In the following subsections we present several approaches for calculating the updating payoffs. Initially, we discuss the conventional approach of the expected payoffs and afterwards we present our newly introduced approach.

2.1 Updating Payoffs based on the expected payoffs

The expected payoffs are the conventional payoffs used in the updating stage. They are defined as the mean payoff of an individual in a well-mixed population that engages in an infinitely repeated games with all other population members. In an infinitely repeated game the payoff of a reactive strategy can explicitly be calculated using a markovian approach . Namely, assume two reactive strategies $s_1 = (y_1, p_1, q_1)$ and $s_2 = (y_2, p_2, q_2)$, their play can be defined as a Markov process with the transition matrix M,

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

Note that y_1 and y_2 are omitted since they have no effect in the long run of the game. The stationary vector $\mathbf{v}$, which is the solution to $\mathbf{v}M = \mathbf{v}$, combined with the payoff vector $\mathbf{u}$, yields the game stage outcome for each strategy,

$$\langle \mathbf{v}(s_1, s_2), \mathbf{u} \rangle \quad \text{and} \quad \langle \mathbf{v}(s_2, s_1), \mathbf{u} \rangle.$$

In the case of low mutation there can be only one type of mutant strategy in the population. So in a population of size N, there will be k mutants and $N - k$ residents, whose strategies we denote respectively as $s_M = (y_M, p_M, q_M)$ and $s_R = (y_R, p_R, q_R)$. The expected payoffs of a resident (π_R) and for a mutant (π_M) are give by,

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

The number of mutants in the population increases if a resident adopts the strategy of a mutant, and decreases if a mutant adopts the strategy of a resident. The probabilities that the number of mutants decreases and increases, λ_k^- and λ_k^+, are now explicitly defined as,

$$\lambda_k^- = \rho(\pi_M, \pi_R) \quad \text{and} \quad \lambda_k^+ = \rho(\pi_R, \pi_M).$$

Simulation Results based on the expected payoffs

We can simulate the evolutionary process described by Algorithm 1 when individuals use reactive strategies and update their strategies based on their expected payoffs. We performed two independent runs of the process where we differed the benefit of cooperation b. The results are shown in Figure 2.

It is observed that a higher value of benefit results in a more cooperative population. For a low benefit the resident population cooperates on average 52% of the time. For a a high benefit the average cooperation increases to 98%. For both a low and a high benefit the resident population consists either of defectors or conditional cooperators. A conditional cooperator, otherwise known as a generous tit for tat (GTFT) strategy, always cooperates if the co-player cooperated ($p \approx 1$) and cooperates with a probability q if the co-player defected. The conditional cooperator strategy adopted by the population differs between the two simulations. A higher value of benefit results to a higher tolerance to defection.

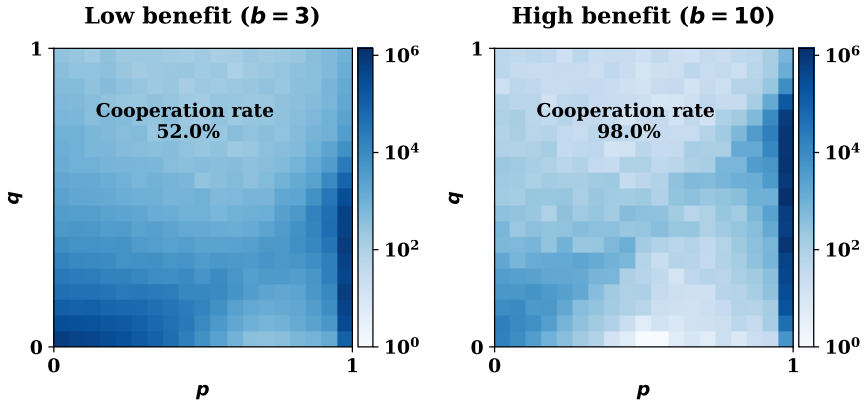

Figure 2: Evolutionary dynamics under expected payoffs with low (left) and high (right) benefit. We run the simulations for $T = 10^7$ time steps. For each time step, we have recorded the current resident population (y, p, q). Since in the case of the expected payoffs the individuals interact for an infinite number of times $\delta \to 1$, 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. In both cases players update based on their expected payoffs (A) In the case where $b = 3$, the resident population either consists of defectors (with $p \approx q \approx 0$) or of conditional cooperators for which $p \approx 1$ and $q \leq 1 - \frac{1}{3} = 0.7$. (B) In the case where $b = 10$, the resident population typically applies a strategy for which $p \approx 1$ and $q \leq 1 - \frac{1}{10} = 0.9$. Parameters: $N = 100, c = 1, \beta = 1$.

Invasion Analysis based on the expected payoffs

We can explicitly calculate the value of q which evolves. Namely, we will calculate how easily a single ALLD mutant can invade into a resident population with strategy GTFT. In that case, GTFT$= (1, 1, q)$, ALLD$= (0, 0, 0)$, and $k = 1$.

When two GTFT players interact in an infinitely repeated game, their respective probabilities for each of the four outcomes in the last round simplify to,

$$\mathbf{v}(\text{GTFT}, \text{ALLD}) = (1, 0, 0, 0).$$

On the other hand, if an ALLD player interacts with a GTFT player, the respective probabilities become,

$$\mathbf{v}(\text{ALLD}, \text{GTFT}) = (0, q, 0, (1-q)).$$

Using the above we can define the payoffs of a GTFT individual (resident) and of the ALLD individual (mutant) follows,

$$\pi_{\text{GTFT}} = \frac{N-2}{N-1}(b-c) - \frac{qc}{N-1} \quad \text{and} \quad \pi_{\text{ALLD}} = bq.$$

As a consequence, we can calculate the ratio of transition probabilities as,

$$\frac{\lambda^+}{\lambda^-} = \frac{\rho(\pi_{\text{GTFT}}, \pi_{\text{ALLD}})}{\rho(\pi_{\text{ALLD}}, \pi_{\text{GTFT}})} = \frac{e^{-\beta\left(\frac{(N-2)(b-c)}{N-1} - q(b+\frac{c}{N-1})\right)} + 1}{e^{-\beta\left(bq - \frac{b(N-2)}{N-1} - \frac{c(N-2-q)}{N-1}\right)} + 1}$$

In particular, in the limit of strong selection $\beta \to \infty$ and large populations $N \to \infty$, we obtain that the ratio is than smaller to 1 if $q \leq 1 - \frac{c}{b}$. Thus, ALLD is disfavored to invade if $q \leq 1 - \frac{c}{b}$. For $q = 1 - \frac{c}{b}$ the probability that the number of mutants increase by one equals the probability that the mutant goes extinct.

2.2 Updating Payoffs based on the last last m round(s) payoffs of n interaction(s)

The aim of this work is to explore the effect that the updating payoffs have on the cooperation rate of the evolved population. To this end we introduce a new method calculating the updating payoffs. Compared to the literature where a infinite number of rounds and a well mixed population are assumed, we constrain the number of turns and number of turns.

Namely, in the expected payoffs the payoff of a pair depends on the average payoff they received over an infinite number of turns. In our approach, the payoff of a pair depends on the average payoffs they received in the last m turns. Furthermore, in expected payoffs it is assumed that a player interacts with every member of the population, whereas in our approach with consider one n interactions of the player with other members. Here we present results for four cases; ($m = 1$ and $n = 1$), ($m = 2$ and $n = 1$), ($m = 1$ and $n = 2$), and ($m = 2$ and $n = 2$).

2.2.1 Updating Payoffs based on the last round payoff of one interaction ($m = 1$ and $n = 1$)

In this case an individual updates her/his strategy based on the last payoff they received against one other member of the population. Initially, we define the probability that a reactive strategy receives the payoff $u \in \mathcal{U}$ in the very last round of the game against another reactive strategy (Proposition 1).

Proposition 1. *Consider a repeated game, with continuation probability δ, between players with reactive strategies $s_1 = (y_1, p_1, q_1)$ and $s_2 = (y_2, p_2, q_2)$ respectively. Then the probability that the s_1 player receives the payoff $u \in \mathcal{U}$ in the very last round of the game is given by $v_u(s_1, s_2)$, as given by Equation (6).*

$$
\begin{aligned}
v_r(s_1, s_2) &= (1-\delta)\frac{y_1 y_2}{1-\delta^2 l_1 l_2} + \delta\frac{\Big(q_1 + l_1\big((1-\delta)y_2 + \delta q_2\big)\Big)\Big(q_2 + l_2\big((1-\delta)y_1 + \delta q_1\big)\Big)}{(1-\delta l_1 l_2)(1-\delta^2 l_1 l_2)} \times r, \\
v_s(s_1, s_2) &= (1-\delta)\frac{y_1 \bar{y}_2}{1-\delta^2 l_1 l_2} + \delta\frac{\Big(q_1 + l_1\big((1-\delta)y_2 + \delta q_2\big)\Big)\Big(\bar{q}_2 + \bar{r}_2\big((1-\delta)y_1 + \delta p_1\big)\Big)}{(1-\delta l_1 l_2)(1-\delta^2 l_1 l_2)} \times s, \\
v_t(s_1, s_2) &= (1-\delta)\frac{\bar{y}_1 y_2}{1-\delta^2 l_1 l_2} + \delta\frac{\Big(\bar{q}_1 + \bar{r}_1\big((1-\delta)y_2 + \delta p_2\big)\Big)\Big(q_2 + l_2\big((1-\delta)y_1 + \delta q_1\big)\Big)}{(1-\delta l_1 l_2)(1-\delta^2 l_1 l_2)} \times t, \\
v_p(s_1, s_2) &= (1-\delta)\frac{\bar{y}_1 \bar{y}_2}{1-\delta^2 l_1 l_2} + \delta\frac{\Big(\bar{q}_1 + \bar{r}_1\big((1-\delta)y_2 + \delta p_2\big)\Big)\Big(\bar{q}_2 + \bar{r}_2\big((1-\delta)y_1 + \delta p_1\big)\Big)}{(1-\delta l_1 l_2)(1-\delta^2 l_1 l_2)} \times p.
\end{aligned}
\tag{6}
$$

In these expressions, we have used the notation $l_i := p_i - q_i$, $\bar{y}_i = 1 - y_i$, $\bar{q}_i := 1 - q_i$, and $\bar{l}_i := \bar{p}_i - \bar{q}_i = -l_i$ for $i \in \{1, 2\}$.

Note that in the proposition we here we focus on the case of the donation game/prisoner's dilemma but the result applies to any 2×2 symmetric game.

Proof. Given a play between two reactive strategies with continuation probability δ. The outcome at turn t is given by,

$$
(1-\delta)\mathbf{v_0}\sum \delta^t M^{(t)}, \tag{7}
$$

where $\mathbf{v_0}$ 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,

$$
\begin{aligned}
(1-\delta)\mathbf{v_0}\sum \delta^t M^{(t)} &= (1-\delta)(\mathbf{v_0} + \delta \mathbf{v_0} M + \delta^2 \mathbf{v_0} M^2 + \dots) \\
&= (1-\delta)\mathbf{v_0}(1 + \delta M + \delta^2 M^2 + \dots) \text{ using standard formula for geometric series} \\
&= (1-\delta)\mathbf{v_0}(I_4 - \delta M)^{-1}
\end{aligned}
$$

where $(1-\delta)\mathbf{v_0}(I_4 - \delta M)^{-1}$ is vector $\in R^4$ and it the probabilities for being in any of the outcomes CC, CD, DC, DD in the last round. Combining this with the payoff vector u and some algebraic manipulation we derive to the Equation 6. □

At each step of the evolutionary process we choose a role model and a learner to update the population. We consider the case where both the role model and the learner estimate their fitness after interacting with a single member of the population. There are five possible pairings at each step. They interact with other with a probability $\frac{1}{N-1}$, and thus they do not interact with other with a probability $1 - \frac{1}{N-1}$. In the latter case, each of them can interact with either a mutant or a resident. Both of them interact with a mutant with a probability $\frac{(k-1)(k-2)}{(N-2)(N-3)}$ and both interact with a resident with a probability $\frac{(N-k-1)(N-k-2)}{(N-2)(N-3)}$. The last two possible pairings are that either of them interacts with a resident whilst the other interacts with a mutant, and this happens with a probability $\frac{(N-k-1)(k-1)}{(N-2)(N-3)}$.

Given the possible pairings and Proposition 1, we define the probability that the respective last round payoffs of two players s_1, s_2 are given by u_1 and u_2 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 $\frac{1}{(N-1)}$. 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.

The probability that the number of mutants increases, and decreases respectively, by one is now given

by,

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

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

In this expression, $\frac{(N-k)}{N}$ is the probability that the randomly chosen learner is a resident, and $\frac{k}{N}$ 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_R and u_M that the two player may have received in their respective last rounds. We use $x(u_R, u_M)$ to denote the probability that the randomly chosen resident obtained a payoff of u_R in the last round of his respective game, and that the mutant obtained a payoff of u_M.

Simulation Results based on the last round payoff of one interaction

We simulate the evolutionary process given that individuals now use the the last round payoff of one interaction to update their strategies. The results are shown in Figure 3. Similarly to the results of the expected payoffs, a higher benefit results in a more cooperative population. However we note that the cooperation rate of the evolved populations are lower compared to the case of expected payoffs, and that the cooperation rate increases less between the two simulations. The evolving both consist of defectors and conditional cooperators. The evolved qs are smaller which also results to a less cooperative population. In between the two simulations $q \approx \frac{1}{2}$. In the invasion analysis we will demonstrate that in the case of the the last round payoff of one interaction, a conditional player needs to have $q \leq \frac{1}{2}$ to avoid being invaded by a defector. In both simulations $q \leq \frac{1}{2}$, the higher benefit pushes the evolved population closer to the bounty.

Invasion Analysis based on the last round payoff of one interaction

We once again calculate how easily a single ALLD mutant can invade into a resident population with strategy GTFT. When two GTFT players interact in the game, their respective probabilities for each of the four outcomes in the last round simplify to,

$$v_r(GTFT, GTFT) = 1, \quad v_t(GTFT, GTFT) = 0,$$
$$v_s(GTFT, GTFT) = 0, \quad v_p(GTFT, GTFT) = 0.$$

On the other hand, if an ALLD player interacts with a GTFT player, the respective probabilities according

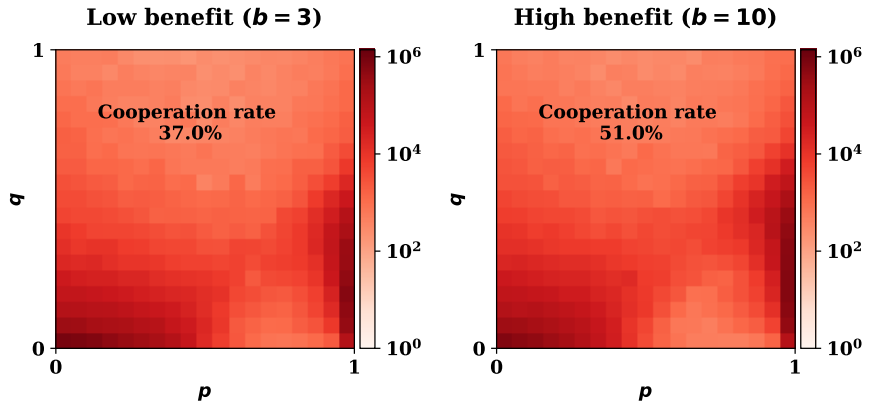

Figure 3: Evolutionary dynamics under one interaction last round payoffs with low (left) and high (right) benefit. The cooperation rates in both cases are less compared to the expected payoffs. Moreover, increase is less when the benefit is increased to ten. In the cases of low and high benefit case the resident population either consists of defectors (with $p \approx q \approx 0$) or of conditional cooperators for which $p \approx 1$ and $q \leq \frac{1}{2}$. Parameters: $N = 100, c = 1, \beta = 1$.

to Eq. 6 become

$$v_r(ALLD, GTFT) = 0, \qquad v_s(ALLD, GTFT) = 0,$$
$$v_t(ALLD, GTFT) = 1 - \delta + \delta q, \qquad v_p(ALLD, GTFT) = \delta(1 - q).$$

As a consequence, we obtain the following probabilities $x(u_1, u_2)$ that the payoff of a randomly chosen GTFT player is u_1 and that the payoff of the ALLD player is u_2,

$$\begin{aligned} x(r,t) &= \frac{N-2}{N-1} \cdot (1 - \delta + \delta q) \\ x(r,r) &= \frac{N-2}{N-1} \cdot \delta(1-q) \\ x(s,r) &= \frac{1}{N-1} \cdot (1 - \delta + \delta q) \\ x(p,p) &= \frac{1}{N-1} \cdot \delta(1-q) \end{aligned}$$

We now calculate the ratio of transition probabilities as

$$\frac{\lambda^+}{\lambda^-} = \frac{\frac{N-2}{N-1}\left(\frac{\delta(1-q)}{1+e^{-\beta(b-c)}} + \frac{\delta q-\delta+1}{e^{\beta c}+1}\right) + \frac{1}{N-1}\left(\frac{\delta(1-q)}{2} + \frac{\delta q-\delta+1}{1+e^{-\beta(-b-c)}}\right)}{\frac{N-2}{N-1}\left(\frac{\delta(1-q)}{1+e^{-\beta(-b+c)}} + \frac{\delta q-\delta+1}{1+e^{-\beta c}}\right) + \frac{1}{N-1}\left(\frac{\delta(1-q)}{2} + \frac{\delta q-\delta+1}{1+e^{-\beta(b+c)}}\right)}$$

In particular, in the limit of strong selection $\beta \to \infty$ and large populations $N \to \infty$, we obtain

$$\frac{\lambda^+}{\lambda^-} = \frac{1-\delta+\delta q}{\delta(1-q)}$$

This ratio is smaller than 1 (such that ALLD is disfavored to invade) if $q < 1 - 1/(2\delta)$. For infinitely repeated games, $\delta \to 1$, this condition becomes $q < 1/2$ (for $q = 1/2$, the payoff of the ALLD player is $r > r$ for half of the time, and it is $p < r$ for the other half. The probability that the number of mutants increase by one equals the probability that the mutant goes extinct).

2.2.2 Updating Payoffs based on the last round payoff of two interactions ($m = 1$ and $n = 2$)

Here, in the updating phase we choose one role model and one learner, however now, we consider the case where both the role model and the learner estimate their updating payoff after interacting with two members of the population. There are stages in the pairing process and there are twenty four possible pairings.

In the first stage the role model and the learner are matched together with a probability $\frac{1}{N-1}$ or not with a probability $(1 - \frac{1}{N-1})$. If they were matched then in the second stage of pairings there are only four possible outcomes; both of them interact with a mutant with a probability $\frac{(k-1)(k-2)}{(N-2)(N-3)}$ and both interact with a resident with a probability $\frac{(N-k-1)(N-k-2)}{(N-2)(N-3)}$. The last two possible pairings are that either of them interacts with a resident whilst the other interacts with a mutant, and this happens with a probability $\frac{(N-k-1)(k-1)}{(N-2)(N-3)}$.

In the later case, where they were not matched in the first stage, there are four possible outcomes; both of them interact with a mutant and both interact with a resident, either of them interacts with a resident whilst the other interacts with a mutant. For each of the above pairings of the first stage there are five possible; they interact with each other, both of them interact with a mutant and both interact with a resident, either of them interacts with a resident whilst the other interacts with a mutant.

The new possible pairings change how we define the probability that the respective last round payoffs of two players s_1, s_2 are given by u_1 and u_2. The probability denoted as $\bar{x}(u_1, u_2)$ is given by,

$$\begin{aligned}
x(u_1,u_2) = \frac{1}{N-1}\cdot v_{u_1}(s_1,s_2)\cdot 1_{(u_1,u_2)\in\mathcal{U}_F^2}\cdot A + \left(1-\frac{1}{N-1}\right)[\\
v_{u_1}(s_1,s_2)_{u_2}(s_2,s_2)\tfrac{(k-2)(k-1)}{(N-2)(N-3)}\left(\tfrac{1}{N-2}\cdot v_{u_1}(s_1,s_2)\cdot 1_{(u_1,u_2)\in\mathcal{U}_F^2} + (1-\tfrac{1}{N-2})[B_1+B_2+B_3+B_4]\right) +\\
v_{u_1}(s_1,s_2)_{u_2}(s_2,s_1)\tfrac{(k-1)(N-k-1)}{(N-2)(N-3)}\left(\tfrac{1}{N-2}\cdot v_{u_1}(s_1,s_2)\cdot 1_{(u_1,u_2)\in\mathcal{U}_F^2} + (1-\tfrac{1}{N-2})[C_1+C_2+C_3+C_4]\right) +\\
v_{u_1}(s_1,s_1)_{u_2}(s_2,s_2)\tfrac{(k-1)(N-k-1)}{(N-2)(N-3)}\left(\tfrac{1}{N-2}\cdot v_{u_1}(s_1,s_2)\cdot 1_{(u_1,u_2)\in\mathcal{U}_F^2} + (1-\tfrac{1}{N-2})[D_1+D_2+D_3+D_4]\right) +\\
v_{u_1}(s_1,s_1)_{u_2}(s_2,s_1)\tfrac{(N-k-2)(N-k-1)}{(N-2)(N-3)}\left(\tfrac{1}{N-2}\cdot v_{u_1}(s_1,s_2)\cdot 1_{(u_1,u_2)\in\mathcal{U}_F^2} + (1-\tfrac{1}{N-2})[E_1+E_2+E_3+E_4]\right)\Big]
\end{aligned} \tag{11}$$

$$
\begin{aligned}
A \;&= \left(1-\tfrac{1}{N-1}\right)\Big[\tfrac{k-1}{N-2}\tfrac{k-2}{N-3}v_{u_1}(s_1,s_2)v_{u_2}(s_2,s_2)+\tfrac{k-1}{N-2}\tfrac{N-k-1}{N-3}v_{u_1}(s_1,s_2)v_{u_2}(s_2,s_1)+\\
&\tfrac{N-k-1}{N-2}\tfrac{k-1}{N-3}v_{u_1}(s_1,s_1)v_{u_2}(s_2,s_2)+\tfrac{N-k-1}{N-2}\tfrac{N-k-2}{N-3}v_{u_1}(s_1,s_1)v_{u_2}(s_2,s_1)\Big]\\
B_1 &= \tfrac{(k-3)(k-2)}{(N-3)(N-4)}v_{u_1}(s_1,s_2)v_{u_2}(s_2,s_2) \quad B_2 = \tfrac{(k-2)(N-k-1)}{(N-3)(N-4)}v_{u_1}(s_1,s_1)v_{u_2}(s_2,s_2)\\
B_3 &= \tfrac{(k-2)(N-k-1)}{(N-3)(N-4)}v_{u_1}(s_1,s_2)v_{u_2}(s_2,s_1) \quad B_4 = \tfrac{(N-k-2)(N-k-1)}{(N-3)(N-4)}v_{u_1}(s_1,s_1)v_{u_2}(s_2,s_1)\\
C_1 &= \tfrac{(k-3)(k-1)}{(N-3)(N-4)}v_{u_1}(s_1,s_2)v_{u_2}(s_2,s_2) \quad C_2 = \tfrac{(k-1)(N-k-1)}{(N-3)(N-4)}v_{u_1}(s_1,s_1)v_{u_2}(s_2,s_2)\\
C_3 &= \tfrac{(k-2)(N-k-2)}{(N-3)(N-4)}v_{u_1}(s_1,s_2)v_{u_2}(s_2,s_1) \quad C_4 = \tfrac{(N-k-2)^2}{(N-3)(N-4)}v_{u_1}(s_1,s_1)v_{u_2}(s_2,s_1)\\
D_1 &= \tfrac{(k-2)^2}{(N-3)(N-4)}v_{u_1}(s_1,s_2)v_{u_2}(s_2,s_2) \quad D_2 = \tfrac{(k-2)(N-k-2)}{(N-3)(N-4)}v_{u_1}(s_1,s_1)v_{u_2}(s_2,s_2)\\
D_3 &= \tfrac{(k-1)(N-k-1)}{(N-3)(N-4)}v_{u_1}(s_1,s_2)v_{u_2}(s_2,s_1) \quad D_4 = \tfrac{(N-k-3)(N-k-1)}{(N-3)(N-4)}v_{u_1}(s_1,s_1)v_{u_2}(s_2,s_1)\\
E_1 &= \tfrac{(k-2)(k-1)}{v}{}_{u_1}(s_1,s_2)v_{u_2}(s_2,s_2) \quad E_2 = \tfrac{(k-1)(N-k-2)}{(N-3)(N-4)}v_{u_1}(s_1,s_1)v_{u_2}(s_2,s_2)\\
E_3 &= \tfrac{(k-1)(N-k-2)}{(N-3)(N-4)}v_{u_1}(s_1,s_2)v_{u_2}(s_2,s_1) \quad E_4 = \tfrac{(N-k-3)(N-k-2)}{(N-3)(N-4)}v_{u_1}(s_1,s_1)v_{u_2}(s_2,s_1)
\end{aligned}
\tag{12}
$$

The first term on the right side corresponds to the case that the learner and the role model happened to be matched during the first stage of pairing, followed by them being paired with another member of the population on the second stage. The second terms corresponds to the case that the learner and the role model interact with a mutant with a probability ($\frac{(k-2)(k-1)}{(N-2)(N-3)}$). In the seconds stage, they can either interact with each other $\frac{1}{N-2}$ or not $(1-\frac{1}{N-2})$. If they do not interact with each other, then each of the following can happen: both of them interact with a mutant with a probability $\frac{(k-3)(k-2)}{(N-4)(N-3)}$ and both interact with a resident with a probability $\frac{(N-k-1)(N-k-2)}{(N-3)(N-4)}$. The last two possible pairings are that either of them interacts with a resident whilst the other interacts with a mutant, and this happens with a probability $\frac{(N-k-2)(k-1)}{(N-4)(N-3)}$. The rest of the cases follow the same pattern.

The probability that the number of mutants increases, and decreases respectively, by one is now given by,

$$\lambda_k^+ = \frac{N-k}{N} \cdot \frac{k}{N} \cdot \sum_{u_R,u_M \in \mathcal{U}} \bar{x}(u_R,u_M) \cdot \rho(u_R,u_M), \tag{13}$$

$$\lambda_k^- = \frac{N-k}{N} \cdot \frac{k}{N} \cdot \sum_{u_R,u_M \in \mathcal{U}} \bar{x}(u_R,u_M) \cdot \rho(u_M,u_R). \tag{14}$$

Simulation Results based on the last round payoff of two interactions

We simulate the evolutionary process when individuals update their strategies based on the last round payoffs from two interactions with other members of the population, Figure 3. For both a low and a high benefit value the average cooperation rate has increased compared to the the last round against a single other member of the population. In the case of low benefit the evolved population's cooperation rate is almost the same as in the case of expected payoffs. However, the cooperation rate does not increase as much as in for the expected payoffs for $b = 10$.

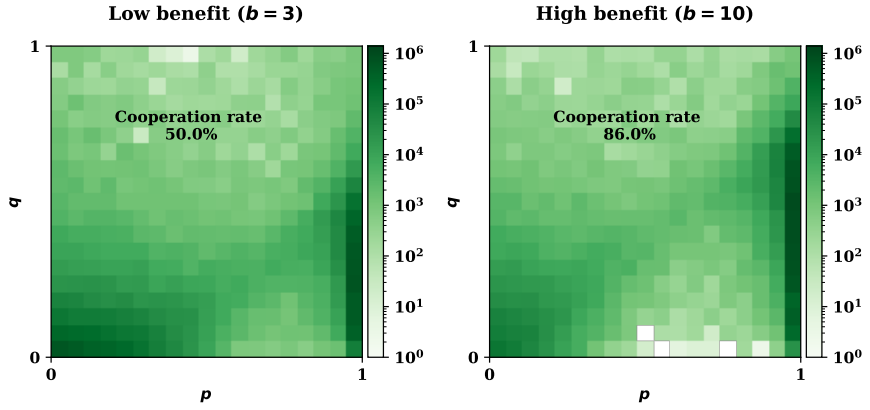

Figure 4: Evolutionary dynamics under two interactions last round payoffs with low (left) and high (right) benefit. In the cases of low and high benefit case the resident population either consists of defectors (with $p \approx q \approx 0$) or of conditional cooperators for which $p \approx 1$ and $q \leq \frac{\sqrt{2}}{2}$. Parameters: $N = 100, c = 1, \beta = 1$.

Invasion Analysis based on the last round payoff of two interactions

In a similar fashion we can calculate the threshold of q for which a GTFT player can not be invaded by an ALLD mutant. The ratio of transition probabilities is given by,

$$\frac{\lambda^+}{\lambda^-} = \frac{\frac{N-3}{N-1}\left(\frac{\delta^2(1-q)^2}{1+e^{-\beta(-b+c)}} + \frac{2\delta(1-q)(\delta q-\delta+1)}{1+e^{-\beta\left(-\frac{b}{2}+c\right)}} + \frac{(\delta q-\delta+1)^2}{1+e^{-\beta c}}\right) + \frac{2}{N-1}\left(\frac{\delta^2(1-q)^2}{1+e^{-\beta\left(-\frac{b}{2}+\frac{c}{2}\right)}} + \frac{\delta(1-q)(\delta q-\delta+1)}{1+e^{-\frac{\beta c}{2}}} + \frac{\delta(1-q)(\delta q-\delta+1)}{1+e^{-\beta c}} + \frac{(\delta q-\delta+1)^2}{1+e^{-\beta\left(\frac{b}{2}+c\right)}}\right)}{\frac{N-3}{N-1}\left(\frac{\delta^2(q-1)^2}{1+e^{-\beta(b-c)}} + \frac{2\delta(1-q)(\delta q-\delta+1)}{1+e^{-\beta\left(\frac{b}{2}-c\right)}} + \frac{(\delta q-\delta+1)^2}{e^{\beta c}+1}\right) + \frac{2}{N-1}\left(\frac{\delta^2(q-1)^2}{1+e^{-\beta\left(\frac{b}{2}-\frac{c}{2}\right)}} + \frac{\delta(1-q)(\delta q-\delta+1)}{e^{\beta c}+1} + \frac{\delta(1-q)(\delta q-\delta+1)}{e^{\frac{\beta c}{2}}+1} + \frac{(\delta q-\delta+1)^2}{1+e^{-\beta\left(-\frac{b}{2}-c\right)}}\right)} \tag{15}$$

In the limit of strong selection $\beta \to \infty$ and large populations $N \to \infty$, we obtain

$$\frac{\lambda^+}{\lambda^-} = \begin{cases} \frac{(\delta q - \delta + 1)^2}{\delta^2 (1-q)^2 + 2\delta(1-q)(\delta q - \delta + 1)} & \frac{b}{2} > c \\ \frac{-\delta q + \delta - 1}{\delta(q-1)} & \frac{b}{2} = c \\ -\frac{(\delta q - \delta - 1)(\delta q - \delta + 1)}{\delta^2 (q-1)^2} & \frac{b}{2} < c \end{cases} \tag{16}$$

Note that in this case the relationship of the payoffs, cost and benefit, have an effect of the q. Namely if $\frac{b}{2}$ is higher than the cost c. In the case where $\frac{b}{2} = c$ the result is the same as in the case of $m = n = 1$. For $\frac{\lambda^+}{\lambda^-} < 1$:

$$\begin{cases} q \in \{\frac{\delta - 1 + \frac{\sqrt{2}}{2}}{\delta}\} & \frac{b}{2} > c \\ q = \frac{\delta - \frac{1}{2}}{\delta} & \frac{b}{2} = c \\ q \in \{\frac{\delta - \frac{\sqrt{2}}{2}}{\delta}, \ \frac{\delta + \frac{\sqrt{2}}{2}}{\delta}\} & \frac{b}{2} < c \end{cases} \tag{17}$$

There are two distinguished cases. For $\frac{b}{2} > c$ the ration is smaller than if $q < \frac{\delta - 1 + \frac{\sqrt{2}}{2}}{\delta}$. For infinitely repeated games, $\delta \to 1$, this condition becomes $q < \frac{\sqrt{2}}{2}$. In the case of $\frac{b}{2} < c$ there are two possible roots, however, we assume repeated games that are repeated for a large number of turn such as $\delta \to 1$. The condition then becomes $q < 1 - \frac{\sqrt{2}}{2}$.

2.3 Updating Payoffs based on the last two rounds payoff of one interactions ($m = 2$ and $n = 1$)

We extend Proposition 1 to n rounds. The result is given by Proposition 2.

Proposition 2. *Consider a repeated game, with continuation probability δ, between players with reactive strategies $s_1 = (y_1, p_1, q_1)$ and $s_2 = (y_2, p_2, q_2)$ respectively. Let $\tilde{\mathcal{U}}$ denote the set of feasible payoffs in the last n rounds, and let $\tilde{\mathbf{u}}$ be the corresponding payoff vector. Then the probability that the s_1 player receives the payoff $u \in \tilde{\mathcal{U}}$ in the very last two rounds of the game is given by,*

$$\langle \tilde{\mathbf{v}}(s_1, s_2), \tilde{\mathbf{u}} \rangle, \text{ where } \tilde{\mathbf{v}} \in R^{4^n} \text{ is given by }, \tag{18}$$

$$\tilde{\mathbf{v}}(s_1, s_2) = (1-\delta) w_{a_1,a_2} \delta^2 \left[\mathbf{v_0} (I_4 - \delta M)^{-1} \right]_{a_1,a_2}, \quad w_{a_1,a_2} \in M \ \forall \ a_1, a_2 \in \{1,2,3,4\}. \tag{19}$$

Here consider the case of $n = 2$. The framework remains the same as in section 2.2.1, the only difference is the game stage payoffs.

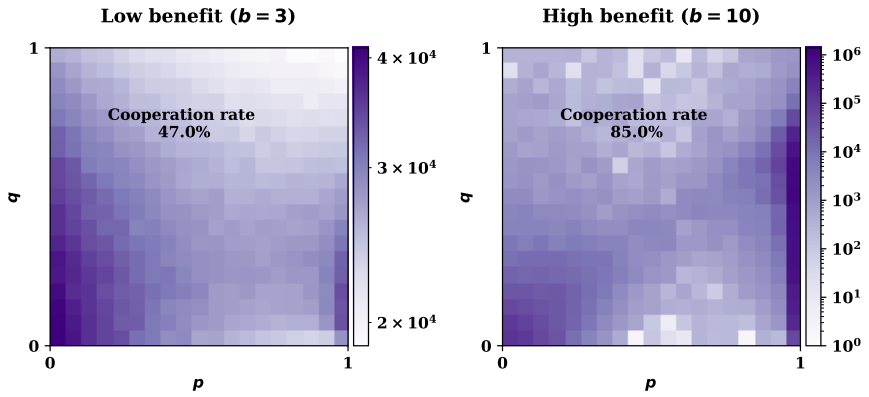

Figure 5: Evolutionary dynamics under two interactions last round payoffs with low (left) and high (right) benefit. In the cases of low and high benefit case the resident population either consists of defectors (with $p \approx q \approx 0$) or of conditional cooperators for which $p \approx 1$ and $q \leq \frac{1}{2}$. Parameters: $N = 100, c = 1, \beta = 1$.

Simulation Results based on the last two rounds payoff of one interaction

Invasion Analysis based on the last two rounds payoff of one interaction

$$\frac{\lambda^+}{\lambda^-} = \frac{\frac{\delta^2(N-2)}{N-1}\left(\frac{\delta(1-q)^2}{1+e^{-\beta(-b+c)}}+\frac{q(\delta q-\delta+1)}{1+e^{-\beta c}}+\frac{(1-q)(2\delta q-\delta+1)}{1+e^{-\beta\left(-\frac{b}{2}+c\right)}}\right)+\frac{1}{N-1}\left(\frac{\delta(1-q)^2}{2}+\frac{q(\delta q-\delta+1)}{1+e^{-\beta(b+c)}}+\frac{(1-q)(2\delta q-\delta+1)}{1+e^{-\beta\left(\frac{b}{2}+\frac{c}{2}\right)}}\right)}{\frac{\delta^2(N-2)}{N-1}\left(\frac{\delta(1-q)^2}{1+e^{-\beta(b-c)}}+\frac{q(\delta q-\delta+1)}{e^{\beta c}+1}+\frac{(1-q)(2\delta q-\delta+1)}{1+e^{-\beta\left(\frac{b}{2}-c\right)}}\right)+\frac{1}{N-1}\left(\frac{\delta(1-q)^2}{2}+\frac{q(\delta q-\delta+1)}{1+e^{-\beta(-b-c)}}+\frac{(1-q)(2\delta q-\delta+1)}{1+e^{-\beta\left(-\frac{b}{2}-\frac{c}{2}\right)}}\right)} \tag{20}$$

In the limit of strong selection $\beta \to \infty$ and large populations $N \to \infty$, we obtain

$$\frac{\lambda^+}{\lambda^-} = \begin{cases} \frac{q(\delta q-\delta+1)}{\delta(1-q)^2+(1-q)(2\delta q-\delta+1)} & \frac{b}{2} > c \\ -\frac{\delta q-\delta+q+1}{(\delta+1)(q-1)} & \frac{b}{2} = c \\ \frac{q(\delta q-\delta+1)+(1-q)(2\delta q-\delta+1)}{\delta(1-q)^2} & \frac{b}{2} < c \end{cases} \tag{21}$$

For $\frac{\lambda^+}{\lambda^-} < 1$:

$$\begin{cases} q \in \{\frac{\delta-\sqrt{\delta^2+1}-1}{2\delta},\ \frac{\delta+\sqrt{\delta^2+1}-1}{2\delta}\} & \frac{b}{2} > c \\ q = \frac{\delta}{\delta+1} & \frac{b}{2} = c \\ q \in \{1-\frac{\sqrt{2}}{2\sqrt{\delta}},\ 1+\frac{\sqrt{2}}{2\sqrt{\delta}}\} & \frac{b}{2} < c \end{cases} \tag{22}$$

2.4 Updating Payoffs based on the last two rounds payoff of two interactions ($m = 2$ and $n = 2$)

The last case for which we present results in this work is the case of $m = 2$ and $n = 2$, that is, an individual considers the last two rounds payoffs he/she received against two different members of the population. The

probability that the number of mutants increases, and decreases respectively, by one is now given by, are based on the approached on section 2.2.2 and 2.3. In this case we only present only the results on the numerical simulations which are given in Figure 6.

The evolved cooperation rates are between the other cases and the expected payoffs. For a low benefit we get results similar to the case where one the last round was considered, however, in the high benefit case the evolved cooperation rate increases to 91% which is higher then any of the newly introduced approaches. Though we do no demonstrate or method of more cases, we assume, that as n, and m increase the results tend to the expected payoffs.

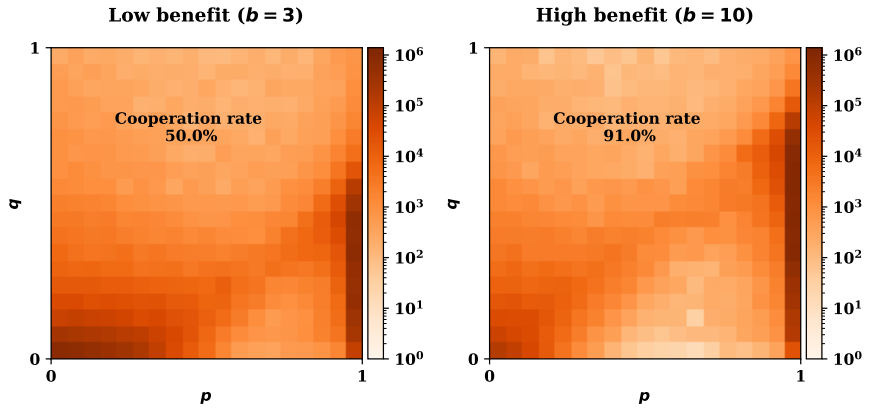

Figure 6: Evolutionary dynamics under two interactions last round payoffs with low (left) and high (right) benefit. In the cases of low and high benefit case the resident population either consists of defectors (with $p \approx q \approx 0$) or of conditional cooperators for which $p \approx 1$ and $q \leq \frac{1}{2}$. Parameters: $N = 100, c = 1, \beta = 1$.

3 Expected and Last Round Updating Payoffs for Memory One Strategies

So far we have only considered the case where members can adopt reactive strategies during the game phase. In order to demonstrate that our results hold for higher memory strategies here we present numerical results in the case of memory-one strategies.

Memory-one strategies consider the outcome of the previous round to decide on an action. There are four possible outcome in each round; $(C, C), (C, D), (D, C), (D, D)$. A memory-one strategy s can be written as a five-dimensional vector $s = (y, p_1, p_2, p_3, p_4)$. The parameter y is the probability that the strategy opens with a cooperation and p_1, p_2, p_3, p_4 are the probabilities that the strategy cooperates for each of the possible outcomes of the last round.

We perform four separate simulations where we differ the updating payoff and the benefit of cooperation b. The results for a low value of benefit are given in Figure 7, and for a high benefit in Figure 8.

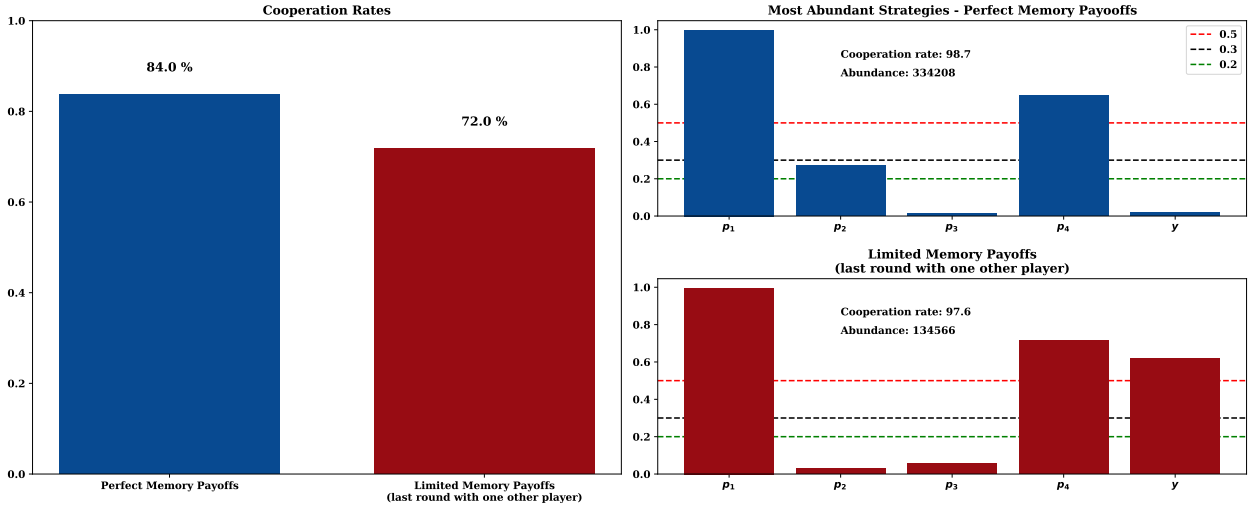

Figure 7: Evolutionary dynamics results for memory-one strategies for low benefit. We perform two independent simulations. In one simulation individuals use expected payoffs and in the other the last round one interacts when they update their strategies. We run each simulation for $T = 10^8$ time steps. For each time step, we have recorded the current resident population, who is now of the form (y, p_1, p_2, p_3, p_4) In the left panel we report the cooperation rates for each simulation. It can be shown than even for memory-one strategies expected payoffs result in a more cooperative population. The right panel reports the most abundant strategy of each simulation. Abundance is the number of mutants a strategy can repel before being invaded. The most abundant strategies have some similarities, namely, $p_1 \approx 1$, $p_3 \approx 0$ and $p_4 > \frac{1}{2}$. There are also differences, in the latter case a strategy is more likely to open with a cooperation and their tolerance to a (C, D) outcome is almost zero. A difference between the strategies is their abundance. In the expected payoffs case a strategy can repel a way greater number of mutants. In the case of last round payoffs strategies become less robust. Parameters: $N = 100, c = 1, b = 3, \beta = 1$.

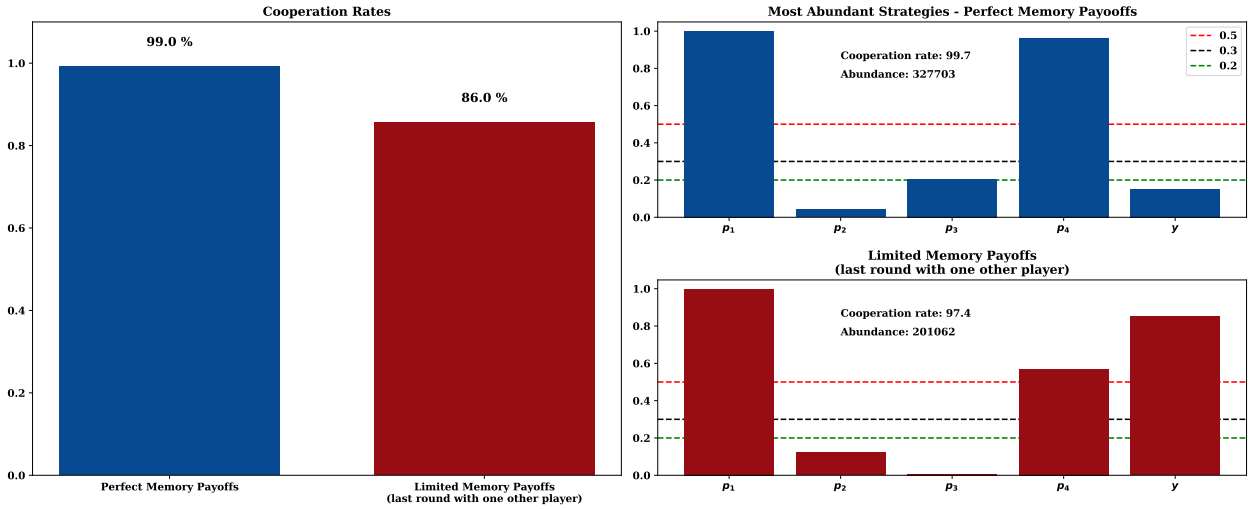

Figure 8: Evolutionary dynamics results for memory-one strategies for high benefit. We perform two independent simulations. In the case of high benefit expected payoffs again result in a more cooperative population. The right panel reports the most abundant strategy of each simulation. Abundance is the number of mutants a strategy can repel before being invaded. For the expected payoffs he most abundant is that of win-stay lose-shift. However in the later case the most abundant strategy is a strategy with no tolerance to one defection, and it cooperates with a probability .5 after a mutual defection. In the expected payoffs case strategies are more robust. Parameters: $N = 100, c = 1, b = 10, \beta = 1$.