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 the paper we explore the effect of the updating payoffs on the evolved populations. We used as an example the prisoner's dilemma. We focused mainly on two approaches for calculating the updating payoffs which we referred to as the perfect memory and the limited memory approaches. In Section 2 we describe the perfect memory approach and in Section 3 we present the limited memory approach. In the limited memory approach we assume that individuals update based on their last round payoff against one other member of the population. The framework can easily be extended to consider more rounds, more interactions or both. In Sections 4-6 we present each of this extensions and for each a special case. In Section 7 we present numerical results of the pairwise comparison process using the several updating payoff approaches we have presented. In Sections 8 and 9 we verify the main result of the paper when we no longer assume (i) that individuals use reactive strategies but instead memory one strategies (ii) low mutation.

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 infinitely 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** 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.

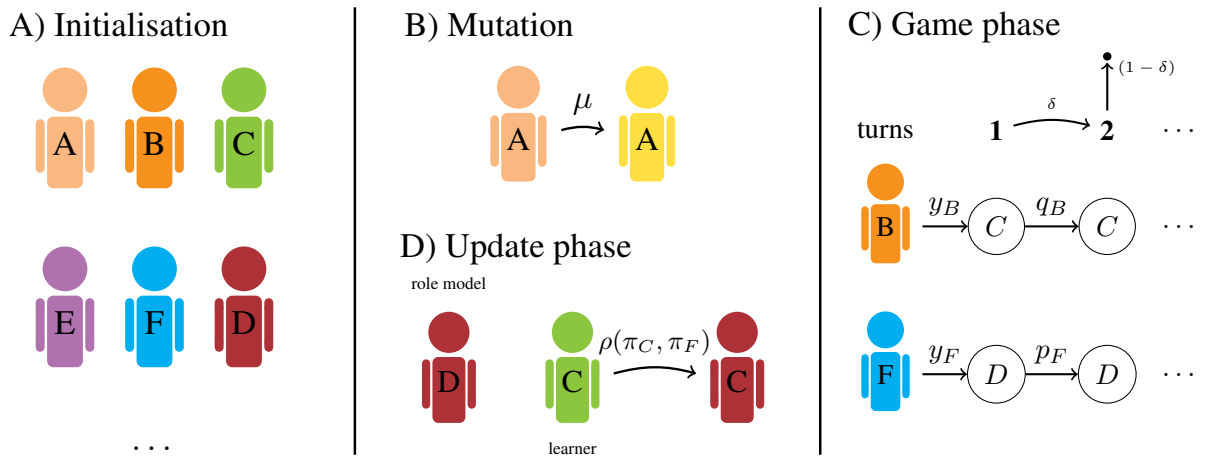

Figure 1: Pairwise comparison process phases. A) Initialisation. The process begins with a finite population where each member is assigned a given strategy. Each color represents a different strategy, and the members are labelled by letters. **B) Mutation phase.** An individual is selected (in the example 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,

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

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 strategy;
    fixation probability ← φ_M;
    if φ_M > 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 (Section 9).

1.2 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 person 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 sections 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 approaches.

2 Updating Payoffs based on the expected payoffs (perfect memory)

The expected payoffs are the conventional payoffs used in the updating stage. We refer to this approach as the perfect memory approach. The expected payoffs 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. The game between two reactive strategies $s_1 = (y_1, p_1, q_1)$ and $s_2 = (y_2, p_2, q_2)$ can be described by a Markovian process [1] with the transition matrix M,

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

The stationary vector $\mathbf{v}(s_1, s_2)$ is the solution to $\mathbf{v}(s_1, s_2)M = \mathbf{v}(s_1, s_2)$. In an infinitely repeated game the game stage outcome for each strategy is given by,

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

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,

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

$\frac{N-k-1}{N-1}$ is the probability that a resident meets another resident and $\frac{k}{N-1}$ is the probability that a resident meets a mutant. Likewise, $\frac{N-k}{N-1}$ is the probability that a mutant interacts with a resident and $\frac{k-1}{N-1}$ the probability that the mutants interacts with a mutant.

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,

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

Invasion Analysis

Let's assume that individuals use their expected payoffs at the updating phase. We can calculate how easily a single defecting mutant (ALLD) can invade into a resident population of conditional cooperators, otherwise known as generous tit for tat (GTFT), players. Let GTFT$= (1, 1, q)$, ALLD$= (0, 0, 0)$, and $k = 1$. When two GTFT players interact in an infinitely repeated game, the stationary distribution $\mathbf{v}(\text{GTFT}, \text{ALLD})$ simplifies 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.

This result will become important when we analyse the behaviour that emerges from the evolutionary process. We will repeat this "invasion analysis" for the rest of the updating payoffs cases.

3 Updating Payoffs based on the Last Round Payoff of One Interaction (limited memory)

In the expected payoffs case the payoff of a pair depends on the average payoff they received over an infinite number of turns. In the limited memory payoffs, the payoff of a pair depends on the average payoffs they received in the last turn. Furthermore, in expected payoffs it is assumed that a player interacts with every member of the population, whereas in the limited memory approach a player has one interaction. 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 Eq. (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. In this case both the role model and the learner estimate their fitness after interacting with a single member of the population, and so there are five possible pairings at each step. They interact with each other with a probability $\frac{1}{N-1}$, and 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 players 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.

Invasion Analysis

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

$$\begin{aligned} v_r(GTFT, GTFT) = 1, \quad v_t(GTFT, GTFT) = 0, \\ v_s(GTFT, GTFT) = 0, \quad v_p(GTFT, GTFT) = 0. \end{aligned}$$

On the other hand, if an ALLD player interacts with a GTFT player, the respective probabilities according to Eq. 6 become

$$\begin{aligned} v_r(ALLD, GTFT) &= 0, & v_s(ALLD, GTFT) &= 0, \\ v_t(ALLD, GTFT) &= 1 - \delta + \delta q, & v_p(ALLD, GTFT) &= \delta(1-q). \end{aligned}$$

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).

4 Updating Payoffs based on the last round payoff of n interactions

In the limited memory payoffs we assume that an individual recalls the last round they received against one member of the population. This is an edge case. The framework of the limited memory can be generalised such that an individual considers m rounds and of n interactions. Here we discuss the case that the update depends on the last round of n interactions.

At each step of the evolutionary process the role model and the learner now participate in n matches. We need to define the probability that for these matches they are paired with a mutant, with a resident or with each other. We assume that each pair is unique, so the resident and role model can be matched together only once at each step. A representation of the process is given in Figure.

In the case of $n = 1$ there are five possible pairs, however, the number of possible pairs increases non linearly as we increase the number of possible interactions. We demonstrate this case for $n = 2$, namely, for when the role model and the learner have two interactions.

Special case: Last Round Payoff of Two Interactions

For $n = 2$ there are two stages of matching and there are twenty four possible pairs. 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 together then in the second stage 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 pairs 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 pairs of the first stage there are five possible pairs in the second stage; 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 pairs 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 new probability denoted as $\tilde{x}(u_1, u_2)$ is given by,

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

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

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 $\frac{(N-k-2)(k-1)}{(N-4)(N-3)}$, and so on.

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

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

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

Invasion Analysis

In a similar fashion we can calculate the condition for which a population of GTFT players can not be invaded by an ALLD mutant. Using the new formulation we obtain the following probabilities $\bar{x}(u_1,u_2)$ that the payoff of a randomly chosen GTFT player is u_1 and that the payoff of the ALLD player is u_2,

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

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 the following cases,

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

We note that the relationship between the cost and benefit have an effect on how generous a conditional cooperator must be to avoid invasion. In the case of $\frac{b}{2} = c$ the result remains the same expression as in the case of $m = n = 1$. For the other two cases we show that for $\frac{\lambda^+}{\lambda^-} < 1$,

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

For $\frac{b}{2} > c$ the ratio is smaller for $q < \left\{ \frac{\delta-1-\frac{\sqrt{2}}{2}}{\delta}, \frac{\delta-1+\frac{\sqrt{2}}{2}}{\delta} \right\}$, however, $\frac{\delta-1-\frac{\sqrt{2}}{2}}{\delta}$ is not a feasible root since it's always smaller than 1, and thus $q < \frac{\delta-1+\frac{\sqrt{2}}{2}}{\delta}$. 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. For repeated games that are repeated for a large number of turn such as $\delta \to 1$ the condition then becomes $q < 1 - \frac{\sqrt{2}}{2}$.

5 Updating Payoffs based on the Last m Rounds Payoffs of One Interaction

The second generalised case of the limited memory payoffs we discuss in that of individuals updating based on their last m rounds payoffs with one member. To this end we define the probability that a reactive strategy receives the payoffs $u \in \tilde{\mathcal{U}}$, for $\tilde{\mathcal{U}} = \{\underbrace{rrr\ldots r}_{m}, \underbrace{rrr\ldots s}_{m}, \ldots, \underbrace{ppp\ldots p}_{m}\}$, in the last m rounds against another reactive strategy (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,*

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

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

The probability that the number of mutants increases, and decreases respectively remains the same as in

section **??** (Eq. 13). The frameworks differ in the game stage payoffs,

$$\lambda_k^+ = \frac{N-k}{N} \cdot \frac{k}{N} \cdot \sum_{\tilde{u}_R, \tilde{u}_M \in \tilde{\mathcal{U}}} \tilde{x}(\tilde{u}_R, \tilde{u}_M) \cdot \rho(\tilde{u}_R, \tilde{u}_M), \tag{20}$$

$$\lambda_k^- = \frac{N-k}{N} \cdot \frac{k}{N} \cdot \sum_{\tilde{u}_R, \tilde{u}_M \in \tilde{\mathcal{U}}} \tilde{x}(\tilde{u}_R, \tilde{u}_M) \cdot \rho(\tilde{u}_M, \tilde{u}_R). \tag{21}$$

Special case: Last Round Payoff of Two Interactions

Invasion Analysis

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

$$\frac{\lambda^+}{\lambda^-} = \frac{\frac{\delta^2(N-2)}{N-1}\left(\frac{\delta(1-q)^2}{1+e^{-\beta(-b+c)}}+\frac{q(\delta q-\delta+1)}{1+e^{-\beta c}}+\frac{(1-q)(2\delta q-\delta+1)}{1+e^{-\beta\left(-\frac{b}{2}+c\right)}}\right)+\frac{1}{N-1}\left(\frac{\delta(1-q)^2}{2}+\frac{q(\delta q-\delta+1)}{1+e^{-\beta(b+c)}}+\frac{(1-q)(2\delta q-\delta+1)}{1+e^{-\beta\left(\frac{b}{2}+\frac{c}{2}\right)}}\right)}{\frac{\delta^2(N-2)}{N-1}\left(\frac{\delta(1-q)^2}{1+e^{-\beta(b-c)}}+\frac{q(\delta q-\delta+1)}{e^{\beta c}+1}+\frac{(1-q)(2\delta q-\delta+1)}{1+e^{-\beta\left(\frac{b}{2}-c\right)}}\right)+\frac{1}{N-1}\left(\frac{\delta(1-q)^2}{2}+\frac{q(\delta q-\delta+1)}{1+e^{-\beta(-b-c)}}+\frac{(1-q)(2\delta q-\delta+1)}{1+e^{-\beta\left(-\frac{b}{2}-\frac{c}{2}\right)}}\right)} \tag{22}$$

In the limit of strong selection $\beta \to \infty$ and large populations $N \to \infty$ we obtain three expressions depending on the cost-benefit relationship. We note that for $\frac{b}{2} = c$ the result remains the same expression as in the case of $m = n = 1$.

$$\frac{\lambda^+}{\lambda^-} = \begin{cases} -\frac{q(\delta q-\delta+1)}{(q-1)(\delta q+1)} & \frac{b}{2} > c \\ -\frac{\delta q-\delta+q+1}{(\delta+1)(q-1)} & \frac{b}{2} = c \\ -\frac{\delta q^2-2\delta q+\delta-1}{\delta(q-1)^2} & \frac{b}{2} < c \end{cases} \tag{23}$$

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

$$\begin{cases} q \in \left\{\frac{\delta-\sqrt{\delta^2+1}-1}{2\delta},\ \frac{\delta+\sqrt{\delta^2+1}-1}{2\delta}\right\} & \frac{b}{2} > c \\ q \in \left\{1-\frac{\sqrt{2}}{2\sqrt{\delta}},\ 1+\frac{\sqrt{2}}{2\sqrt{\delta}}\right\} & \frac{b}{2} < c \end{cases} \tag{24}$$

For $\frac{b}{2} > c$ the ratio is smaller for $q < \left\{\frac{\delta-\sqrt{\delta^2+1}-1}{2\delta}, \frac{\delta+\sqrt{\delta^2+1}-1}{2\delta}\right\}$, however, the first root is not a feasible root since it's always smaller than 1, and thus $q < \frac{\delta+\sqrt{\delta^2+1}-1}{2\delta}$. 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. For repeated games that are

repeated for a large number of turn such as $\delta \to 1$ the condition then becomes $q < 1 - \frac{\sqrt{2}}{2}$.

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

Finally a possible extension to the limited memory framework is to consider that thee number of rounds and number of interactions increase. For this case we need to consider a combination of the methods we presented in Section 4 and Section 5.

7 Simulation Results on the Pairwise Comparison Process

We simulate the evolutionary process described in Algorithm 1 for the different updating mechanisms we have described in Sections 2-6. For each approach we performed an independent run of the process and for each time step we recorded the current resident population (y, p, q). The results are shown in Figure 2. We observe that in most cases the resident population consists either of defectors or conditional cooperators. A conditional cooperator always cooperates if the co-player cooperated ($p \approx 1$) and cooperates with a probability q if the co-player defected. The most abundant conditional cooperators in each simulation differ as a result of the updating payoffs. More specifically, in order for a resident population of conditional cooperators to avoid being invaded they need to adopt a different value of q. For each method we have discussed this under the invasion analysis subsection.

In the cases of perfect memory the resident population adopts a $q \leq 1 - \frac{c}{b} = 0.9$. In the limited memory case the generosity of a conditional cooperator is independently of the benefit lower than $\frac{1}{2}$. The rest of the cases also condition on the cost benefit relationship. In these simulations the cost of cooperation is set to 1 and the benefit to 10. As a result, in the case of two interactions $q \leq \frac{\delta - 1 + \frac{\sqrt{2}}{2}}{\delta} = 0.7068$, in the case of two rounds $q \leq \frac{\delta + \sqrt{\delta^2 + 1} - 1}{2\delta} = 0.7069$, and in the last case $q \leq 1 - \frac{c}{b} = 0.9$. The higher tolerance to defection results to a more cooperative population. As a result the expected payoffs allow for the most cooperative population. Between the limited memory approaches we observe a big difference in the cases of one or more information. We hypothesis that as we allow for more information the results will tend to the case of perfect memory.

8 Expected and Last Round Updating Payoffs for Memory One Strategies

So far we have assumed that individuals can adopt reactive strategies. To demonstrate that our results hold for higher memory strategies here we present results for the expected payoffs, and the last round payoff when members use 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

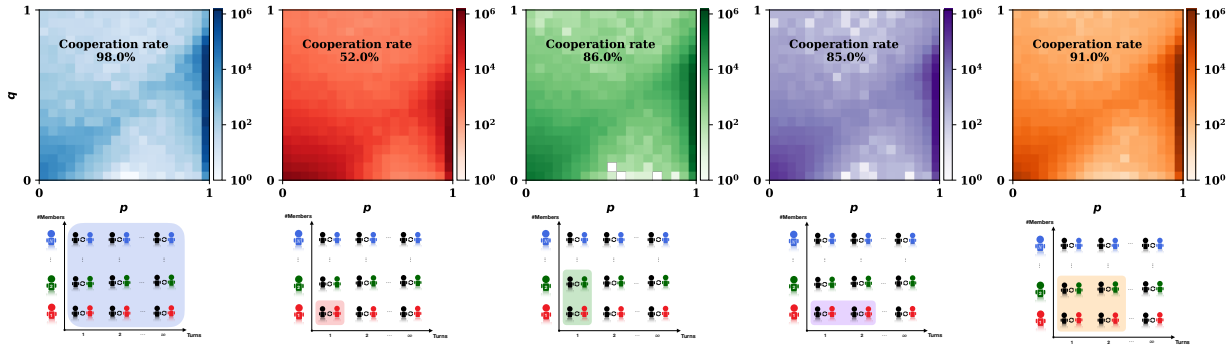

Figure 2: Evolutionary dynamics with difference updating approach. From left to right, we present result on the following updating payoffs cases; the expected payoffs (perfect memory), the last round payoff from one interaction (limited memory), the last round payoff from two interactions, the last two rounds payoffs from one interaction, the last two rounds payoffs from two interactions. We run each simulation for $T = 10^7$ time steps. For each time step we recorded the current resident population (y, p, q). Since $\delta \to 1$ 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.

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 3, and for a high benefit in Figure 4. We verify that even when individuals are allowed to use memory-one strategies, the cooperation rate is higher in the perfect memory approach compared to the limited memory.

9 Expected and Last Round Updating Payoffs for High Mutation ($\mu \neq 0$)

In this section we evaluate the main result of this work for $\mu \neq 0$. Namely, we explore the evolved population when individuals use perfect and limited updating payoff memory for different values of μ. We perform five independent runs of the pairwise process described in Section 1, and at each time step we record the average player $\bar{s} = (\bar{y}, \bar{p}, \bar{q})$. The average cooperation of the resident population for different values of mutation are shown in Figure 5. The cooperation rate in the case of perfect memory is always higher compared to the limited memory regardless of the mutation value. For mutation value of 1 the processes become random and this results to a cooperation rate of $\frac{1}{2}$ in both simulations.

References

[1] Martin Nowak and Karl Sigmund. Game-dynamical aspects of the prisoner's dilemma. *Applied Mathematics and Computation*, 30(3):191–213, 1989.

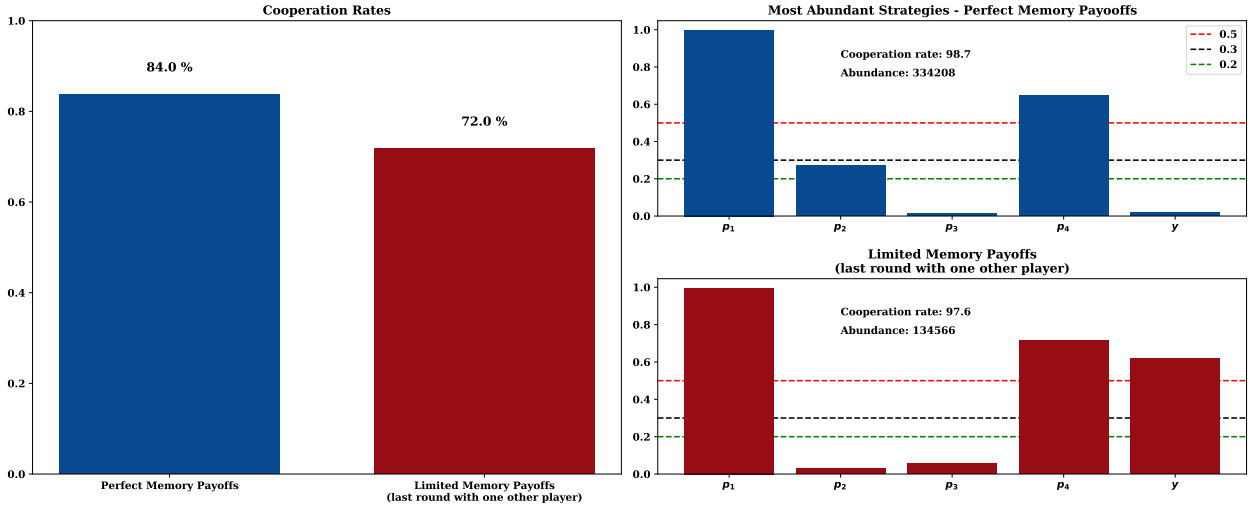

Figure 3: Evolutionary dynamics results for memory-one strategies for low benefit. We perform two independent simulations. In one simulation individuals use expected payoffs and in the other the last round one interacts when they update their strategies. We run each simulation for $T = 10^8$ time steps. 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 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: $N = 100, c = 1, b = 3, \beta = 1$.

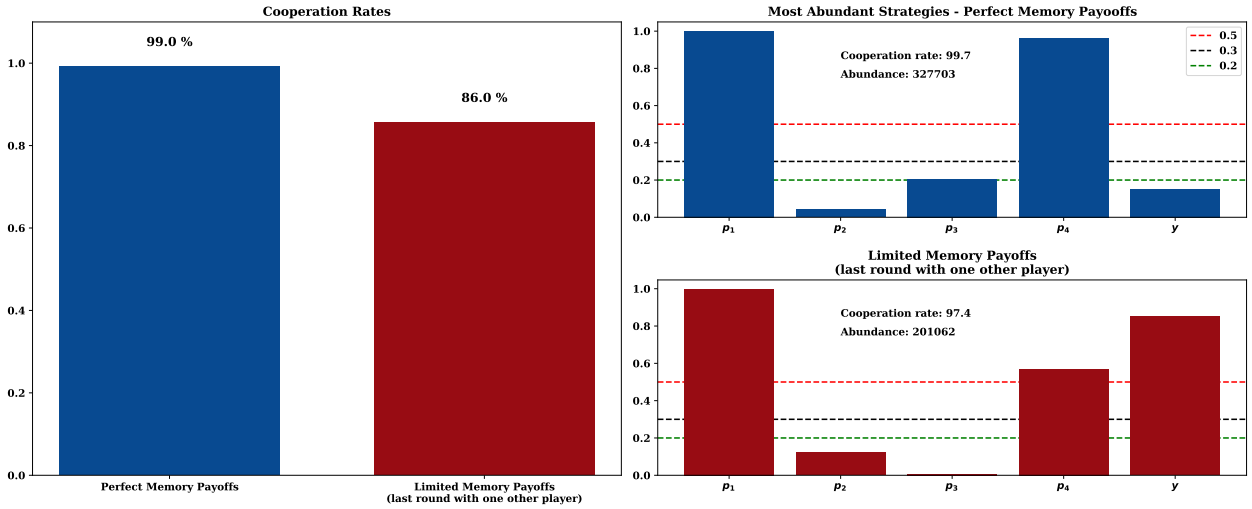

Figure 4: Evolutionary dynamics results for memory-one strategies for high benefit. We perform two independent simulations. In the case of high benefit expected payoffs again result in a more cooperative population. The right panel reports the most abundant strategy of each simulation. Abundance is the 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 0.5 after a mutual defection. In the expected payoffs case strategies are more robust. Parameters: $N = 100, c = 1, b = 10, \beta = 1$.

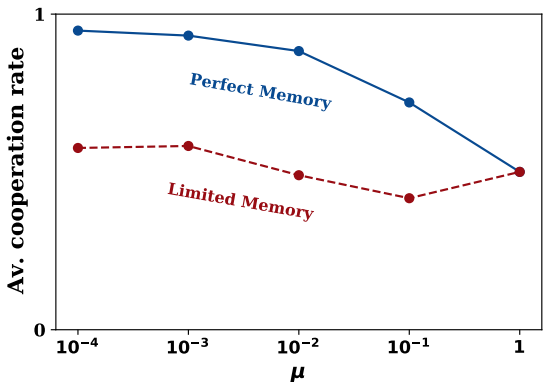

Figure 5: Evolutionary dynamics results for perfect and limited memory for different mutation values. We perform five independent simulations. Simulations are run for $T = 4 \times 10^7$ time steps for each parameter. In each time step we introduce a new mutant with a probability μ, and we then select two random players to serve as the role model and the learner. The learner adopts the strategy of the role model with a probability $\rho(\pi_L, \pi_{RM})$ where the updating payoffs depend on the method. In the case of perfect memory the expected payoffs are used and in the case of limited memory the last round payoff against one opponent. We plot the average cooperation rate within the resident population for each value of μ. For $\mu = 1$ the process becomes random and so the cooperation rates are 0.5. For the rest of the mutation values the perfect memory payoffs once again overestimate the evolved cooperation, confirming the results of low mutation. Parameters: $N = 100, c = 1, b = 10, \beta = 1$.