Electronic supplementary material

Evolution of reciprocity with limited payoff memory

Nikoleta E. Glynatsi, Alex McAvoy, Christian Hilbe

This document provides further details on our methods and derivations, and it contains additional simulation results. Section 1 summarizes the model. In particular, we provide further details on our implementation of the evolutionary dynamics, and our use of the rare-mutations limit. In Section 2 we derive analytical results for the various settings we consider. These settings differ in what kind of payoff information individuals take into account when updating their strategies. In the perfect memory setting, individuals take into account all their interactions against all co-players. In the limited memory setting, they only consider the very last round of their very last interaction. In addition, we describe several model extensions in which the amount of information taken into account is in between these two extremes. Finally, Section 3 presents further simulation results. In particular, we confirm that our main results continue to hold (*i*) when players use memory-1 strategies instead of reactive strategies, and (*ii*) when mutations are no longer rare.

1 Description of the model

Summary of the model. As described in the main text, we study cooperative behavior in a population of size N. The dynamics unfolds on two time-scales. The short time scale describes the game dynamics. Here individuals are randomly matched to interact in a repeated prisoner's dilemma. Each round individuals can choose whether to cooperate (C) or defect (D). In the most general setting, the resulting one-shot payoffs can be summarized by the payoff matrix

$$\begin{array}{c} \\ \text{cooperate} \\ \text{defect} \end{array} \begin{array}{c} \begin{array}{cc} \text{cooperate} & \text{defect} \end{array} \\ \begin{pmatrix} R & S \\ T & P \end{pmatrix} \end{array}. \tag{1}$$

Here, R is interpreted as the reward for mutual cooperation, S is the sucker's payoff, T is the temptation, and P is the punishment payoff [1]. Throughout this work, we parametrize these payoffs as $R = b - c$, $S = -c$, $T = b$, and $P = 0$, where b and c are the benefit and cost of cooperation, respectively, with $b > c > 0$. After each round, players learn their co-player's previous action. Then the game continues for another round with

probability δ. Players make their decisions whether to cooperate in any given round based on their reactive strategies $\mathbf{s} = (y, p, q)$. The entry y determines a player's first-round cooperation probability. The other entries p and q determine the player's cooperation probability in all subsequent rounds. The probability is p if the co-player cooperated in the previous round, and it is q if the co-player defected. On the short time scale, the players' strategies are fixed.

The long time scale describes the evolutionary dynamics. Here, players are allowed to update their strategies based on the payoffs they yield. We model these strategy updates with a pairwise comparison process [2]. This process assumes that at regular time intervals, one player is randomly selected from the population. We refer to this player as the 'learner' (L). The learner is then given an opportunity to update its strategy. There are two possibilities for how this update may occur. With probability μ, the player's strategy mutates randomly. In that case, the player's new strategy is drawn uniformly from the space of all reactive strategies $[0, 1]^3$. With probability $1-\mu$, the player engages in a pairwise comparison. In that case, the player randomly picks another individual from the population (referred to as the 'role model', R). The learner adopts the role model's strategy with a probability ρ given by

$$\rho(\pi_L, \pi_R) = \frac{1}{1+e^{-\beta(\pi_\mathrm{R}-\pi_\mathrm{L})}}. \tag{2}$$

The parameter β is the selection strength. It determines how important payoff differences are for the learner's decision to imitate the role model. The variables π_R and π_L refer to the relevant payoffs of the role model and the learner, respectively. The exact value of these payoffs depend on the players' memory. We say players have *perfect memory* when π_R and π_L are given by the players' expected payoffs (across all rounds and across all possible co-players). We say players have *limited memory* when π_R and π_L are given by the players' realized payoff in the very last round of the game with their very last interaction partner. In addition, we consider several model extensions in which individuals have memory capacities in between these two extremes. We provide a detailed description of these different settings and the resulting payoffs in Section 2.

Evolutionary simulations for the rare-mutations limit. To simulate the evolutionary dynamics of the pairwise comparison process, it is sometimes useful to assume that mutations are rare, $\mu \to 0$. In that case, whenever a mutant strategy appears, it either fixes in the population or goes extinct before the next mutant appears. As a result, at any given time there are at most two different strategies present in the population [3–5]. This assumptions makes computations more efficient, and it makes some of the results easier to interpret. In the following, we describe our implementation of the process in the rare-mutations limit in more detail.

Initially, the process starts with a population where all members use the same strategy (referred to as the resident strategy `R`). Then one individual adopts a mutant strategy selected uniformly at random from the set

Algorithm 1: Evolutionary process in the limit of rare mutations

$N \leftarrow$ population size;
resident $\leftarrow$ starting resident;
while $t <$ *maximum number of steps* **do**
  mutant $\leftarrow$ random strategy;
  fixation probability $\leftarrow \varphi_{\mathrm{M}}$;
  if $\varphi_M >$ *random:* $i \to [0,1]$ **then**
    resident $\leftarrow$ mutant;
  end
end

of feasible strategies. The fixation probability φ_{M} of the mutant strategy can be calculated explicitly [6],

$$\varphi_{\mathrm{M}} = \frac{1}{1 + \sum_{i=1}^{N-1} \prod_{k}^{i} \frac{\lambda_k^-}{\lambda_k^+}}. \tag{3}$$

Here, the index k corresponds to the current number of players with the mutant strategy (mutants). The variables λ_k^-, λ_k^+ are the probabilities that the number of mutants decreases or increases within a single updating step. These probabilities depend on the probability that a mutant and a resident are chosen as the learner and the resident, respectively. In addition, they depend on the respective switching probability ρ, as described by Eq. (2). We specify the exact values of λ_k^-, λ_k^+ for each memory-setting in the next section.

Depending on the fixation probability φ_{M}, the mutant strategy either fixes (becomes the new resident) or goes extinct. Afterwards, another random mutant strategy is introduced into the population. We iterate this elementary population updating process for a large number of mutant strategies. At each step, we record the current resident strategy, and the resulting average cooperation rate.

We consider this limit of rare-mutations throughout the main text. The respective process is summarised by Algorithm 1. In Section 3 we present additional simulation results to show that our qualitative results continue to hold when the mutation rate is strictly bounded away from zero.

2 Analytical results

In the following, we discuss our different memory settings in more detail. We discuss five cases explicitly. In these cases, updating occurs (*i*) based on average payoffs based on all interactions (perfect memory), (*ii*) based on the last round of one interaction (limited memory), (*iii*) based on the last round of two interactions, (*iv*) based on the last two rounds of one interaction, and (*v*) based on the last two rounds of two interactions. In each case, we consider the case that there are only two strategies present in the population (a resident and a mutant strategy). We first derive how likely it is that a learner (of any type) assigns a given payoff $\pi_{\rm L}$ to itself, and a payoff of $\pi_{\rm R}$ to the role model. This allows us to derive explicit expressions for λ_k^-/λ_k^+, and hence for the mutant's fixation probability according to Eq. (3). Based on these expressions we can characterize under which conditions cooperation is stochastically stable.

2.1 Perfect payoff memory

Computing the ratio λ_k^-/λ_k^+. The case of perfect payoff memory corresponds to the classical case considered in the previous literature. Here, individuals update their strategies based on the expected payoffs, based on all rounds and all possible interaction partners. When players use reactive strategies (or more generally, strategies with finite memory), these expected payoffs can be computed explicitly, based on a Markov chain approach [7]. To this end, consider two players with strategies $\sigma_1 = (y_1, p_1, q_1)$ and $\sigma_2 = (y_2, p_2, q_2)$, respectively. In each round t of the game, player 1 may get one of the four possible payoffs R, S, T, or P, as described by the general payoff matrix (1). Let $\mathbf{v}(t) = \big(v_R(t), v_S(t), v_T(t), v_P(t)\big)$ denote the respective probability distribution of observing one of these four outcomes. This probability distribution can be computed recursively. Using the shortcut notation $\bar{z}=1-z$ for any $z\in[0,1]$, we get for the initial round

$$\mathbf{v}_0 := \mathbf{v}(0) = (y_1y_2,\ y_1\bar{y}_2,\ \bar{y}_1y_2,\ \bar{y}_1\bar{y}_2). \tag{4}$$

Given $\mathbf{v}(t)$, we can compute $\mathbf{v}(t+1)$ as

$$\mathbf{v}(t+1) = \mathbf{v}(t)\cdot M, \tag{5}$$

where M is the transition matrix of the process,

$$M = \begin{bmatrix} p_1p_2 & p_1\left(1-p_2\right) & p_2\left(1-p_1\right) & \left(1-p_1\right)\left(1-p_2\right) \\ p_2q_1 & q_1\left(1-p_2\right) & p_2\left(1-q_1\right) & \left(1-p_2\right)\left(1-q_1\right) \\ p_1q_2 & p_1\left(1-q_2\right) & q_2\left(1-p_1\right) & \left(1-p_1\right)\left(1-q_2\right) \\ q_1q_2 & q_1\left(1-q_2\right) & q_2\left(1-q_1\right) & \left(1-q_1\right)\left(1-q_2\right) \end{bmatrix}. \tag{6}$$

Based on this recursion, we can compute how often player 1 receives one of the four payoffs R, S, T, P on average (across all possible realizations of games among the two players). This average distribution $\mathbf{v}$ is

$$\mathbf{v} := (1-\delta)\sum_{t=0}^{\infty}\delta^t\mathbf{v}(t) = (1-\delta)\mathbf{v_0}\sum_{t=0}^{\infty}\delta^t M^t = (1-\delta)\mathbf{v_0}(I_4-\delta M)^{-1}, \tag{7}$$

where I_4 is the 4×4 identity matrix. Based on this general formula, the four entries of $\mathbf{v}=(v_R,v_S,v_T,v_P)$ can be computed explicitly. Using the shortcut notation $r_i := p_i - q_i$, we obtain

$$\begin{aligned}
v_R &= (1-\delta)\frac{y_1y_2}{1-\delta^2r_1r_2} + \delta\frac{\Big(q_1+r_1\big((1-\delta)y_2+\delta q_2\big)\Big)\Big(q_2+r_2\big((1-\delta)y_1+\delta q_1\big)\Big)}{(1-\delta r_1r_2)(1-\delta^2r_1r_2)},\\
v_S &= (1-\delta)\frac{y_1\bar{y}_2}{1-\delta^2r_1r_2} + \delta\frac{\Big(q_1+r_1\big((1-\delta)y_2+\delta q_2\big)\Big)\Big(\bar{q}_2+\bar{r}_2\big((1-\delta)y_1+\delta p_1\big)\Big)}{(1-\delta r_1r_2)(1-\delta^2r_1r_2)},\\
v_T &= (1-\delta)\frac{\bar{y}_1y_2}{1-\delta^2r_1r_2} + \delta\frac{\Big(\bar{q}_1+\bar{r}_1\big((1-\delta)y_2+\delta p_2\big)\Big)\Big(q_2+r_2\big((1-\delta)y_1+\delta q_1\big)\Big)}{(1-\delta r_1r_2)(1-\delta^2r_1r_2)},\\
v_P &= (1-\delta)\frac{\bar{y}_1\bar{y}_2}{1-\delta^2r_1r_2} + \delta\frac{\Big(\bar{q}_1+\bar{r}_1\big((1-\delta)y_2+\delta p_2\big)\Big)\Big(\bar{q}_2+\bar{r}_2\big((1-\delta)y_1+\delta p_1\big)\Big)}{(1-\delta r_1r_2)(1-\delta^2r_1r_2)}.
\end{aligned} \tag{8}$$

Using this distribution $\mathbf{v}$, we compute player 1's expected payoff as the weighted average

$$\pi(\sigma_1,\sigma_2) = R\,v_R + S\,v_S + T\,v_T + P\,v_P. \tag{9}$$

After these preparations, consider now a population with k mutants and $N-k$ residents, whose strategies we denote by $\sigma_{\mathtt{M}} = (y_{\mathtt{M}},p_{\mathtt{M}},q_{\mathtt{M}})$ and $\sigma_{\mathtt{R}} = (y_{\mathtt{R}},p_{\mathtt{R}},q_{\mathtt{R}})$, respectively. Assuming that population members are matched randomly (or equivalently, that they interact with all other population members), the resulting expected payoffs of residents and mutants are

$$\begin{aligned}
\pi_{\mathtt{R}}(k) &= \frac{N-k-1}{N-1}\cdot\pi(\sigma_{\mathtt{R}},\sigma_{\mathtt{R}}) + \frac{k}{N-1}\cdot\pi(\sigma_{\mathtt{R}},\sigma_{\mathtt{M}}),\\
\pi_{\mathtt{M}}(k) &= \frac{N-k}{N-1}\cdot\pi(\sigma_{\mathtt{M}},\sigma_{\mathtt{R}}) + \frac{k-1}{N-1}\cdot\pi(\sigma_{\mathtt{M}},\sigma_{\mathtt{M}}).
\end{aligned} \tag{10}$$

The number of mutants in the population decreases in a single time step if a mutant is chosen to be the learner and adopts the strategy of a resident. Similarly, it increases if a resident is the learner and adopts the strategy

of a mutant. The respective transition probabilities are

$$\lambda_k^- = \frac{N-k}{N}\frac{k}{N-1}\rho\big(\pi_{\mathrm{M}}(k),\, \pi_{\mathrm{R}}(k)\big) \quad \text{and} \quad \lambda_k^+ = \frac{k}{N}\frac{N-k}{N-1}\rho\big(\pi_{\mathrm{R}}(k),\, \pi_{\mathrm{M}}(k)\big).$$

For ρ as defined by Eq. (2), the ratio of these two transition probabilities simplifies to

$$\frac{\lambda_k^-}{\lambda_k^+} = \frac{\rho\big(\,\pi_{\mathrm{M}}(k),\, \pi_{\mathrm{R}}(k)\,\big)}{\rho\big(\,\pi_{\mathrm{R}}(k),\, \pi_{\mathrm{M}}(k)\,\big)} = e^{-\beta\left(\pi_{\mathrm{M}}(k)-\pi_{\mathrm{R}}(k)\right)}. \tag{11}$$

Based on these ratios for each k, we can compute the mutant's fixation probability by Eq. (3).

Invasion Analysis. As an application of this formalism, we can compute when cooperation is stochastically stable in the perfect information setting. To this end, suppose there is only a single mutant, $k = 1$. The residents adopt Generous Tit-for-Tat, $\mathtt{GTFT} = (1, 1, q)$ and the mutant adopts $\mathtt{ALLD} = (0, 0, 0)$. When two GTFT players interact, the resulting average distribution according to Eq. (8) simplifies to

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

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

$$\mathbf{v}(\mathtt{ALLD}, \mathtt{GTFT}) = \big(0,\ 0,\ 1-\delta+\delta q,\ \delta(1-q)\big).$$

Based on Eq. (10), we can compute the strategies' expected payoffs as

$$\pi_{\mathtt{GTFT}} = \frac{N-2}{N-1}(b-c) - \frac{1}{N-1}\big(1-\delta+\delta q\big)c \quad \text{and} \quad \pi_{\mathtt{ALLD}} = \big(1-\delta+\delta q\big)b.$$

As a consequence, we can calculate the corresponding ratio of transition probabilities according to Eq. (11),

$$\frac{\lambda_1^-}{\lambda_1^+} = e^{-\beta\left((1-\delta+\delta q)(b+\frac{c}{N-1}) - \frac{N-2}{N-1}(b-c)\right)}$$

By definition, cooperation is stochastically stable if this ratio exceeds one, which is equivalent to

$$q < 1 - \frac{1}{\delta}\cdot\frac{b+(N-1)c}{(N-1)b+c}.$$

For such a strategy to be feasible we require $q > 0$, which implies $\delta > \big(b + (N-1)c\big)/\big((N-1)b+c\big)$. In particular, in the limit of large populations $N \to \infty$, we obtain that cooperation is stochastically stable if $q < 1 - c/(\delta b)$. The minimum continuation probability for such a strategy to exist is $\delta > c/b$. In this way, we recover the classical conditions for cooperation to be feasible under direct reciprocity [8–10].

2.2 Limited payoff 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 payoff memory, the payoff of a pair depends on the average payoffs they received in the last turn. Moreover, in expected payoffs it is assumed that a player interacts with every member of the population whereas in the limited payoff 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. (12).*

$$
\begin{aligned}
v_r(s_1,s_2) &= (1-\delta)\frac{y_1 y_2}{1-\delta^2 r_1 r_2} + \delta\frac{\Big(q_1 + r_1\big((1-\delta)y_2 + \delta q_2\big)\Big)\Big(q_2 + r_2\big((1-\delta)y_1 + \delta q_1\big)\Big)}{(1-\delta r_1 r_2)(1-\delta^2 r_1 r_2)},\\
v_s(s_1,s_2) &= (1-\delta)\frac{y_1 \bar{y}_2}{1-\delta^2 r_1 r_2} + \delta\frac{\Big(q_1 + r_1\big((1-\delta)y_2 + \delta q_2\big)\Big)\Big(\bar{q}_2 + \bar{r}_2\big((1-\delta)y_1 + \delta p_1\big)\Big)}{(1-\delta r_1 r_2)(1-\delta^2 r_1 r_2)},\\
v_t(s_1,s_2) &= (1-\delta)\frac{\bar{y}_1 y_2}{1-\delta^2 r_1 r_2} + \delta\frac{\Big(\bar{q}_1 + \bar{r}_1\big((1-\delta)y_2 + \delta p_2\big)\Big)\Big(q_2 + r_2\big((1-\delta)y_1 + \delta q_1\big)\Big)}{(1-\delta r_1 r_2)(1-\delta^2 r_1 r_2)},\\
v_p(s_1,s_2) &= (1-\delta)\frac{\bar{y}_1 \bar{y}_2}{1-\delta^2 r_1 r_2} + \delta\frac{\Big(\bar{q}_1 + \bar{r}_1\big((1-\delta)y_2 + \delta p_2\big)\Big)\Big(\bar{q}_2 + \bar{r}_2\big((1-\delta)y_1 + \delta p_1\big)\Big)}{(1-\delta r_1 r_2)(1-\delta^2 r_1 r_2)}.
\end{aligned}
\tag{12}
$$

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,

$$
(1-\delta)\mathbf{v_0}\sum \delta^t M^{(t)}, \tag{13}
$$

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,

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

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 12. □

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,

$$\begin{aligned}x(u_1,u_2) = \quad & \frac{1}{N-1}\cdot v_{u_1}(s_1,s_2)\cdot 1_{(u_1,u_2)\in\mathcal{U}_F}\\ &+\left(1-\frac{1}{N-1}\right)\Big[\tfrac{k-1}{N-2}\tfrac{k-2}{N-3}v_{u_1}(s_1,s_2)v_{u_2}(s_2,s_2)+\tfrac{k-1}{N-2}\tfrac{N-k-1}{N-3}v_{u_1}(s_1,s_2)v_{u_2}(s_2,s_1)\\ &\qquad +\tfrac{N-k-1}{N-2}\tfrac{k-1}{N-3}v_{u_1}(s_1,s_1)v_{u_2}(s_2,s_2)+\tfrac{N-k-1}{N-2}\tfrac{N-k-2}{N-3}v_{u_1}(s_1,s_1)v_{u_2}(s_2,s_1)\Big].\end{aligned} \tag{14}$$

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, $(u_1,\ u_2) \in \mathcal{U}_F := \big\{(r,\ r),\ (s,\ t),\ (t,\ s),\ (p,\ p)\big\}$, 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,

$$\lambda_k^+ = \frac{N-k}{N} \cdot \frac{k}{N} \cdot \sum_{u_R, u_M \in \mathcal{U}} x(u_R, u_M) \cdot \rho(u_R, u_M), \tag{15}$$

$$\lambda_k^- = \frac{N-k}{N} \cdot \frac{k}{N} \cdot \sum_{u_R, u_M \in \mathcal{U}} x(u_R, u_M) \cdot \rho(u_M, u_R). \tag{16}$$

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,

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

On the other hand, if an ALLD player interacts with a GTFT player, the respective probabilities according to Eq. 12 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,

$$
\begin{aligned}
x(r,t) =& \frac{N-2}{N-1}\cdot(1-\delta+\delta q)\\
x(r,r) =& \frac{N-2}{N-1}\cdot\delta(1-q)\\
x(s,r) =& \frac{1}{N-1}\cdot(1-\delta+\delta q)\\
x(p,p) =& \frac{1}{N-1}\cdot\delta(1-q)\\
x(u_1,u_2) =& 0 \text{ for all other payoff pairs } (u_1,u_2).
\end{aligned}
$$

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).

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

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 each of the matches they are paired with a mutant, with a resident or with each other. We assume that each pair is unique, so for example the resident and role model can be matched together only once at each step. A representation of the process is given in Figure 1. 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.

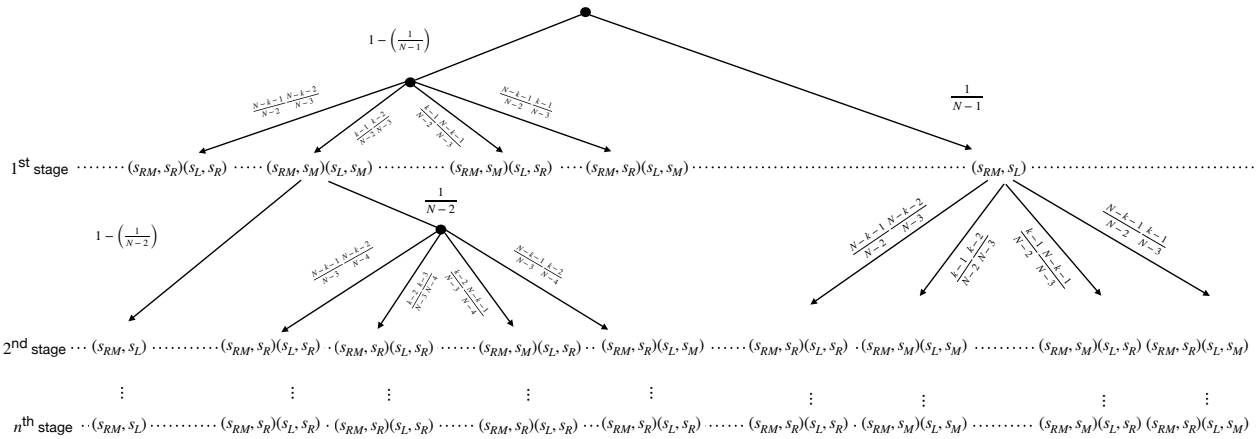

Figure 1: Tree diagram of the possible pairs when the learner and the role model have n interactions. In the diagram (s_i, s_j) represent a possible pair. We denote the role model as s_{RM} and the learner as s_L, and a member of the population that plays a mutant strategy is denoted as s_M and a resident strategy as s_R. The role model and the learner need to be paired with n other members of the population (including each other). We break down the process into stages. For the first pair (stage one) the role model and the mutant can be paired together (with probability $\frac{1}{N-1}$) or not (with probability $1-\left(\frac{1}{N-1}\right)$). In the case that they are not paired together, both can be paired with a mutant $\frac{(k-1)(k-2)}{(N-2)(N-3)}$, with a resident $\frac{(N-k-1)(N-k-2)}{(N-2)(N-3)}$, or one is paired with a mutant whilst the other is paired with a resident $\frac{(N-k-1)(k-1)}{(N-2)(N-3)}$. There are five possible pairs for stage one as we have already discussed in Section 2.2. In the second stage we need to consider if the role model and the learner interacted with each other. If yes, then they can not interact with each other again, and at each stage there are at most four possible pairs; they interact with mutants, with residents, or the one interacts with a mutant whilst the second interacts with a resident. In the case that they were not paired in stage 1 there are five possible pairs. These pairs are the same as in stage 1, however now the probabilities differ. Namely, let's assume that in the first stage both were matched with a mutant. They interact with each other with a probability $\frac{1}{N-2}$ or not with a probability $1-\left(\frac{1}{N-2}\right)$. In the latter case there are four possible pairs. Both can be paired with a mutant $\frac{(k-1)(k-2)}{(N-3)(N-4)}$, with a resident $\frac{(N-k-1)(N-k-2)}{(N-3)(N-4)}$, or one is paired with a mutant whilst the other is paired with a resident $\frac{(N-k-1)(k-2)}{(N-3)(N-4)}$. The process continues following the same logic until it reaches the n^{th} stage.

2.3.1 Last Round Payoff of Two Interactions

Here the learner and role model consider the last round payoff of two interactions. At each time step of the evolutionary process there are twenty four possible pairs. At the first stage there are fiver possible pairs. For four of these pairs there are five possible pairs in the second stage ($4 \times 5 = 20$), and for the last one there are four possible pairs in the second stage (20 + 4 = 24).

We assume that player s_1 receives the payoff u_{11} with their first interaction and u_{12} with their second. Respectively, player s_2 receives u_{21} and u_{22}. We define the probability that the respective last round payoffs of the two players s_1, s_2 are given by (u_{11}, u_{11}) and (u_{21}, u_{22}) as $x^2((u_{11}, u_{11}), (u_{21}, u_{22}))$ given by,

$$
\begin{aligned}
x^2((u_{11},u_{12}),(u_{21},u_{22})) = \frac{1}{N-1} \cdot v_{u_{11}}(s_1,s_2) \cdot 1_{(u_{11},u_{21}) \in \mathcal{U}_F} \cdot A + \left(1 - \frac{1}{N-1}\right) [\\
v_{u_{11}}(s_1,s_2)_{u_{21}}(s_2,s_2) \tfrac{(k-2)(k-1)}{(N-2)(N-3)} \left(\tfrac{1}{N-2} \cdot v_{u_{11}}(s_1,s_2) \cdot 1_{(u_{11},u_{21}) \in \mathcal{U}_F} + (1 - \tfrac{1}{N-2})[B_1 + B_2 + B_3 + B_4] \right) + \\
v_{u_{11}}(s_1,s_2)_{u_{21}}(s_2,s_1) \tfrac{(k-1)(N-k-1)}{(N-2)(N-3)} \left(\tfrac{1}{N-2} \cdot v_{u_{11}}(s_1,s_2) \cdot 1_{(u_{11},u_{21}) \in \mathcal{U}_F} + (1 - \tfrac{1}{N-2})[C_1 + C_2 + C_3 + C_4] \right) + \\
v_{u_{11}}(s_1,s_1)_{u_{21}}(s_2,s_2) \tfrac{(k-1)(N-k-1)}{(N-2)(N-3)} \left(\tfrac{1}{N-2} \cdot v_{u_{11}}(s_1,s_2) \cdot 1_{(u_{11},u_{21}) \in \mathcal{U}_F} + (1 - \tfrac{1}{N-2})[D_1 + D_2 + D_3 + D_4] \right) + \\
v_{u_{11}}(s_1,s_1)_{u_{21}}(s_2,s_1) \tfrac{(N-k-2)(N-k-1)}{(N-2)(N-3)} \left(\tfrac{1}{N-2} \cdot v_{u_{11}}(s_1,s_2) \cdot 1_{(u_{11},u_{21}) \in \mathcal{U}_F} + (1 - \tfrac{1}{N-2})[E_1 + E_2 + E_3 + E_4] \right)]
\end{aligned}
\tag{17}
$$

$$
\begin{aligned}
A \;&= \left(1-\tfrac{1}{N-1}\right)\Big[\tfrac{k-1}{N-2}\tfrac{k-2}{N-3}v_{u_{21}}(s_1,s_2)v_{u_{22}}(s_2,s_2)+\tfrac{k-1}{N-2}\tfrac{N-k-1}{N-3}v_{u_{21}}(s_1,s_2)v_{u_{22}}(s_2,s_1)+\\
&\quad \tfrac{N-k-1}{N-2}\tfrac{k-1}{N-3}v_{u_{21}}(s_1,s_1)v_{u_{22}}(s_2,s_2)+\tfrac{N-k-1}{N-2}\tfrac{N-k-2}{N-3}v_{u_{21}}(s_1,s_1)v_{u_{22}}(s_2,s_1)\Big]\\
B_1 &= \tfrac{(k-3)(k-2)}{(N-3)(N-4)}v_{u_{21}}(s_1,s_2)v_{u_{22}}(s_2,s_2) \quad B_2 = \tfrac{(k-2)(N-k-1)}{(N-3)(N-4)}v_{u_{21}}(s_1,s_1)v_{u_{22}}(s_2,s_2)\\
B_3 &= \tfrac{(k-2)(N-k-1)}{(N-3)(N-4)}v_{u_{21}}(s_1,s_2)v_{u_{22}}(s_2,s_1) \quad B_4 = \tfrac{(N-k-2)(N-k-1)}{(N-3)(N-4)}v_{u_{21}}(s_1,s_1)v_{u_{22}}(s_2,s_1)\\
C_1 &= \tfrac{(k-3)(k-1)}{(N-3)(N-4)}v_{u_{21}}(s_1,s_2)v_{u_{22}}(s_2,s_2) \quad C_2 = \tfrac{(k-1)(N-k-1)}{(N-3)(N-4)}v_{u_{21}}(s_1,s_1)v_{u_{22}}(s_2,s_2)\\
C_3 &= \tfrac{(k-2)(N-k-2)}{(N-3)(N-4)}v_{u_{21}}(s_1,s_2)v_{u_{22}}(s_2,s_1) \quad C_4 = \tfrac{(N-k-2)^2}{(N-3)(N-4)}v_{u_{21}}(s_1,s_1)v_{u_{22}}(s_2,s_1)\\
D_1 &= \tfrac{(k-2)^2}{(N-3)(N-4)}v_{u_{21}}(s_1,s_2)v_{u_{22}}(s_2,s_2) \quad D_2 = \tfrac{(k-2)(N-k-2)}{(N-3)(N-4)}v_{u_{21}}(s_1,s_1)v_{u_{22}}(s_2,s_2)\\
D_3 &= \tfrac{(k-1)(N-k-1)}{(N-3)(N-4)}v_{u_{21}}(s_1,s_2)v_{u_{22}}(s_2,s_1) \quad D_4 = \tfrac{(N-k-3)(N-k-1)}{(N-3)(N-4)}v_{u_{21}}(s_1,s_1)v_{u_{22}}(s_2,s_1)\\
E_1 &= \tfrac{(k-2)(k-1)}{v}{}_{u_{21}}(s_1,s_2)v_{u_{22}}(s_2,s_2) \quad E_2 = \tfrac{(k-1)(N-k-2)}{(N-3)(N-4)}v_{u_{21}}(s_1,s_1)v_{u_{22}}(s_2,s_2)\\
E_3 &= \tfrac{(k-1)(N-k-2)}{(N-3)(N-4)}v_{u_{21}}(s_1,s_2)v_{u_{22}}(s_2,s_1) \quad E_4 = \tfrac{(N-k-3)(N-k-2)}{(N-3)(N-4)}v_{u_{21}}(s_1,s_1)v_{u_{22}}(s_2,s_1)
\end{aligned}
\tag{18}
$$

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, 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 pairs are that either of them interacts with a resident whilst the other interacts with a mutant, and this happens with a probability $\frac{(N-k-2)(k-1)}{(N-4)(N-3)}$, and so on.

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

$$\lambda_k^+ = \frac{N-k}{N} \cdot \frac{k}{N} \cdot \sum_{u_{R1},u_{M1} \in \mathcal{U}} \sum_{u_{R2},u_{M2} \in \mathcal{U}} x^2((u_{R1},u_{R2}),(u_{M1},u_{M2})) \cdot \rho\left(\frac{u_{R1}+u_{R2}}{2}, \frac{u_{M1}+u_{M2}}{2}\right), \tag{19}$$

$$\lambda_k^- = \frac{N-k}{N} \cdot \frac{k}{N} \cdot \sum_{u_{R1},u_{M1} \in \mathcal{U}} \sum_{u_{R2},u_{M2} \in \mathcal{U}} x^2((u_{R1},u_{R2}),(u_{M1},u_{M2})) \cdot \rho\left(\frac{u_{M1}+u_{M2}}{2}, \frac{u_{R1}+u_{R2}}{2}\right). \tag{20}$$

We use $x^2((u_{R1},u_{R2}),(u_{M1},u_{M2}))$ to denote the probability that the randomly chosen resident obtained a payoff of u_{R1} in the last round of their first respective game and u_{R1} in the last round of their second game. Similarly, that the mutant obtained a payoff of u_{M1} from their first interaction and u_{M2} from the second. Note that we assume that the players compare their average payoff of their two interaction when considering adopting the others strategy. For example the mutant adopts the resident's strategy with a probability $\rho\left(\frac{u_{R1}+u_{R2}}{2}, \frac{u_{M1}+u_{M2}}{2}\right)$.

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 $x^2((u_{11},u_{12}),(u_{21},u_{22}))$ that the payoff of a randomly chosen GTFT player is (u_{11},u_{12}) and that the payoff of the ALLD player is (u_{21},u_{22}),

$$\begin{aligned}
x^2((r,r),(t,t)) &= \frac{N-3}{N-1}(\delta q-\delta+1)^2 & x^2((r,r),(t,p)) &= -\frac{N-3}{N-1}\delta(q-1)(\delta q-\delta+1)\\
x^2((r,r),(p,t)) &= -\frac{N-3}{N-1}\delta(q-1)(\delta q-\delta+1) & x^2((r,r),(p,p)) &= \frac{N-3}{N-1}\delta^2(q-1)^2\\
x^2((r,s),(t,t)) &= \frac{1}{N-1}(\delta q-\delta+1)^2 & x^2((r,s),(p,t)) &= -\frac{1}{N-1}\delta(q-1)(\delta q-\delta+1)\\
x^2((r,p),(t,p)) &= -\frac{1}{N-1}\delta(q-1)(\delta q-\delta+1) & x^2((r,p),(p,p)) &= \frac{1}{N-1}\delta^2(q-1)^2\\
x^2((s,r),(t,t)) &= \frac{1}{N-1}(\delta q-\delta+1)^2 & x^2((s,r),(t,p)) &= -\frac{1}{N-1}\delta(q-1)(\delta q-\delta+1)\\
x^2((p,r),(p,t)) &= -\frac{1}{N-1}\delta(q-1)(\delta q-\delta+1) & x^2((p,r),(p,p)) &= \frac{1}{N-1}\delta^2(q-1)^2
\end{aligned}$$

$x^2((u_{11},u_{12}),(u_{21},u_{22})) = 0$ for all other payoff pairs $((u_{11},u_{12}),(u_{21},u_{22}))$.

The ratio of transition probabilities is given by,

$$\frac{\lambda^+}{\lambda^-}=\frac{\frac{N-3}{N-1}\left(\frac{\delta^2(1-q)^2}{1+e^{-\beta(-b+c)}}+\frac{2\delta(1-q)(\delta q-\delta+1)}{1+e^{-\beta\left(-\frac{b}{2}+c\right)}}+\frac{(\delta q-\delta+1)^2}{1+e^{-\beta c}}\right)+\frac{2}{N-1}\left(\frac{\delta^2(1-q)^2}{1+e^{-\beta\left(-\frac{b}{2}+\frac{c}{2}\right)}}+\frac{\delta(1-q)(\delta q-\delta+1)}{1+e^{-\frac{\beta c}{2}}}+\frac{\delta(1-q)(\delta q-\delta+1)}{1+e^{-\beta c}}+\frac{(\delta q-\delta+1)^2}{1+e^{-\beta\left(\frac{b}{2}+c\right)}}\right)}{\frac{N-3}{N-1}\left(\frac{\delta^2(q-1)^2}{1+e^{-\beta(b-c)}}+\frac{2\delta(1-q)(\delta q-\delta+1)}{1+e^{-\beta\left(\frac{b}{2}-c\right)}}+\frac{(\delta q-\delta+1)^2}{e^{\beta c}+1}\right)+\frac{2}{N-1}\left(\frac{\delta^2(q-1)^2}{1+e^{-\beta\left(\frac{b}{2}-\frac{c}{2}\right)}}+\frac{\delta(1-q)(\delta q-\delta+1)}{e^{\beta c}+1}+\frac{\delta(1-q)(\delta q-\delta+1)}{e^{\frac{\beta c}{2}}+1}+\frac{(\delta q-\delta+1)^2}{1+e^{-\beta\left(-\frac{b}{2}-c\right)}}\right)} \tag{21}$$

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

$$\frac{\lambda^+}{\lambda^-}=\begin{cases}-\frac{q(\delta q-\delta+1)}{(q-1)(\delta q+1)} & \frac{b}{2}>c\\ \frac{-\delta q+\delta-1}{\delta(q-1)} & \frac{b}{2}=c\\ -\frac{\delta q^2-2\delta q+\delta-1}{\delta(q-1)^2} & \frac{b}{2}<c\end{cases} \tag{22}$$

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 $\frac{\lambda^+}{\lambda^-}<1$,

$$\begin{cases}q<\left\{\frac{\delta-1-\frac{\sqrt{2}}{2}}{\delta},\frac{\delta-1+\frac{\sqrt{2}}{2}}{\delta}\right\} & \frac{b}{2}>c\\ q<\left\{\frac{\delta-\frac{\sqrt{2}}{2}}{\delta},\frac{\delta+\frac{\sqrt{2}}{2}}{\delta}\right\} & \frac{b}{2}<c\end{cases} \tag{23}$$

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 $q<1-\frac{\sqrt{2}}{2}$.

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

The second generalised case of the limited memory payoffs we discuss is that of individuals updating based on their last m rounds payoffs with one member. Let $\mathcal{U}^m=\{\underbrace{rrr\dots r}_{m},\underbrace{rrr\dots s}_{m},\dots,\underbrace{ppp\dots p}_{m}\}$ be the set of feasible payoffs of the last m rounds. The probability that a reactive strategy receives the payoffs $u\in\mathcal{U}^m$ 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 $\mathcal{U}^m$ denote the set of feasible payoffs in the last m rounds, and let $\mathbf{u}^m$ be the corresponding payoff vector. Then the probability that the s_1 player receives the payoff $u\in\mathcal{U}^m$ in the very last two rounds of the game is given by,*

$$\langle \mathbf{v}^m(s_1, s_2), \mathbf{u}^m \rangle, \text{ where } \mathbf{v}^m \in R^{4^m} \text{ is given by }, \tag{24}$$

$$\mathbf{v}^m(s_1, s_2) = (1-\delta) w_{a_1,a_2} \delta^2 \left[\mathbf{v_0} (I_4 - \delta M)^{-1} \right]_{a_1,a_2}, \quad w_{a_1,a_2} \in M \; \forall \; a_1, a_2 \in \{1,2,3,4\}. \tag{25}$$

2.4.1 Last Round Payoff of Two Interactions

In the special case of $m = 2$ the stationary distribution $\mathbf{v}^2(s_1, s_2)$ is sixteen dimension instead of four dimensional where $v_1(s_1, s_2)$ is the long term probability that s_1 and s_2 mutually cooperated in the last two rounds. Let the feasible payoffs a strategy can receive be $\mathcal{U}^2 = \{rr, rs, rt, rp, sr, \ldots, pp\}$. The role model and the learner interact with only one other member and so the probability x remains the same as in Section 2.2.

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

$$\lambda_k^+ = \frac{N-k}{N} \cdot \frac{k}{N} \cdot \sum_{u_R, u_M \in \mathcal{U}^2} x(u_R, u_M) \cdot \rho\left(\frac{u_R}{2}, \frac{u_M}{2}\right), \tag{26}$$

$$\lambda_k^- = \frac{N-k}{N} \cdot \frac{k}{N} \cdot \sum_{u_R, u_M \in \mathcal{U}^2} x(u_R, u_M) \cdot \rho\left(\frac{u_M}{2}, \frac{u_R}{2}\right). \tag{27}$$

We assume again that players consider their average payoff of the last two turns when deciding to imitate another strategy or not.

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,

$$\begin{aligned} v_{rr}(GTFT, GTFT) =& \delta^2, \text{ and} \\ v_i(GTFT, GTFT) =& 0 \text{ for all other } i \in \mathcal{U}^2. \end{aligned}$$

On the other hand, if an ALLD player interacts with a GTFT player, the respective probabilities according to Eq. 12 become

$$v_{tt}(ALLD, GTFT) = \delta^2 q\left(\delta q - \delta + 1\right), \qquad v_{tp}(ALLD, GTFT) = -\delta^2\left(q-1\right)\left(\delta q - \delta + 1\right),$$
$$v_{pt}(ALLD, GTFT) = \delta^3 q\left(1-q\right), \qquad v_{pp}(ALLD, GTFT) = \delta^3\left(q-1\right)^2 \text{ and}$$
$$v_i(ALLD, GTFT) = 0 \text{ for all other } i \in \mathcal{U}^2.$$

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,

$$x(rr,tt) = \frac{N-2}{N-1}\delta^4 q\left(N-2\right)\left(\delta q - \delta + 1\right) \qquad x(rr,tp) = -\frac{N-2}{N-1}\delta^4\left(N-2\right)\left(q-1\right)\left(\delta q - \delta + 1\right)$$
$$x(rr,pt) = -\frac{N-2}{N-1}\delta^5 q\left(N-2\right)\left(q-1\right) \qquad x(rr,pp) = \frac{N-2}{N-1}\delta^5\left(N-2\right)\left(q-1\right)^2$$
$$x(ss,tt) = \frac{1}{N-1}\delta^2 q\left(\delta q - \delta + 1\right) \qquad x(sp,tp) = -\frac{1}{N-1}\delta^2\left(q-1\right)\left(\delta q - \delta + 1\right)$$
$$x(ps,pt) = -\frac{1}{N-1}\delta^3 q\left(q-1\right) \qquad x(pp,pp) = \frac{1}{N-1}\delta^3\left(q-1\right)^2$$
$$x(u_1,u_2) = 0 \text{ for all other payoff pairs } (u_1,u_2).$$

We now calculate the ratio of transition probabilities as

$$\frac{\lambda^+}{\lambda^-} = \frac{\frac{\delta^2(N-2)}{N-1}\left(\frac{\delta(1-q)^2}{1+e^{-\beta(-b+c)}} + \frac{q(\delta q-\delta+1)}{1+e^{-\beta c}} + \frac{(1-q)(2\delta q-\delta+1)}{1+e^{-\beta\left(-\frac{b}{2}+c\right)}}\right) + \frac{1}{N-1}\left(\frac{\delta(1-q)^2}{2} + \frac{q(\delta q-\delta+1)}{1+e^{-\beta(b+c)}} + \frac{(1-q)(2\delta q-\delta+1)}{1+e^{-\beta\left(\frac{b}{2}+\frac{c}{2}\right)}}\right)}{\frac{\delta^2(N-2)}{N-1}\left(\frac{\delta(1-q)^2}{1+e^{-\beta(b-c)}} + \frac{q(\delta q-\delta+1)}{e^{\beta c}+1} + \frac{(1-q)(2\delta q-\delta+1)}{1+e^{-\beta\left(\frac{b}{2}-c\right)}}\right) + \frac{1}{N-1}\left(\frac{\delta(1-q)^2}{2} + \frac{q(\delta q-\delta+1)}{1+e^{-\beta(-b-c)}} + \frac{(1-q)(2\delta q-\delta+1)}{1+e^{-\beta\left(-\frac{b}{2}-\frac{c}{2}\right)}}\right)} \tag{28}$$

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 $m = n = 1$.

$$\frac{\lambda^+}{\lambda^-} = \begin{cases} -\frac{q(\delta q-\delta+1)}{(q-1)(\delta q+1)} & \frac{b}{2} > c \\ -\frac{\delta q-\delta+q+1}{(\delta+1)(q-1)} & \frac{b}{2} = c \\ -\frac{\delta q^2-2\delta q+\delta-1}{\delta(q-1)^2} & \frac{b}{2} < c \end{cases} \tag{29}$$

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

$$\begin{cases} q \in \left\{ \frac{\delta - \sqrt{\delta^2+1}-1}{2\delta},\ \frac{\delta+\sqrt{\delta^2+1}-1}{2\delta} \right\} & \frac{b}{2} > c \\ q \in \left\{ 1 - \frac{\sqrt{2}}{2\sqrt{\delta}},\ 1 + \frac{\sqrt{2}}{2\sqrt{\delta}} \right\} & \frac{b}{2} < c \end{cases} \tag{30}$$

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 $q < 1 - \frac{\sqrt{2}}{2}$.

Note that as $\delta \to 1$ the condition for which condition cooperators can avoid invasion is the same for the case of $m = 2$ and $n = 2$.

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

The final extension to the limited memory framework we consider is that of increasing the number of rounds and the number of interactions. For this case we need to consider a combination of the methods we presented in Section 2.3 and Section 2.4. As an example consider the case of $m = n = 2$. The probability that the number of mutants increases, and decreases respectively, by one is now given by,

$$\lambda_k^+ = \frac{N-k}{N} \cdot \frac{k}{N} \cdot \sum_{u_{R1},u_{M1} \in \mathcal{U}^2} \sum_{u_{R2},u_{M2} \in \mathcal{U}^2} x^2((u_{R1},u_{R2}),(u_{M1},u_{M2})) \cdot \rho\left(\frac{u_{R1}+u_{R2}}{2}, \frac{u_{M1}+u_{M2}}{2} \right), \tag{31}$$

$$\lambda_k^- = \frac{N-k}{N} \cdot \frac{k}{N} \cdot \sum_{u_{R1},u_{M1} \in \mathcal{U}^2} \sum_{u_{R2},u_{M2} \in \mathcal{U}^2} x^2((u_{R1},u_{R2}),(u_{M1},u_{M2})) \cdot \rho\left(\frac{u_{M1}+u_{M2}}{2}, \frac{u_{R1}+u_{R2}}{2} \right). \tag{32}$$

Though we do not carry any further analytical exploration of this case, in the next section we present simulation results when the updating payoffs of the pairwise process depend on the limited memory framework with $m = n = 2$, as well as the rest of the cases we have discussed so far.

3 Further simulation results

3.1 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 2.1-2.5. 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 in 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 jump in the cooperation rate when we allow for more information (in the form of interactions or rounds). We hypothesise that as we allow for more information the results will tend to the case of perfect memory.

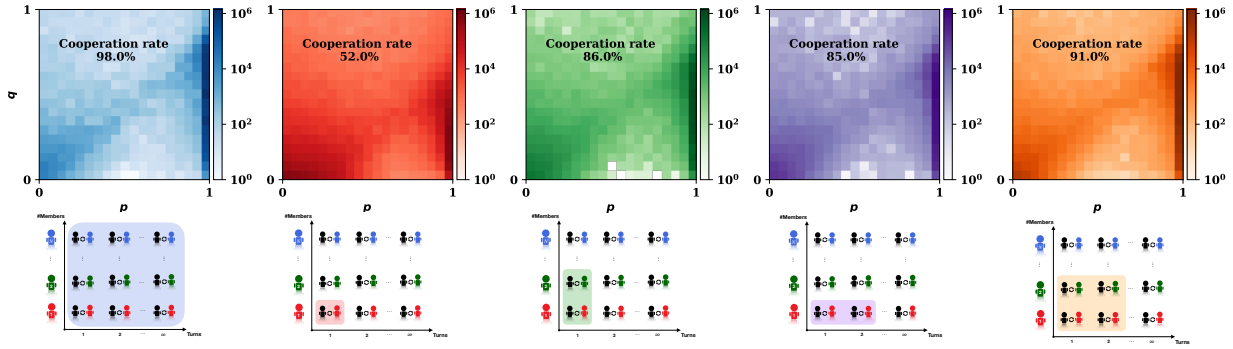

Figure 2: Evolutionary dynamics with difference updating approach. From left to right, we present result on the following updating payoffs cases; the expected payoffs (perfect memory), the last round payoff from one interaction (limited memory), the last round payoff from two interactions, the last two 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.

3.2 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 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.

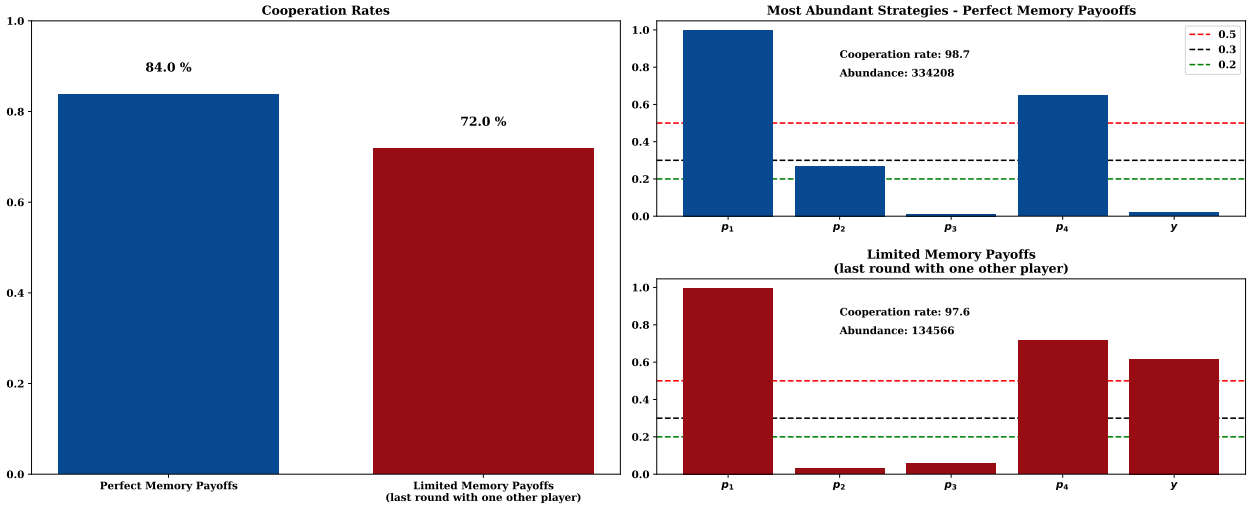

Figure 3: Evolutionary dynamics results for memory-one strategies for low benefit. We perform two independent simulations. In one simulation individuals use expected payoffs and in the other the last round one interacts when they update their strategies. We run each simulation for $T=10^8$ time steps.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

3.3 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

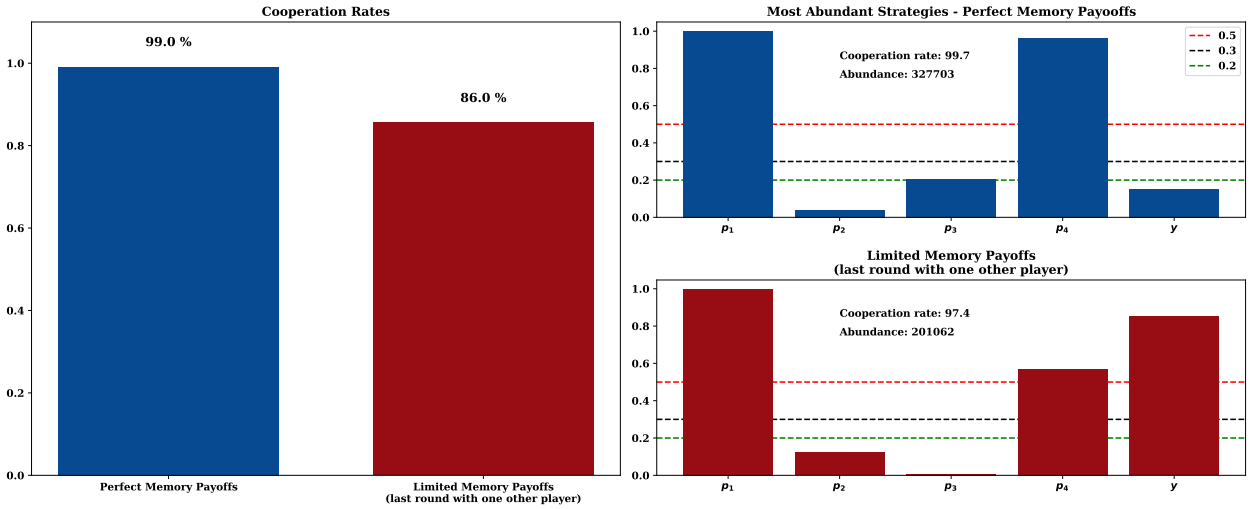

Figure 4: Evolutionary dynamics results for memory-one strategies for high benefit. We perform two independent simulations. In the case of high benefit expected payoffs again result in a more cooperative population. The right panel reports the most abundant strategy of each simulation. Abundance is the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

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.

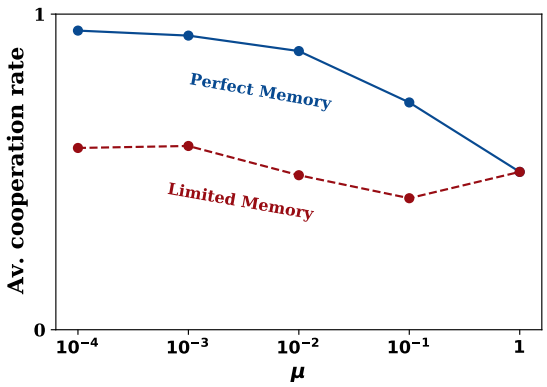

Figure 5: Evolutionary dynamics results for perfect and limited memory for different mutation values. We perform five independent simulations. Simulations are run for $T = 4 \times 10^7$ time steps for each parameter. In each time step we introduce a new mutant with a probability μ, and we then select 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: $N = 100, c = 1, b = 10, \beta = 1$.

References

[1] Axelrod, R. & Hamilton, W. D. The evolution of cooperation. *Science* **211**, 1390–1396 (1981).

[2] Traulsen, A., Pacheco, J. M. & Nowak, M. A. Pairwise comparison and selection temperature in evolutionary game dynamics. *Journal of theoretical biology* **246**, 522–529 (2007).

[3] Fudenberg, D. & Imhof, L. A. Imitation processes with small mutations. *Journal of Economic Theory* **131**, 251–262 (2006).

[4] Wu, B., Gokhale, C. S., Wang, L. & Traulsen, A. How small are small mutation rates? *Journal of Mathematical Biology* **64**, 803–827 (2012).

[5] McAvoy, A. Comment on "Imitation processes with small mutations". *J. Econ. Theory* **159**, 66–69 (2015).

[6] Nowak, M. A., Sasaki, A., Taylor, C. & Fudenberg, D. Emergence of cooperation and evolutionary stability in finite populations. *Nature* **428**, 646–650 (2004).

[7] Sigmund, K. *The calculus of selfishness* (Princeton University Press, 2010).

[8] Molander, P. The optimal level of generosity in a selfish, uncertain environment. *Journal of Conflict Resolution* **29**, 611–618 (1985).

[9] Nowak, M. A. & Sigmund, K. Tit for tat in heterogeneous populations. *Nature* **355**, 250–253 (1992).

[10] Schmid, L., Chatterjee, K., Hilbe, C. & Nowak, M. A unified framework of direct and indirect reciprocity. *Nature Human Behaviour* **5**, 1292–1302 (2021).

CH: Adjust symbols in main text – for role model, strategy, etc