Evolution of cooperation among individuals with limited payoff memory

Nikoleta E. Glynatsi, Christian Hilbe, Alex McAvoy

Abstract

1 Introduction

Evolutionary game theory ???? describes the evolutionary dynamics of populations consisting of different types of interacting individuals. In the last two decades the interest of the field has shifted to the analysis of stochastic finite population dynamics in preference to traditional approaches such as the replicator dynamics ??. This is especially true in the topic of cooperation ???. In stochastic evolutionary dynamics disadvantageous mutants have a small yet non-zero probability to reach fixation and this can lead to fundamental changes in the results. As shown in ?, using such dynamics a single cooperative strategy can invade a population of defectors without any special modifications in comparison to traditional approaches that one has to consider spatial structure ? or when payoffs are subject to aggregate shocks ?.

Stochastic evolutionary dynamics model a finite population. Each time step of the process consists of three phases; (1) the *mutation phase* (2) the *game phase* (3) the *update phase*. In the mutation phase one individual from the population 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. The updating phase depends on the process. Two classes of finite stochastic processes have been used extensively: (i) fitness-based processes in which an individual chosen proportional to fitness reproduces and the offspring replaces a randomly chosen individual ? or (ii) *pairwise comparison processes* in which a pair of individuals is chosen and subsequently one of these individuals may adopt the strategy of the other ?.

Similar to all other theoretical models, finite stochastic processes rest on a set of assumptions. Namely, in the game phase it is assumed that players use strategies with finite memory ??. For example, in a repeated game between two individuals the actions of the players in each turn are often determined by the action of the co-players in the previous turn. This assumption is common as it allows for an explicit calculation of the players' payoffs ?. In addition, it is assumed that individuals play many times and with all other players before reproduction takes place. So that in the updating phase the updating payoffs, and subsequently the fitness, of individuals is given by the average payoffs of an infinitely repeated game against the mean distribution of types in the population.

The above two assumptions create a curious inconsistency; a player with a finite memory in the game stage can recall infinite information regarding their interactions and each interaction's outcomes in the update stage. Previous work has explored the effects of constraining the interactions an individual has. This can be done by letting individuals have a stochastic number of interactions **???**, a small number of interactions, or even in the extreme case a single interaction **?**. However, no previous work considers that not only the interactions are limited but also the information regarding each interaction. To this end, we propose a framework in which individuals, similar to the decisions at each turn, estimate their fitness based on a minimum of information.

We first consider two extreme scenarios, the classical scenario where an individual has a perfect memory of all their interactions and the alternative scenario of limited memory where individuals update their strategies only based on the very last payoff they obtained. We demonstrate the effects of limited updating payoff memory using a pairwise comparison process and the well studied game of the prisoner's dilemma. We observe that individuals with limited memory tend to adopt less generous strategies and they achieve less cooperation when interacting in a prisoner's dilemma. We obtain similar results when we consider that individuals update their strategies based on more information. More specifically, up to the last two payoffs they obtained when interacting with up to two different members of the population.

2 Model Setup

A pairwise comparison process starts with assigning all individuals of the population the same strategy. Each elementary time step of the process consists of the mutation phase, the game phase and the update phase. In the game phase individuals are matched in pairs and that they participate in a repeated 2 person donation game; a special case of the prisoner's dilemma. 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,

$$\begin{array}{cc} & \begin{array}{cc}\text{cooperate} & \text{defect}\end{array} \\ \begin{array}{c}\text{cooperate}\\ \text{defect}\end{array} & \left(\begin{array}{cc} b-c & -c \\ b & 0 \end{array}\right). \end{array} \tag{1}$$

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 repeated games there are infinite many strategies, however, similar to the literature we will assume that individuals use 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. The play between a pair of reactive strategies ($s_1 = (y_1, p_1, q_1)$, $s_2 = (y_2, p_2, q_2)$) can be model as a Markov process with the transition matrix M,

$$M = \begin{bmatrix} p_1p_2 & p_1\left(1-p_2\right) & p_2\left(1-p_1\right) & \left(1-p_1\right)\left(1-p_2\right) \\ p_2q_1 & q_1\left(1-p_2\right) & p_2\left(1-q_1\right) & \left(1-p_2\right)\left(1-q_1\right) \\ p_1q_2 & p_1\left(1-q_2\right) & q_2\left(1-p_1\right) & \left(1-p_1\right)\left(1-q_2\right) \\ q_1q_2 & q_1\left(1-q_2\right) & q_2\left(1-q_1\right) & \left(1-q_1\right)\left(1-q_2\right) \end{bmatrix} \tag{2}$$

and the stationary vector $\mathbf{v}(s_1, s_2)$ which is the solution to $\mathbf{v}(s_1, s_2) \times M = \mathbf{v}(s_1, s_2)$.

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,

$$\rho(\pi_L, \pi_{RM}) = \frac{1}{1+e^{-\beta(\pi_{RM}-\pi_L)}}. \tag{3}$$

π_L and π_{RM} are the updating payoffs/fitness 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.

For the results presented here we assume that mutations are rare ($\mu \to 0$). In fact, so rare that only two different strategies can be present in the population at any given time. However, in the Supplementary Information Section 4 we show in that the main result hold for $\mu \neq 0$. 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, 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,

$$\varphi_M = \frac{1}{1 + \sum\limits_{i=1}^{N-1} \prod\limits_{k}^{i} \frac{\lambda_k^-}{\lambda_k^+}}, \tag{4}$$

where λ_k^-, λ_k^+ are the probabilities that the number of mutants decreases and increases respectively, and k is the number of mutants. 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 probabilities λ_k^- and λ_k^+ depend on the updating payoffs of the mutant and the resident strategies. In the next section we present how they are calculated in the cases of perfect and limited memory.

2.1 Updating payoffs based Perfect and Limited Memory

Perfect Memory

In the perfect memory case an individual updates based on the average payoff against each other member of the population, otherwise as expected payoffs. In an infinitely repeated game ($\delta \to 1$) the payoff of a reactive strategy s_1 against the reactive strategy s_2 can explicitly be calculated using the stationary vector $\mathbf{v}$ and the payoffs' vector as $\langle \mathbf{v}(s_1, s_2), \mathbf{u}\rangle$. In a population of size N there are k mutants and $N-k$ residents. We denote the strategies of a mutant and of a resident 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 of a mutant (π_M) are given by,

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

The probabilities that the number of mutants decreases and increases, λ_k^- and λ_k^+, in the perfect memory case are defined as,

$$\lambda_k^- = \rho(\pi_M, \pi_R) \quad \text{and} \quad \lambda_k^+ = \rho(\pi_R, \pi_M). \tag{6}$$

Limited Memory

In this case of limited memory we initially define the probability that a reactive strategy receives the payoff $u \in \mathcal{U}$ in the very last round of the game. This is given by Proposition **??** (see Supplementary Information Section 2.2.1 for proof).

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

$$
\begin{aligned}
v_r(s_1,s_2) &= (1-\delta)\frac{y_1 y_2}{1-\delta^2 l_1 l_2} + \delta\frac{\Big(q_1 + l_1\big((1-\delta)y_2 + \delta q_2\big)\Big)\Big(q_2 + l_2\big((1-\delta)y_1 + \delta q_1\big)\Big)}{(1-\delta l_1 l_2)(1-\delta^2 l_1 l_2)} \times r,\\
v_s(s_1,s_2) &= (1-\delta)\frac{y_1 \bar{y}_2}{1-\delta^2 l_1 l_2} + \delta\frac{\Big(q_1 + l_1\big((1-\delta)y_2 + \delta q_2\big)\Big)\Big(\bar{q}_2 + \bar{r}_2\big((1-\delta)y_1 + \delta p_1\big)\Big)}{(1-\delta l_1 l_2)(1-\delta^2 l_1 l_2)} \times s,\\
v_t(s_1,s_2) &= (1-\delta)\frac{\bar{y}_1 y_2}{1-\delta^2 l_1 l_2} + \delta\frac{\Big(\bar{q}_1 + \bar{r}_1\big((1-\delta)y_2 + \delta p_2\big)\Big)\Big(q_2 + l_2\big((1-\delta)y_1 + \delta q_1\big)\Big)}{(1-\delta l_1 l_2)(1-\delta^2 l_1 l_2)} \times t,\\
v_p(s_1,s_2) &= (1-\delta)\frac{\bar{y}_1 \bar{y}_2}{1-\delta^2 l_1 l_2} + \delta\frac{\Big(\bar{q}_1 + \bar{r}_1\big((1-\delta)y_2 + \delta p_2\big)\Big)\Big(\bar{q}_2 + \bar{r}_2\big((1-\delta)y_1 + \delta p_1\big)\Big)}{(1-\delta l_1 l_2)(1-\delta^2 l_1 l_2)} \times p.
\end{aligned}
\tag{7}
$$

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* $i \in \{1,2\}$.

In the case of limited payoffs memory both the role model and the learner estimate their fitness after interacting with a single member of the population. At each time step there are five possible pairings. They interact with each other with a probability $\frac{1}{N-1}$, and thus they do not interact with other with a probability $1-\frac{1}{N-1}$. In the latter case, each of them can interact with either a mutant or a resident. Both of them interact with a mutant with a probability $\frac{(k-1)(k-2)}{(N-2)(N-3)}$ and both interact with a resident with a probability $\frac{(N-k-1)(N-k-2)}{(N-2)(N-3)}$. The last two possible pairings are that either of them interacts with a resident whilst the other interacts with a mutant, and this happens with a probability $\frac{(N-k-1)(k-1)}{(N-2)(N-3)}$. Thus we 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 as $x(u_R, u_M)$.

$$
\begin{aligned}
x(u_1,u_2) = \quad & \frac{1}{N-1} \cdot v_{u_1}(s_1,s_2) \cdot 1_{(u_1,u_2)\in\mathcal{U}_F^2}\\
& + \left(1-\frac{1}{N-1}\right) \Big[\tfrac{k-1}{N-2}\tfrac{k-2}{N-3} v_{u_1}(s_1,s_2) v_{u_2}(s_2,s_2) + \tfrac{k-1}{N-2}\tfrac{N-k-1}{N-3} v_{u_1}(s_1,s_2) v_{u_2}(s_2,s_1)\\
& \qquad + \tfrac{N-k-1}{N-2}\tfrac{k-1}{N-3} v_{u_1}(s_1,s_1) v_{u_2}(s_2,s_2) + \tfrac{N-k-1}{N-2}\tfrac{N-k-2}{N-3} v_{u_1}(s_1,s_1) v_{u_2}(s_2,s_1) \Big].
\end{aligned}
\tag{8}
$$

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 $\frac{1}{(N-1)}$. The probability that the number of mutants increases and decreases by one in the case of limited memory are now given by,

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

In this expression, $\frac{(N-k)}{N}$ is the probability that the randomly chosen learner is a resident, and $\frac{k}{N}$ is the probability that the role model is a mutant. The sum corresponds to the total probability that the learner adopts the role model's strategy over all possible payoffs u_R and u_M that the two player may have received in their respective last rounds.

Simulation Results

In order to account for the effect of the updating payoffs we simulate the evolutionary process and record which strategies the players adopt over time based on perfect memory payoffs and limited memory payoffs. In the case of perfect memory payoffs we use Eq **??** when we estimate the probability that a mutant fixes, and Eq **??** in the case on limited memory. We run two independent runs for each approach and we vary the value of benefit b. Figure **??** shows the simulation results. It depicts the evolving conditional cooperation probabilities p and q. The discount factor δ is comparably high, thus we do not report the opening move y as it is a transient effect. The upper panel corresponds to the standard scenario considered in the literature, it considers players who use expected payoffs to update their strategies. The bottom panel shows the scenario considered herein, in which players update their strategies based on their last round's payoff.

The figure suggests that when updating is based on perfect memory payoffs players tend to be more generous and more cooperative. The q-values of the resident strategies are on average higher in the case of perfect memory. Thus, players will occasionally forgive a defection more often if their fitness depends on interacting with every member of the population, and this effect becomes more notable as the value of the benefit increases. In the case of limited memory the resident strategies are less forgiving and their forgiveness is independent of the values of benefit. In the Supplementary Information Sections 2.1 and 2.2.1 we show that q-values of a resident strategy have to be less than $1 - c/b$ in the perfect memory case and less than 0.5 in the limited memory case. The generosity for the perfect memory case is always higher for a benefit value higher than 2, and it increases as the benefit value increases. This leads to a more cooperative population. We calculate the average cooperation rate for each simulation which is the average cooperation rate within the resident population. In the case of the perfect memory payoffs the average cooperation rate is strictly higher than that of the last round payoffs, and the difference is most for $b = 10$. The average cooperation of resident strategies drops from 97% to 57%. This indicates that the expected payoffs overestimate the evolved cooperation.

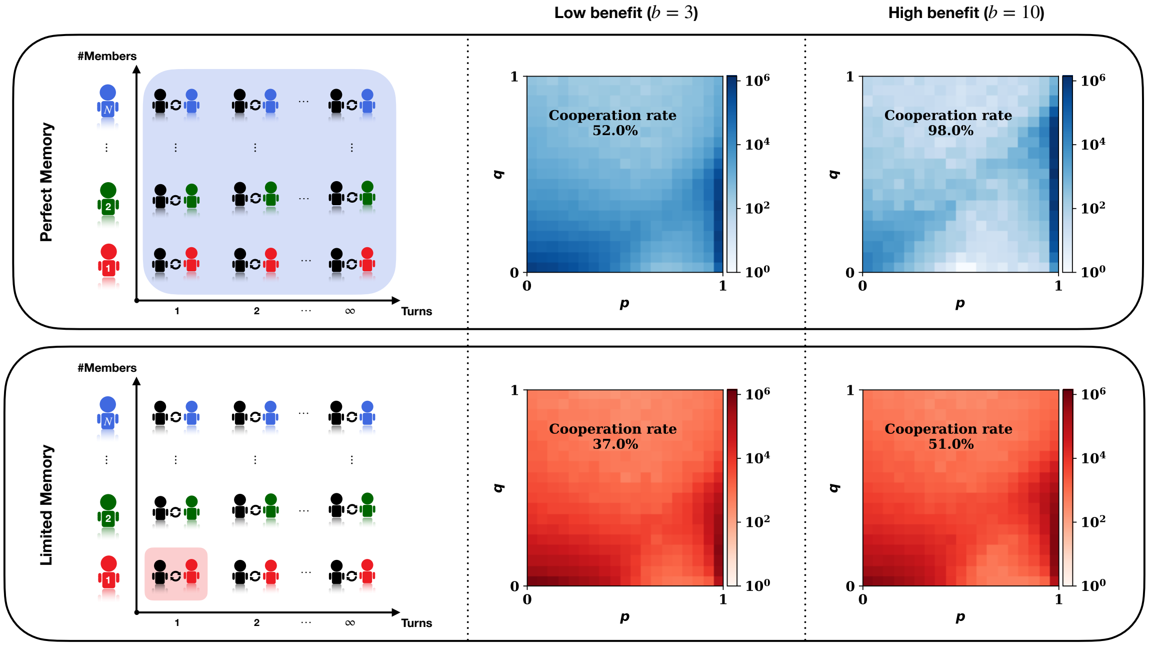

Figure 1: Evolutionary dynamics under perfect and limited payoff memory. (**Schematic illustrations**) On the left panels we show schematic illustrations of the perfect memory and the limited memory cases. The shaded background denotes the game phase information that an individual considered when updating strategies. In the case of perfect memory the entire region is shaded and in the case of limited memory only one turn with a single member of the population. (**Simulations**) We have run four simulations of the pairwise comparison process for $T = 10^7$ time steps. For each time step, we have recorded the current resident population (y, p, q). Since simulations are run for a relatively high continuation probability of $\delta = 0.999$, we do not report the players' initial cooperation probability y. The graphs show how often the resident population chooses each combination (p, q) of conditional cooperation probabilities in the subsequent rounds. We also report the evolved cooperation rate which is calculated as the average cooperation rate within the resident population. (**Perfect Memory**) In the case of low benefit the resident population either consists of defectors (with $p \approx q \approx 0$) or of conditional cooperators. Conditional cooperators, or otherwise known as generous tit for tat, are a set of strategies that always cooperate following a cooperation ($p \approx 1$) and cooperate with a probability q given that the co-player has defected. q denotes the generosity of a player. The resident population applies a conditional cooperator strategy for which $q \leq 1{-}c/b{=}0.67$. This is true also for the case of high benefit. In this case the population mainly consists of conditional cooperators of the form ($p \approx 1, q \leq 1 - 1/10 = 0.9$). In the Supplementary Information Section 2.1 we show that a conditional cooperator needs to be of the form ($p \approx 1, q \leq 1 - c/b$) to not be invaded by defecting strategies. A higher generosity in the population results in a higher average cooperation rate. The average cooperation rate increases from 52% for b to 98% for $b = 10$. (**Limited Memory**) When players update their strategies based on their realized payoffs in the last round, there are two different predominant behaviors regardless of the benefit value. The resident population either consists of defectors (with $p \approx q \approx 0$) or of conditional cooperators. The maximum level of q consistent with stable cooperation is somewhat smaller compared to the perfect memory setting, $q < 0.5$. Namely, in the Supplementary Information Section 2.2.1 we show that regardless of the value of benefit, in the case of limited payoff memory, an individual needs a generosity smaller of 0.5 to repel defectors. The plots are fairly similar regardless of the benefit, and the evolved cooperation rate only slightly increases from 37% to 51%. Parameters: $N{=}100$, $c{=}1$, $\beta{=}1$, $\delta{=}0.999$.

We further explore the effects of the cooperation benefit and the strength of selection on the evolved generosity q and cooperation rate (Figure **??**). Figure **??A** suggests that perfect memory always yields a higher cooperation rate. We observe that the cooperation rate increases as the value of the benefit gets higher, whereas in comparison for the limited memory payoffs, the cooperation rate remains unchanged (as we would have expected) at approximately 50% once $b = 5$. From Figure **??B** we observe that for weak selection, $\beta < 1$, the two methods yield similar results, however, as β increases there is variation in the evolving populations. In the case of expected payoffs the resident populations become more cooperative as β increases, whereas in the case of limited memory payoffs, the resident populations become more defective.

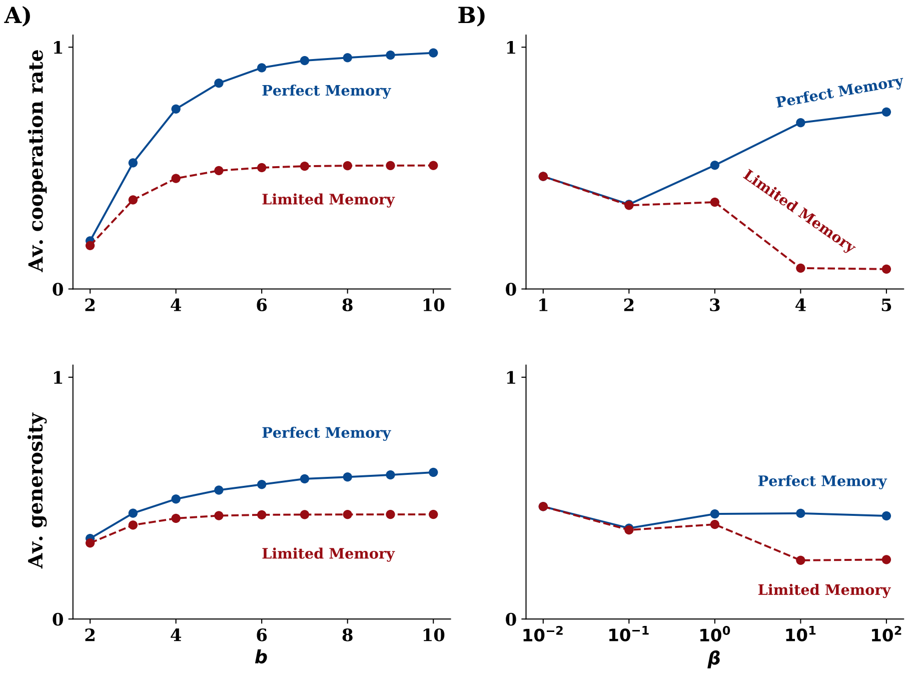

Figure 2: The evolution of cooperation and generosity for different values of benefit (A) and strength of selection (B). We report the average cooperation and the average reciprocity. The average cooperation rate is the average cooperation rate within the resident population. For the average reciprocity we select the residents that have a $p \approx 1$ and we take the average of their cooperation probability q. (**A**) We vary the benefit of cooperation b. In all cases, perfect memory updating payoffs appear to overestimate the average cooperation rate the population achieves. As expected in the case of limited memory the average generosity over the different values of benefit remains the same ($q \approx 0.5$), and as a result so does the average cooperation. (**B**) We vary the selection strength β. For weak selection, $\beta < 1$, the two methods yield similar results. However, as β increases in the case of limited memory payoffs the resident populations become more defective. Unless explicitly varied, the parameters of the simulation are $N = 100$, $b = 3$, $c = 1$, $\beta = 1$, $\delta = 0.99$. Simulations are run for $T = 5 \times 10^7$ time steps for each parameter combination.

2.2 Updating Payoffs based on More Memory

So far we have explored the difference between the expected payoffs and the last round payoffs. In order to explore further the effect of limited memory we allow individuals to remember more. More precisely, up to two interactions and up to the last two rounds. In total we present results for three more updating payoffs. These are the payoffs when individuals consider the last two rounds with another member of the population, the last round with two members of the population, and the last two rounds with two members of the population. Similar to the last round payoff, we use simulations and record which strategies the players adopt over time based

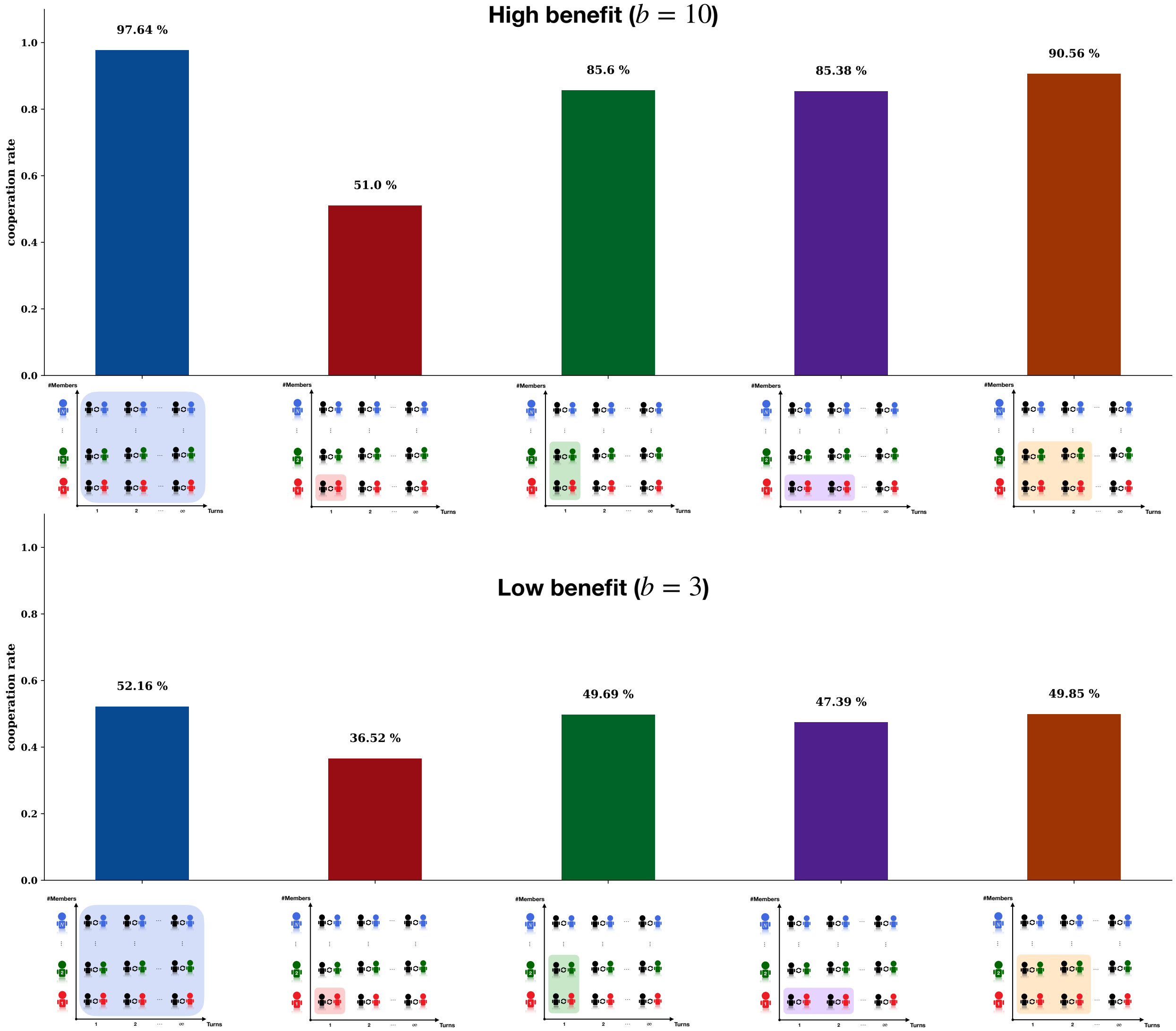

Figure 3: Average cooperation rates for different updating payoffs. (From right to left) updating Payoffs based on perfect memory, limited memory, the last round payoff of two interactions, the last two rounds payoff of one interaction and the last two round payoffs of two interactions.(**High Benefit**) (**Low Benefit**) We vary the selection strength β. In all cases, stochastic payoff evaluation tends to reduce the evolving cooperation rates. Unless explicitly varied, the parameters of the simulation are $N = 100$, $b = 3$, $c = 1$, $\beta = 1$, $\delta = 0.99$. Simulations are run for $T = 5 \times 10^7$ time steps for each parameter combination.