Evolution of cooperation among individuals with limited payoff memory

Nikoleta E. Glynatsi, Christian Hilbe, Alex McAvoy

Abstract

Repeated games are vastly used in evolutionary game theory to explain cooperation in a variety of environments. Although existing models have shaped our understanding of human cooperation they often work with idealized assumptions. In an evolutionary process, individuals imitate other individuals of the population based on their fitness. It is commonly assumed that an individual computes their fitness after interacting with a representative sample of the population and remembering all the interactions they participated in. In real life we do not always remember all our interactions, instead we rather recall our most recent ones.

Here we introduce a framework that allows individuals to estimate their fitness based on a minimum of social information. We explore the difference between our framework and the classical framework using computer simulations. The simulations show that in the prisoner's dilemma, individuals with limited memory tend to adopt less generous strategies and they achieve less cooperation than in the classical scenario.

1 Introduction

Evolutionary game theory [1–4] 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 [5, 6]. This is especially true in the topic of cooperation [7–9]. 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 [3], 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 [10] or when payoffs are subject to aggregate shocks [11].

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 [3] 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 [12].

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 [13, 14]. 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 [15]. 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. Either by letting individuals have a stochastic number of interactions [16–18] or by only allowing for a single interaction [19]. 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 payoff memory using a pairwise comparison process and the well studied game of the prisoner's dilemma. We observe that individuals with limited payoff 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. Thenceforth 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 given by,

$$
\begin{array}{cc}
& \begin{array}{cc} \text{cooperate} & \text{defect} \end{array} \\
\begin{array}{c} \text{cooperate} \\ \text{defect} \end{array} & \begin{pmatrix} b - c & -c \\ b & 0 \end{pmatrix}.
\end{array}
\tag{1}
$$

Let $\mathbf{u} = (b - c, -c, b, 0)$ be the round 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 $s_1 = (y_1, p_1, q_1)$ and $s_2 = (y_2, p_2, q_2)$ can be model as a Markov process with the transition matrix M,

$$
M = \begin{bmatrix}
p_1 p_2 & p_1 (1 - p_2) & p_2 (1 - p_1) & (1 - p_1)(1 - p_2) \\
p_2 q_1 & q_1 (1 - p_2) & p_2 (1 - q_1) & (1 - p_2)(1 - q_1) \\
p_1 q_2 & p_1 (1 - q_2) & q_2 (1 - p_1) & (1 - p_1)(1 - q_2) \\
q_1 q_2 & q_1 (1 - q_2) & q_2 (1 - q_1) & (1 - q_1)(1 - q_2)
\end{bmatrix}
\tag{2}
$$

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 9 we show that the main result holds 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 as,

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

where k is the number of mutants and λ_k^-, λ_k^+ are the probabilities that the number of mutants decreases and increases respectively. Depending on the fixation probability ϕ_M the mutant either fixes (becomes the new resident) or goes extinct. Regardless, in the next 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 fitness of the mutant and the resident strategies. In the next section we present how fitness is calculated in the cases of perfect and limited payoff memory.

2.1 Fitness based on Perfect and Limited Payoff Memory

In the perfect payoff memory case an individual updates based on the average payoff against each other member of the population, otherwise known as expected payoffs. The payoff of a reactive strategy s_1 against the reactive strategy s_2 in an infinitely repeated game ($\delta \to 1$) is explicitly calculated 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. Let $s_M = (y_M, p_M, q_M)$ and $s_R = (y_R, p_R, q_R)$ denote the strategies of a mutant and a resident, the expected payoffs π_R and 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 payoff memory case are defined as,

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

In the case of limited payoff 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 1 (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 (7).*

$$v_r(s_1, s_2) = (1-\delta)\frac{y_1 y_2}{1-\delta^2 l_1 l_2} + \delta \frac{\Big(q_1 + l_1\big((1-\delta)y_2 + \delta q_2\big)\Big)\Big(q_2 + l_2\big((1-\delta)y_1 + \delta q_1\big)\Big)}{(1-\delta l_1 l_2)(1-\delta^2 l_1 l_2)},$$

$$v_s(s_1, s_2) = (1-\delta)\frac{y_1 \bar{y}_2}{1-\delta^2 l_1 l_2} + \delta \frac{\Big(q_1 + l_1\big((1-\delta)y_2 + \delta q_2\big)\Big)\Big(\bar{q}_2 + \bar{r}_2\big((1-\delta)y_1 + \delta p_1\big)\Big)}{(1-\delta l_1 l_2)(1-\delta^2 l_1 l_2)},$$

$$(7)$$

$$v_t(s_1, s_2) = (1-\delta)\frac{\bar{y}_1 y_2}{1-\delta^2 l_1 l_2} + \delta \frac{\Big(\bar{q}_1 + \bar{r}_1\big((1-\delta)y_2 + \delta p_2\big)\Big)\Big(q_2 + l_2\big((1-\delta)y_1 + \delta q_1\big)\Big)}{(1-\delta l_1 l_2)(1-\delta^2 l_1 l_2)},$$

$$v_p(s_1, s_2) = (1-\delta)\frac{\bar{y}_1 \bar{y}_2}{1-\delta^2 l_1 l_2} + \delta \frac{\Big(\bar{q}_1 + \bar{r}_1\big((1-\delta)y_2 + \delta p_2\big)\Big)\Big(\bar{q}_2 + \bar{r}_2\big((1-\delta)y_1 + \delta p_1\big)\Big)}{(1-\delta l_1 l_2)(1-\delta^2 l_1 l_2)}.$$

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 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 $\frac{(N-k-1)(k-1)}{(N-2)(N-3)}$. We define 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 $x(u_R, u_M)$.

5

$$x(u_R, u_M) = \frac{1}{N-1} \cdot v_{u_R}(s_R, s_M) \cdot 1_{(u_R, u_M) \in \mathcal{U}_F^2}$$

$$+ \left(1 - \frac{1}{N-1}\right) \left[\tfrac{k-1}{N-2} \tfrac{k-2}{N-3} v_{u_R}(s_R, s_M) v_{u_M}(s_M, s_M) + \tfrac{k-1}{N-2} \tfrac{N-k-1}{N-3} v_{u_R}(s_R, s_M) v_{u_M}(s_M, s_R) \right.$$

$$\left. + \tfrac{N-k-1}{N-2} \tfrac{k-1}{N-3} v_{u_R}(s_R, s_R) v_{u_M}(s_M, s_M) + \tfrac{N-k-1}{N-2} \tfrac{N-k-2}{N-3} v_{u_R}(s_R, s_R) v_{u_M}(s_M, s_R) \right].$$

$$(8)$$

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 payoff memory are now given by,

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

$$(9)$$

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.

3 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 and limited payoff memory. We performed two independent runs for each approach varying the value of benefit b. Figure 1 depicts the evolving conditional cooperation probabilities p and q (since the discount factor δ is comparably high we do not report the opening move y). The upper panel of Figure 1 corresponds to the standard scenario considered in the literature, and 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 payoff memory players tend to be more generous and more cooperative. The q-values of the resident strategies are on average higher in this case, thus, players will occasionally forgive a defection more often if their fitness depends on interacting with every member of the population. 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 regardless of the

values of benefit. In analysis carried out in the Supplementary Information Section 3 we show that a resident population of conditional cooperators with $q > \frac{1}{2}$ can not avoid invasion by a defecting strategy. This holds independently of the cost and benefit values.

Higher q values lead 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 payoff memory the average cooperation rate is strictly higher than that of the last round payoffs. The difference between the cooperation rate is the highest at $b = 10$.
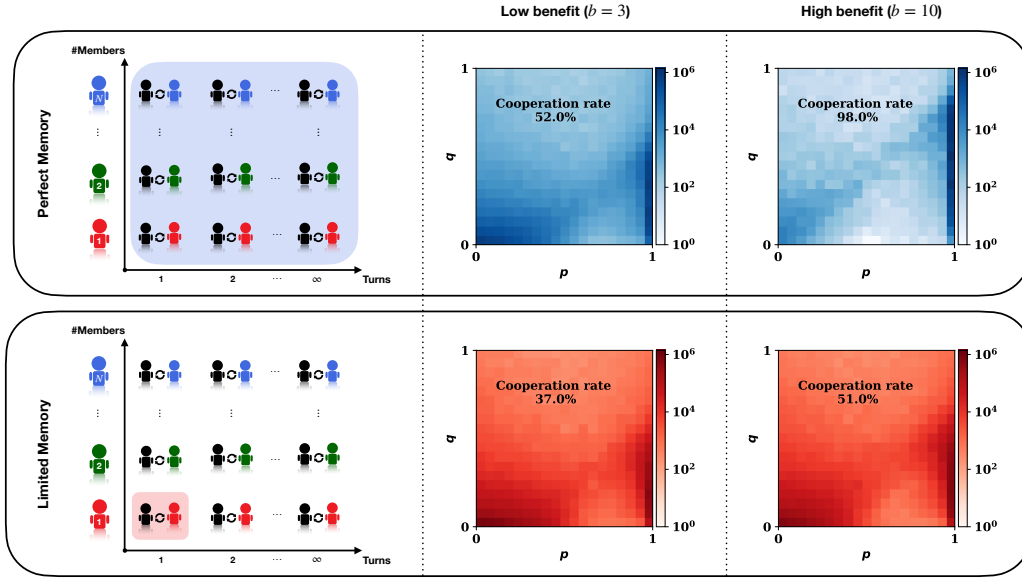

Figure 1: Evolutionary dynamics under perfect and limited payoff memory. (**Schematic illustrations**) On the left panels we show schematic illustrations of the perfect memory and the limited memory cases. The shaded background denotes the game phase information that an individual considers when updating strategies. In the case of perfect payoff memory the entire region is shaded and in the case of limited payoff 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 of the process we record 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 plots 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 $q \leq 1 - c/b = 0.67$. In the case of high benefit 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 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 $b = 3$ 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. Namely, in the Supplementary Information Section 3 we show that regardless of the value of benefit a conditional cooperator need to be of the form $q < \frac{1}{2}$ to not be invaded by defectors. The evolved cooperation rate only slightly increases from 37% ($b = 3$) to 51% ($b = 10$). 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 on the cooperation rate (Figure 2). Figure 2**A** suggests that perfect memory always yields a higher cooperation rate. In this case the cooperation rate increases as the benefit increases. In comparison for limited payoff memory the cooperation rate remains unchanged at approximately 50% once $b = 5$. From Figure 2**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, whereas in the case of limited payoff memory, the resident populations become more defective.
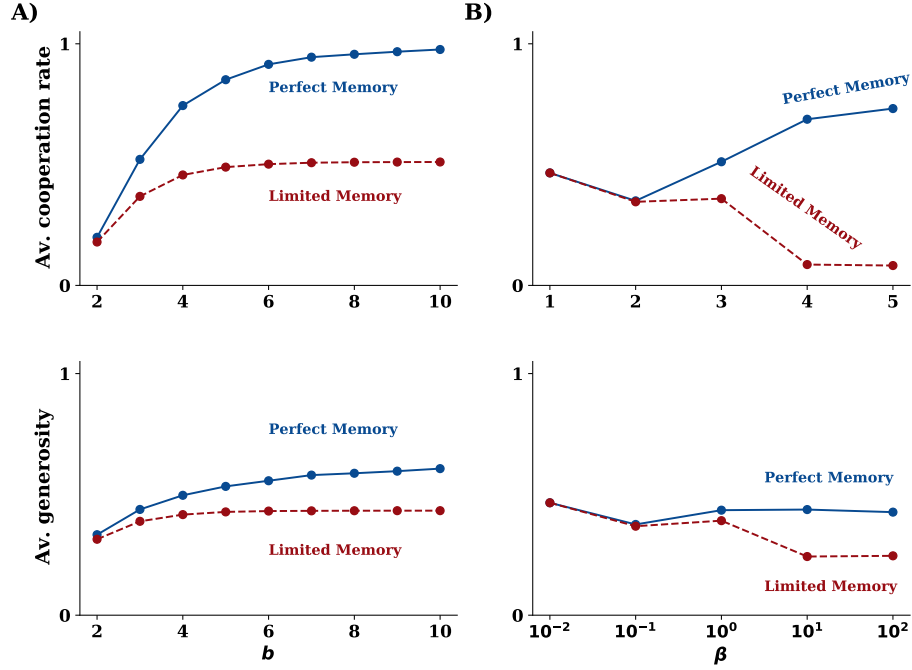

Figure 2: The evolution of cooperation and generosity for different values of benefit (A) and strength of selection (B). We report the average cooperation and the average reciprocity. The average cooperation rate is the average cooperation rate within the resident population. For the average reciprocity 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.

Up to this point we have discussed only two edge cases. The framework of limited payoff memory can be easily extended to allow individuals to observe more rounds, to interact with more members or both (see Supplementary Information Sections 4-6). To further understand the effect of limited payoff memory we consider the cases of updating based on the last round with two members of the population, the last two rounds with another member of the population, and the last two rounds with two members of the population.

Similar to the last round payoff. We perform numerical simulations for each of the fitness approaches. In Figure 3 we report the cooperation rates for low and high values of benefit.
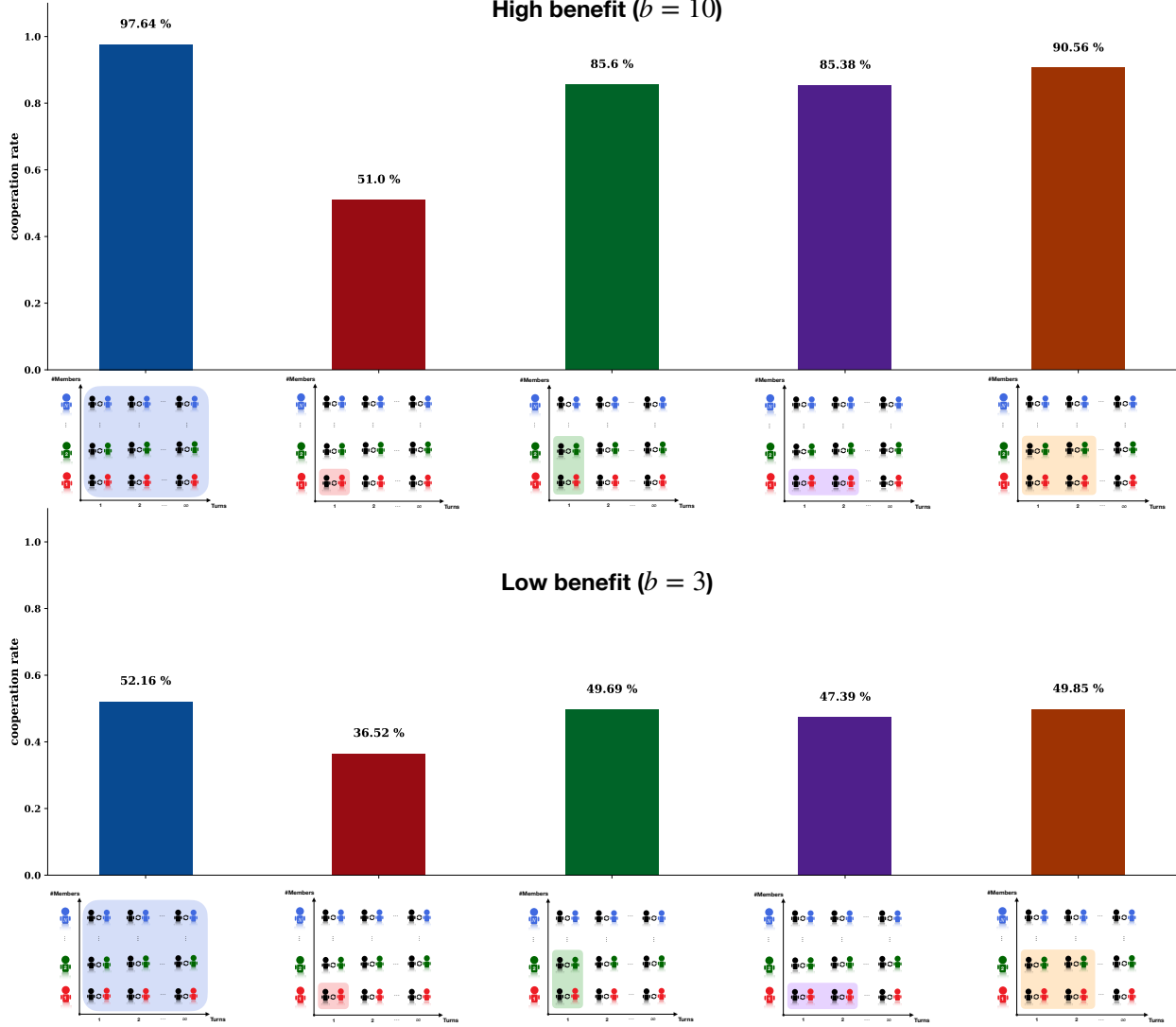

Figure 3: Average cooperation rates for different updating payoffs. From left to right, we present result on the following updating payoffs cases; (a) the expected payoffs (perfect memory), (b) the last round payoff from one interaction (limited memory), (c) the last round payoff from two interactions, (d) the last two rounds payoffs from one interaction, (e) the last two rounds payoffs from two interactions. For the updating payoffs (c) and (d) we have carried out an analysis which has shown that conditional cooperators should adopt a q value smaller than $\frac{\delta - 1 + \frac{\sqrt{2}}{2}}{\delta}$ and $\frac{\delta + \sqrt{\delta^2 + 1} - 1}{2\delta}$ respectively. Note that as $\delta \to 1$ both right hand sides tend to $\frac{\sqrt{2}}{2}$. Regardless the cooperation rate between for case (c) is slightly higher. In case (d) the cooperation rate is the second highest hinting that as we allow for more information the closer we move to the perfect payoff memory. 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.

4 Conclusions

Cooperation can be seen at odd, why is it that we choose to help others at a personal cost? In spite of all the selfish genes', animal and human communities show signs of altruism and cooperation [20–22]. Evolutionary game theoretical models have helped us shape our understanding of the evolution of cooperation. In fact, the evolution of cooperation constitutes such a major focus of the field that evolutionary game theory seems to be reduced to the evolution of cooperation [23]. Previous evolutionary models often feature a curious inconsistency. When modeling how individuals make decisions in each round, these models assume that players only remember the last round. However, when modeling how individuals update their strategies over time, individuals are assumed to have perfect memory.

Here, we have explored how robust cooperation is as models deviate from the perfect memory assumption. We considered a special case of the prisoner's dilemma, the donation game. The prisoner's dilemma is one of the most applied types of social games when we discuss the evolution of cooperation. We showed that when the last round payoff is used instead of the expected payoffs, cooperation struggles to evolve when only a minimum of social information is used at the updating stage. The resident strategies were both less cooperative and less generous. This effect was only intensified we increase the benefit and the strength of selection independently. The results showed that as each parameter was being increased the difference between the cooperation rates were widening. This indicates that cooperative players benefit from being able to interact with everyone in the population.

We extended our approach and presented results on different limited memory payoffs. The analysis showed cooperation can even evolve when we allow individual to consider more information.

5 Acknowledgements

References

[1] John Maynard Smith. *Evolution and the Theory of Games*. Cambridge university press, 1982.

[2] Josef Hofbauer, Karl Sigmund, et al. *Evolutionary games and population dynamics*. Cambridge university press, 1998.

[3] Martin A Nowak, Akira Sasaki, Christine Taylor, and Drew Fudenberg. Emergence of cooperation and evolutionary stability in finite populations. *Nature*, 428(6983):646–650, 2004.

[4] Christoph Hauert and György Szabó. Game theory and physics. *American Journal of Physics*, 73(5):405–414, 2005.

[5] Josef Hofbauer, Peter Schuster, and Karl Sigmund. A note on evolutionary stable strategies and game dynamics. *Journal of Theoretical Biology*, 81(3):609–612, 1979.

[6] Peter D Taylor and Leo B Jonker. Evolutionary stable strategies and game dynamics. *Mathematical biosciences*, 40(1-2):145–156, 1978.

[7] Christian Hilbe, Martin A Nowak, and Karl Sigmund. Evolution of extortion in iterated prisoner's dilemma games. *Proceedings of the National Academy of Sciences*, 110(17):6913–6918, 2013.

[8] Christian Hilbe, Štěpán Šimsa, Krishnendu Chatterjee, and Martin A Nowak. Evolution of cooperation in stochastic games. *Nature*, 559(7713):246–249, 2018.

[9] Nikoleta E Glynatsi and Vincent A Knight. Using a theory of mind to find best responses to memory-one strategies. *Scientific Reports*, 10(1):17287, 2020.

[10] Martin A Nowak and Robert M May. Evolutionary games and spatial chaos. *Nature*, 359(6398):826–829, 1992.

[11] Drew Fudenberg and Christopher Harris. Evolutionary dynamics with aggregate shocks. *Journal of Economic Theory*, 57(2):420–441, 1992.

[12] Arne Traulsen, Jorge M Pacheco, and Martin A Nowak. Pairwise comparison and selection temperature in evolutionary game dynamics. *Journal of theoretical biology*, 246(3):522–529, 2007.

[13] Martin A Nowak and Karl Sigmund. Tit for tat in heterogeneous populations. *Nature*, 355(6357):250–253, 1992.

[14] Seung Ki Baek, Hyeong-Chai Jeong, Christian Hilbe, and Martin A Nowak. Comparing reactive and memory-one strategies of direct reciprocity. *Scientific reports*, 6(1):1–13, 2016.

[15] Karl Sigmund. *The calculus of selfishness*. Princeton University Press, 2010.

[16] Carlos P Roca, José A Cuesta, and Angel Sánchez. Time scales in evolutionary dynamics. *Physical review letters*, 97(15):158701, 2006.

[17] Angel Sánchez and José A Cuesta. Altruism may arise from individual selection. *Journal of theoretical biology*, 235(2):233–240, 2005.

[18] Arne Traulsen, Martin A Nowak, and Jorge M Pacheco. Stochastic payoff evaluation increases the temperature of selection. *Journal of theoretical biology*, 244(2):349–356, 2007.

[19] Benno Woelfing and Arne Traulsen. Stochastic sampling of interaction partners versus deterministic payoff assignment. *Journal of Theoretical Biology*, 257(4):689–695, 2009.

[20] Manfred Milinski. Tit for tat in sticklebacks and the evolution of cooperation. *nature*, 325(6103):433–435, 1987.

[21] Benjamin Kerr, Margaret A Riley, Marcus W Feldman, and Brendan JM Bohannan. Local dispersal promotes biodiversity in a real-life game of rock–paper–scissors. *Nature*, 418(6894):171–174, 2002.

[22] Gerald G Carter, Damien R Farine, Rachel J Crisp, Julia K Vrtilek, Simon P Ripperger, and Rachel A Page. Development of new food-sharing relationships in vampire bats. *Current Biology*, 30(7):1275–1279, 2020.

[23] Arne Traulsen and Nikoleta E. Glynats. The future of theoretical evolutionary game theory. *Philosophical Transactions B*, 2022.