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. The framework of evolutionary game theory has been applied in explaining the behavior of sticklebacks **?**, the rock-scissors-paper cycles in bacterial populations **?** and food sharing in vampire bats **?** colonies.

Traditional approaches of evolutionary game theory assume that individuals meet each other at random in an infinitely large well-mixed population; an example of such an approach is the replicator dynamics. The replicator dynamics describes how the abundance of strategic types in a population changes based on their fitness. In this deterministic formulation, individuals with higher fitness increase in abundance and ultimately, the system reaches a stable fixed point in which the population may consist either of a single type or of a mixture of different types. The works of **?????** have shown that constraining the population to be finite can lead to fundamental changes due to stochastic effects. Due to these stochastic effects disadvantageous mutants have a small, yet non-zero probability to reach fixation in a finite population.

Two classes of such 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 **?** *moran process*; (ii) *pairwise comparison processes* in which a pair of individuals is chosen, and where subsequently one of these individuals may adopt the strategy of the other **?**.

A pairwise comparison process considers a finite population of fixed size. In the population at a time step different types of individuals can exist. In the simplest of cases each individual can be of one of two types, A and B. The state of the population is thus characterized by the number k of individuals of type A. The interaction between the two types of individuals is described by the functions π^i_{A} and π^i_{B}.

At each step an individual change their type based. Two individuals, A and B, are selected at random. The individual chosen for reproduction A replaces B with probability ρ, which depends on the fitness difference

$\pi_{\mathrm{A}}^{i} - \pi_{\mathrm{B}}^{i}$ between the two individuals. The composition of the population can only change if both individuals are of different types. The probability ρ follows the Fermi function,

$$\rho(\pi_{\mathrm{A}}^{i}, \pi_{\mathrm{B}}^{i}) = \frac{1}{1+\exp^{-\beta(\pi_{\mathrm{A}}^{i}-\pi_{\mathrm{B}}^{i})}}. \tag{1}$$

The parameter β denotes the intensity of selection.

The fitness of the individuals $(\pi_{\mathrm{A}}^{i}, \pi_{\mathrm{B}}^{i})$ is identified with the payoff resulting from the underline game. The evolution of cooperation remains one of the greatest problems for biological and social sciences. Cooperation is the action of choosing to help others at one's own expense. The standard game of formulating such situation is the Prisoner's Dilemma **????**, in which two players can choose to cooperate or to defect. The players are offered a certain payoff, R, for mutual cooperation and a lower payoff, P, for mutual defection. If one player cooperates while the other defects, then the cooperator gets the lowest payoff, S, while the defector gains the highest payoff, T. Thus, the payoffs have the following property $T > R > P > S$, making defection the dominant strategy in the non-repeated game.

$$\begin{array}{cc} & \begin{array}{cc}\text{cooperate} & \text{defect}\end{array} \\ \begin{array}{c}\text{cooperate} \\ \text{defect}\end{array} & \begin{pmatrix} R & S \\ T & P \end{pmatrix} \end{array} \tag{2}$$

In literature, the payoffs of A and B individuals only on the fraction of both types in the population. If there are k A individuals and $N-k$ B individuals. In the context of the prisoner's dilemma the payoffs and assuming that A and B follow simple strategies where they cooperate and defect unconditionally respective, then the A and B individuals have fitness of

$$\begin{aligned} \pi_{\mathrm{A}}^{i} &= \frac{N-k-1}{N-1} \cdot R + \frac{k}{N-1} \cdot S, \\ \pi_{\mathrm{B}}^{i} &= \frac{N-k}{N-1} \cdot T + \frac{k-1}{N-1} \cdot P. \end{aligned} \tag{3}$$

respectively while self-interactions are excluded. In literature these are the expected payoffs. For $\beta = 1$, fitness equals payoff. This scenario describes "strong selection". For $\beta \ll 1$, the payoff only provides a small perturbation to the overall fitness of an individual, a limit known as weak selection.

In the evolutionary game theory literature it is usually assumed that players use strategies with finite memory **??**. This assumption is common as it allows for an explicit calculation of the players' payoffs **?**. Players' payoffs are calculated in the limit of interactions assuming that the two strategies meet several times.

Customarily, most evolutionary game studies make the additional assumption that individuals play many times and with all other players before reproduction takes place, so that payoffs, equivalently fitness, are given by the mean distribution of types in the population. This hypothesis, selection occurs much more slowly than the interaction between individuals and more that remember all the interactions they participated in. Although recent experimental studies show that this may not always be the case in biology [12– 14], it is clear that in cultural evolution or social learning the time scale of selection is much closer to the time scale of interaction. The effects of this mixing of scales cannot be disregarded [15], and then it is natural to ask about the consequences of the above assumption and the effect of relaxing it. The work of **?**.

In this work we ask studies that make use of this classical framework report that cooperation can substantially evolve. However, the inconsistency in the memory size of players leads us to question the robustness of our understanding of cooperation. 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 and the alternative scenario where individuals update their strategies only based on the very last payoff they obtained. 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. We extend our approach to the rest of the symmetric 2×2 games.

The remainder of the paper is organized as follows. In section **??** we describe the model. In section **??** we present the results of the simulations, and in section **??** we outline the main conclusions.

2 Model Setup

In the following, we consider a well mixed population of fixed size. In each step of t

We study the transmission of strategies with a frequency-dependent birth-death process26 in a finite population of size n. In each time step, two randomly chosen individuals compare their payoffs and one of them can switch to the other one's strategy. This process can be interpreted as a model for social learning, whereby successful strategies spread, and, occasionally, random strategy exploration introduces novel strategies (corresponding to mutations in biological models).

We consider a population of N players [1] where N is even and mutations are sufficiently rare. Therefore, at any point in time there are at most two different strategies present in the population; a *resident* strategy and a *mutant* strategy. To describe how strategies spread we use a pairwise comparison process **?**. Each step of the evolutionary process consists of two stages, a game stage and an updating stage.

In the game stage each individual is randomly matched with some other individual in the population. They engage in a match where each subsequent turn occurs with a fixed probability δ. At each turn players choose independently to either cooperate (C) or to defect (D), and the payoffs of the turn depend on both

[1]The terms "player" and "individual" are used interchangeably here.

their decisions. If both players cooperate they receive the reward payoff R, whereas if both defect they receive the punishment payoff P. If one cooperates but the other defects, the defector receives the temptation payoff T, whereas the cooperator receives the sucker's payoff S. We denote the feasible payoff of each turn as $\mathcal{U} = \{R, S, T, P\}$. We assume that individuals use *reactive strategies* to make decisions in each turn. Reactive strategies are a set of memory-one strategies that only take into account the previous action of the opponent. They can be written explicitly as a vector in $\mathbb{R}_3$, more specifically, a reactive strategy s is given by $s = (y, p, q)$. The parameter y is the probability that the strategy opens with a cooperation and p, q are the probabilities that the strategy cooperates given that the opponent cooperated and defected equivalently.

In the updating stage, two players are randomly drawn from the population, a 'learner' and a 'role model'. Given the learner's payoff $u_L \in \mathcal{U}$ and the role model's payoff $u_{RL} \in \mathcal{U}$, the learner adopts the role model's strategy with probability,

$$\rho(u_L, u_{RM}) = \frac{1}{1+\exp^{-\beta(u_{RM}-u_L)}}. \tag{4}$$

where $\beta \geq 0$ is the strength of selection. For small values of β the imitation probability is independent of the strategies of the involved players. As the value of β increases, the more likely it is that the learner adopts only strategies that yield a higher payoff. Conventionally the updating payoffs of the learner and the role model are based on their expected payoffs. A player's expected payoff is the mean payoff the player yields after engaging in matches of multiple turns with each member of the population. At each match a player bases their next turn decision only on the previous action of the opponent, however, the same player bases their expected payoffs on the outcomes of all their matches. Thus, a player is assumed to have limited and perfect memory at the same time. We propose a new a set of updating payoffs where it is also assumed that a player has limited memory. We referee to these as the limited memory payoffs.

The evolutionary step is repeated until either the mutant strategy goes extinct, or until it fixes in the population. If the mutant fixes in the population then the mutant strategy becomes the new resident strategy. After either outcome we introduce a new mutant strategy uniformly chosen from all reactive strategies at random, and we set the number of mutants to 1. This process of mutation and fixation/extinction is then iterated many times.

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 (i) the expected payoffs (ii) the limited memory payoffs. We compare the difference in the cooperation rate within the resident population for the two approaches. To account for the various types of social behaviour we also present results on multiple social dilemmas.