Prisoner's dilemma with stochastic payoffs

Christian Hilbe, Nikoleta E. Glynatsi

1 Introduction

Evolutionary game theory is a mathematical framework for modeling evolution in biological, social and economical systems. The success of individuals is not constant, but depends on the composition of the population. Individuals interact according to rules of the game and their payoff is translated into reproductive fitness.

Traditionally, evolutionary game studies work with idealized assumptions when calculating fitness. As stated in [1], it is commonly assumed 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. Initially, this implies that selection occurs much more slowly than the interaction between individuals, even experimental studies show that this may not always be the case in biology. Based on 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 work of [1] relaxed this assumption and showed that rapid selection affects evolutionary dynamics in such a dramatic way that for some games it even changes the stability of equilibria.

Furthermore, the second assumption implies that individuals have the time to interact with with all other players. Moreover, this implies that individuals when update their strategies over time, individuals are assumed to have perfect memory. However, when modeling how individuals make decisions in each round, these models assume that players only remember the last round [2–4]. The aim of this work is too relax this assumption, and to further the work of [1] by also considering strategies that interact to the past interactions.

In literature, when individuals update their strategies over time they interact with all other players so that the are given by the mean distribution of types in the population. These payoffs are referred to as the *expected* payoffs of individuals. We contrast this standard scenario with an alternative scenario where individuals update their strategies when remembering several of their past interactions but not all of them. We refer to these as the *stochastic* payoffs of individuals.

Evolutionary game dynamics have extensively been studied from mostly deterministic models based on rate equations to stochastic individual-based models. These more sophisticated models make use of different rules to determine how successful strategies spread. As stated in [5] there are two classes have been used extensively, *fitness-based processes* in which an individual chosen proportional to fitness reproduces and the offspring replaces a randomly chosen individual and *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. This adoption occurs with a probability that depends on the payoff of both individuals, such that better players are more likely to be imitated than those who do worse.

In this we work we follow a pairwise comparison processes. We consider a well mixed population of N individuals. In our model mutation happens rarely and such each individual can be of one of the two types, either of the *resident* or either of the *mutant*.

Regarding the game, we the important case of symmetric 2×2 games, in which the payoffs are given by the following matrix:

$$\begin{pmatrix} R & S \\ T & P \end{pmatrix} \tag{1}$$

These cases include the Snowdrift [6], Stag Hunt [7], Harmony [8] and the Prisoner's Dilemma games. These results are presented in Section 3.

Moreover, we will specifically consider the donation game where each player can cooperate by providing a benefit b to the other player at their cost c, with $0 < c < b$. Then, $T = b, R = b - c, S = -c, P = 0$, and matrix (**??**) is give by:

$$\begin{pmatrix} b - c & -c \\ b & 0 \end{pmatrix} \tag{2}$$

These results are presented in Section 2. The relative influence of the game is controlled by an external parameter, the so called intensity of selection β.

The space of strategies in repeated is infinite. To continue with our evolutionary analysis, we assume herein that individuals at most make use of simple *reactive strategies*. Reactive strategy are a set of memory-one strategies that only take into account the previous action of the opponent. An example of a reactive strategy is Tit For Tat. Reactive strategies can be written explicitly as a vector $\in \mathbb{R}_3$. More specifically a reactive strategy s is given by $s = (y, p, q)$ where 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 individuals of our population do not interact for a given number of turns, on the contrast, we consider that each match has a probability δ to continue following each turn.

2 Expected and stochastic payoffs in the donation game

Figure 1 shows simulation results for the above described process. Figure 1 depicts the evolving conditional cooperation probabilities p and q (assuming that the discount factor δ and the benefit b are comparably high). The left panel corresponds to the standard scenario considered in the previous literature. It considers players who use expected payoffs to update their strategies. The right 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 expected payoffs, players tend to be more generous (their q-values are higher on average). In addition, players are generally more cooperative.

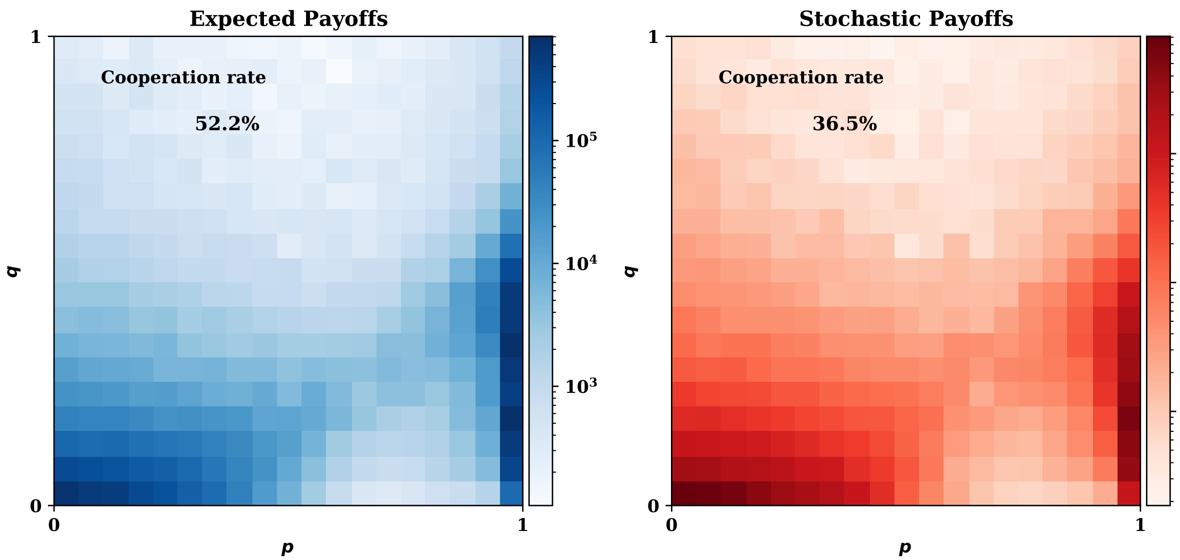

Figure 1: Evolutionary dynamics under expected payoffs and stochastic payoffs. We have run two simulations of the evolutionary process described in Section 4 for $T=10^7$ time steps. For each time step, we have recorded the current resident population (y, p, q). Since simulations are run for a relatively 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. (**A**) If players update based on their expected payoffs, the resident population typically applies a strategy for which $p \approx 1$ and $q \leq 1 - c/b = 0.9$. The cooperation rate within the resident population (averaged over all games and over all time steps) is close to 100%. (**B**) When players update their strategies based on their realized payoffs in the last round, there are two different predominant behaviors. The resident population either consists of defectors (with $p \approx q \approx 0$) or of conditional cooperators. In the latter case, the maximum level of q consistent with stable cooperation is somewhat smaller compared to the expected-payoff setting, $q < 0.5$. Also the resulting cooperation rate is smaller. On average, players cooperate roughly in half of all rounds. Parameters: $N = 100$, $b = 3$, $c = 1$, $\beta = 1$, $\delta = 0.999$.

The difference in the cooperation rates becomes more and more obvious as the temptation to defect increases, Figure 2.

We explored how the evolving cooperation rates change as we vary the benefit b and the selection strength β (Figure 3). In all cases, we find that the two scenarios yield similar cooperation rates when the respective

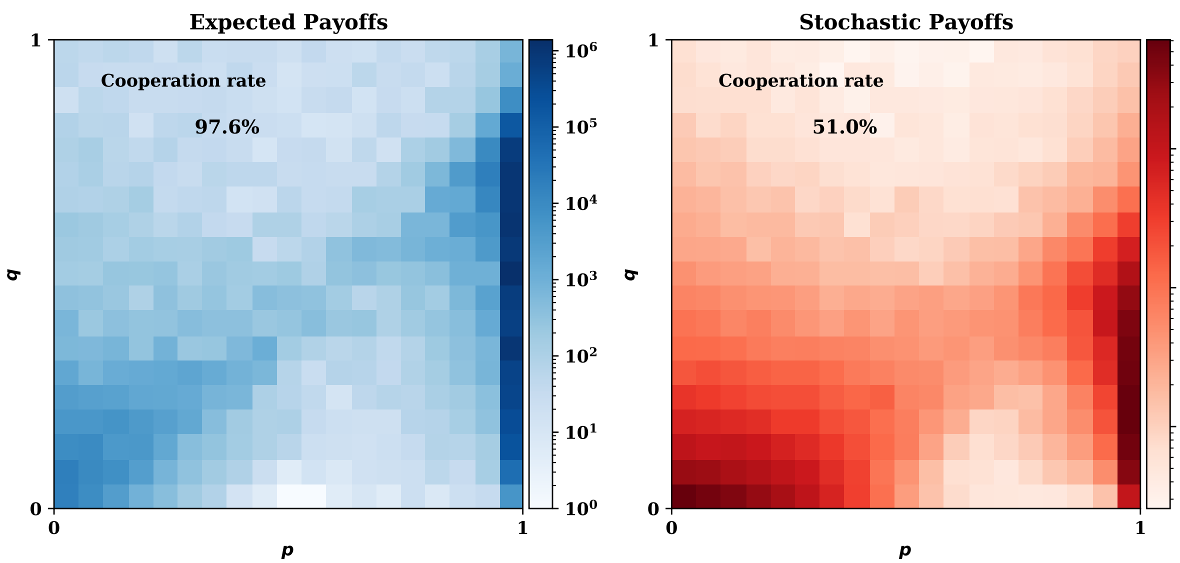

Figure 2: Evolutionary dynamics under expected payoffs and stochastic payoffs. Parameters: $N = 100$, $b = 10$, $c = 1$, $\beta = 1$, $\delta = 0.999$.

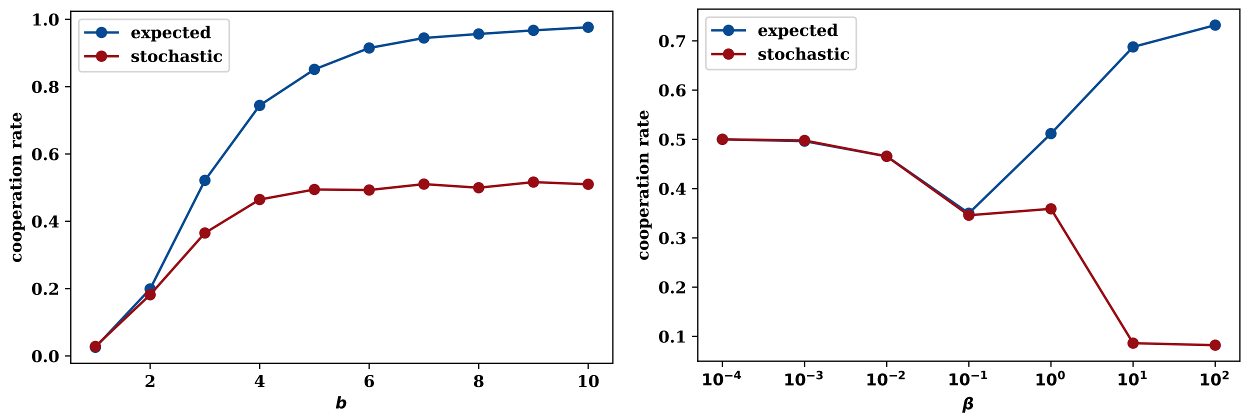

Figure 3: The evolution of cooperation for different parameter values. While the previous figures depict the evolutionary outcome for fixed parameter values, here we vary the benefit of cooperation b and the strength of selection β. In all cases, stochastic payoff evaluation tends to reduce the evolving 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 $T = 5 \times 10^6$ time steps for each parameter combination.

parameters are small. Once there is a high benefit to cooperation, strong selection, or a high expected number of rounds, updating based on expected payoffs yields higher cooperation rates.

3 Expected and stochastic payoffs in 2×2 games

4 Methods

In evolution context we consider a population of N players, where N is even, and where mutations are sufficiently rare. At any point in time the there are at most two different strategies are present in the population. Suppose there are $N - k$ players who use the strategy $s_1 = (y_1, p_1, q_1)$, whereas k players use the strategy $s_2 = (y_2, p_2, q_2)$. We refer to these two player types as 'residents' and 'mutants', respectively.

Each step of the evolutionary process consists of two stages, a game stage and an updating stage.

1. In the game stage, each player is randomly matched with some other player in the population to interact in one instance of the IPD.

2. In the updating stage, two players are randomly drawn from the population, a 'learner' and a 'exemplar'. Given that the learner's payoff in the last round is $u_L \in \mathcal{U}$ and that the exemplar's last round's payoff $u_E \in \mathcal{U}$, we assume the learner adopts the role model's strategy with probability

$$\rho(u_L, u_E) = \frac{1}{1+\exp\left[-\beta(u_E - u_L)\right]}. \tag{3}$$

where $\beta \geq 0$ corresponds to the strength of selection.

We iterate this basic evolutionary step until either the mutant strategy goes extinct, or until it fixes in the population (in which case the mutant strategy becomes the new resident strategy). After either outcome, we introduce a new mutant strategy $s'_2 = (y'_2, p'_2, q'_2)$ (uniformly chosen from all reactive strategies at random), and we set the number of mutants to $k = 1$. This process of mutation and fixation/extinction is then iterated many times.

We compare this process for what we defined as **stochastic payoff evaluation** with the analogous process where players update their strategies with respect to their **expected** payoffs,

$$\begin{aligned}\pi_1 &= \frac{N-k-1}{N-1} \cdot \langle \mathbf{v}(s_1, s_1), \mathbf{U} \rangle + \frac{k}{N-1} \cdot \langle \mathbf{v}(s_1, s_2), \mathbf{U} \rangle, \\ \pi_2 &= \frac{N-k}{N-1} \cdot \langle \mathbf{v}(s_2, s_1), \mathbf{U} \rangle + \frac{k-1}{N-1} \cdot \langle \mathbf{v}(s_2, s_2), \mathbf{U} \rangle.\end{aligned} \tag{4}$$

In the limit of no discounting, $\delta \to 1$, this process based on expected payoffs has been considered in [**?**].

We define **stochastic payoff** as the average payoff $u \in \mathcal{U}$ a player receives in the last n rounds of the game given that they interact with m players.

Case $n = m = 1$.

Initially, consider the situation where $n = m = 1$. The player's stochastic payoff is what they receive in the last round against a single opponent. There only four possible outcomes for the last round, those are CC, CD, DC, DD. Consider two players with reactive strategies $S_1 = (y_1, p_1, q_1)$ and $S_2 = (y_2, p_2, q_2)$ who interact in a repeated prisoner's dilemma with continuation probability δ, the probability that are in each of the four possible states in the last round is given by:

$$\mathbf{v}(s_1, s_2) = \Big(\mathbf{v}_R(s_1, s_2), \mathbf{v}_S(s_1, s_2), \mathbf{v}_T(s_1, s_2), \mathbf{v}_P(s_1, s_2)\Big). \tag{5}$$

where,

$$
\begin{aligned}
\mathbf{v}_R(S_1,S_2) &= (1-\delta)\frac{y_1 y_2}{1-\delta^2 r_1 r_2} + \delta \frac{\Big(q_1 + r_1\big((1-\delta)y_2 + \delta q_2\big)\Big)\Big(q_2 + r_2\big((1-\delta)y_1 + \delta q_1\big)\Big)}{(1-\delta r_1 r_2)(1-\delta^2 r_1 r_2)}, \\
\mathbf{v}_S(S_1,S_2) &= (1-\delta)\frac{y_1 \bar{y}_2}{1-\delta^2 r_1 r_2} + \delta \frac{\Big(q_1 + r_1\big((1-\delta)y_2 + \delta q_2\big)\Big)\Big(\bar{q}_2 + \bar{r}_2\big((1-\delta)y_1 + \delta p_1\big)\Big)}{(1-\delta r_1 r_2)(1-\delta^2 r_1 r_2)}, \\
\mathbf{v}_T(S_1,S_2) &= (1-\delta)\frac{\bar{y}_1 y_2}{1-\delta^2 r_1 r_2} + \delta \frac{\Big(\bar{q}_1 + \bar{r}_1\big((1-\delta)y_2 + \delta p_2\big)\Big)\Big(q_2 + r_2\big((1-\delta)y_1 + \delta q_1\big)\Big)}{(1-\delta r_1 r_2)(1-\delta^2 r_1 r_2)}, \\
\mathbf{v}_P(S_1,S_2) &= (1-\delta)\frac{\bar{y}_1 \bar{y}_2}{1-\delta^2 r_1 r_2} + \delta \frac{\Big(\bar{q}_1 + \bar{r}_1\big((1-\delta)y_2 + \delta p_2\big)\Big)\Big(\bar{q}_2 + \bar{r}_2\big((1-\delta)y_1 + \delta p_1\big)\Big)}{(1-\delta r_1 r_2)(1-\delta^2 r_1 r_2)}.
\end{aligned}
\tag{6}
$$

Proof. Assume a repeated prisoner's dilemma between two reactive strategies. Given the continuation probability δ, probability that the game ends in the after the first round $(1 - \delta)$ and the expected distribution of the four outcomes in the very first round is $\mathbf{v_0}$ defined as. Following the first round the, the outcome of the next rounds with a probability δ is M such that,

$$\ldots$$

It can shown that, $!(1-\delta)\mathbf{v_0}(I_4 - \delta M)^{-1}$ and with some algebraic manipulation we derive to Equation 6.

□

References

[1] Carlos P. Roca, José A. Cuesta, and Angel Sánchez. Time scales in evolutionary dynamics. *Phys. Rev. Lett.*, 97:158701, Oct 2006.

[2] 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.

[3] 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.

[4] Wenxing Ye, Weiying Feng, Chen Lü, and Suohai Fan. Memory-based prisoner's dilemma game with conditional selection on networks. *Applied Mathematics and Computation*, 307:31–37, 2017.

[5] Bin Wu, Benedikt Bauer, Tobias Galla, and Arne Traulsen. Fitness-based models and pairwise comparison models of evolutionary games are typically different—even in unstructured populations. *New Journal of Physics*, 17(2):023043, 2015.

[6] Robert Sugden et al. *The economics of rights, co-operation and welfare*. Springer, 2004.

[7] Brian Skyrms. The stag hunt. In *Proceedings and Addresses of the American Philosophical Association*, volume 75, pages 31–41. JSTOR, 2001.

[8] Amir N Licht. Games commissions play: 2x2 games of international securities regulation. *Yale J. Int'l L.*, 24:61, 1999.