

Prisoner's dilemma with stochastic payoffs

January 13, 2021

1 Games dynamics between reactive strategies

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 . We consider the vector

$$\mathbf{v}(S_1, S_2) = \left(v_R(S_1, S_2), v_S(S_1, S_2), v_T(S_1, S_2), v_P(S_1, S_2) \right) := (1-\delta)\mathbf{v}_0(I_4 - \delta M)^{-1}. \quad (1)$$

Here, \mathbf{v}_0 denotes the expected distribution of the four outcomes in the very first round, I_4 is the 4×4 identity matrix, and M is the transition matrix of the Markov chain. The entries of \mathbf{v} can be calculated explicitly,

$$\begin{aligned} v_R(S_1, S_2) &= (1-\delta) \frac{y_1 y_2}{1-\delta^2 r_1 r_2} + \delta \frac{\left(q_1 + r_1((1-\delta)y_2 + \delta q_2) \right) \left(q_2 + r_2((1-\delta)y_1 + \delta q_1) \right)}{(1-\delta r_1 r_2)(1-\delta^2 r_1 r_2)}, \\ v_S(S_1, S_2) &= (1-\delta) \frac{y_1 \bar{y}_2}{1-\delta^2 r_1 r_2} + \delta \frac{\left(q_1 + r_1((1-\delta)y_2 + \delta q_2) \right) \left(\bar{q}_2 + \bar{r}_2((1-\delta)y_1 + \delta p_1) \right)}{(1-\delta r_1 r_2)(1-\delta^2 r_1 r_2)}, \\ v_T(S_1, S_2) &= (1-\delta) \frac{\bar{y}_1 y_2}{1-\delta^2 r_1 r_2} + \delta \frac{\left(\bar{q}_1 + \bar{r}_1((1-\delta)y_2 + \delta p_2) \right) \left(q_2 + r_2((1-\delta)y_1 + \delta q_1) \right)}{(1-\delta r_1 r_2)(1-\delta^2 r_1 r_2)}, \\ v_P(S_1, S_2) &= (1-\delta) \frac{\bar{y}_1 \bar{y}_2}{1-\delta^2 r_1 r_2} + \delta \frac{\left(\bar{q}_1 + \bar{r}_1((1-\delta)y_2 + \delta p_2) \right) \left(\bar{q}_2 + \bar{r}_2((1-\delta)y_1 + \delta p_1) \right)}{(1-\delta r_1 r_2)(1-\delta^2 r_1 r_2)}. \end{aligned} \quad (2)$$

In these expressions, we have used the notation $r_i := p_i - q_i$, $\bar{y}_i = 1 - y_i$, $\bar{q}_i := 1 - q_i$, and $\bar{r}_i := \bar{p}_i - \bar{q}_i = -r_i$ for $i \in \{1, 2\}$. Let $\mathcal{U} = \{R, S, T, P\}$ denote the set of feasible payoffs in each round, and let $\mathbf{u} = (R, S, T, P)$ be the corresponding payoff vector. Then one can show the following result.

Proposition 1. *Consider a repeated prisoner's dilemma 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. (2).*

2 Pairwise imitation dynamics under stochastic payoff evaluation

2.1 Basic setup

In the following we consider a population of size N , where N is even. We assume mutations are sufficiently rare such that at any point in time 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. In the game stage, each player is randomly matched with some other player in the population to interact in one instance of the repeated prisoner’s dilemma. In the updating stage, two players are randomly drawn from the population, a ‘learner’ and a ‘role model’. Given that the learner’s payoff in the last round is $u_L \in \mathcal{U}$ and that the role model’s last round’s payoff $u_M \in \mathcal{U}$, we assume the learner adopts the role model’s strategy with probability

$$\rho(u_L, u_M) = \frac{1}{1 + \exp[-\beta(u_M - u_L)]}. \quad (3)$$

The parameter $\beta \geq 0$ 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 stochastic payoff evaluation with the analogous process where players update their strategies with respect to their *expected* payoffs,

$$\begin{aligned} \pi_1 &= \frac{N-k-1}{N-1} \cdot \langle \mathbf{v}(S_1, S_1), \mathbf{u} \rangle + \frac{k}{N-1} \cdot \langle \mathbf{v}(S_1, S_2), \mathbf{u} \rangle, \\ \pi_2 &= \frac{N-k}{N-1} \cdot \langle \mathbf{v}(S_2, S_1), \mathbf{u} \rangle + \frac{k-1}{N-1} \cdot \langle \mathbf{v}(S_2, S_2), \mathbf{u} \rangle. \end{aligned} \quad (4)$$

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

2.2 Fixation probabilities under stochastic payoff evaluation

Given that $N - k$ players use the resident strategy $S_1 = (y_1, p_1, q_1)$ and that the remaining k players use the mutant strategy $S_2 = (y_2, p_2, q_2)$, the probability that the number of mutants increases by one in one step of the evolutionary process can be written as

$$\lambda_k^+ = \frac{N-k}{N} \cdot \frac{k}{N} \cdot \sum_{u_1, u_2 \in \mathcal{U}} x(u_1, u_2) \cdot \rho(u_1, u_2). \quad (5)$$

In this expression, $(N-k)/N$ is the probability that the randomly chosen learner is a resident, and 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_1 and u_2 that the two player may have received in their respective last rounds. We use $x(u_1, u_2)$ to denote the probability that the randomly chosen resident obtained a payoff of u_1 in the last round of his respective game, and that the mutant obtained a payoff of u_2 . Given that the payoffs are u_1 and u_2 , the imitation probability is then given by $\rho(u_1, u_2)$, as specified by Eq. (3). The probability that the respective payoffs of the players are given by u_1 and u_2 can be calculated as

$$x(u_1, u_2) = \frac{1}{N-1} \cdot v_{u_1}(S_1, S_2) \cdot 1_{(u_1, u_2) \in \mathcal{U}_F^2} + \left(1 - \frac{1}{N-1}\right) \left[\frac{k-1}{N-2} \frac{k-2}{N-3} v_{u_1}(S_1, S_2) v_{u_2}(S_2, S_2) + \frac{k-1}{N-2} \frac{N-k-1}{N-3} v_{u_1}(S_1, S_2) v_{u_2}(S_2, S_1) \right. \\ \left. + \frac{N-k-1}{N-2} \frac{k-1}{N-3} v_{u_1}(S_1, S_1) v_{u_2}(S_2, S_2) + \frac{N-k-1}{N-2} \frac{N-k-2}{N-3} v_{u_1}(S_1, S_1) v_{u_2}(S_2, S_1) \right]. \quad (6)$$

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 $1/(N-1)$. In that case, we note that only those payoff pairs can occur that are feasible in a direct interaction, $(u_1, u_2) \in \mathcal{U}_F^2 := \{(R, R), (S, T), (T, S), (P, P)\}$, 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.

Analogously, we can calculate the probability that the number of mutants decreases by one in one step of the evolutionary process. This probability is

$$\lambda_k^- = \frac{N-k}{N} \cdot \frac{k}{N} \sum_{u_1, u_2 \in \mathcal{U}} x(u_1, u_2) \cdot \rho(u_2, u_1). \quad (7)$$

The fixation probability of the mutant strategy then takes the standard form [2],

$$\varphi = \frac{1}{1 + \sum_{i=1}^{N-1} \prod_k^i \frac{\lambda_k^-}{\lambda_k^+}}. \quad (8)$$

2.3 Invasion analysis of ALLD into GTFT

In the following, we apply the above formalism to calculate how easily a single ALLD mutant can invade into a resident population with strategy GTFT. In that case, $S_1 = (1, 1, q)$, $S_2 = (0, 0, 0)$, and $k = 1$. When two GTFT players interact in the game, their respective probabilities for each of the four outcomes in the last round simplify to

$$v_R(GTFT, GTFT) = 1, \quad v_S(GTFT, GTFT) = 0, \\ v_T(GTFT, GTFT) = 0, \quad v_P(GTFT, GTFT) = 0. \quad (9)$$

On the other hand, if an ALLD player interacts with a GTFT player, the respective probabilities according to Eq. (2) become

$$\begin{aligned} v_R(ALLD, GTFT) &= 0, & v_S(ALLD, GTFT) &= 0, \\ v_T(ALLD, GTFT) &= 1 - \delta + \delta q, & v_P(ALLD, GTFT) &= \delta(1 - q). \end{aligned} \quad (10)$$

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 ,

$$\begin{aligned} x(R, T) &= \frac{N-2}{N-1} \cdot (1 - \delta + \delta q), \\ x(R, P) &= \frac{N-2}{N-1} \cdot \delta(1 - q), \\ x(S, T) &= \frac{1}{N-1} \cdot (1 - \delta + \delta q), \\ x(P, P) &= \frac{1}{N-1} \cdot \delta(1 - q). \end{aligned} \quad (11)$$

$$x(u_1, u_2) = 0 \quad \text{for all other payoff pairs } (u_1, u_2).$$

As a consequence, we can calculate the ratio of transition probabilities as

$$\frac{\lambda_1^+}{\lambda_1^-} = \frac{\frac{N-2}{N-1} \cdot \left(\frac{1 - \delta + \delta q}{1 + \exp[-\beta(T-R)]} + \frac{\delta(1-q)}{1 + \exp[-\beta(P-R)]} \right) + \frac{1}{N-1} \left(\frac{1 - \delta + \delta q}{1 + \exp[-\beta(T-S)]} + \frac{\delta(1-q)}{2} \right)}{\frac{N-2}{N-1} \cdot \left(\frac{1 - \delta + \delta q}{1 + \exp[-\beta(R-T)]} + \frac{\delta(1-q)}{1 + \exp[-\beta(R-P)]} \right) + \frac{1}{N-1} \left(\frac{1 - \delta + \delta q}{1 + \exp[-\beta(S-T)]} + \frac{\delta(1-q)}{2} \right)}. \quad (12)$$

In particular, in the limit of strong selection $\beta \rightarrow \infty$ and large populations $N \rightarrow \infty$, we obtain

$$\frac{\lambda_1^+}{\lambda_1^-} = \frac{1 - \delta + \delta q}{\delta(1 - q)}. \quad (13)$$

This ratio is smaller than 1 (such that ALLD is disfavored to invade) if $q < 1 - 1/(2\delta)$. For infinitely repeated games, $\delta \rightarrow 1$, this condition becomes $q < 1/2$ (for $q = 1/2$, the payoff of the ALLD player is $T > R$ 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).

2.4 Simulation results

Fig. 1– Fig. 3 show simulation results for the above described process. **Fig. 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 (4) to update their strategies. The right panel shows the scenario considered herein,

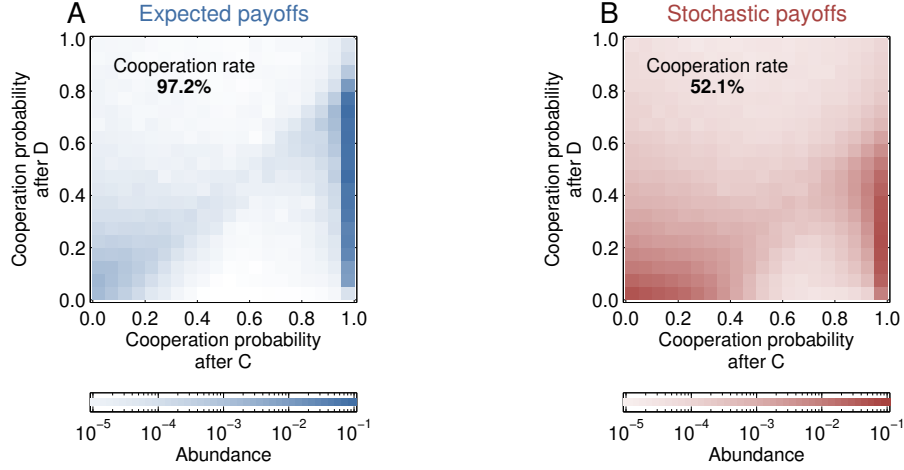


Figure 1: Evolutionary dynamics under expected payoffs (left) and stochastic payoffs (right). We have run two simulations of the evolutionary process described in Section 2.1 for $T = 10^7$ time steps. For each time step, we have recorded the current resident population (y, p, q) . Since simulations 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: $N = 100$, $b = 3$, $c = 1$, $\beta = 1$, $\delta = 0.999$.

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.

To obtain some intuition for this result, we have recorded which mutant strategies typically invade into an ALLD population, and which mutants invade into a GTFT population, for each of the two scenarios (**Fig. 2**). Compared to expected-payoff updating, the stochastic scenario drastically reduces the stability of GTFT. On average, it takes fewer mutants until a GTFT population is successfully invaded. Moreover, successful mutants tend to be less cooperative.

Finally, we have also explored how the evolving cooperation rates change as we vary the benefit b , the selection strength β , and the discount factor δ (**Fig. 3**). In all three cases, we find that the two scenarios yield similar cooperation rates when the respective 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.

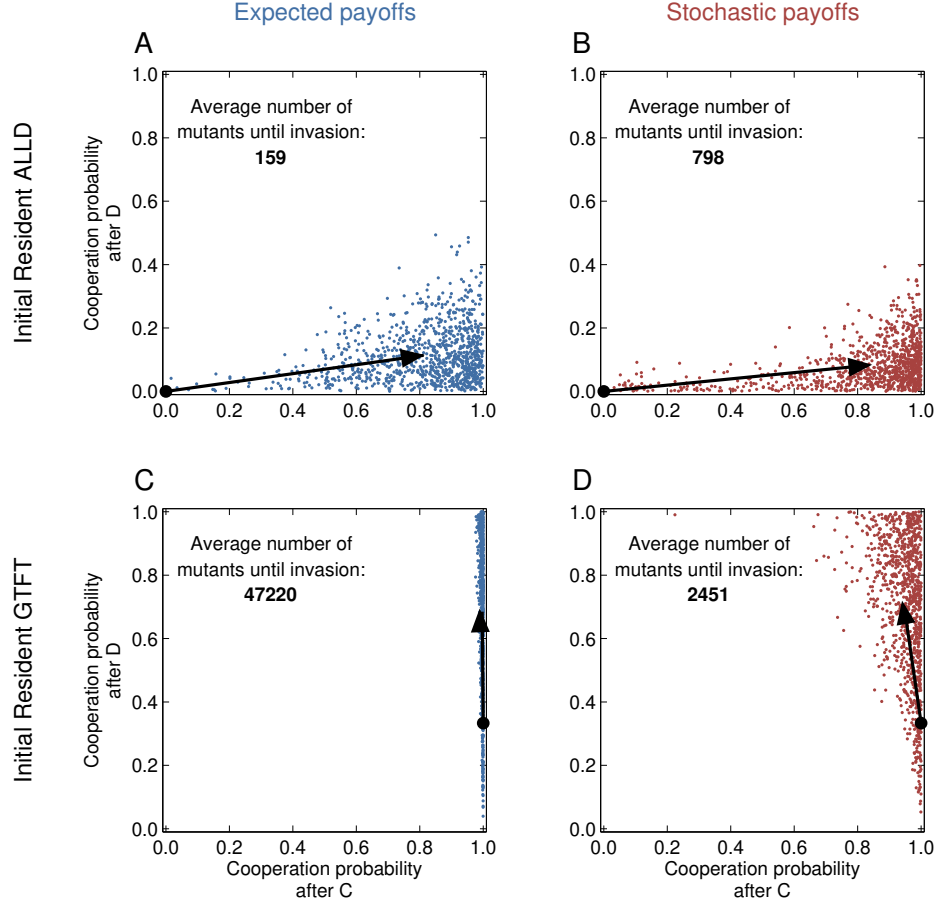


Figure 2: Invasion into ALLD and into GTFT for the two evolutionary scenarios. For this figure, we either consider a resident population of ALLD players (top) or of GTFT players (bottom). In each case, we have run 1,000 independent simulations of the evolutionary process described in Section 2.1. The colored dots within the four panels depict which mutant strategies successfully invaded into the respective resident population. The black arrow indicates the average over all successful mutant strategies. In addition, we have recorded how many mutant strategies need to be introduced into the population until the resident becomes replaced. Stochastic payoffs tend to increase the time until ALLD is invaded, and they reduce the time until GTFT is invaded. Parameters are the same as in **Fig. 1**. Strategies are defined as $ALLD = (0, 0, 0)$ and $GTFT = (1, 1, 1/3)$.

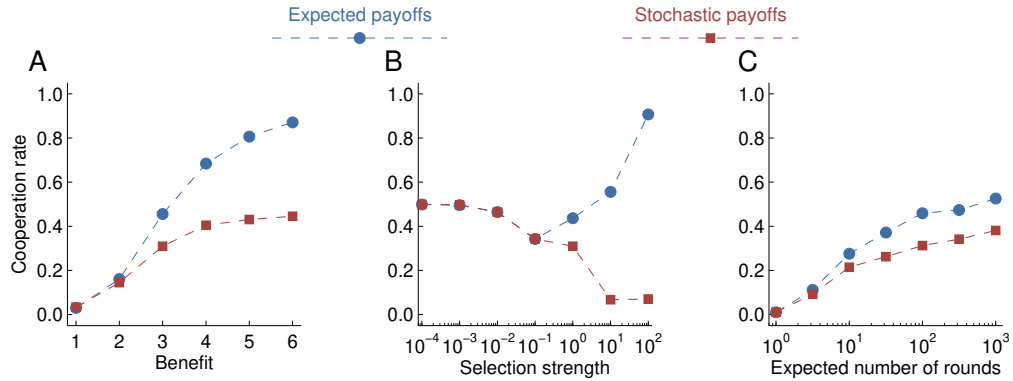


Figure 3: The evolution of cooperation for different parameter values. While the previous figures depict the evolutionary outcome for fixed parameter values, here we vary the benefit of cooperation b , the strength of selection β , and the expected number of rounds, $1/(1-\delta)$. In all cases, stochastic 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.

References

- [1] L. A. Imhof and M. A. Nowak. Stochastic evolutionary dynamics of direct reciprocity. *Proceedings of the Royal Society B*, 277:463–468, 2010.
- [2] M. A. Nowak, A. Sasaki, C. Taylor, and D. Fudenberg. Emergence of cooperation and evolutionary stability in finite populations. *Nature*, 428:646–650, 2004.