Altruistic Prisoners: The Role of Altruism and Discounting in Sustaining Cooperation

Nickolas Geyfman
Undergraduate Researcher, Rutgers University
nickolas.geyfman@rutgers.edu

Abstract

This paper investigates how *altruism* and *discount factors* jointly affect cooperation in a repeated Prisoner's Dilemma (PD) environment using Q-learning agents with a three-step memory. In my model, an agent's effective reward is computed as a weighted combination of its own payoff and its opponent's payoff. I perform a two-dimensional parameter sweep over the altruism parameter, α (ranging from 0.0 to 0.8 in 0.05 increments), and the discount factor, γ (ranging from 0.0 to 0.95 in 0.05 increments). Each (α, γ) pair is evaluated over 50 seeds in a population of 10 agents, using 2000 training episodes, a continuation probability of 0.99, a minimum of 10 rounds per episode, and ring pairing mode.

1 Introduction

The classical Prisoner's Dilemma (PD) demonstrates why rational, self-interested agents might defect even though mutual cooperation would yield a higher collective payoff. In a one-shot PD, mutual defection is the unique Nash equilibrium. However, repeated interactions allow the possibility of cooperation, particularly when agents take into account both future payoffs and the welfare of their partners.

In this work, I modify the standard PD by incorporating an altruism parameter into the reward function, so that each agent i receives an effective reward:

$$r_i = (1 - \alpha_i)p_i + \alpha_i p_j,$$

where p_i is the agent's own payoff and p_j is that of its partner. By systematically varying α and the discount factor γ, I generate a heatmap that captures how these parameters jointly affect the emergence of cooperation.

2 Model and Methodology

2.1 Game Setup

I consider a population of $N = 10$ agents. In each round, agents are randomly paired to play a standard PD with payoffs: Temptation $T = 5$, Reward $R = 3$, Punishment $P = 1$, and Sucker $S = 0$. The effective reward is given by:

$$r_i = (1 - \alpha_i)p_i + \alpha_i p_j,$$

with $\alpha_i \in [0, 1]$ representing the degree of altruism.

2.2 Q-learning Agents with Three-Step Memory

Agents employ a tabular Q-learning algorithm augmented with a three-step memory, meaning that they track the outcomes of the previous three rounds. This extended memory allows them to develop more nuanced strategies in the repeated PD. Agents follow an ϵ-greedy policy with $\epsilon = 0.1$ during training, switching to a greedy policy during evaluation. The Q-values are updated as:

$$Q(s,a) \leftarrow Q(s,a) + \alpha_{\text{learn}}\Big[r_i + \gamma \max_{a'} Q(s',a') - Q(s,a)\Big],$$

with a learning rate $\alpha_{\text{learn}} = 0.1$. Here, the state s encodes the outcomes of the last three rounds.

2.3 Training and Evaluation

Training: Agents interact for 2000 episodes, updating their Q-values.

Evaluation: After training, agents use a frozen Q-table (with $\epsilon = 0$) over 10 evaluation episodes. The final cooperation rate is computed as the fraction of cooperative moves during these episodes.

2.4 Parameter Sweep

I systematically vary:

- **Altruism (α):** from 0.0 to 0.8 in increments of 0.05.
- **Discount Factor (γ):** from 0.0 to 0.95 in increments of 0.05.

For each (α, γ) pair, 50 seeds are run in parallel. The experiments use a continuation probability of 0.99, a minimum of 10 rounds per episode, and ring pairing mode.

3 Experimental Results

My experiments reveal the following trends:

- Purely selfish agents ($\alpha = 0.0$) exhibit minimal cooperation at low γ, with cooperation gradually increasing at higher γ.
- Moderate altruism ($\alpha \approx 0.2$ to 0.4) substantially boosts cooperation, especially when agents discount future payoffs lightly.
- High altruism ($\alpha \geq 0.6$) leads to near-total cooperation across all discount factors.

Instead of providing full numerical matrices, I summarize the results with the heatmap in Figure 1.

3.1 Heatmap

4 Discussion

My findings underscore the synergy between altruism and a long-term outlook:

- When $\alpha = 0.0$, cooperation is minimal except under very high discount factors.

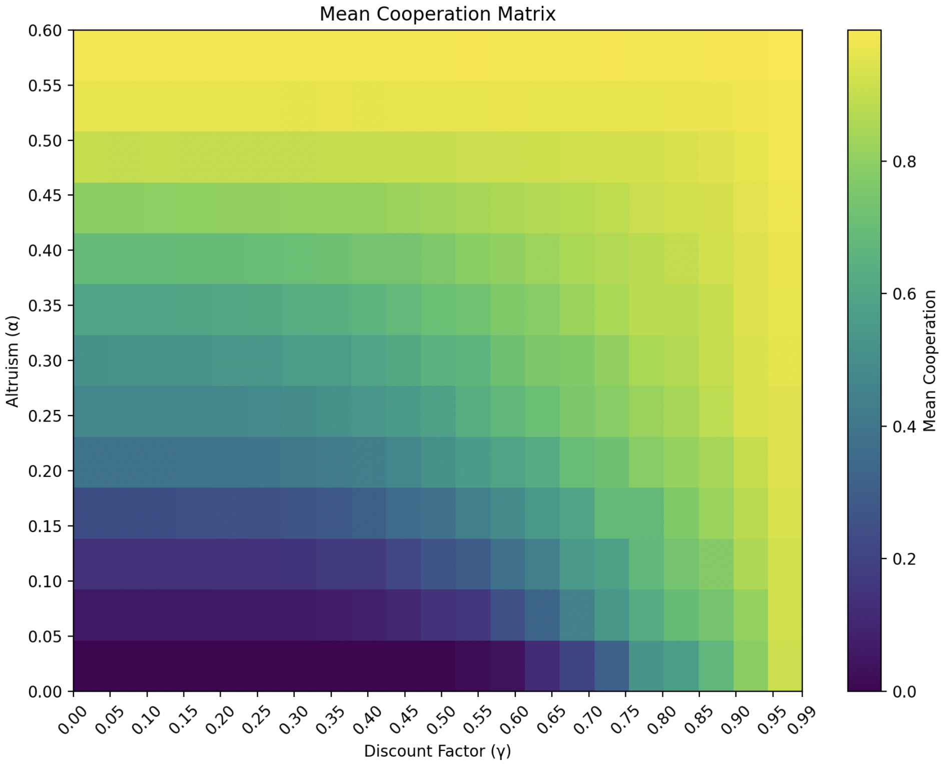

Figure 1: Heatmap of final cooperation rates as a function of α (vertical axis) and γ (horizontal axis). Lighter colors indicate higher cooperation.

- Moderate levels of altruism ($\alpha \approx 0.2$ to 0.4) lead to substantial improvements in cooperation, particularly when agents are future-oriented.
- High altruism nearly guarantees full cooperation regardless of the discount factor.

These results align with predictions from repeated-game theory and related multi-agent reinforcement learning studies [1, 2, 3]. The use of a three-step memory allows agents to leverage richer historical information to sustain cooperation.

5 Conclusions

I have demonstrated that incorporating altruism into Q-learning agents with a three-step memory dramatically enhances cooperation in the repeated Prisoner's Dilemma. The parameter sweep indicates that while purely selfish agents remain largely uncooperative, even moderate altruism combined with a strong future orientation can shift the equilibrium toward full cooperation.

References

References

[1] R. Axelrod, *The Evolution of Cooperation*, Basic Books, 1984.

[2] L. Busoniu, R. Babuška, and B. De Schutter, "A Comprehensive Survey of Multiagent Reinforcement Learning," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 38, no. 2, pp. 156–172, 2008.

[3] Y. Shoham and K. Leyton-Brown, *Multiagent Systems: Algorithmic, Game-Theoretic, and Logical Foundations*, Cambridge University Press, 2009.