An evolutionary approach to Multi Agent Reinforcement Learning,

Learning in Social Dilemmas

B. Mintz, F. Fu (Dartmouth College)

ICIAM, August 2023

Table of Contents

1. Earlier work and Model
2. Results
3. Conclusion

What is R.L.?

Reinforcement Learning is a Machine Learning paradigm inspired by animal psychology that consists of an agent learning the optimal action depending on the state of an environment. The core idea is simple: actions with beneficial consequences will be repeated more.

Multi-Agent R.L.

Most work on R.L. focuses on a single agent learning some task. However in many applications there are multiple individuals, potentially with different objectives.

This further complicates R.L. as the number of agents increases. In particular, the environment becomes less stable, and coordination among agents becomes more difficult, [GD22].

Multi-Agent R.L.

Most work on R.L. focuses on a single agent learning some task. However in many applications there are multiple individuals, potentially with different objectives.

This further complicates R.L. as the number of agents increases. In particular, the environment becomes less stable, and coordination among agents becomes more difficult, [GD22].

The problem we investigate in this work is **Incentive Alignment**: how agents with competing interests learn to work together.

R.L. model: Q-Learning

This is a form of temporal differencing, where agents try to learn the value of each action in a given state by keeping a table of approximate values.

Actions are chosen randomly with an exploration rate / temperature τ, then the values are updated as

$$Q(s_t, a_t) = (1-\alpha)Q(s_t, a_t) + \alpha \left[r_{t-1} + \gamma \max_a Q(s_{t+1}, a) \right]$$

R.L. model: Q-Learning

This is a form of temporal differencing, where agents try to learn the value of each action in a given state by keeping a table of approximate values.

Actions are chosen randomly with an exploration rate / temperature τ, then the values are updated as

$$Q(s_t, a_t) = (1-\alpha)Q(s_t, a_t) + \alpha\left[r_{t-1} + \gamma \max_a Q(s_{t+1}, a)\right]$$

This method has been proven to converge to the optimal policy given "sufficient" updates for each (state, action) pair, and decreasing learning rate, [WD92]. Further, it is shown to have bounded regret, that is, the difference between the actions it chooses and the optimal sequence is bounded, [LP22].

Objective Function 1: Symmetric Two Player, Two Action Games

Each player chooses an action, P or Q, and receives a payoff given by the following table.

Payoffs	P	Q
P	a	b
Q	c	d

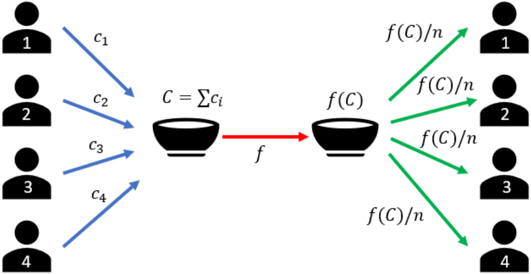
e.g.

	C	D
C	3	1
D	4	2

Of particular interest is a canonical model of conflicting interests, the **Prisoner's Dilemma**, where $b < d < a < c$.

Objective Function 2: Public Goods Games

Each of N individuals each contribute some amount c_i to a pot, which is scaled by some function $f(x)$ then distributed evenly among the players.



Another social dilemma, the **Free-Rider problem**: individuals benefit from contributing less, but this hurts the collective.

R.L. in these games

PD: Learning need not converge to Nash Equilibria, there are many other possibilities depending on the exploration rate, though in these games the NE are attracting for low exploration rates, [KT11, KG12]. It is also possible to observe cyclic dynamics.

PGG: Used to study / explain human behavioral data, multiple models have been proposed, to varying levels of success, [AL04, Cot15, IIO^{+}03]. Also some theoretical work, e.g. [NP15, LM19].

Our extension

Groups often change or consist of agents with different experience levels. In this work, we expand the MARL framework to account for this, studying how the parameters effect pro-social behavior.

Inspired by Evolutionary Game Theory, we investigate population dynamics where agents reproduce and die.

Our extension

Groups often change or consist of agents with different experience levels. In this work, we expand the MARL framework to account for this, studying how the parameters effect pro-social behavior.

Inspired by Evolutionary Game Theory, we investigate population dynamics where agents reproduce and die.

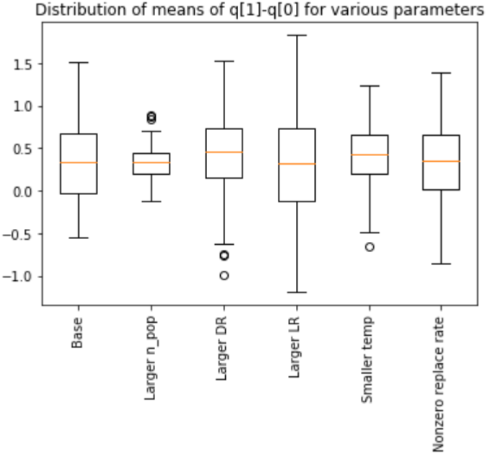
Our model consists of a population of N agents following reinforcement learning: Q-learning with parameters α, γ, and τ. Each time step, a group of k individuals is selected to interact, and receives payoffs from objective function 1 or 2 based on the set of actions they choose. Then with some probability r, one individual is replaced inversely proportional to their payoff.

Table of Contents

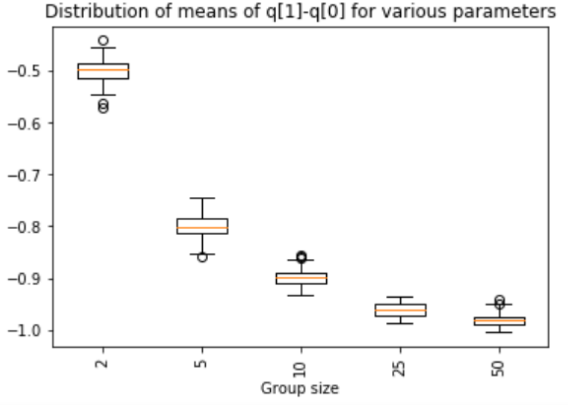
1. Earlier work and Model
2. Results
3. Conclusion

Experiments (PD)

Our model has a large number of parameters that can be adjusted independently. However, most have little effect on the results.

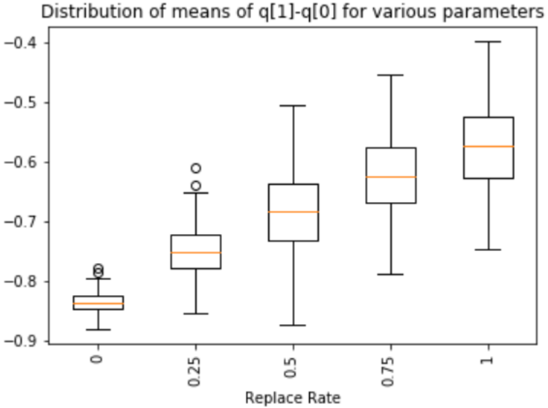


Group size (PGG)



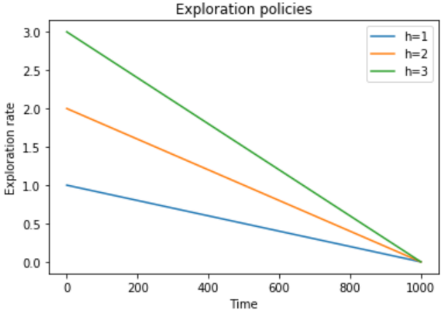
Smaller groups are more cooperative, consistent with the Free-Rider effect. This doesn’t apply to the PD.

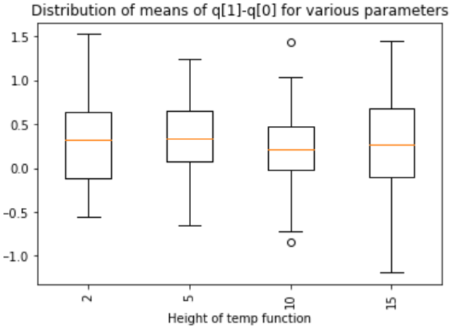
Replacement rate (PGG)



Replacement helps groups contribute more readily. It has little effect in the Prisoner's Dilemma.

Exploration Policies (PD)

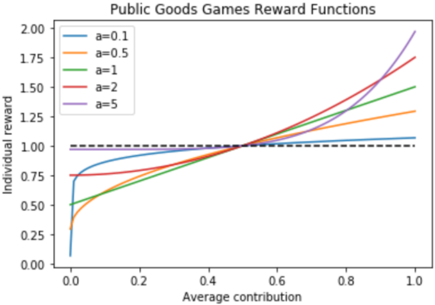


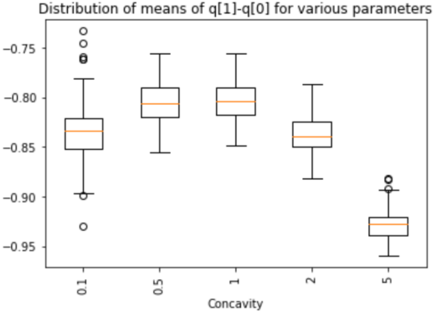


Surprisingly, the exploration policy seems to have little effect on the outcome, for both dilemmas, and for contant temperatures.

Game parameters (PGG)

Here $f(C)/N = C^a + (1 - t^a)$.





Contribution levels are optimal for an intermediate level of concavity in the reward function. The threshold had little effect.

Table of Contents

1 Earlier work and Model

2 Results

3 Conclusion

Summary

- Reinforcement Learning is a powerful Machine Learning framework, but is not sufficiently understood, especially when Multiple Agents are learning simultaneously.
- We extend the MARL framework to include dynamic groups, similar to the transition from Classical to Evolutionary Game Theory.
- Most model parameters have little effect on the dynamics.
- However, cooperation is general improved by smaller groups, and occasionally improved by larger replacement rates.
- Additionally, linear scaling functions are approximately the most effective in promoting contribution to the public goods.

Next Steps

- Allow for higher replacement rates, more than one individual per time step.
- Add a number of rounds, or continuation probability, parameter to tune group stability.
- Perform further evolutionary analysis: allow mutation/inheritance or competition between two types in a finite population.
- Incorporate states into the model (this is a bit subtle, since the groups change frequently, making the environment unstable).
- Use more sophisticated learning methods, such as Frequency-Adjusted Q-learning or WoLF. Could also extend to other Machine Learning techniques such as Neural Networks.

Thank you all for coming to my talk.

Are there any questions?

My website above has these slides.

References I

- Jasmina Arifovic and John Ledyard, *Scaling up learning models in public good games*, Journal of Public Economic Theory **6** (2004), no. 2, 203–238.

- Chenna Reddy Cotla, *Learning in repeated public goods games-a meta analysis*, Available at SSRN 3241779 (2015).

- Sven Gronauer and Klaus Diepold, *Multi-agent deep reinforcement learning: a survey*, Artificial Intelligence Review (2022), 1–49.

- Atsushi Iwasaki, Shuichi Imura, Sobei H Oda, Itsuo Hatono, and Kanji Ueda, *Does reinforcement learning simulate threshold public goods games?: a comparison with subject experiments*, IEICE TRANSACTIONS on Information and Systems **86** (2003), no. 8, 1335–1343.

- Ardeshir Kianercy and Aram Galstyan, *Dynamics of boltzmann q learning in two-player two-action games*, Physical Review E **85** (2012), no. 4, 041145.

References II

Michael Kaisers and Karl Tuyls, *Faq-learning in matrix games: Demonstrating convergence near nash equilibria, and bifurcation of attractors in the battle of sexes.*, Interactive Decision Theory and Game Theory, 2011.

Olof Leimar and John M McNamara, *Learning leads to bounded rationality and the evolution of cognitive bias in public goods games*, Scientific reports **9** (2019), no. 1, 16319.

Stefanos Leonardos and Georgios Piliouras, *Exploration-exploitation in multi-agent learning: Catastrophe theory meets game theory*, Artificial Intelligence **304** (2022), 103653.

Heinrich H Nax and Matjaž Perc, *Directional learning and the provisioning of public goods*, Scientific reports **5** (2015), no. 1, 1–6.

Christopher JCH Watkins and Peter Dayan, *Q-learning*, Machine learning **8** (1992), 279–292.