Beyond Selfishness: Modeling Cooperative AI in Social Dilemmas

Khanh Do[1], Joyce Gill[1], Nicole Moreno Gonzalez[1]

Department of Computer Science, Grinnell College [1]

ABSTRACT

Classical Reinforcement Learning (RL) like Q-learning struggles in multi-agent social dilemmas (MASD) where agents must adapt to each other's changing strategies [1].

CHALLENGES:

• Unstable learning caused by simultaneous policy updates.

• Difficulty achieving cooperation and often settle on sub-optimal Nash-Equilibrium (NE) [2].

NEED:

• Adaptive algorithms that can respond to multi-agent behavior and reach Pareto-Efficient (PE) [2].

• Practical validation of cooperative algorithms in realistic simulations.

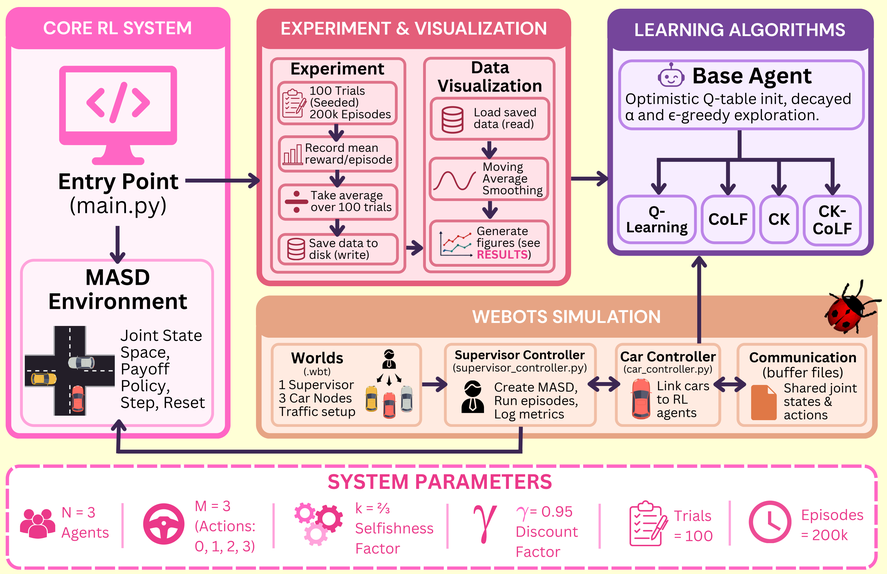
OBJECTIVE:

Reproduce results from the paper *Learning to Cooperate in Multi-Agent Social Dilemmas* and evaluate two Q-learning variants: **Change or Learn Fast (CoLF)** and **Change and Keep (CK)**.

Extend the model to a traffic dilemma setting.

METHODS

Q-Learning	CoLF	CK	CK-CoLF
• Constant learning rate • Assumes non-changing environment	Adjusts learning rates: • Slows down when other agents are unpredictable • Speeds up when stable	Repeat new move so that other agents have time to react before results are recorded	• Variable learning rates (CoLF) • 2-state machine (CK)



RESULTS

CoLF speeds up convergence rate

The Q-Learning baseline performs poorly overall with higher learning rates stabilizing at worse outcomes. However, lower learning rates seem to yield better final payoffs.

CoLF improves the overall payoff outcomes of standard Q-Learning, but it converges much more slowly and does not reliably reach Pareto-Efficient solutions.

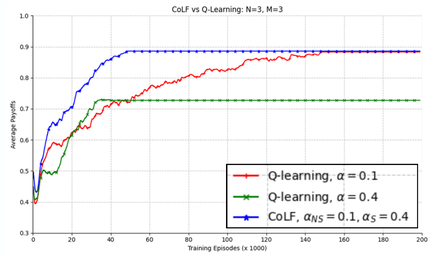


Figure 1: Comparing baseline Q-Learning with CoLF

CK achieves Pareto Efficiency

CK provides a clear improvement over standard Q-learning across variable learning rates, reaching higher payoffs more robustly with lower learning rate.

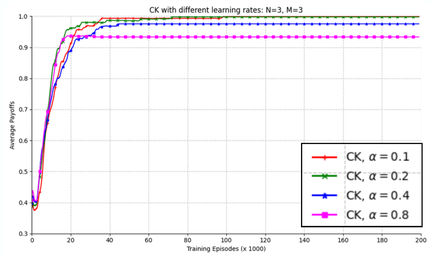


Figure 2: Comparing CK across different learning rates

Hybrid model gets the best of both worlds

CK-CoLF combines the variable learning rates of CoLF and deferred-update mechanism of CK to provide the fastest and most reliable method to achieve PE.

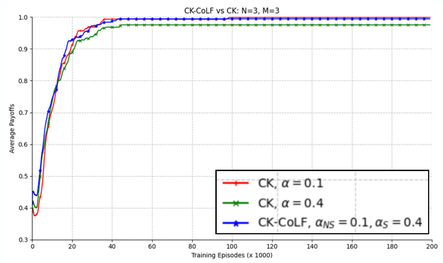


Figure 3: Comparing the hybrid model with CK

Traffic Dilemma Extension

- Preliminary Webots simulation of self-driving in a Traffic Social Dilemma: Uncontrolled 4-way Intersection.
- The CK-Colf world is shown below in three common scenarios at early episodes.

Crash

Figure 4: 3 CK-Colf agents enacting behavior that leads to a 'Crash' outcome. Car agents failed to adjust for speed reduction .

Learning Adjustment

Figure 5: 3 CK-Colf agents starting to learn Speed regulation. The red agent is slower compared to the other agents.

Successfull Crossing

Figure 6: 3 CK-Colf agents chose an efficient adjustment. This led to a 'Successful/ No crash' outcome.

CONCLUSION & FUTURE WORK

Our results show that cooperation is harder when agents only learn from immediate rewards. The best-performing agents reacted to change without becoming unstable. Cooperative AI therefore needs both flexibility and consistency when multiple agents are learning at the same time.

- Make our current Webots simulation fully functional so it can run up to 200k episodes.
- Include SOTA algorithms in the simulation to compare their performance against our current agents.

PROFESSIONAL LINKAGE

Khanh: Learned the skill to do research, read academic papers, build and automate multi-agent systems.

Joyce: I'm starting my PhD at Stanford's EduNLP Lab this fall, so learning about RL was helpful.

Nicole: I'm joining an ITSM team this summer. What I learned building agents will help me automate workflows.

REFERENCES

[1] Muñoz de Cote, et al. (2006). *Learning to Cooperate in Multi-Agent Social Dilemmas.* doi.org/10.1145/1160633.1160770

[2] Banerjee & Sen. (2007). *Reaching pareto-optimality in prisoner's dilemma using conditional joint action learning.* doi.org/10.1016/B978-0-323-91698-1.00014-5

ACKNOWLEDGEMENTS

Thank you to Professor Fernanda Eliott for your guidance, support, and supervision of the project. Thank you to our classmates for providing us with helpful feedback on our project progress presentations.

This project was completed as part of Professor Fernanda Eliott's CSC 261: Artificial Intelligence (Spring 2026).