Introduction

Optimizing Decision-Making in Multi-Agent RL with CPT

- Investigating Multi-Agent Reinforcement Learning (MARL) under Cumulative Prospect Theory (CPT)
- Key motivation: **Aligning autonomous agents with human decision-making biases**
- Key Questions:
    - Do CPT trained agents work follow their utility and probability distortion functions?
    - How do CPT-guided agents optimize strategies in multi-agent games, and how do their behaviors differ from those using traditional utility functions?
    - To what extent do agents adapt their strategies based on the utility functions of counterparties? What emergent dynamics arise in mixed populations of agents?

Background on CPT - Prospect Theory

Developed by Daniel Kahneman and Amos Tversky in 1979. Explains how people make decisions when faced with risk and uncertainty:

- People tend to avoid losses over acquiring equivalent gains (loss aversion).
- People evaluate choices based on relative differences rather than absolute similarities.
- People think in terms of expected utility relative to a reference point.

Background on CPT - Development of CPT

CPT provides a more robust framework for dealing with outcomes that have multiple possible probabilities, avoiding ranking issues and ensuring consistency in decision-making processes through nonlinear functions.

- Probability weighting function - captures the empirical observation that people tend to overweight small probabilities and underweight large probabilities
- Value function - concave for gains and convex for losses

Implementation Strategy

Technical Approach & PyTorch Implementation

- **Policy Gradient Optimization with CPT**
  - CPT-adjusted rewards & probability distortions
  - Model-free learning using policy gradients
- **Implementation Workflow**:
  1. Design neural network for policy representation
  2. Transform rewards using CPT functions
  3. Compute policy gradients using automatic differentiation
  4. Optimize policies using gradient ascent
- **Evaluation**:
  - Multi-agent simulations (PettingZoo, Gym)

Mathematical Formulation

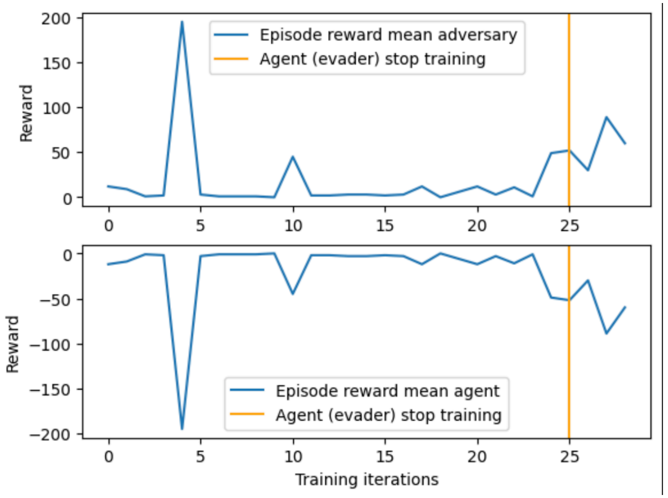
Competitive Environment - Overview

PettingZoo's **Simple Tag** Environment is a basic Multi-Agent Particle Environment (MPE) designed for competition between agents

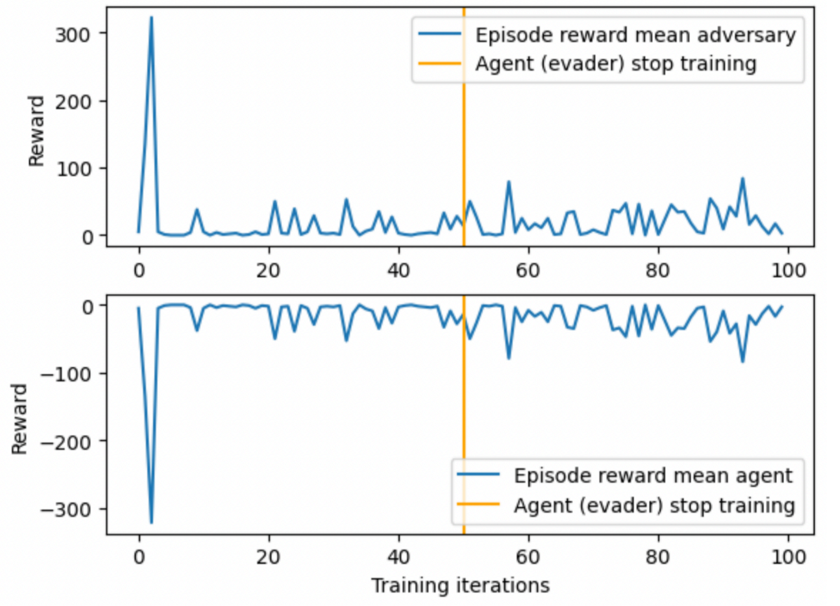
- **Objective**: Predators work to “tag” or catch the prey, while the prey’s goal is to evade capture.
- **Rewards**: Rewards are structured so that predators gain rewards when they successfully tag the prey, and the prey receives a penalty when caught.

Competitive Environment - Rewards

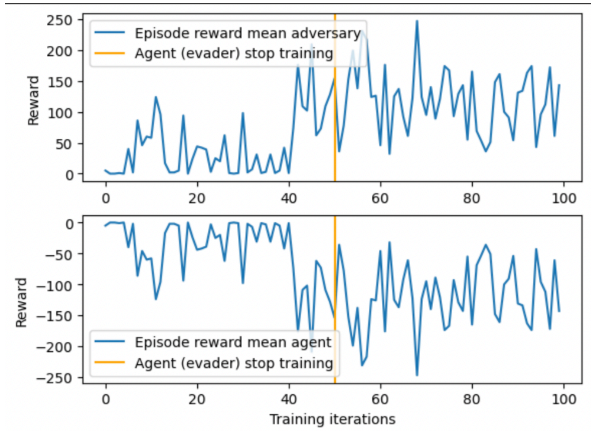
Baseline



Moderate CPT (Risk Seeking)

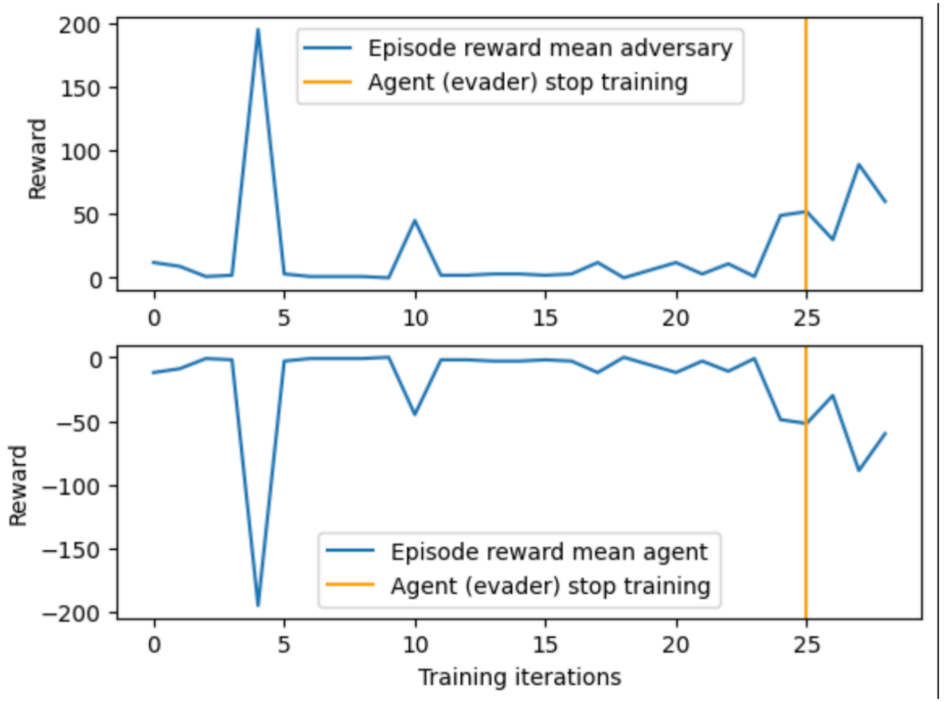


Extreme CPT (Risk Averse)

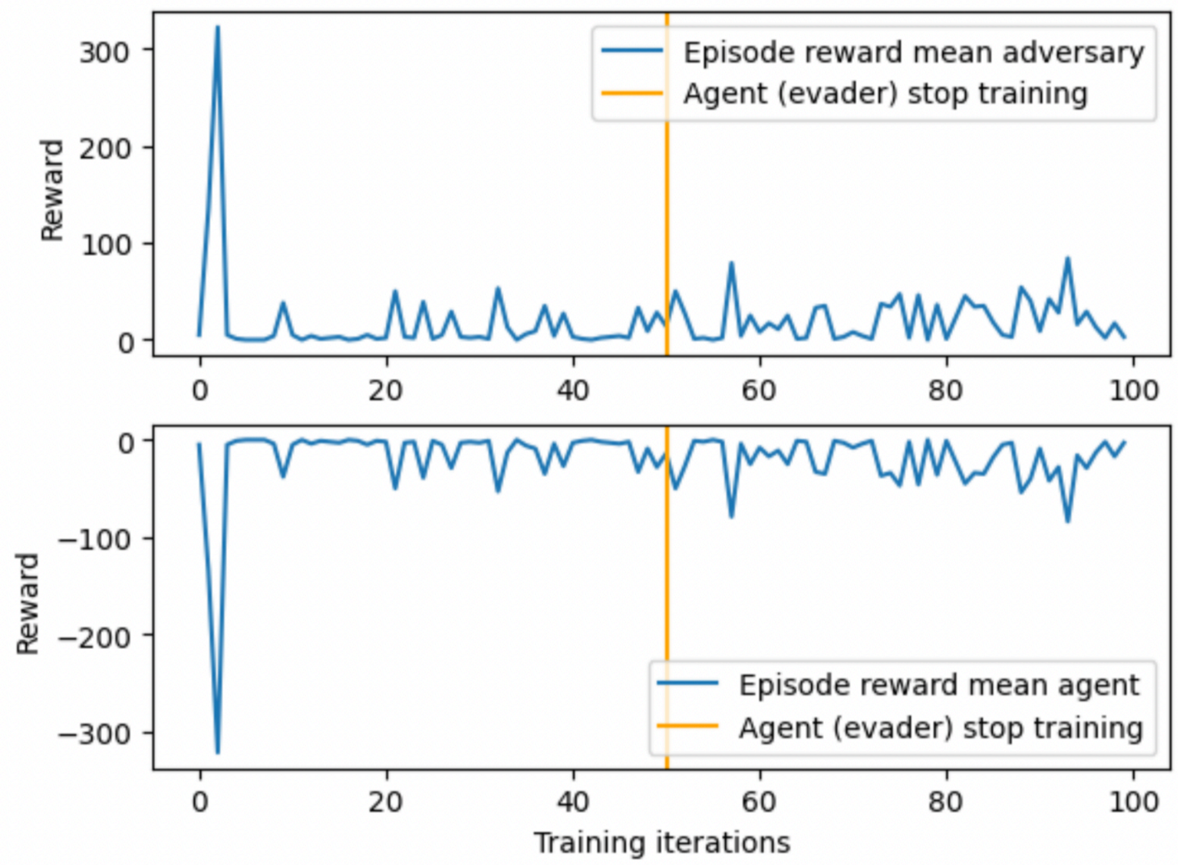


Competitive Environment - Visualization of MPE

Baseline



Moderate CPT (Risk Seeking)



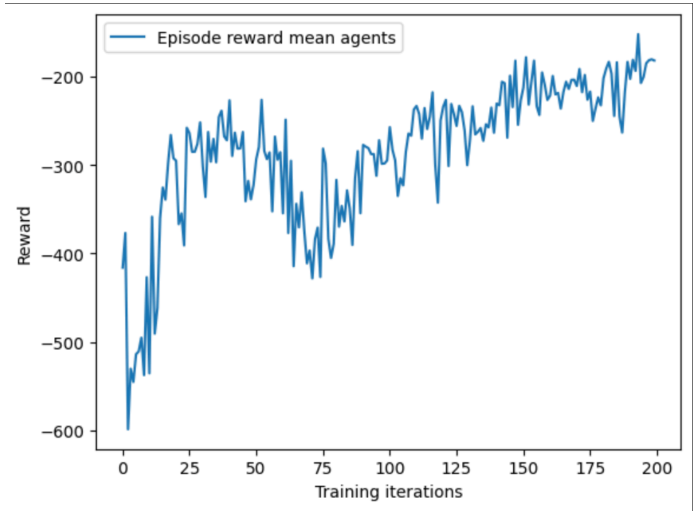
Cooperative Environment - Overview

PettingZoo's **Simple Spread** Environment is a basic Multi-Agent Particle Environment (MPE) designed for semi-collaboration between agents

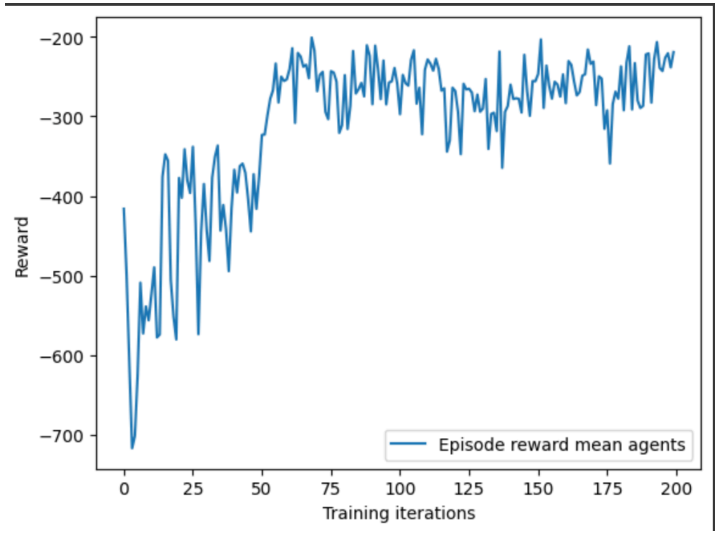
- **Objective**: The agents work cooperatively to cover all the landmarks. Their goal is to position themselves so that each landmark is “covered” by at least one agent, maximizing overall performance.
- **Rewards**: Rewards encourage efficient coverage of landmarks while also penalizing agents for collisions with one another, which promotes coordinated movement and spacing.

Cooperative Environment - Rewards

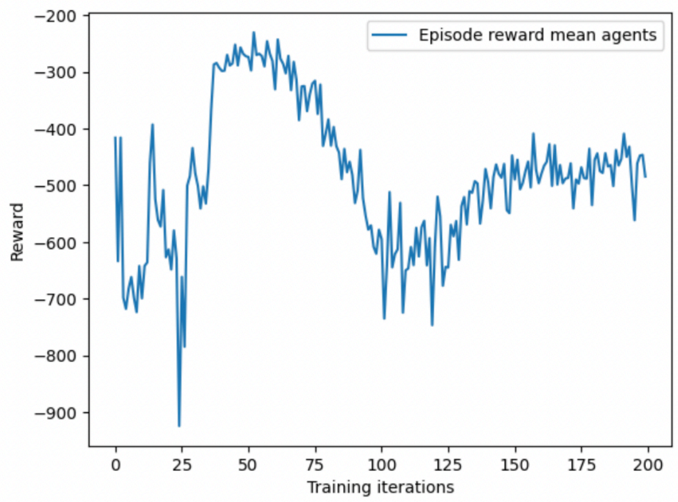
Baseline



Moderate CPT

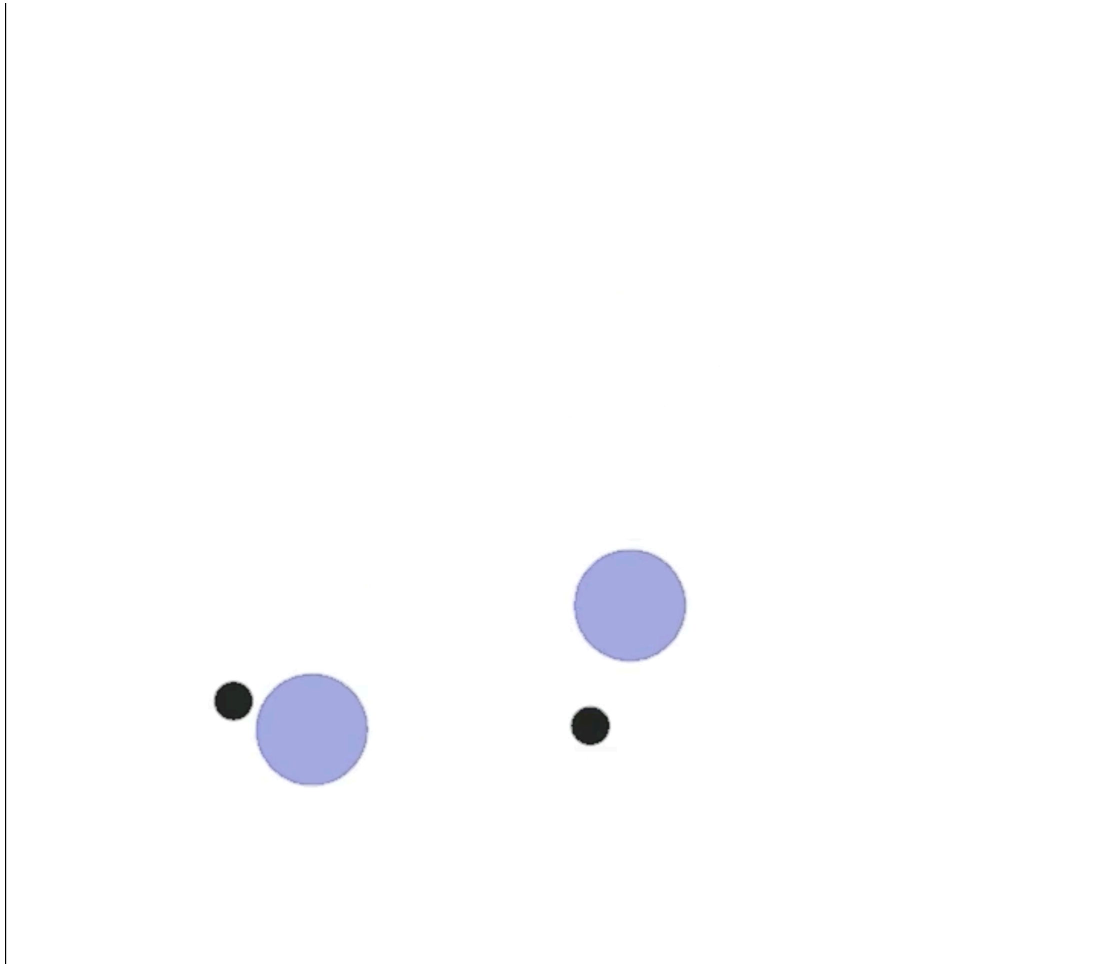


Extreme CPT

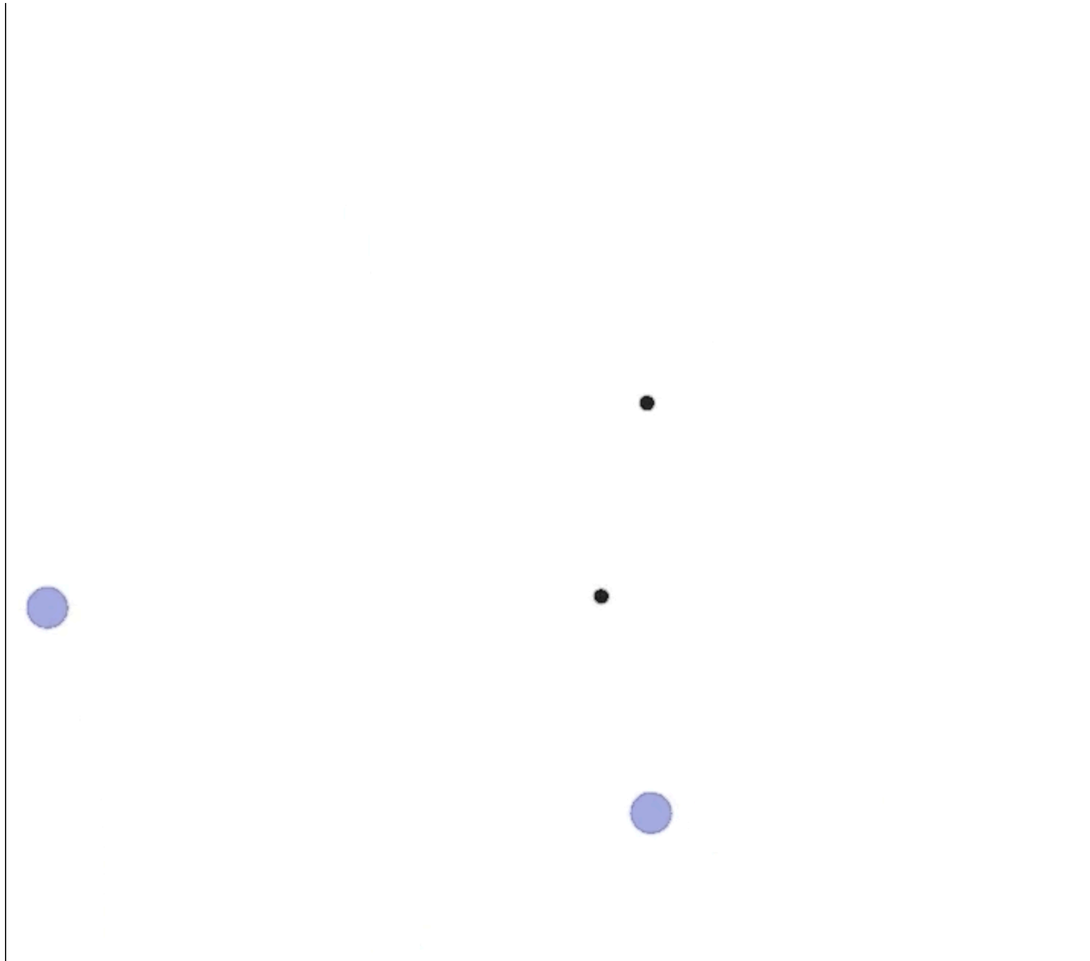


Cooperative Environment - Visualization of MPE

Baseline CPT



Extreme CPT



Next Steps & Challenges

Planned Improvements

- **CPT Integration**
  - Implement probability weighting and value distortions
  - Modify policy updates for CPT-weighted objectives
- **Technical Hurdles**
  - Gradient stability under CPT-induced reward transformations
  - Multi-agent coordination under risk-sensitive behaviors
  - Computational overhead from probability-weighted updates
- **Open Questions**
  - How does CPT impact equilibrium stability?
  - Best strategies for CPT-weighted return approximation?

Conclusion

Summary & Future Directions

- MARL framework successfully implemented, but **CPT integration pending**
- Policy gradient approach chosen for adaptability to nonconvex objectives
- Early results validate **agent learning**, but evaluation metrics need refinement
- **Next Steps**:
  - Incorporate CPT-based distortions
  - Improve training stability & evaluation methods
  - Assess strategic behaviors under CPT in multi-agent environments

Thank You! Questions?