Emergent Adversarial Behaviors via Self-Play in MARL

A Comparative Study of MAPPO and IPPO Algorithms

Charan Reddy Nandyala | Dhanush Chalicheemala | Satyadev Gangineni

University of California, Riverside

Project Overview

🎯 Research Goal

Investigate emergent adversarial behaviors in multi-agent reinforcement learning through competitive self-play mechanisms in predator-prey scenarios.

🔬 Methodology

- Algorithms: MAPPO (centralized training, shared critic) and IPPO (independent learning)
- Self-Play Strategies: Alternating, Population-based, League-based
- Environment: Custom environment (20x20 grid world)

MAPPO: Multi-Agent PPO

Architecture

- Centralized Training, Decentralized Execution (CTDE)
- Shared critic with global state information
- Individual actor networks for each agent
- Parameter sharing among agents of same type

How It Works

1. Collect experiences from all agents
2. Centralized critic evaluates global state
3. Compute advantages using shared value function
4. Update individual actor policies via PPO objective

Advantages

- Superior coordination among agents
- More stable training process
- Better for team-based tasks
- Shared value function reduces variance

Disadvantages

- Requires global state information
- More complex implementation
- Less scalable to many agents
- Coordination overhead in training

Best for: Multi-agent coordination tasks

IPPO: Independent PPO

Architecture

- Fully Independent Learning per agent
- Separate actor-critic networks for each agent
- No information sharing between agents
- Treats other agents as part of environment

How It Works

1. Each agent collects own experiences
2. Each agent has separate critic (value function)
3. Policies updated independently using PPO
4. No coordination or information sharing

Advantages

- Fully decentralized (no global state needed)
- Simpler implementation
- Better scalability to many agents
- Robust to non-stationarity

Disadvantages

- No coordination between agents
- May converge to suboptimal strategies
- Less stable in competitive settings
- Individual credit assignment challenges

Best for: Independent learning, scalability

Self-Play Strategies

What is Self-Play?

Training agents by playing against themselves or other versions, creating a natural curriculum where agents face increasingly skilled opponents.

1. Alternating Self-Play

- Agents alternate between training and being frozen as opponents
- One side trains while the other is fixed, then switch roles
- Pros: Simple, stable | Cons: Sequential learning

2. Population-Based Self-Play

- Maintain diverse agent populations
- Train against random samples from population history
- Pros: Robust, diverse opponents | Cons: Memory intensive

3. League-Based Self-Play

- AlphaStar-inspired competitive training
- Main agents, exploiters, and league members
- Pros: Most sophisticated, prevents overfitting | Cons: Complex, resource-heavy

Results Overview

Key Metrics

- IPPO Predator Win Rate: 96-98%
- MAPPO Predator Win Rate: 45-77%
- Episode Length: 200 steps (maximum)
- Training: 5,000 episodes per configuration

🏆 IPPO Dominance

- Predators achieve 96-98% win rate
- Fast convergence to effective strategies
- Strong adaptation to prey improvements
- Exceptional performance in 3v2 scenarios

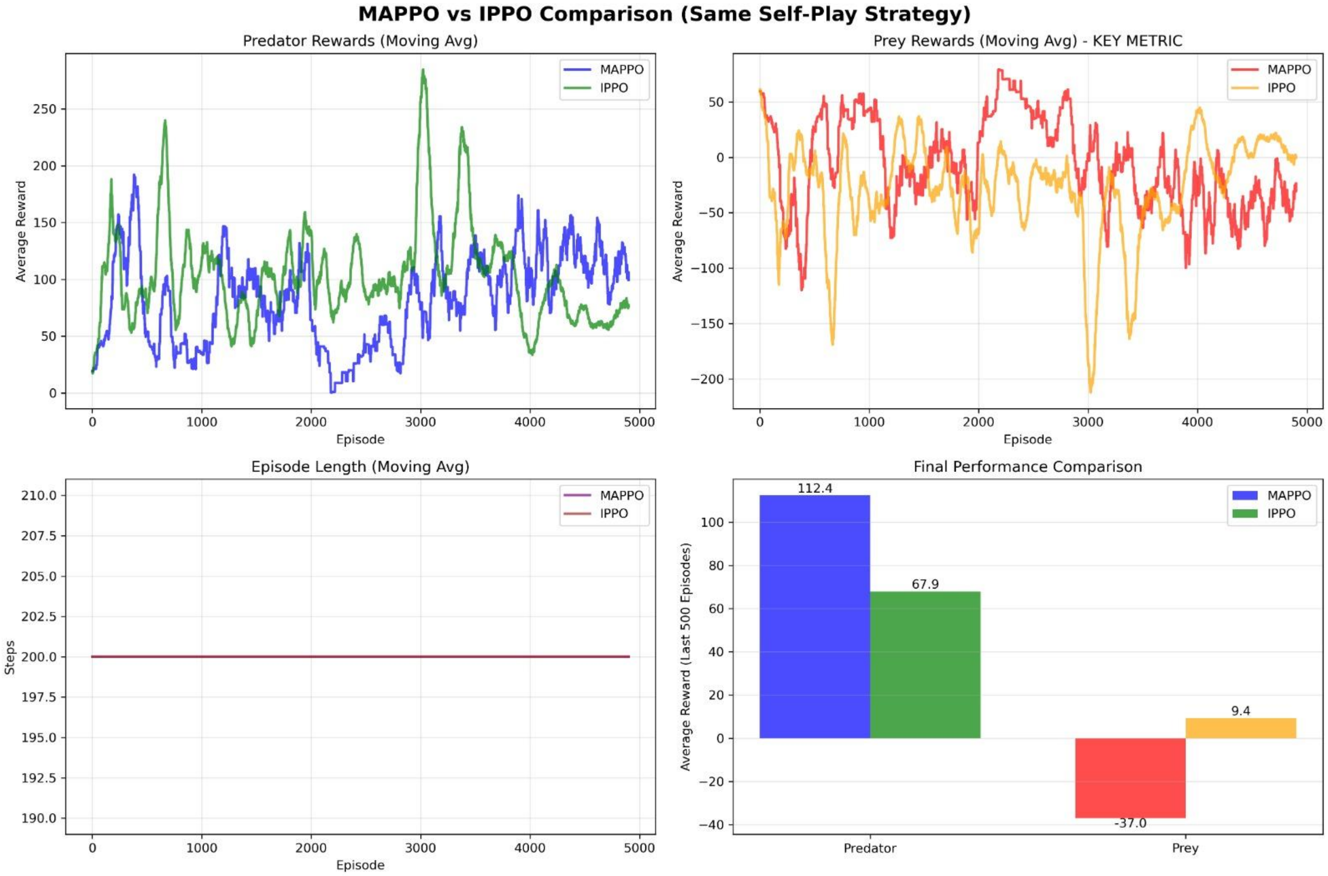
⚖️ MAPPO Balance

- More balanced predator-prey dynamics
- Better coordination among predators
- Superior prey learning and survival rates
- Trade-off: Lower predator win rates but more interesting gameplay

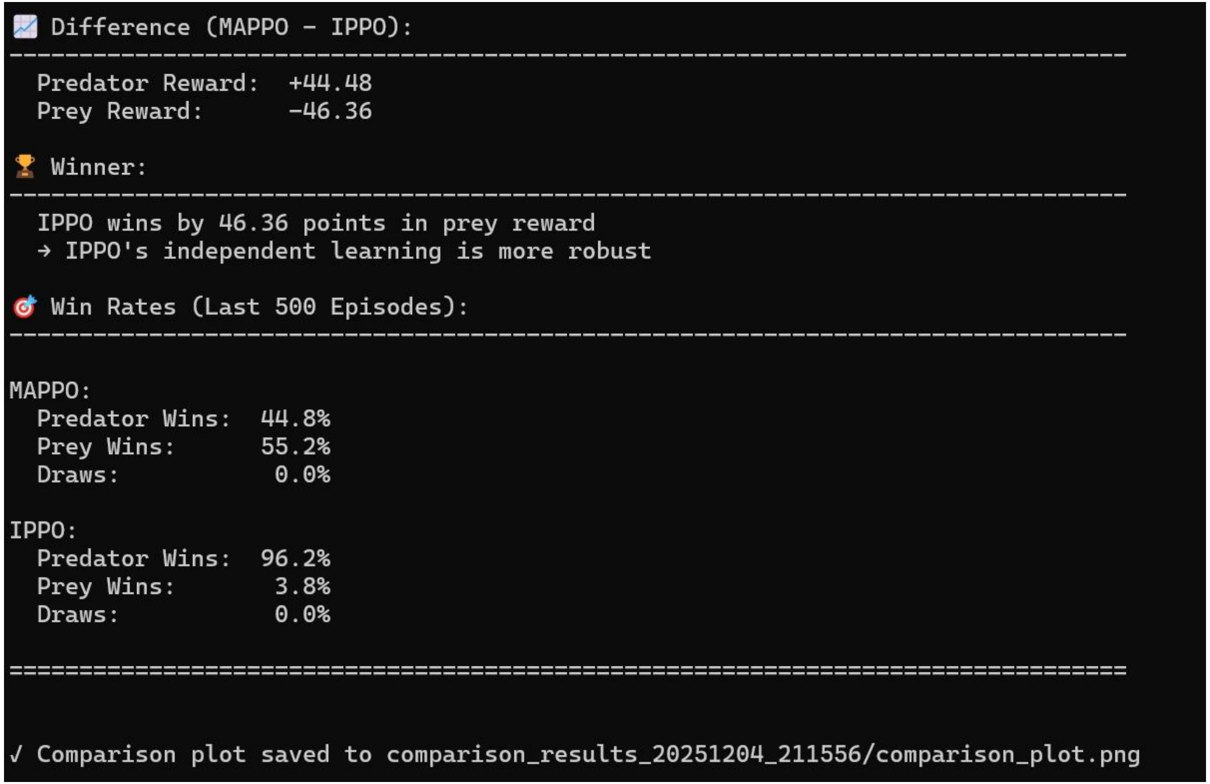
Why IPPO Predators Dominate

1. Independent learning = Direct optimization without coordination overhead
2. Fast adaptation = Quick response to evolving prey strategies
3. Simpler learning signal = Clear credit assignment and reward feedback
4. Numerical advantage = Multiple independent threats overwhelm prey

Results Overview (2v2 Alternating Self play)

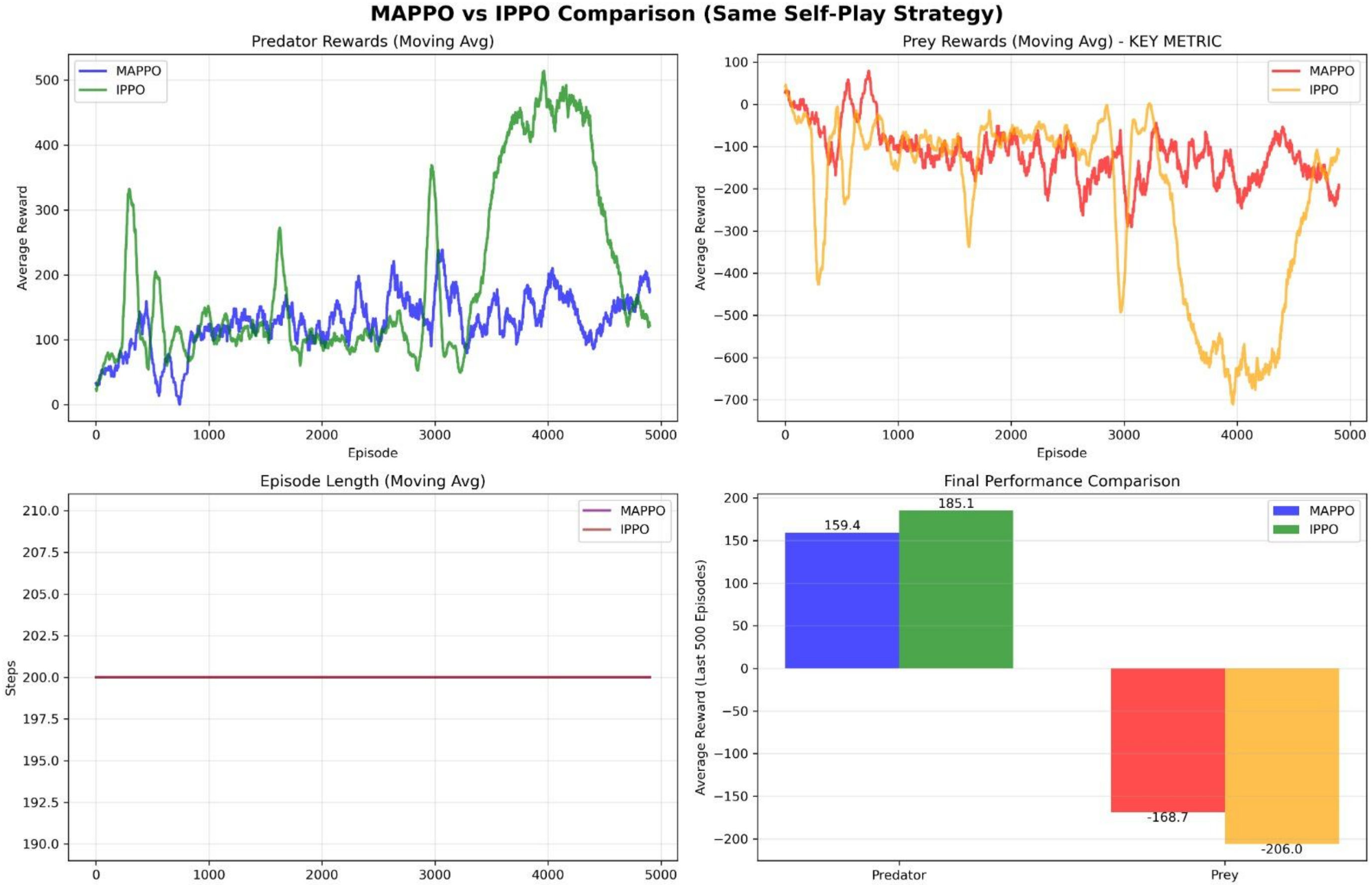


Results Overview (2v2 Alternating Self play)

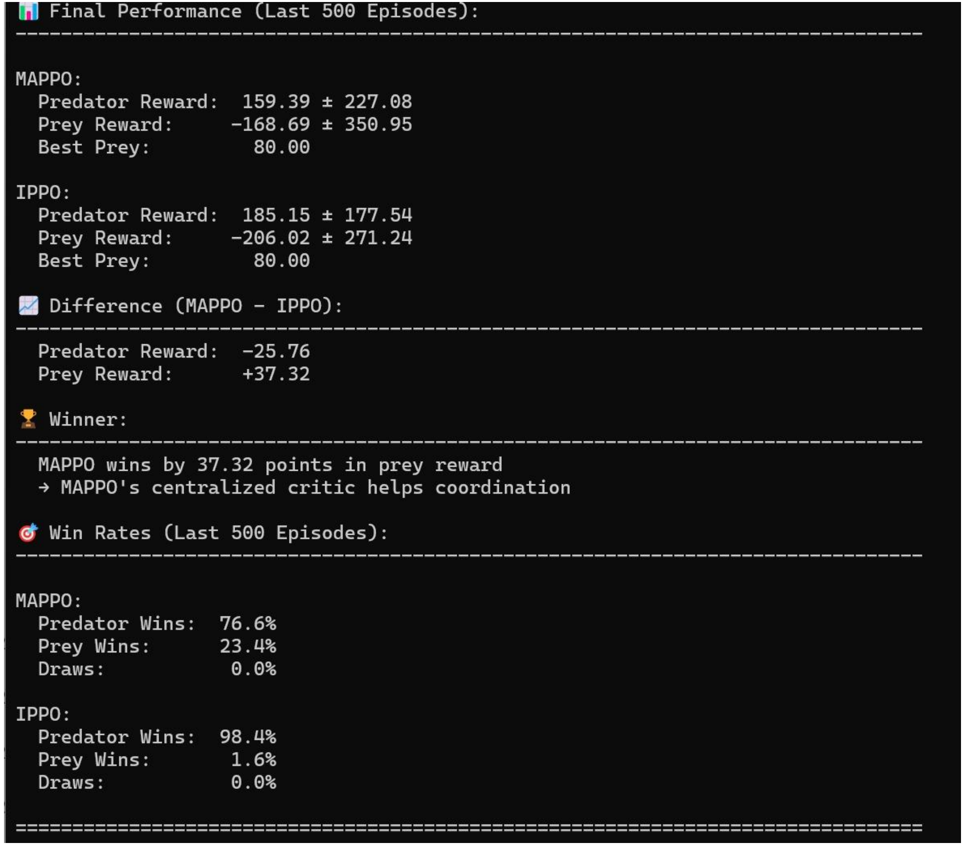


```
📈 Difference (MAPPO - IPPO):
--------------------------------------------------------------------------------
  Predator Reward:   +44.48
  Prey Reward:       -46.36

🏆 Winner:
--------------------------------------------------------------------------------
  IPPO wins by 46.36 points in prey reward
  → IPPO's independent learning is more robust

🎯 Win Rates (Last 500 Episodes):
--------------------------------------------------------------------------------

MAPPO:
  Predator Wins:  44.8%
  Prey Wins:      55.2%
  Draws:           0.0%

IPPO:
  Predator Wins:  96.2%
  Prey Wins:       3.8%
  Draws:           0.0%

================================================================================


✓ Comparison plot saved to comparison_results_20251204_211556/comparison_plot.png
```

Results Overview(3v2 Alternating self play)

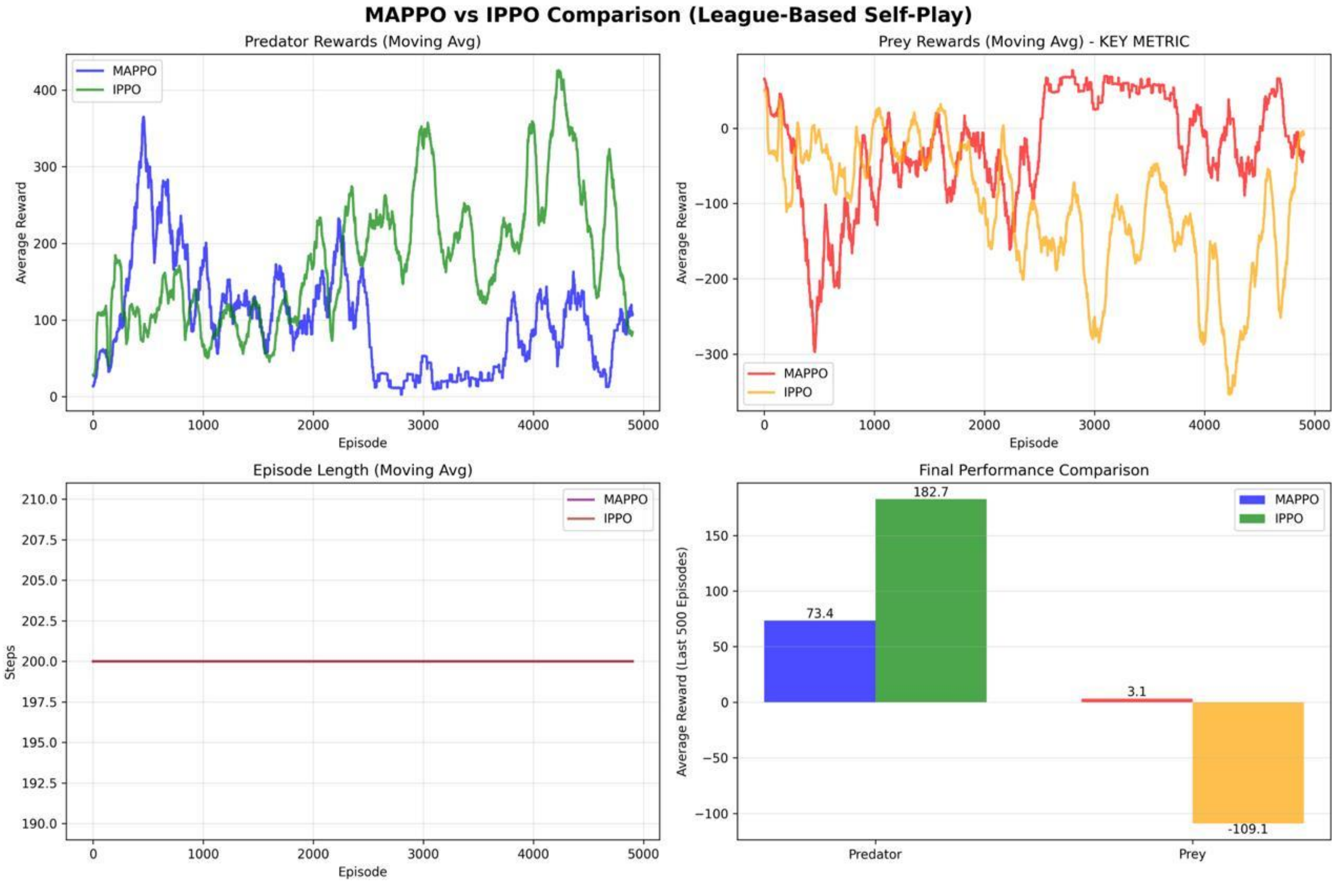


Results Overview(3v2 Alternating self play)

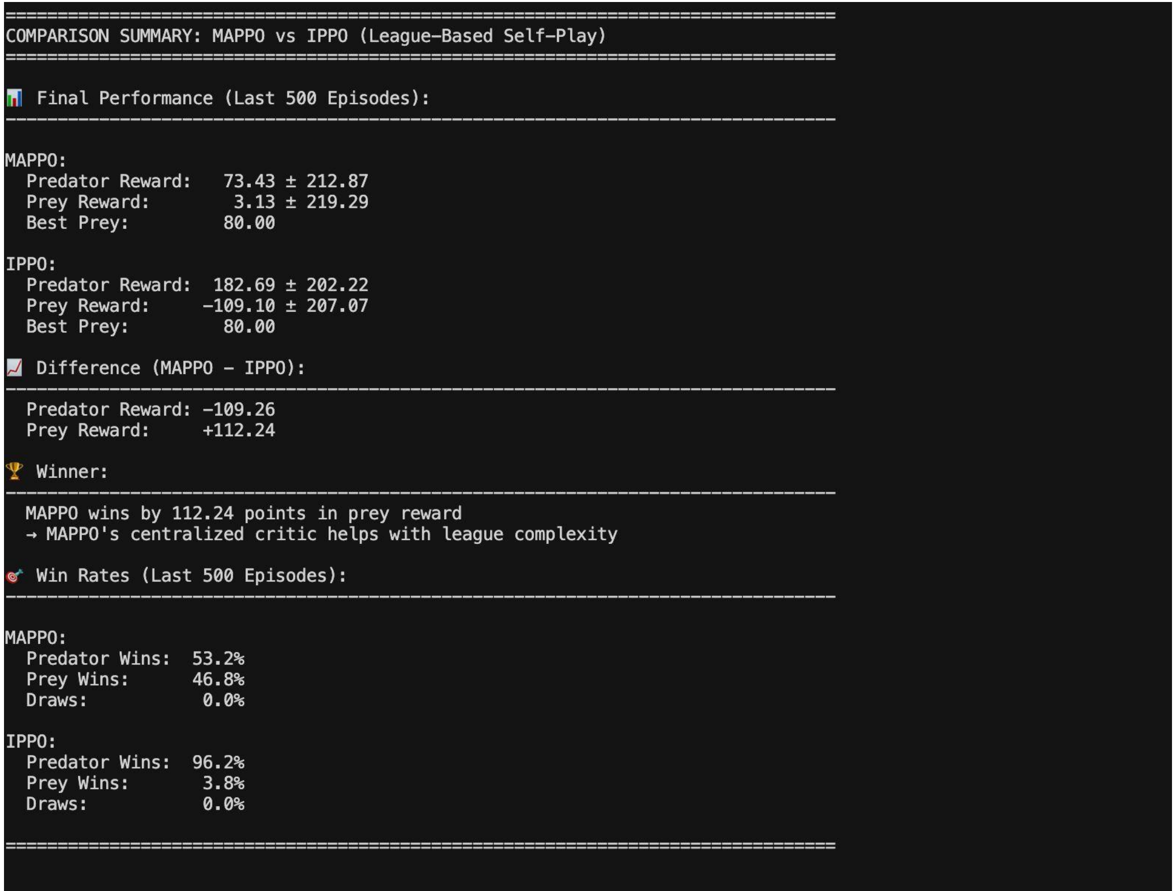


```
📊 Final Performance (Last 500 Episodes):
--------------------------------------------------------------------------------

MAPPO:
  Predator Reward:  159.39 ± 227.08
  Prey Reward:     -168.69 ± 350.95
  Best Prey:         80.00

IPPO:
  Predator Reward:  185.15 ± 177.54
  Prey Reward:     -206.02 ± 271.24
  Best Prey:         80.00

📈 Difference (MAPPO - IPPO):
--------------------------------------------------------------------------------
  Predator Reward:  -25.76
  Prey Reward:      +37.32

🏆 Winner:
--------------------------------------------------------------------------------
  MAPPO wins by 37.32 points in prey reward
  → MAPPO's centralized critic helps coordination

🎯 Win Rates (Last 500 Episodes):
--------------------------------------------------------------------------------

MAPPO:
  Predator Wins:  76.6%
  Prey Wins:      23.4%
  Draws:           0.0%

IPPO:
  Predator Wins:  98.4%
  Prey Wins:       1.6%
  Draws:           0.0%

================================================================================
```

Results Overview(Population based self play)

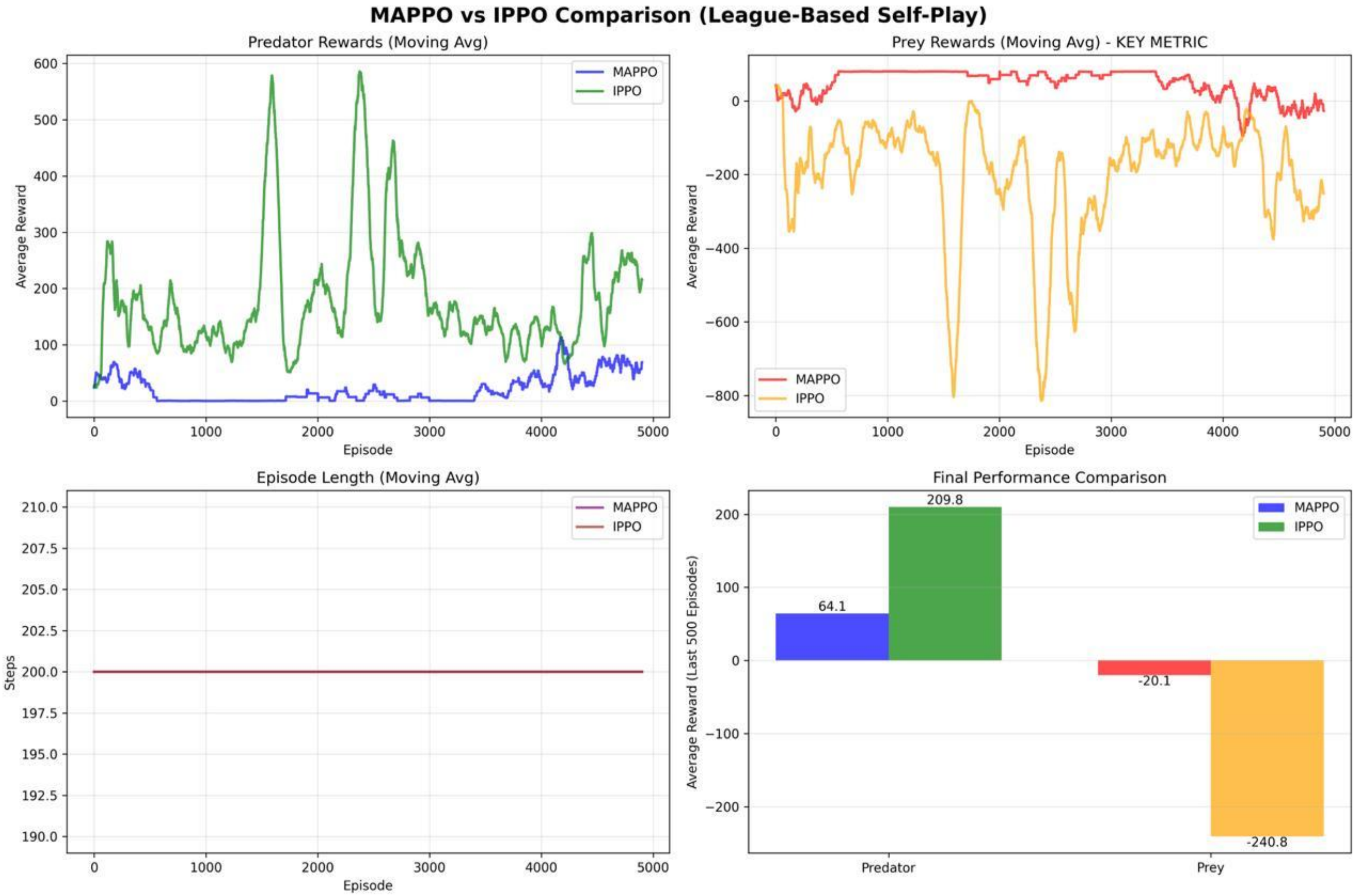


Results Overview(Population based self play)

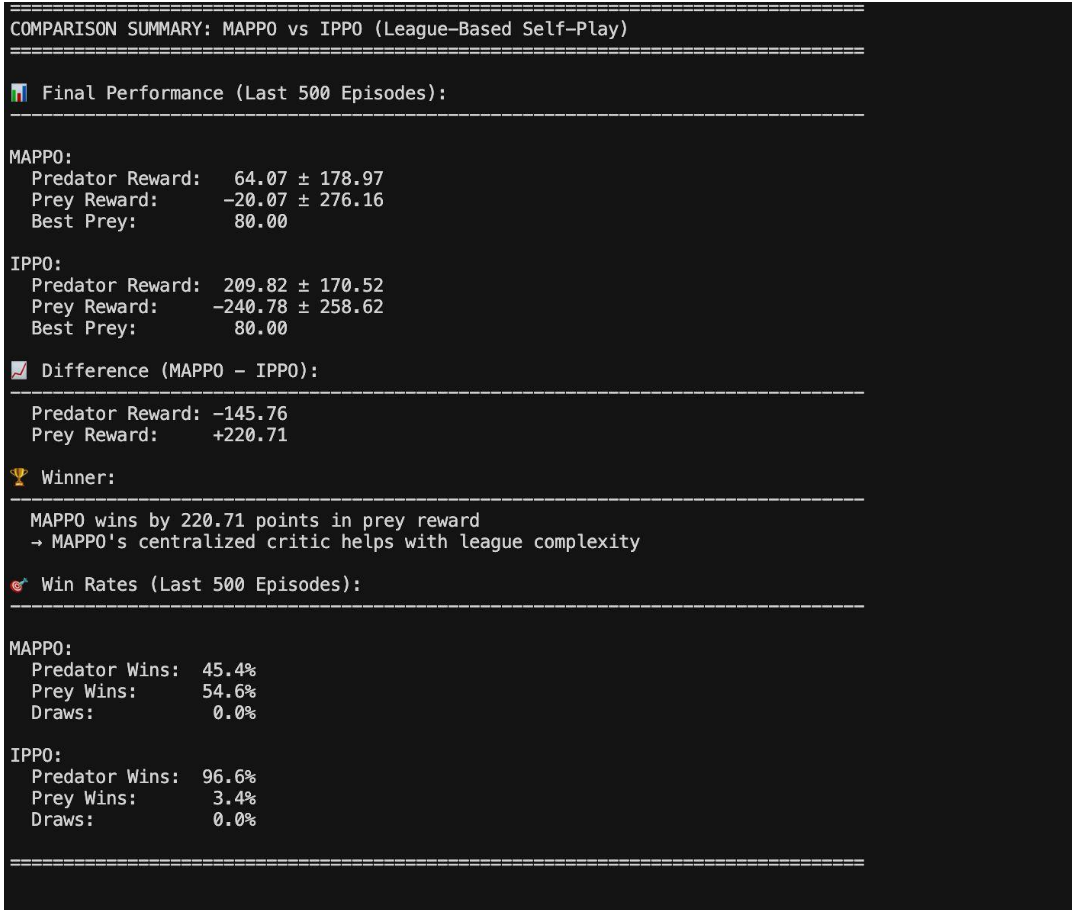


```
================================================================================
COMPARISON SUMMARY: MAPPO vs IPPO (League-Based Self-Play)
================================================================================

📊 Final Performance (Last 500 Episodes):
--------------------------------------------------------------------------------

MAPPO:
  Predator Reward:   73.43 ± 212.87
  Prey Reward:        3.13 ± 219.29
  Best Prey:         80.00

IPPO:
  Predator Reward:  182.69 ± 202.22
  Prey Reward:     -109.10 ± 207.07
  Best Prey:         80.00

📈 Difference (MAPPO - IPPO):
--------------------------------------------------------------------------------
  Predator Reward: -109.26
  Prey Reward:     +112.24

🏆 Winner:
--------------------------------------------------------------------------------
  MAPPO wins by 112.24 points in prey reward
  → MAPPO's centralized critic helps with league complexity

🎯 Win Rates (Last 500 Episodes):
--------------------------------------------------------------------------------

MAPPO:
  Predator Wins:  53.2%
  Prey Wins:      46.8%
  Draws:           0.0%

IPPO:
  Predator Wins:  96.2%
  Prey Wins:       3.8%
  Draws:           0.0%

================================================================================
```

Results Overview(2v2 League Sampling self play)

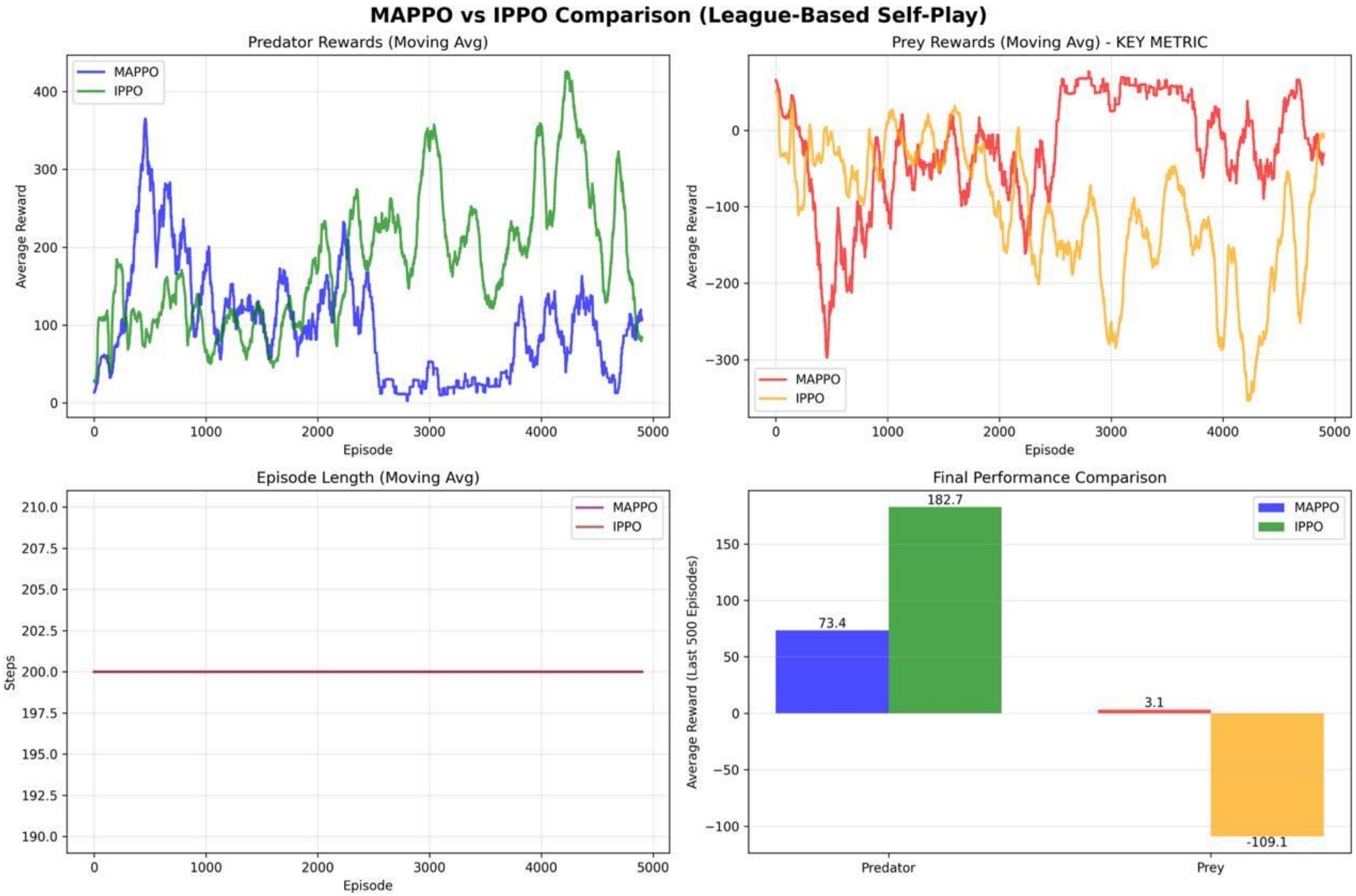


Results Overview(2v2 League Sampling self play)

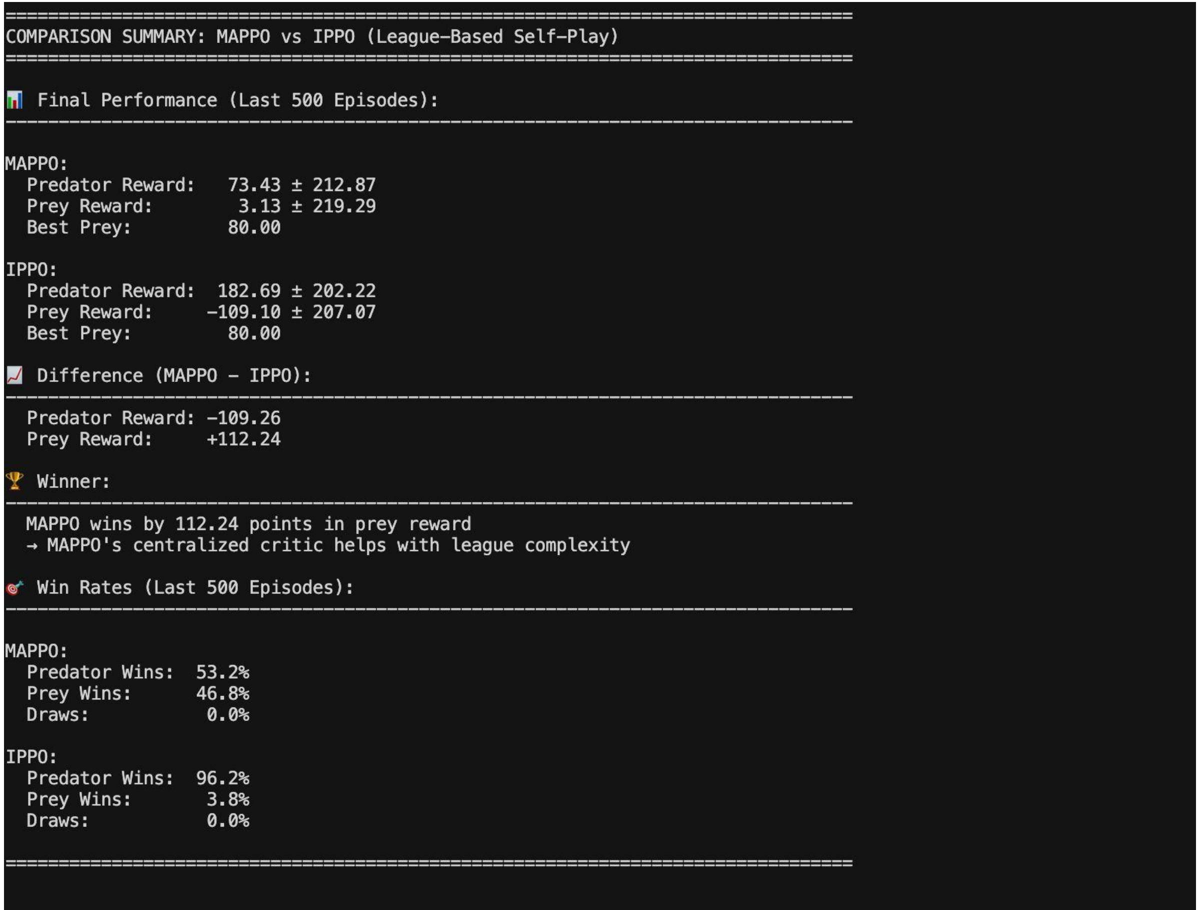


```
================================================================================
COMPARISON SUMMARY: MAPPO vs IPPO (League-Based Self-Play)
================================================================================

📊 Final Performance (Last 500 Episodes):
--------------------------------------------------------------------------------

MAPPO:
  Predator Reward:   64.07 ± 178.97
  Prey Reward:      -20.07 ± 276.16
  Best Prey:         80.00

IPPO:
  Predator Reward:  209.82 ± 170.52
  Prey Reward:     -240.78 ± 258.62
  Best Prey:         80.00

📈 Difference (MAPPO - IPPO):
--------------------------------------------------------------------------------
  Predator Reward: -145.76
  Prey Reward:     +220.71

🏆 Winner:
--------------------------------------------------------------------------------
  MAPPO wins by 220.71 points in prey reward
  → MAPPO's centralized critic helps with league complexity

🎯 Win Rates (Last 500 Episodes):
--------------------------------------------------------------------------------

MAPPO:
  Predator Wins:  45.4%
  Prey Wins:      54.6%
  Draws:           0.0%

IPPO:
  Predator Wins:  96.6%
  Prey Wins:       3.4%
  Draws:           0.0%

================================================================================
```

Results Overview(3v2 League Sampling self play)



Results Overview(3v2 League Sampling self play)



```
================================================================================
COMPARISON SUMMARY: MAPPO vs IPPO (League-Based Self-Play)
================================================================================

📊 Final Performance (Last 500 Episodes):
--------------------------------------------------------------------------------

MAPPO:
  Predator Reward:   73.43 ± 212.87
  Prey Reward:        3.13 ± 219.29
  Best Prey:         80.00

IPPO:
  Predator Reward:  182.69 ± 202.22
  Prey Reward:     -109.10 ± 207.07
  Best Prey:         80.00

📈 Difference (MAPPO - IPPO):
--------------------------------------------------------------------------------
  Predator Reward: -109.26
  Prey Reward:     +112.24

🏆 Winner:
--------------------------------------------------------------------------------
  MAPPO wins by 112.24 points in prey reward
  → MAPPO's centralized critic helps with league complexity

🎯 Win Rates (Last 500 Episodes):
--------------------------------------------------------------------------------

MAPPO:
  Predator Wins:  53.2%
  Prey Wins:      46.8%
  Draws:           0.0%

IPPO:
  Predator Wins:  96.2%
  Prey Wins:       3.8%
  Draws:           0.0%

================================================================================
```

Conclusions

1. MAPPO Shows Better Coordination

- Superior performance in league-based self-play
- Achieves positive prey rewards (+3.1) - best across all strategies
- Centralized critic helps with coordination

2. IPPO Predators Dominate

- Consistently achieve 96-98% win rates
- Independent learning = faster convergence
- But at cost of prey survival (1-4% win rate)

Thank You!

Questions & Discussion