SAND2024-03298C

Sandia National Laboratories

Exceptional service in the national interest

DECENTRALIZED MULTI-AGENT REINFORCEMENT LEARNING FOR INTERCEPTION IN A 3-DIMENSIONAL ENVIRONMENT

Trappett, Crowder, McKenzie, Emmons, Musuvathy, Chance

Sandia National Laboratories

2024 AIAA Defense Forum

U.S. DEPARTMENT OF ENERGY | NNSA National Nuclear Security Administration

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc. for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

SAND2024-03298C

MOTIVATION AND BACKGROUND

- Current lack of multi-agent frameworks for finding agile tactical solutions in defense problems.

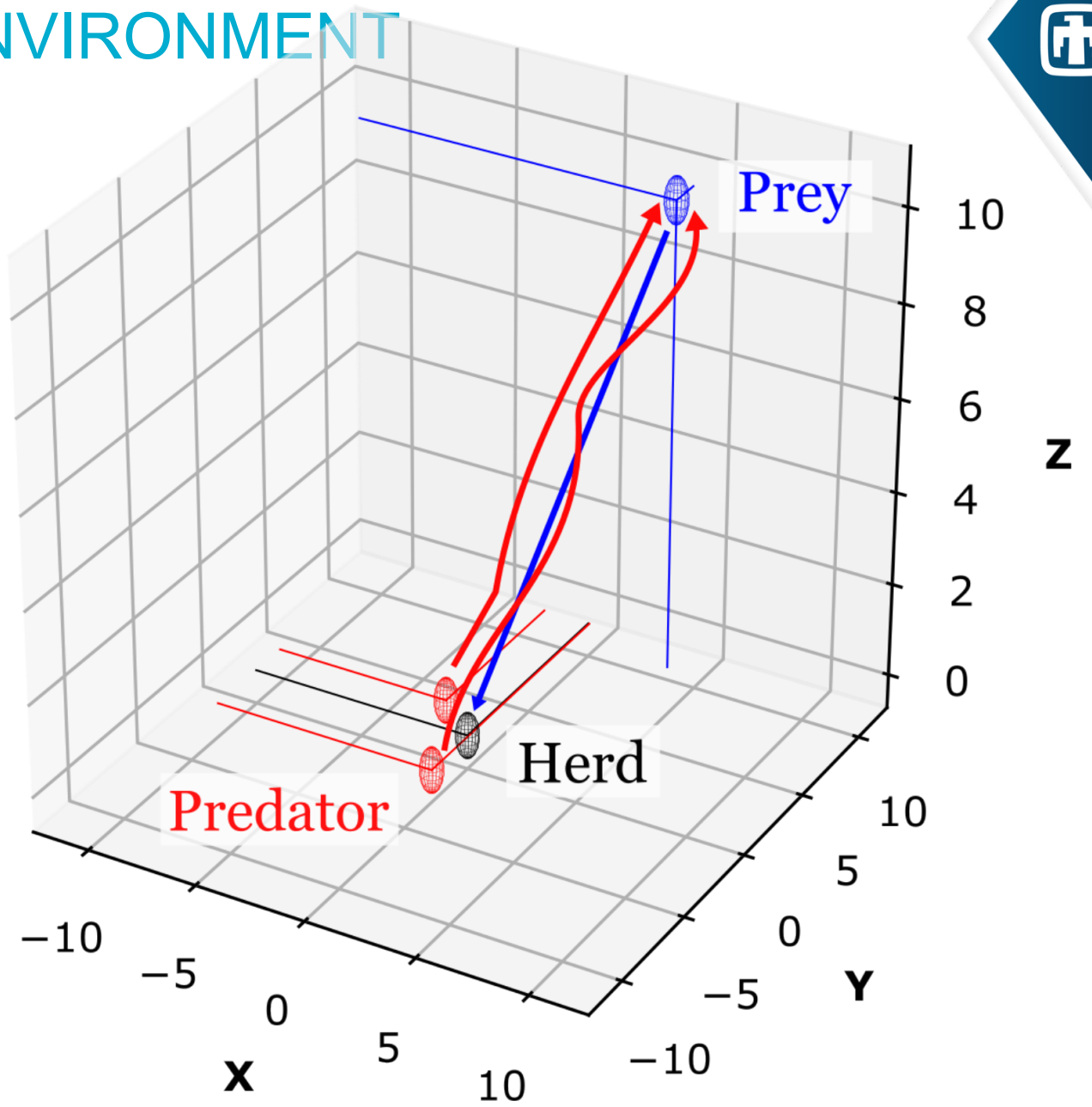
- Multi-Agent Reinforcement Learning (MARL) may be a powerful solution but it is difficult to train.

- We create a novel architecture for improving MARL.
- We characterize attributes of suboptimal MARL convergence.

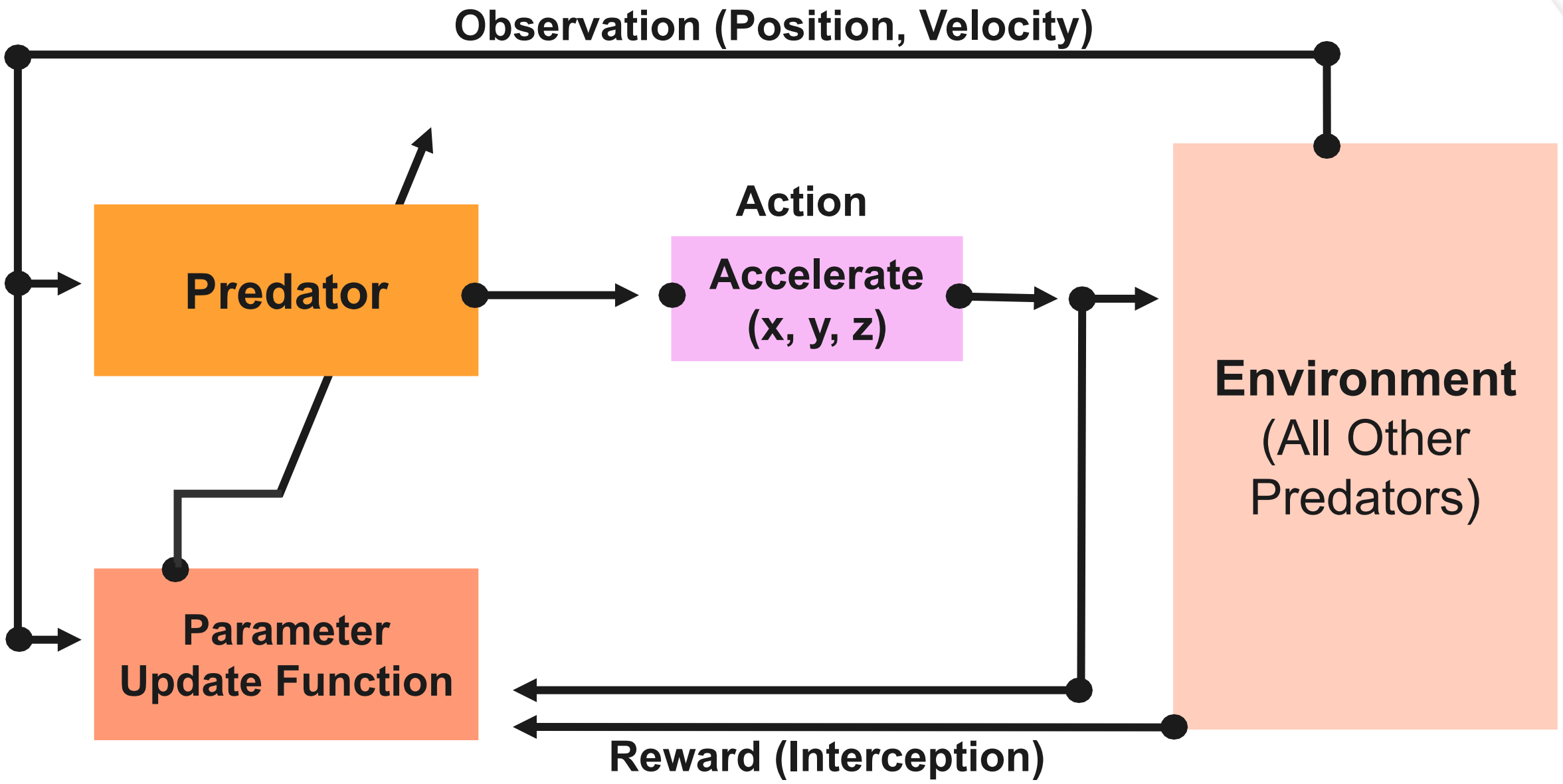
3D PARTICLE INTERCEPTION ENVIRONMENT

Predators hunting a solitary prey moving towards its herd

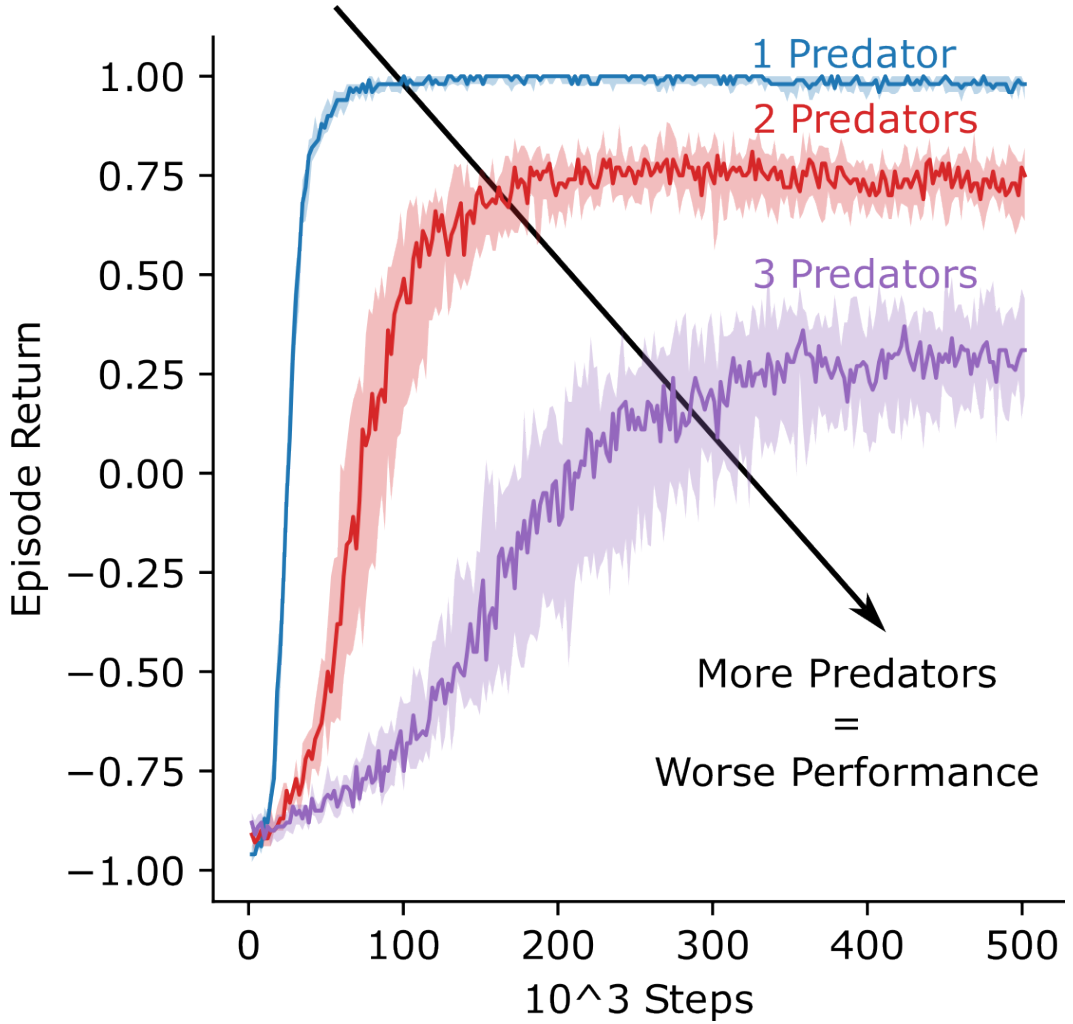
- Predators are trained to catch prey before it reaches the herd.
- Predators can accelerate in any direction, instantaneously.
- Prey moves in a straight line towards herd.
- One or more predators spawn randomly within 2 concentric rings around herd.
- If predators collide, there is no penalty but predators are removed.

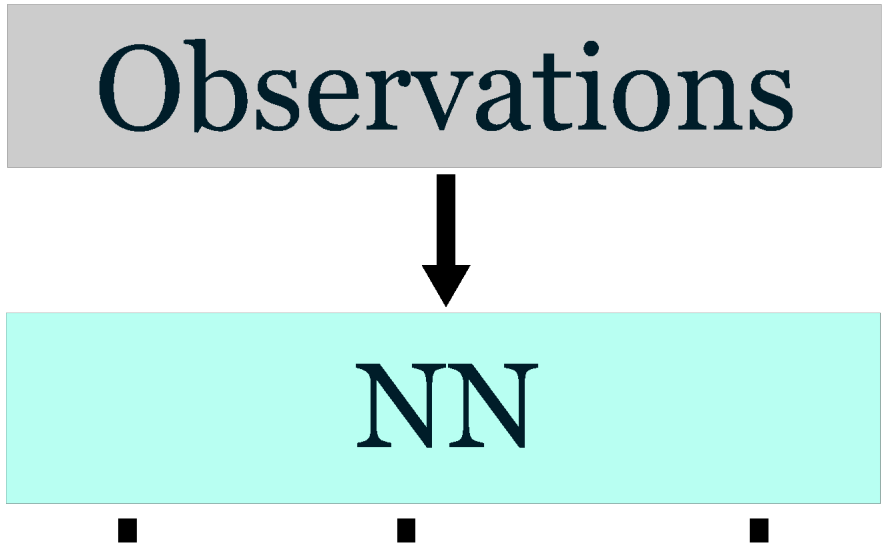


MULTI-AGENT REINFORCEMENT LEARNING



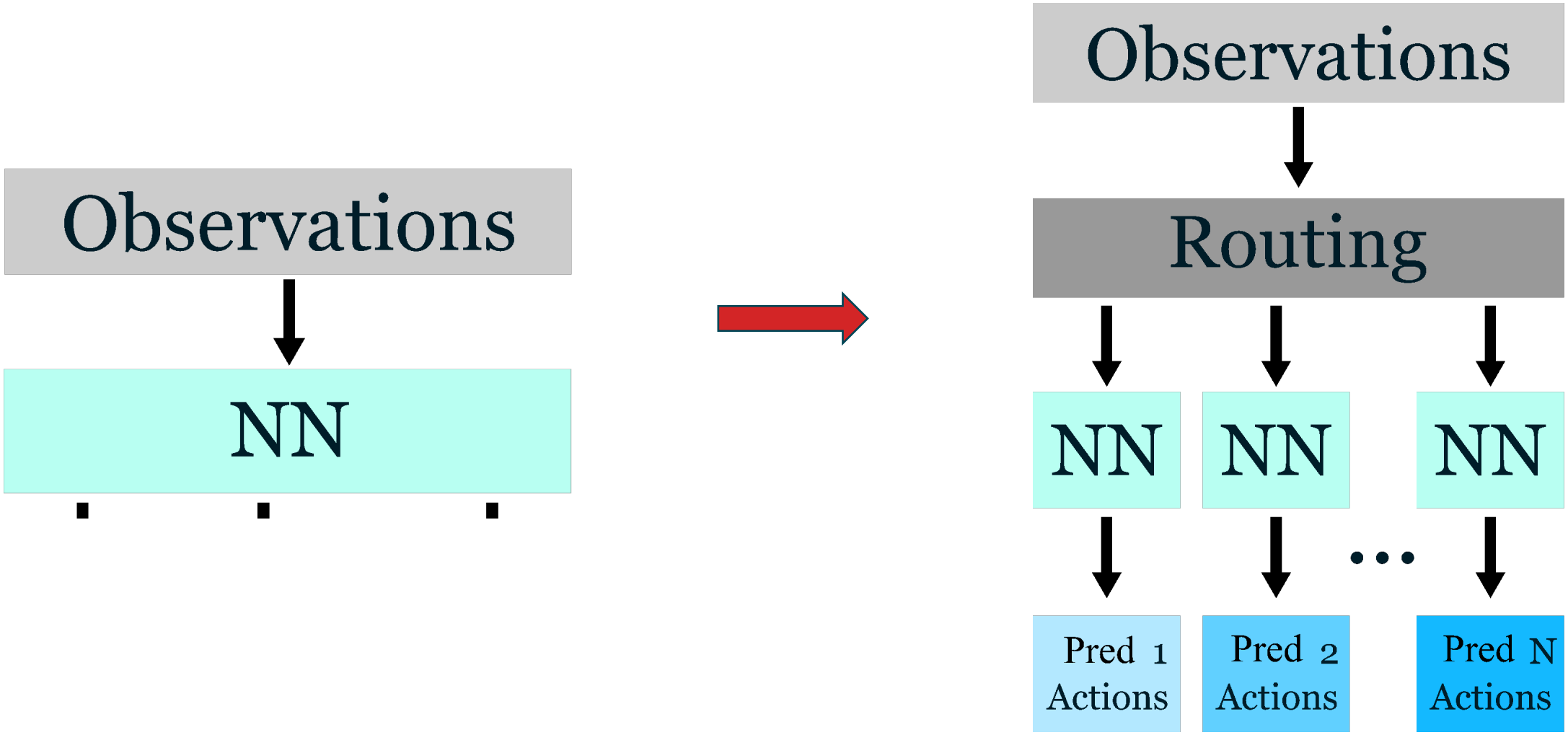
SHARED DEEP NEURAL NETWORK



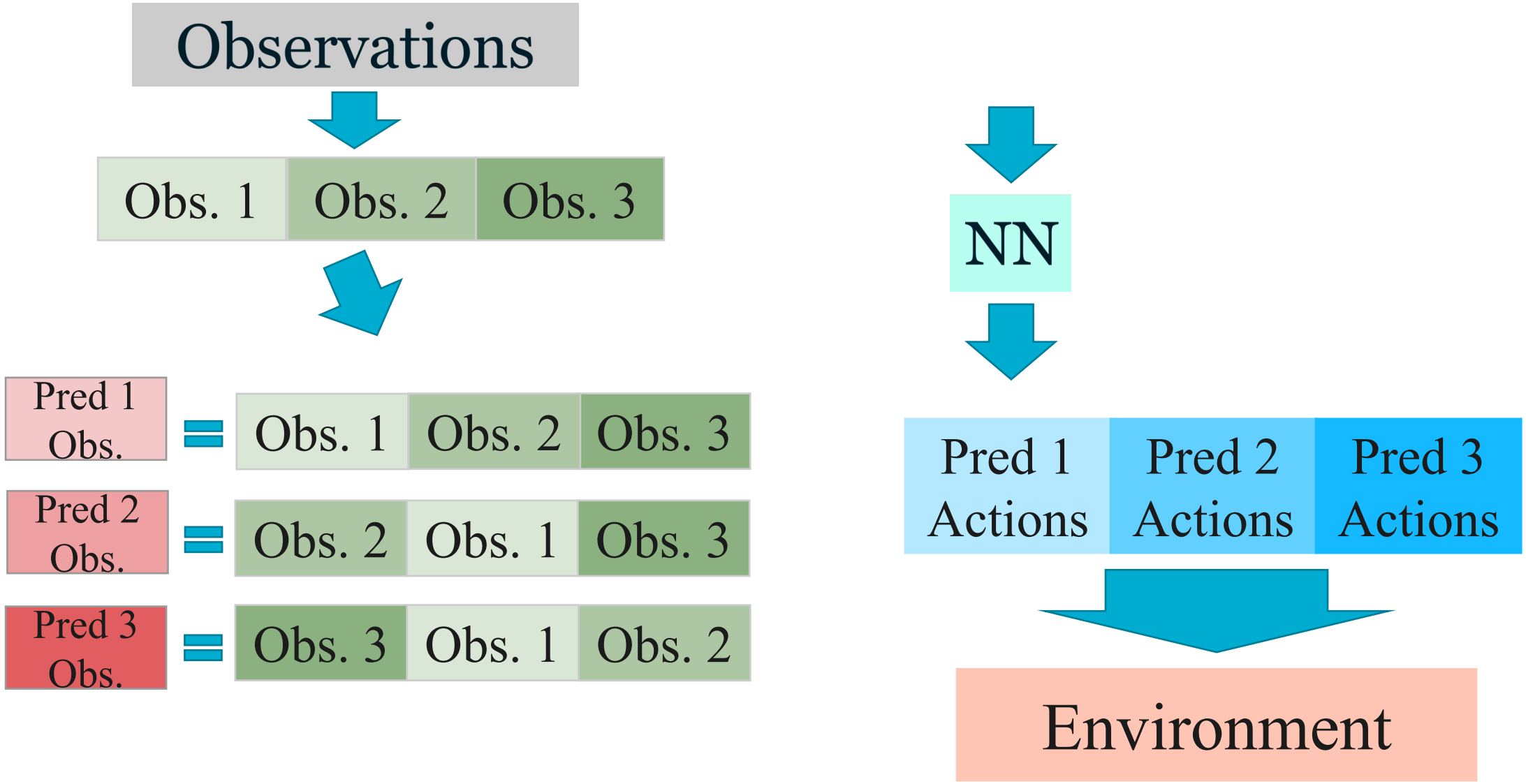


- **Trained using proximal policy optimization (PPO) algorithm**

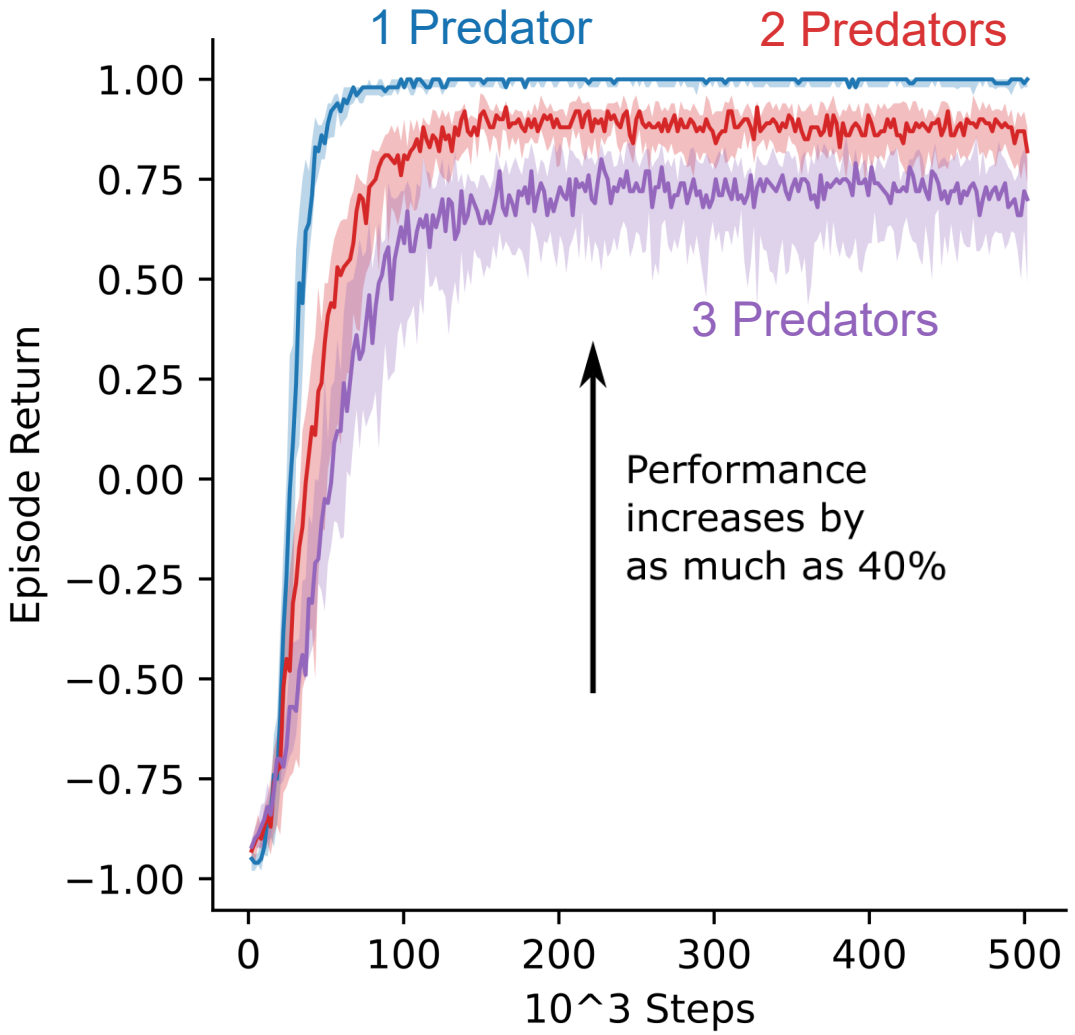
CUSTOM ARCHITECTURE

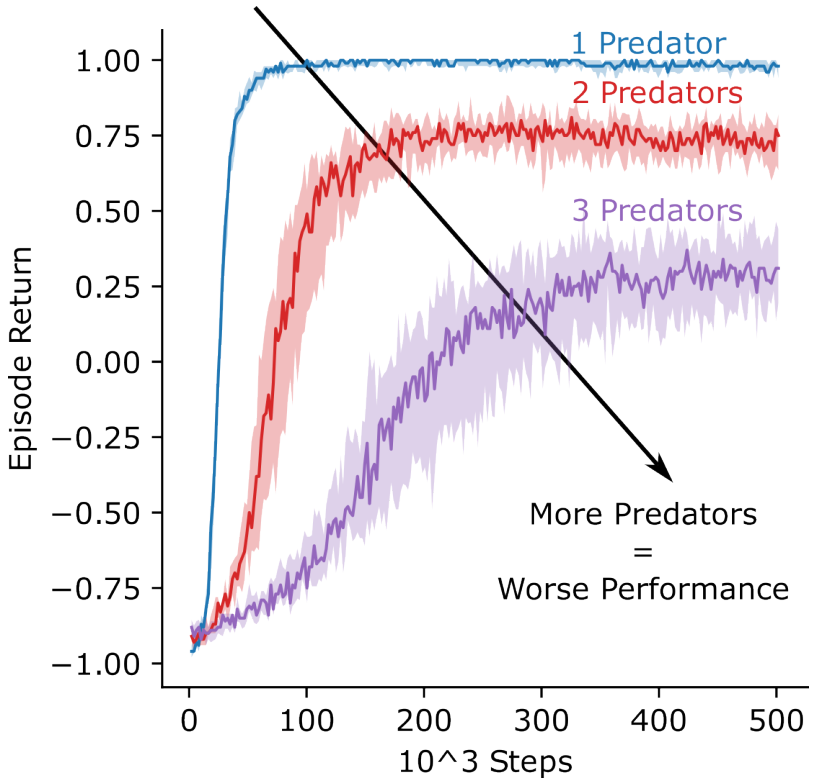


ROUTING


Observations
Obs. 1
Obs. 2
Obs. 3
Pred 1 Obs.
Obs. 1
Obs. 2
Obs. 3
Pred 2 Obs.
Obs. 2
Obs. 1
Obs. 3
Pred 3 Obs.
Obs. 3
Obs. 1
Obs. 2
NN
Pred 1 Actions
Pred 2 Actions
Pred 3 Actions
Environment

PERFORMANCE INCREASE





- **Significant performance improvement**
- **Each predator is controlled with exact neural network copy**
- **Decentralized control without need for communication**

PERFORMANCE ANALYSIS

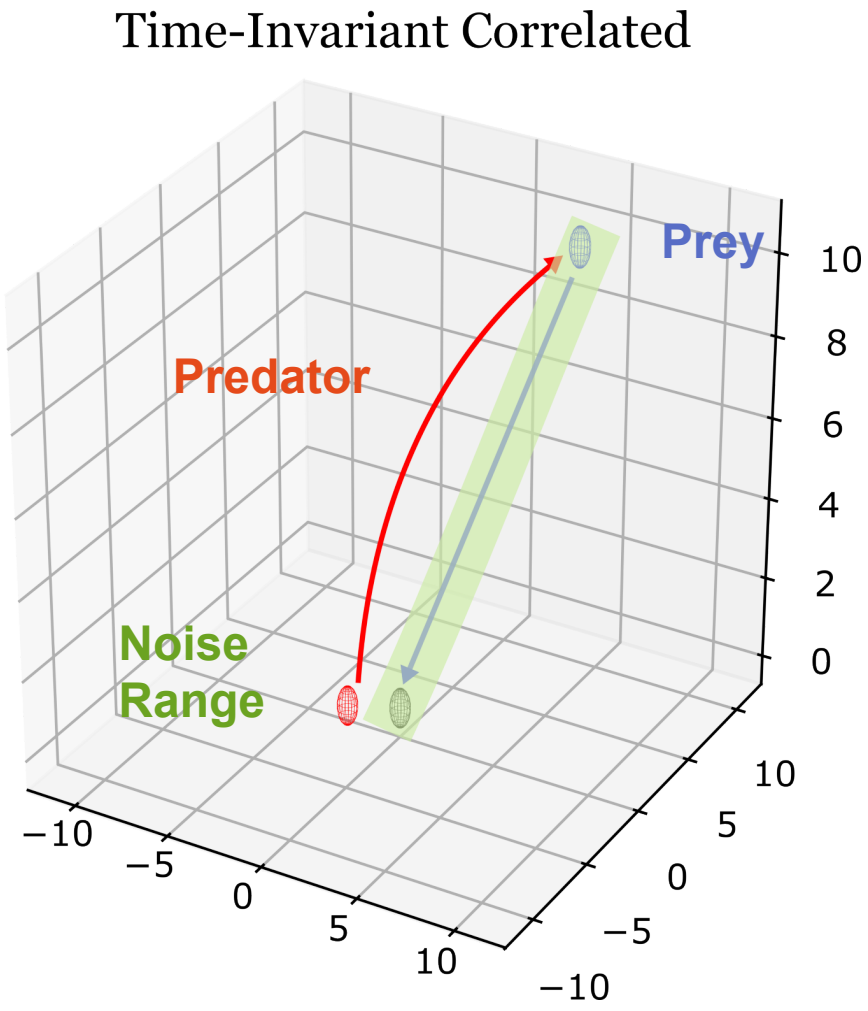
- Complicated Reward Space:

  Collision with Predators

- Larger Observation and Action Spaces:

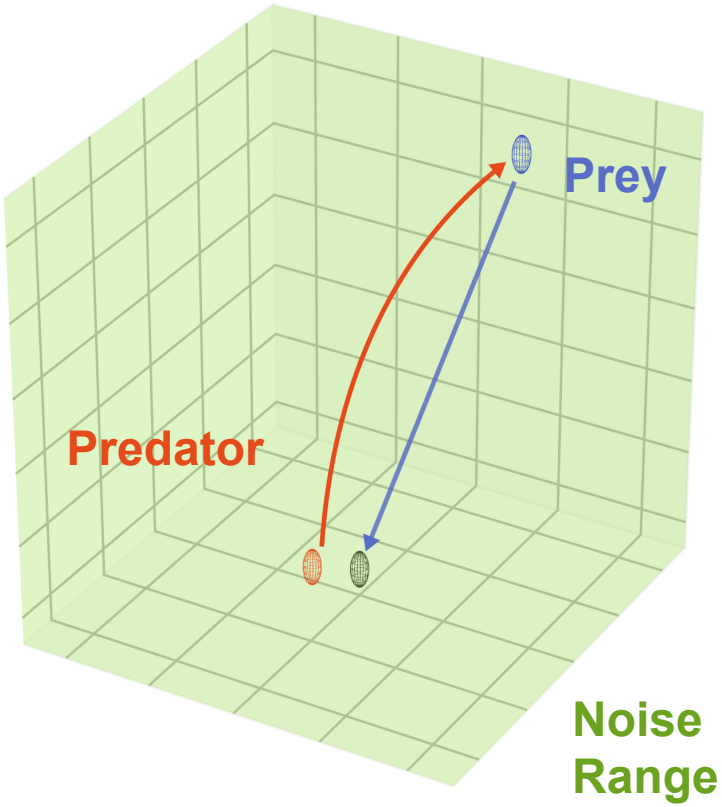
  Curse of Dimensionality

- Time-Varying, Correlated Observations:
  - Predator Correlation: Time-varying
  - Prey Correlation: Time-Invariant
  - Uniform Uncorrelated

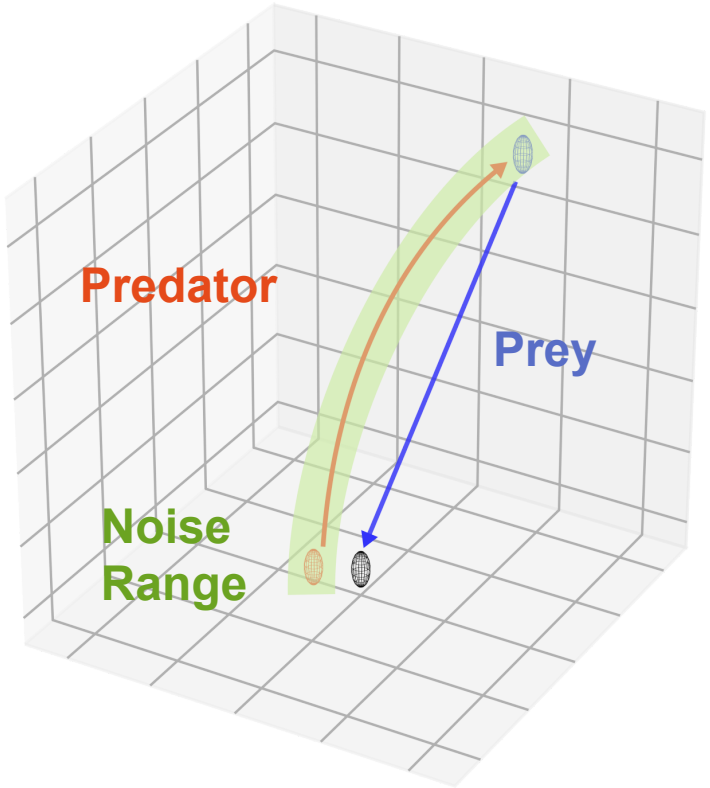


Time-Varying, Correlated Observations

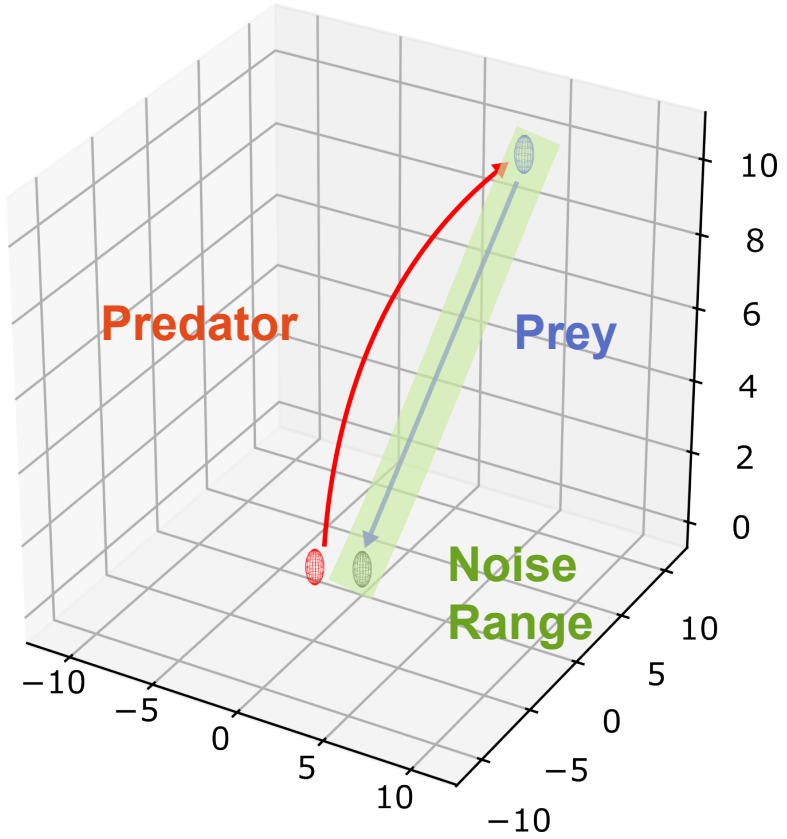
Uniform Uncorrelated



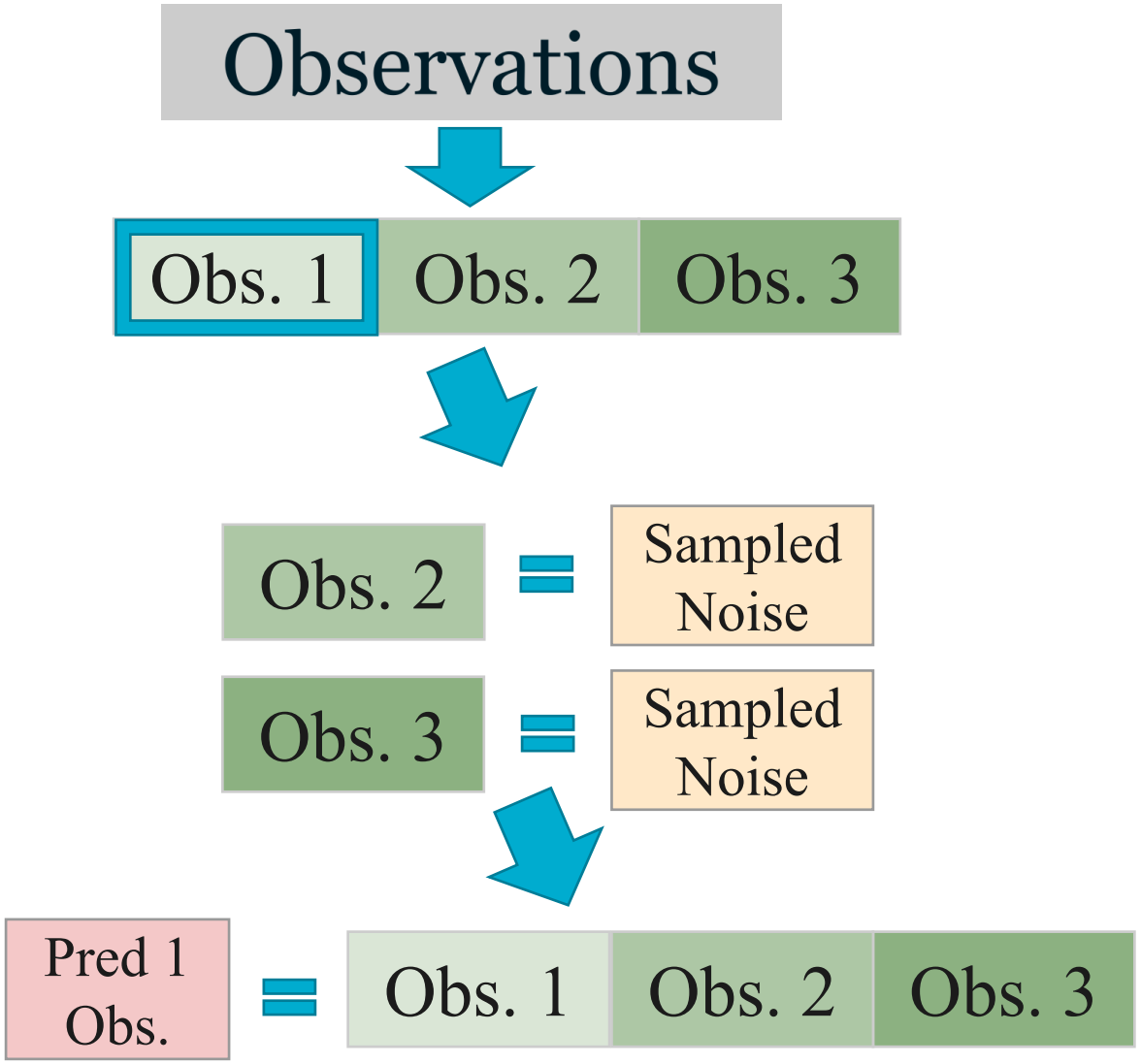
Time-Varying Correlated

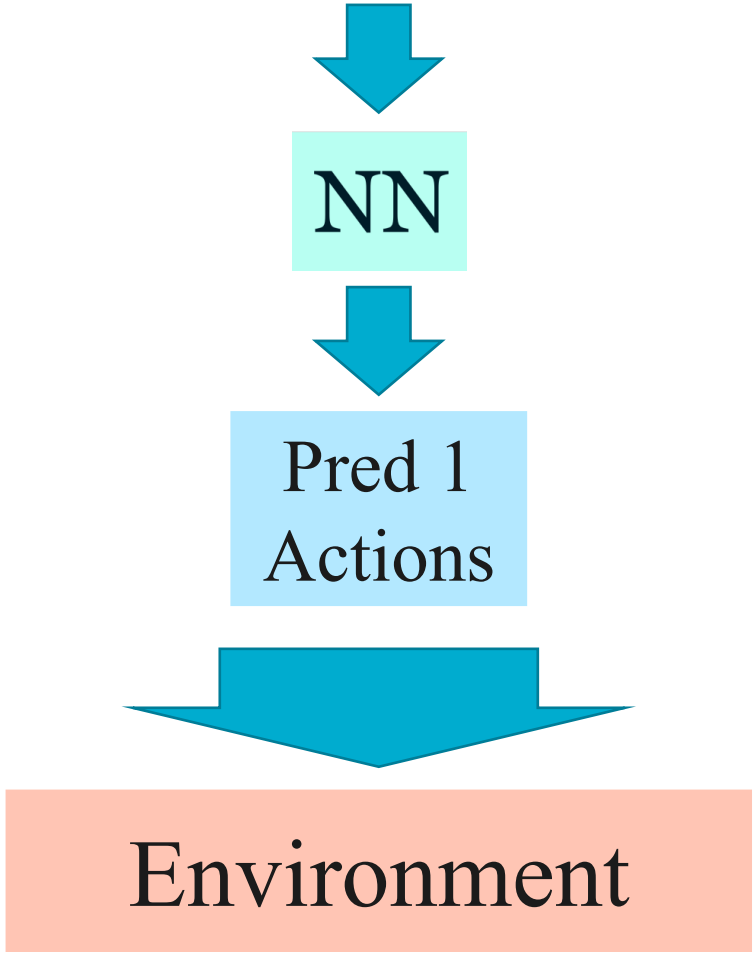


Time-Invariant Correlated

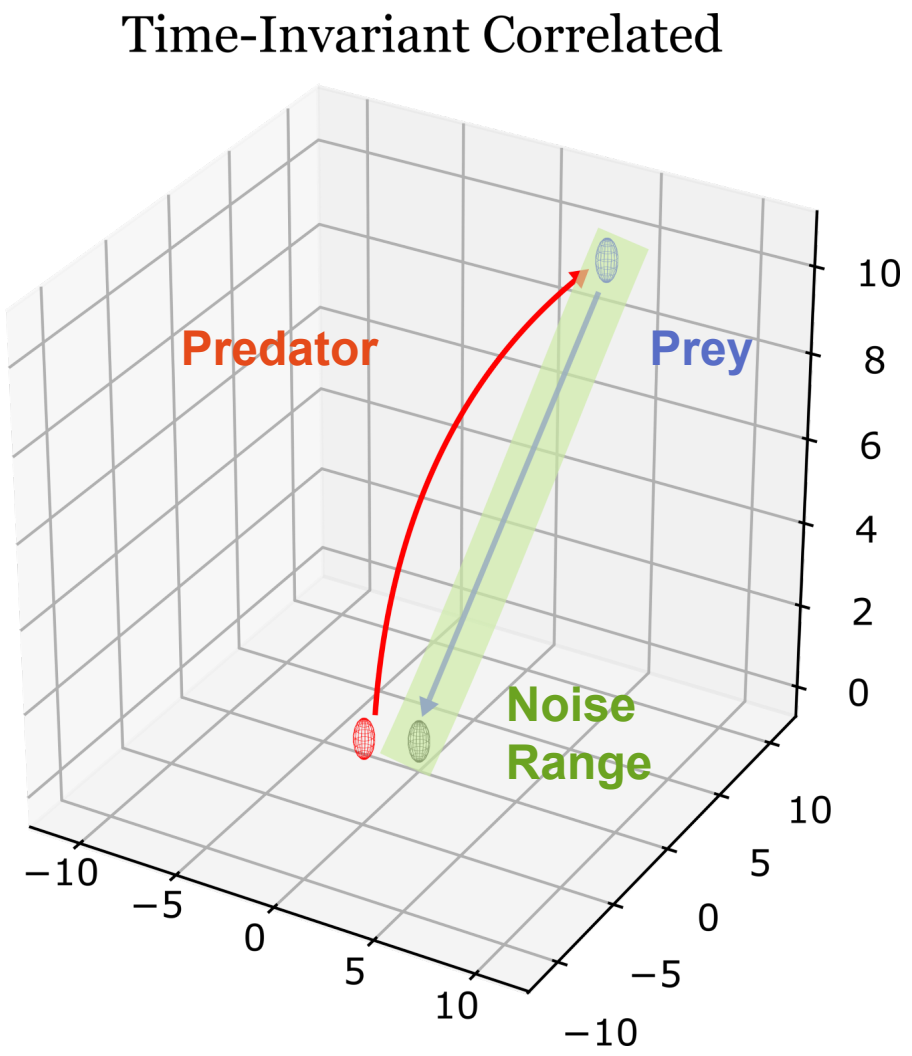


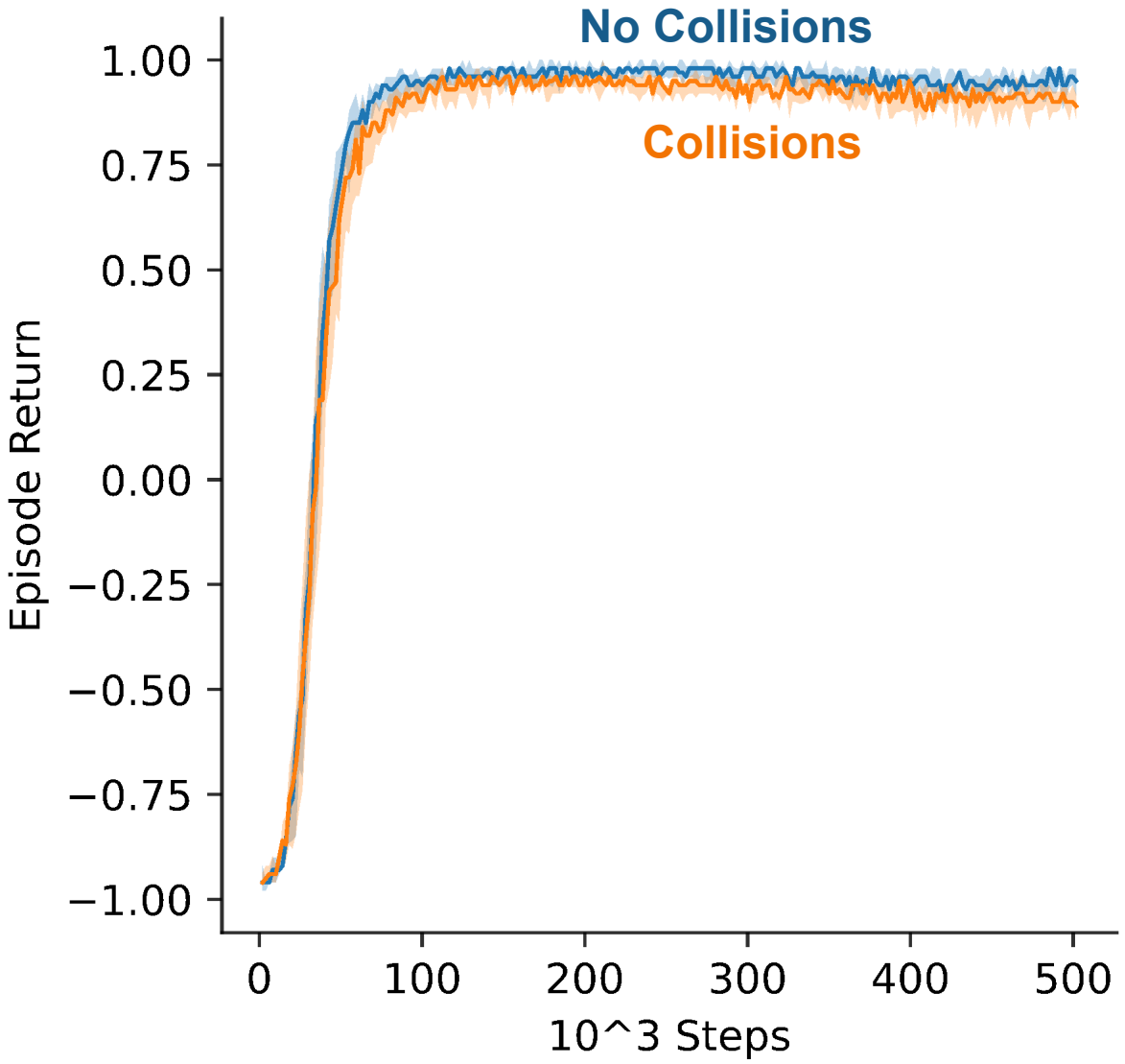
ADDING NOISE TO OBSERVATIONS





COLLISIONS


Time-Invariant Correlated
Predator
Prey
Noise
Range
−10
−5
0
5
10
10
8
6
4
2
0
10
5
0
−5
−10


No Collisions
Collisions
Episode Return
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
0
100
200
300
400
500
10^3 Steps

COLLISIONS

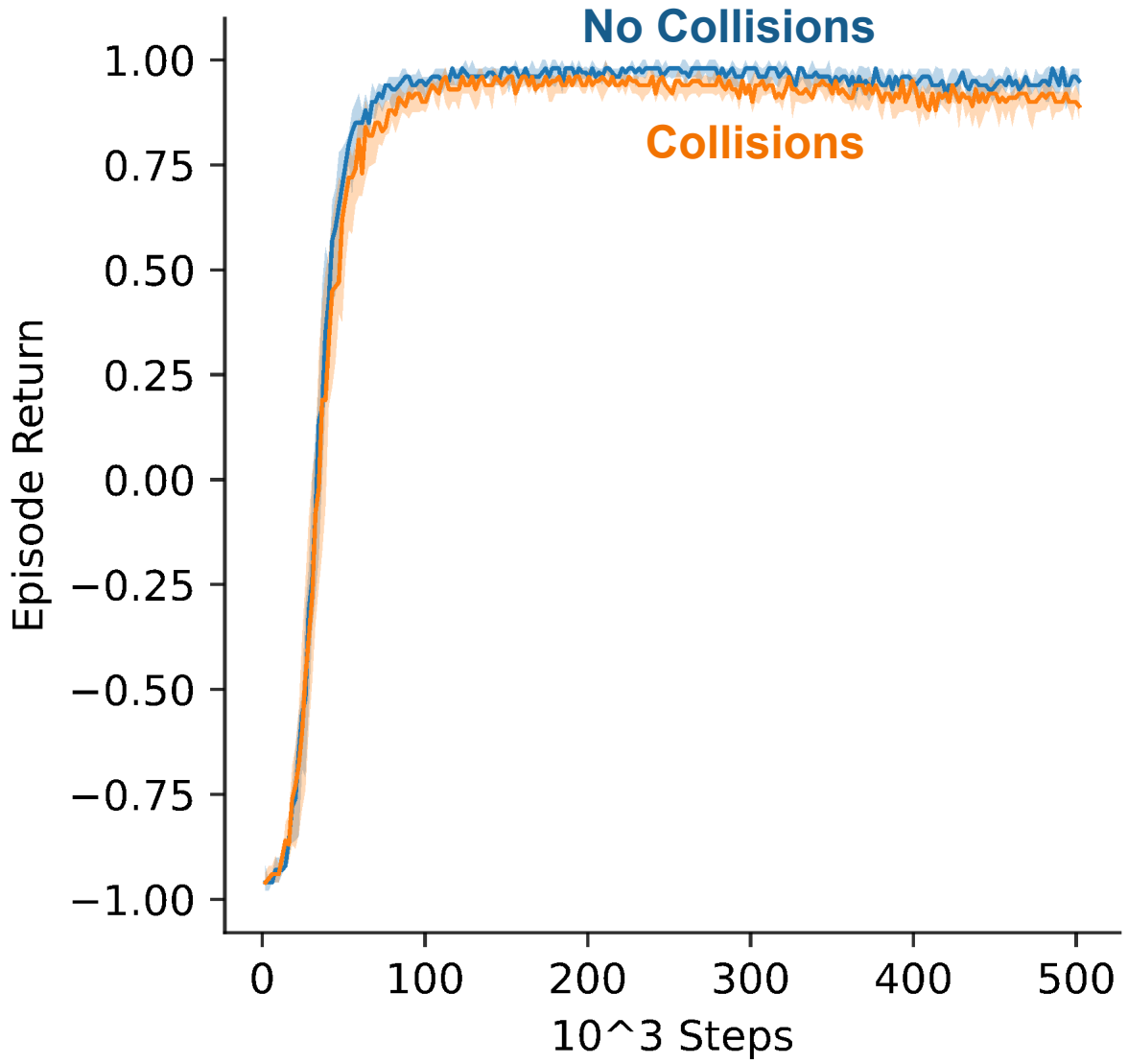
- ~~**Complicated Reward Space:**~~

  ~~**Collision with Predators**~~

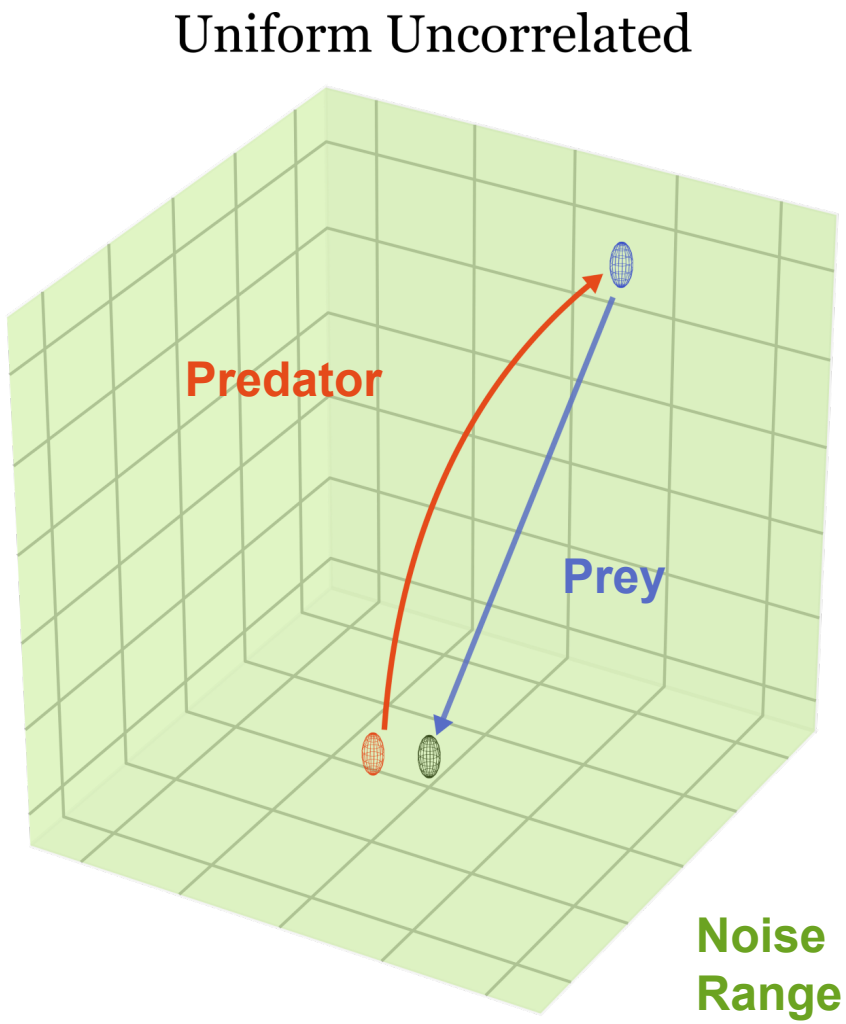
- Larger Observation and Action Spaces:

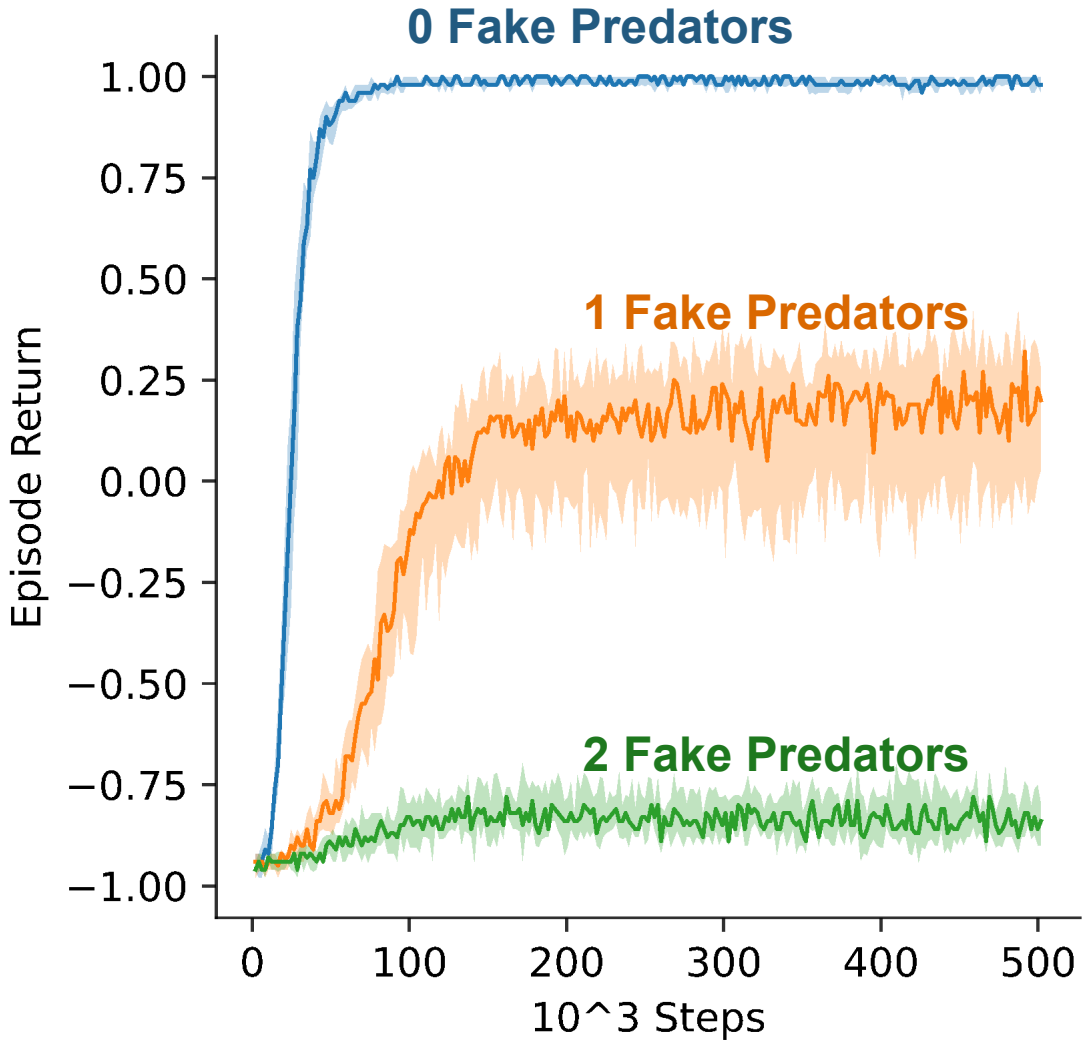
  Curse of Dimensionality

- Time-Varying, Correlated
  - Predator Correlation: Time-varying
  - Prey Correlation: Time-Invariant
  - Uniform Uncorrelated



CURSE OF DIMENSIONALITY





CURSE OF DIMENSIONALITY

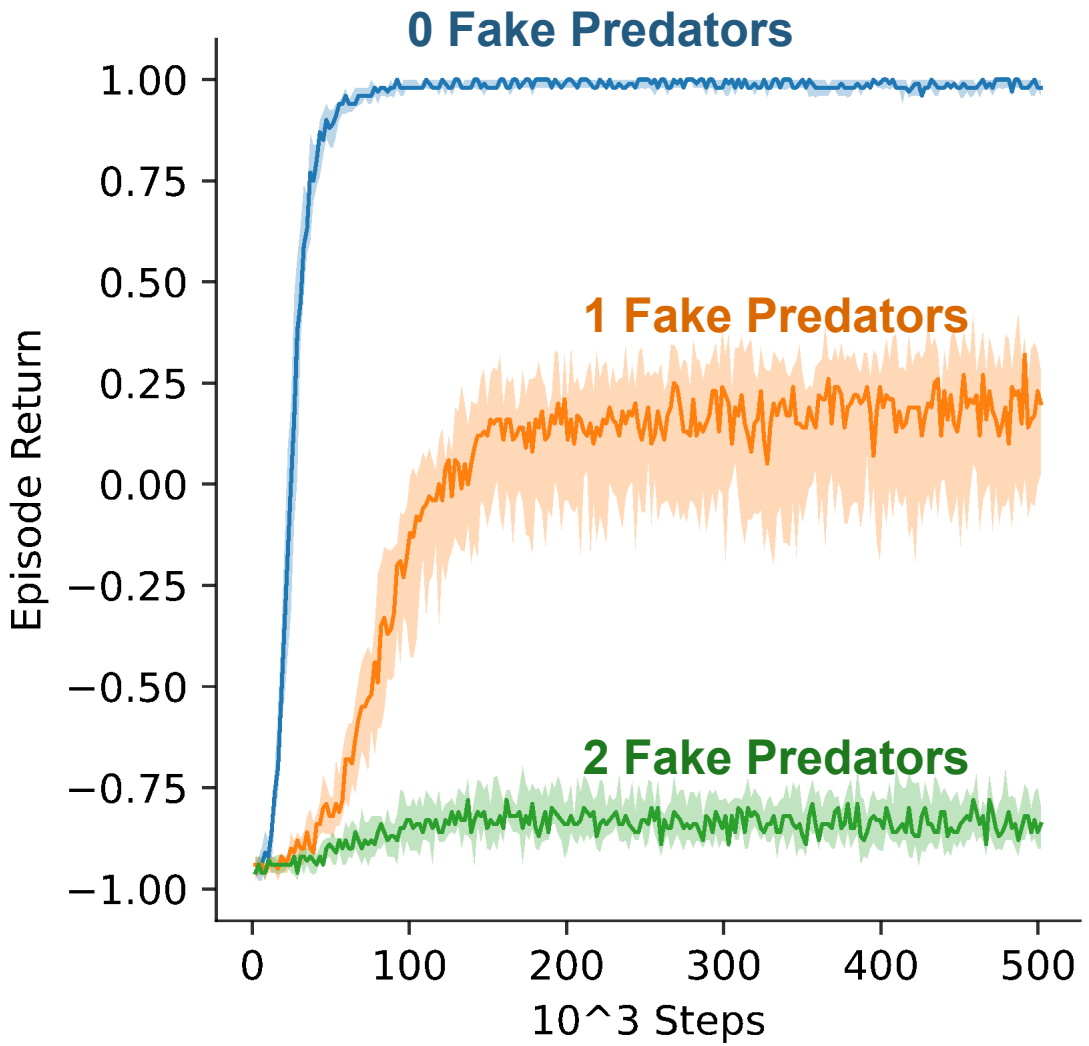
- ~~Complicated Reward Space:~~

  ~~Collision with Predators~~

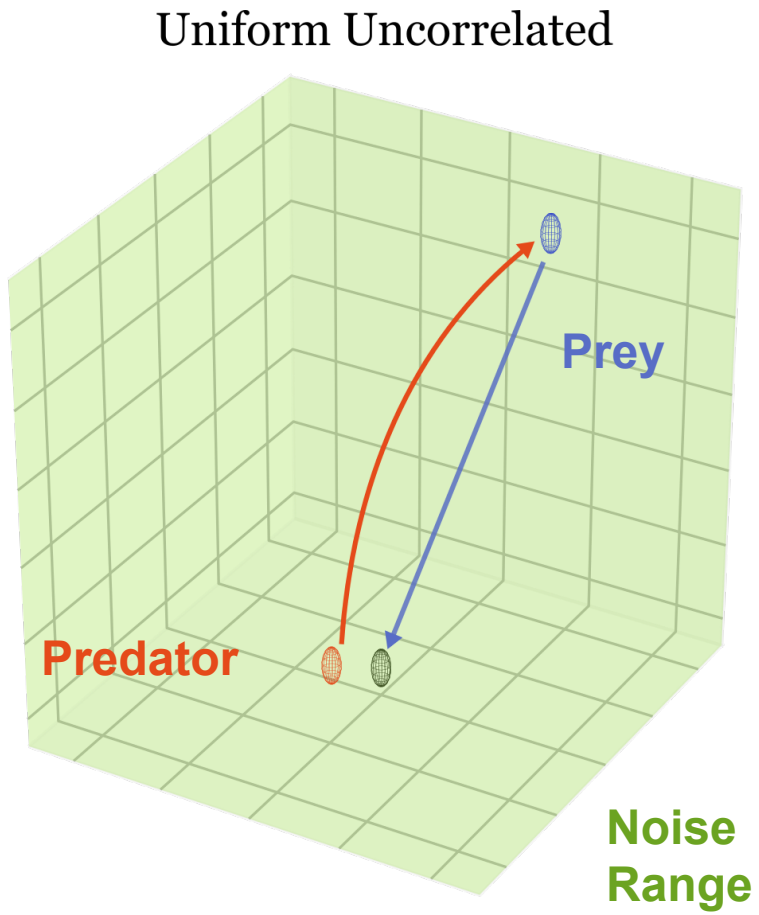
- **Larger Observation and Action Spaces:**

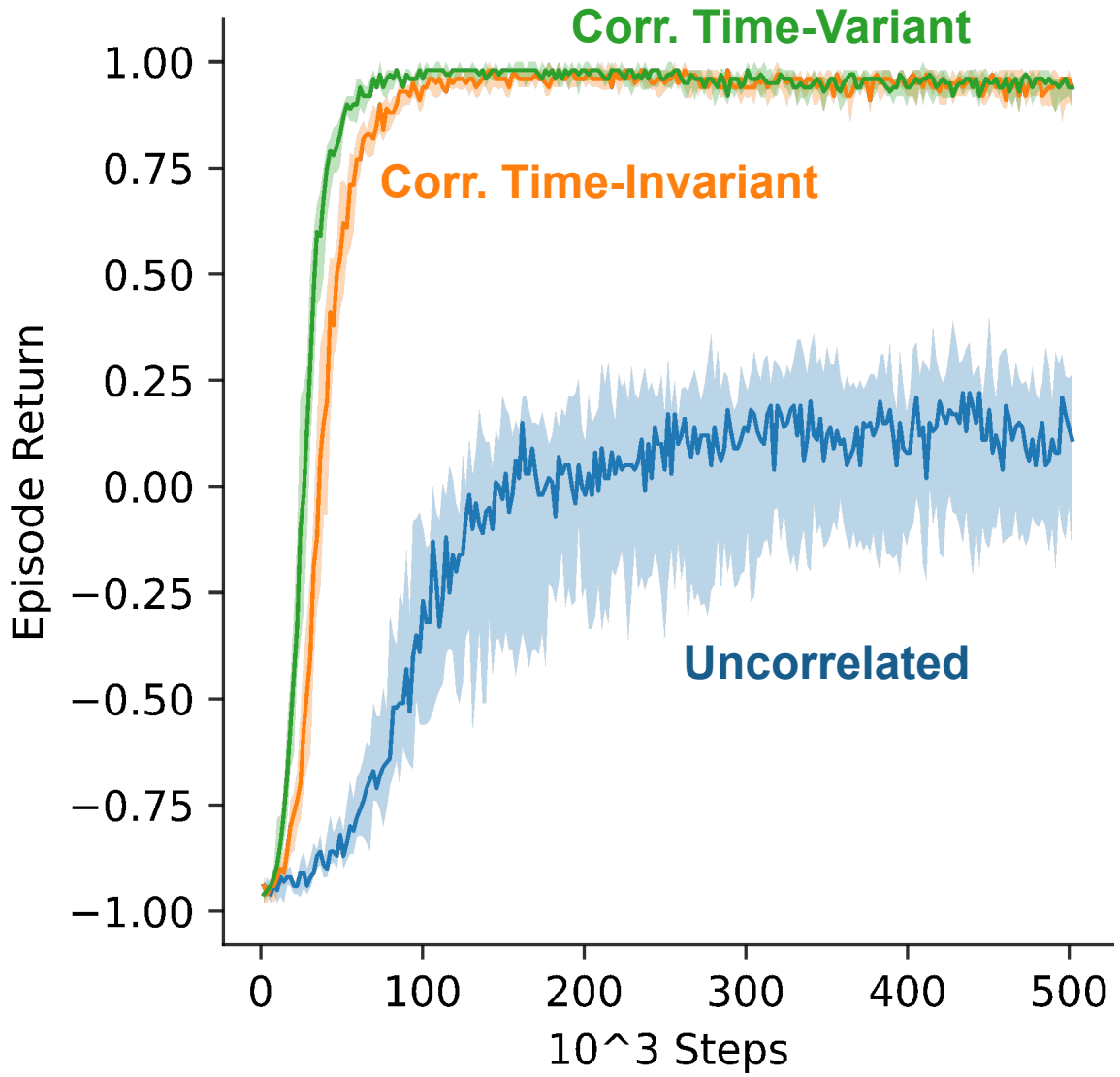
  Curse of Dimensionality

- Time-Varying, Correlated Observations:
  - Predator Correlation: Time-varying
  - Prey Correlation: Time-Invariant
  - Uniform Uncorrelated



NOISE CORRELATION





NOISE CORRELATION

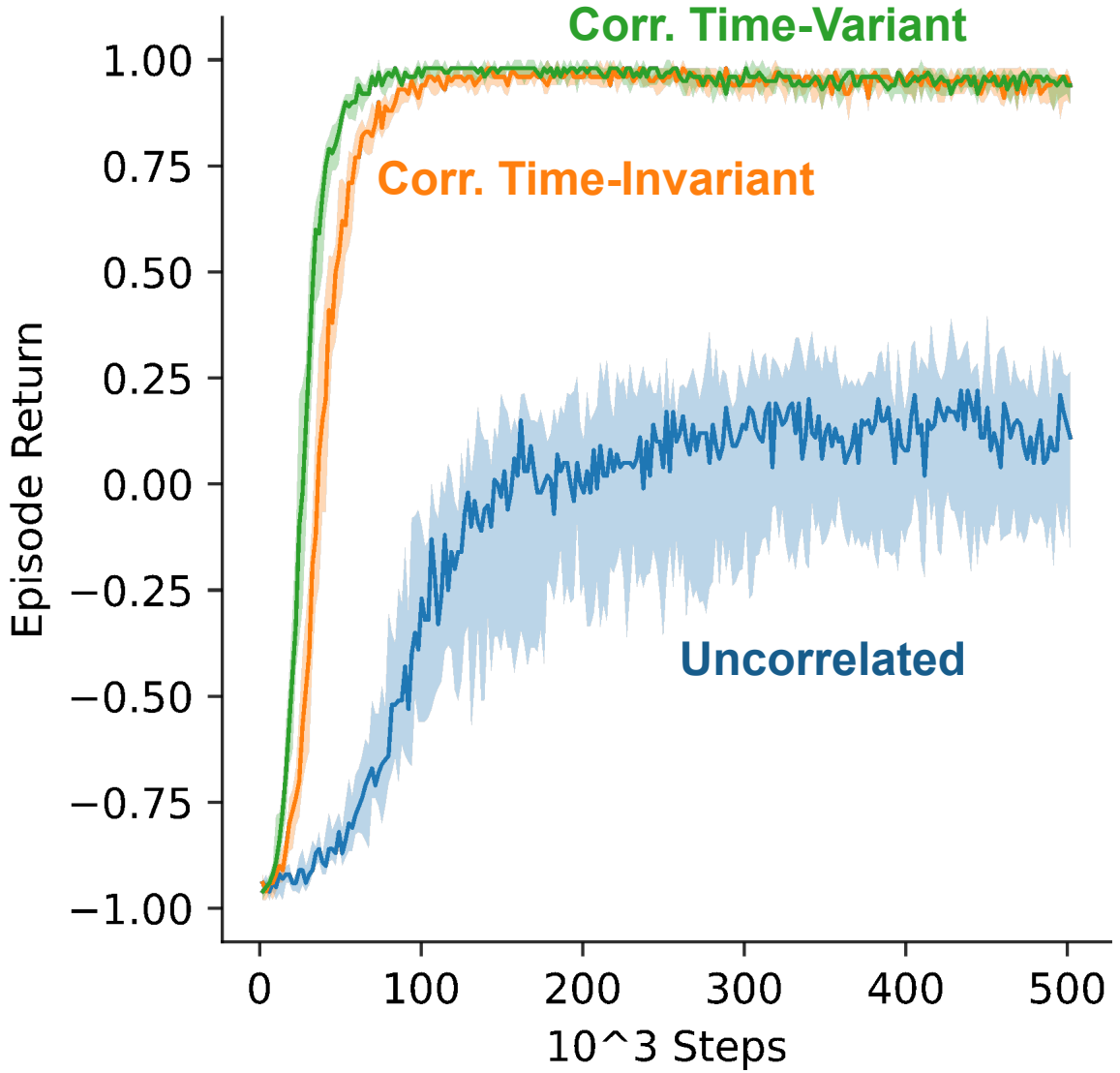
- ~~Complicated Reward Space:~~

  ~~Collision with Predators~~

- Larger Observation and Action Spaces:

  Curse of Dimensionality

- **Time-Varying, Correlated Observations:**
  - ~~**Predator Correlation: Time-varying**~~
  - ~~**Prey Correlation: Time-Invariant**~~
  - **Uniform Uncorrelated**



SUMMARY

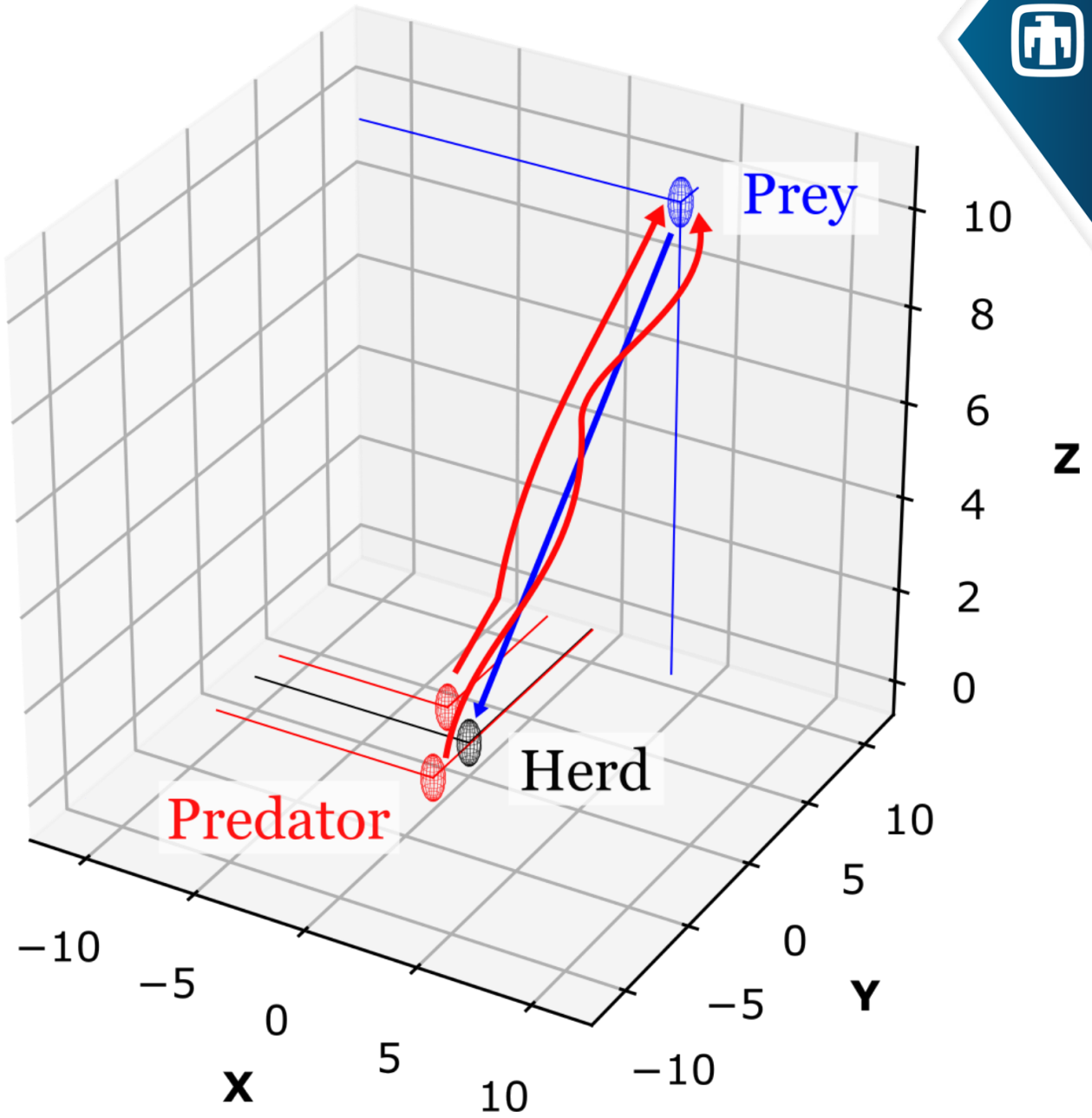
- 40% increase over baseline with our custom MARL architecture in most difficult scenario.

- Increasing the number of predators leads to decreasing performance, demonstrating the curse of dimensionality.

- We showed performance decreases due to uncorrelated noise rather than predator or prey correlated noise.

FUTURE WORK

- Minimize effects of observation noise:
  - Modeling noise explicitly via a model or implicitly with Recurrent Neural Networks.
  - Use regularization, which shrinks parameters towards 0, to help reduce noise impact.
- Explore other RL algorithms with our architecture and their impact on learning speed.
  - Our method allows us to use single-agent algorithms to solve multi-agent problems.
- Increase the scenario difficulty with prey that can learn and evade.



THANK YOU!

LDRD LABORATORY DIRECTED RESEARCH & DEVELOPMENT