Mixed-Initiative Multi-Agent Apprenticeship Learning for Human Training of Multi-Robot Teams

Esmaeil “Esi” Seraj

Neural Information Processing Systems (NeurIPS) 2023

December 2023

Authors: Esmaeil Seraj, Jerry Xiong, Mariah Schrum, Matthew Gombolay

CORE ROBOTICS LAB

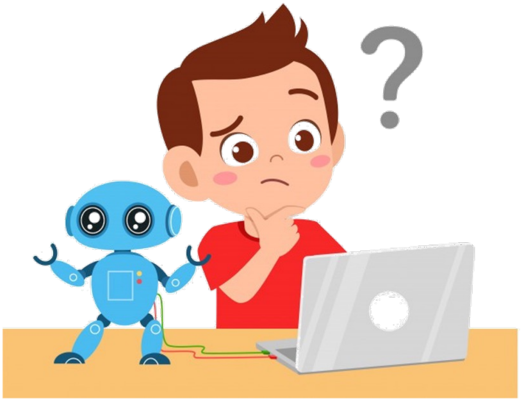
NEURAL INFORMATION PROCESSING SYSTEMS

Georgia Tech | Robotics & Intelligent Machines

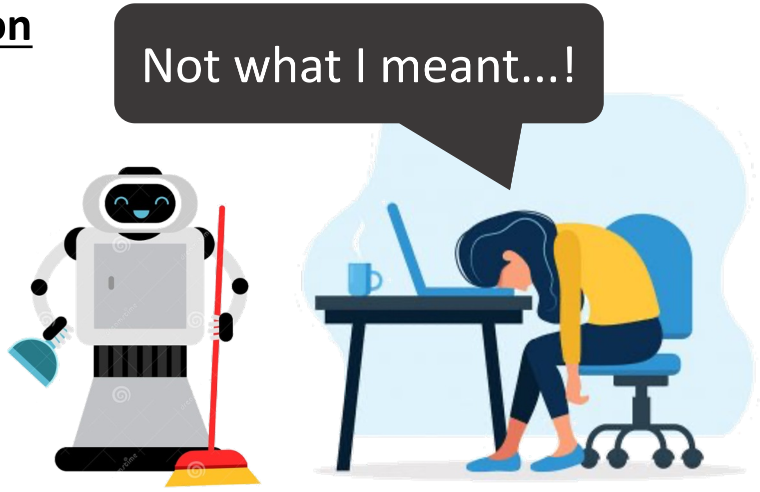
Can We Directly Teach Robots to Coordinate by Showing Them How to?

Learning Multi-Agent Coordination and Collaboration Policies from Expert Human Demonstration

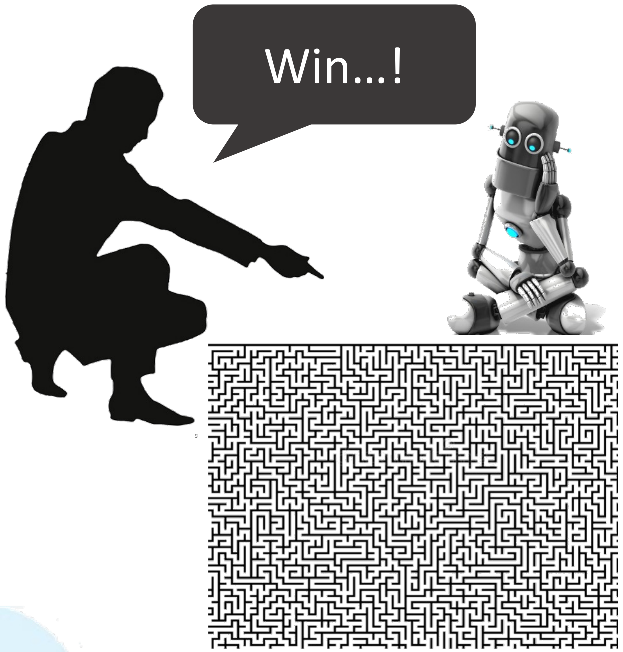
Why Learning from Human Demonstrations?



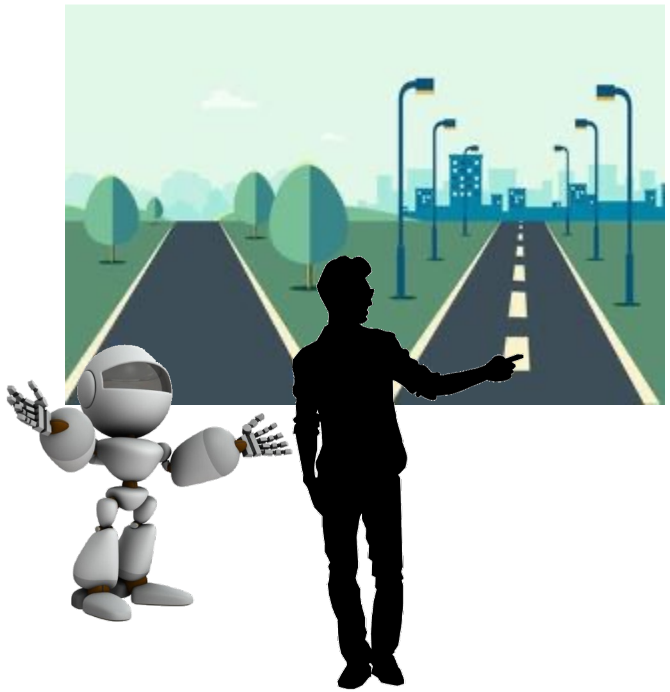
Reward Specification



Reward Expressiveness



Domain Complexity



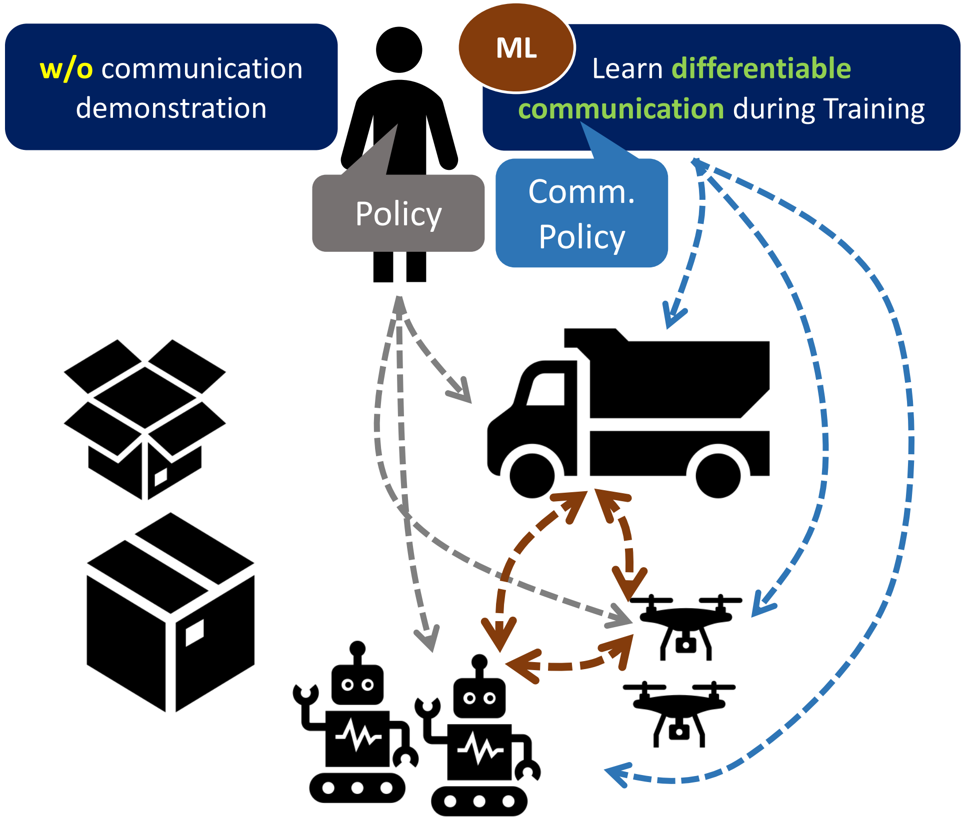
Human’s Preferred Way

Abel, David, et al. "On the Expressivity of Markov Reward." *Advances in Neural Information Processing Systems* 34 (2021).

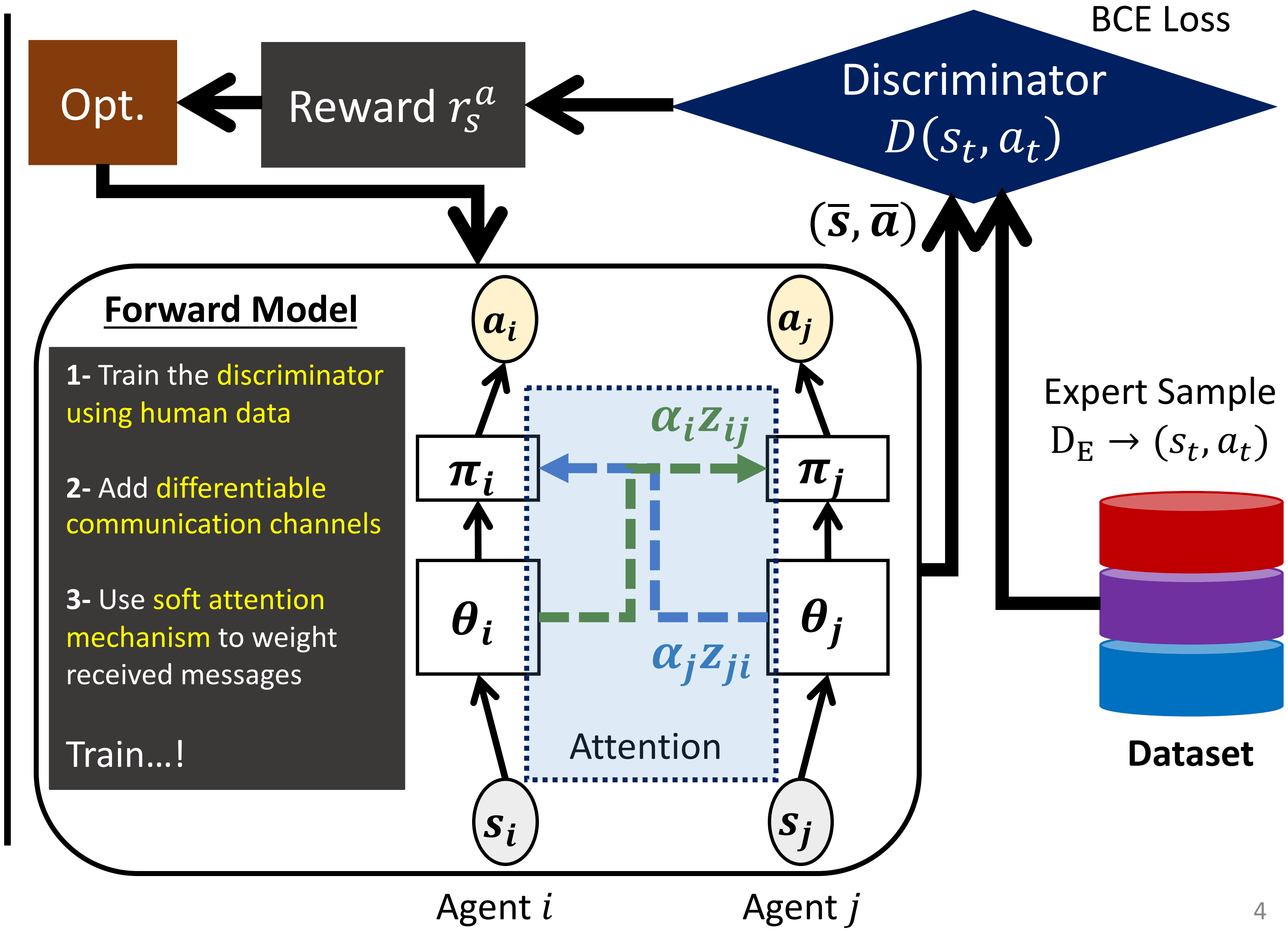
Matignon et al. "Reward function and initial values: Better choices for accelerated goal-directed reinforcement learning." *International Conference on Artificial Neural Networks*. Springer, Berlin, Heidelberg, 2006.

Mixed-Initiative Multi-Agent Apprenticeship Learning (MixTURE) for Human Training of Multi-Robot Teams

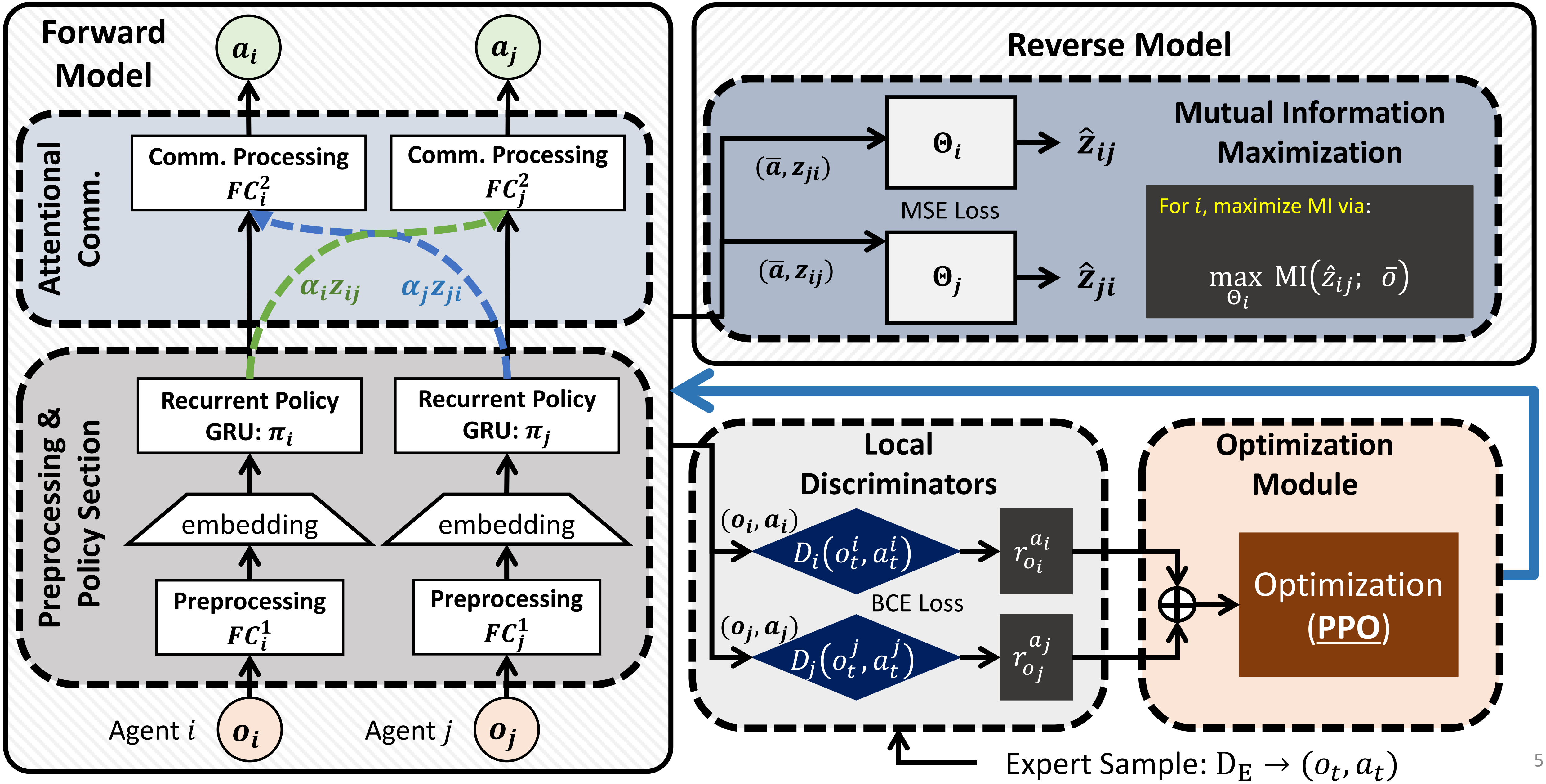
Single Human → Robot Teams



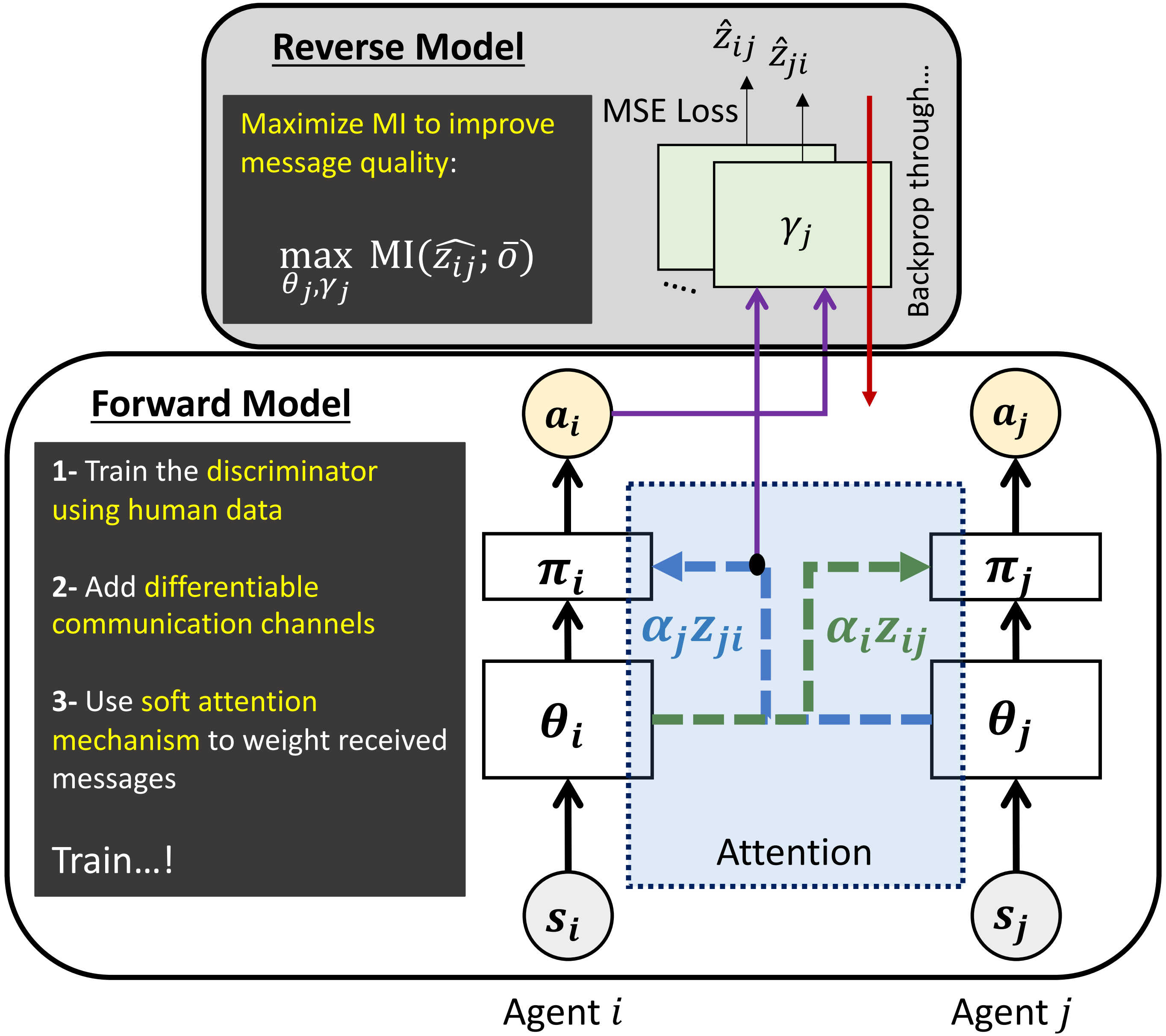
- One human expert can do the job
- Communication will be learned, and heterogeneous interaction is possible
- Much easier to provide demonstration



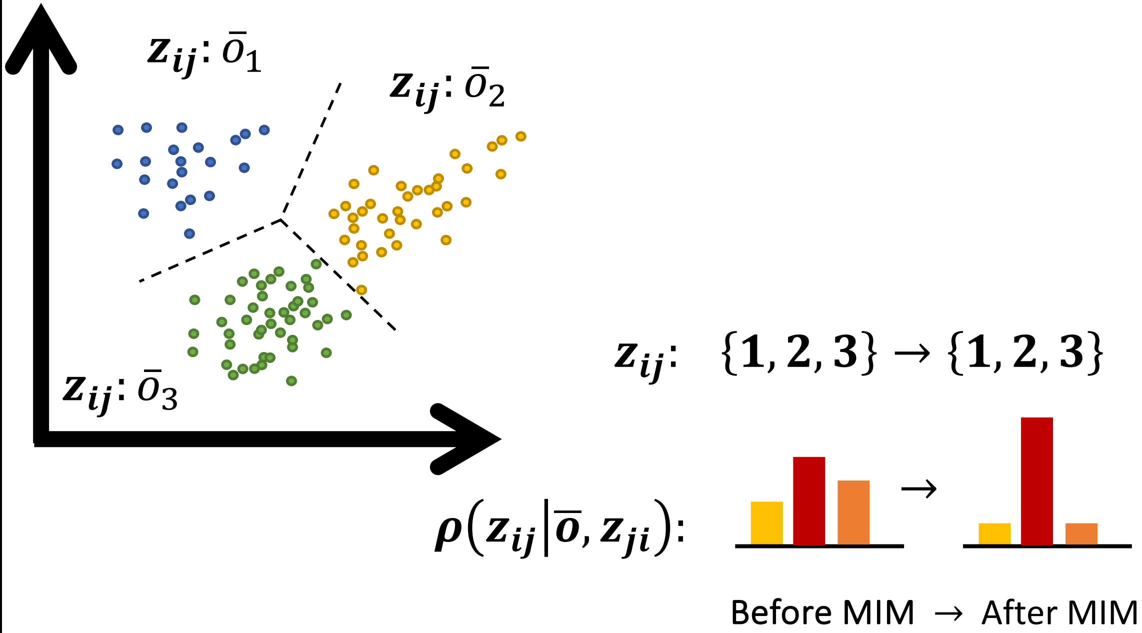
MixTURE: Mixed-Initiative Multi-Agent Apprenticeship Learning



Mutual Information Maximization for Differentiable Communication



- **Improve message distribution given joint observation:**



Make the communication more semantically meaningful based on obs.

Human Subject Study Flow

Recap

- **(RQ1) Can the MixTURE architecture learn useful coordination strategies from synthetic data (models of human experts)?**
  - Evaluate the quality of learned policies against SOTA baselines and ablations to confirm performance and sample efficiency.

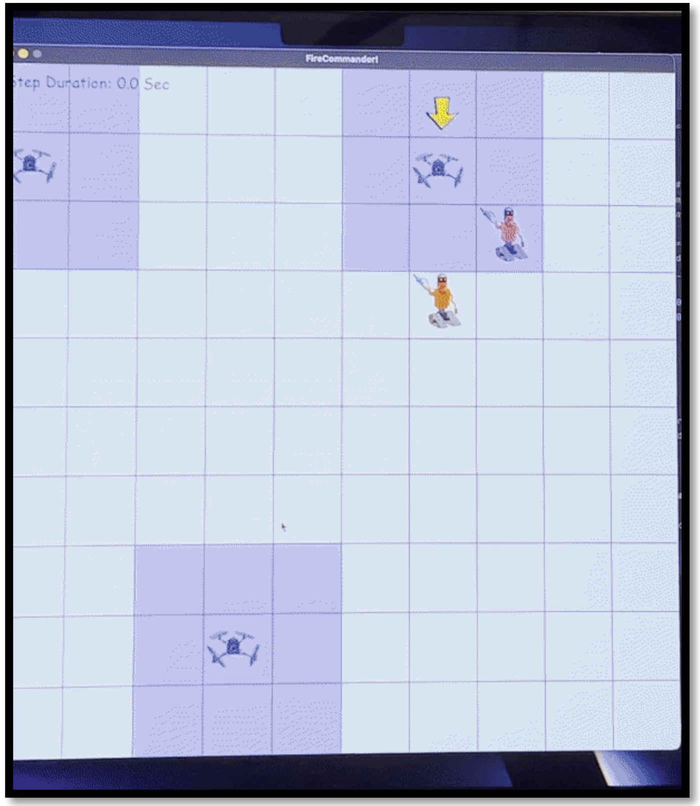
- **(RQ2) Is the MixTURE architecture applicable to learning from real human data?**
  - Evaluate the performance against baseline with expert demonstrated communication.

- **(RQ3) How challenging is it for human experts to provide multi-agent demonstration and does MixTURE alleviate the challenge as compared to classic MA-LfD architectures?**
  - Compare Workload Scores (WS) for cases when a subject uses the MixTURE vs. a classical MA-LfD architecture.
  - Compare System Usability Scores (SUS) for cases when a subject uses the MixTURE vs. a classical MA-LfD architecture.

Human Subject Study Flow

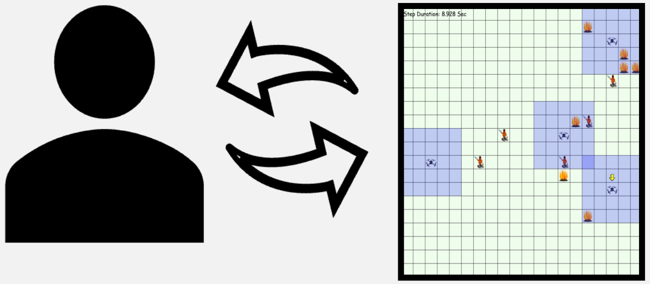
Environment

FireCommander

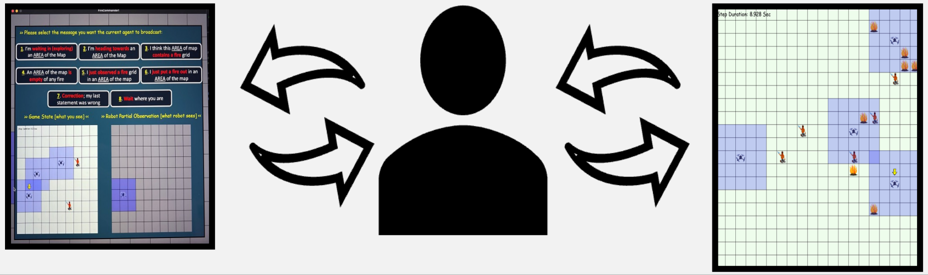


Conditions

1- ***noComm*** Condition: only demonstrate environment actions for each agent



2- ***withComm*** Condition: demonstrate both an environment action and a comm. action (message) to be broadcasted for each agent



Metrics

1- ***Game score***: a function of existing, found, and killed firespots

2- ***Learned policy performance***: deploy learned policies in env.

3- ***Scalability***: number of tasks completed by human

4- ***Time required for demo***

5- ***Workload***

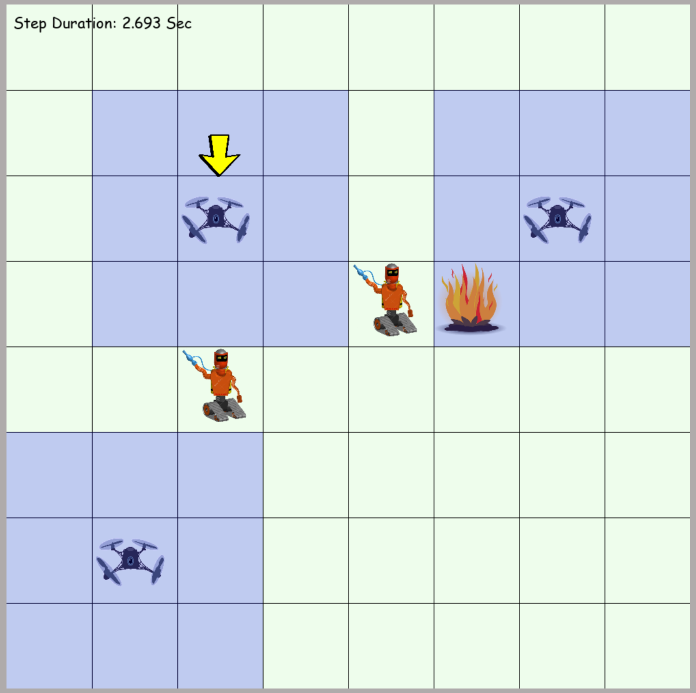
6- ***Usability Score***

55 subjects, *within*-subject study, *GT* students (34.5% female), avg. age of 25 ± 2.6

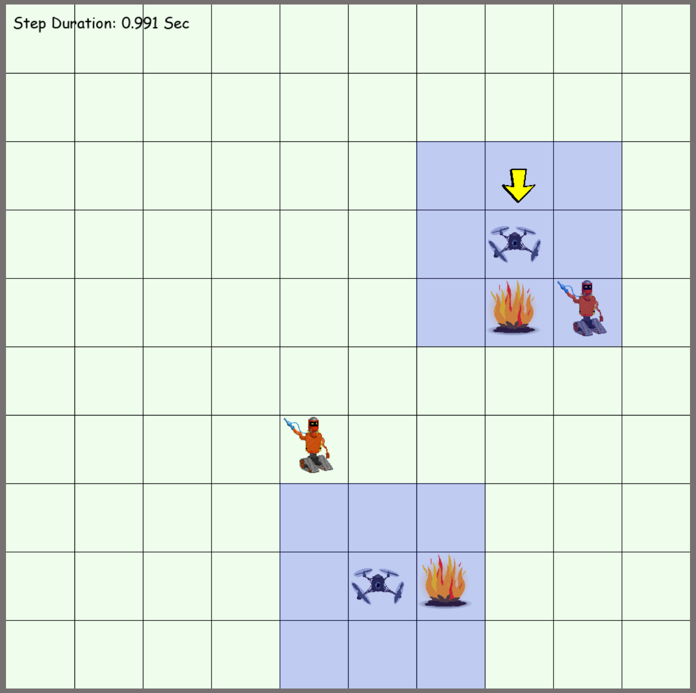
Human-Subject Dataset

- **Baseline Comparison**: Evaluate the learned policy via MixTURE and MA-LfD baselines on real human data.

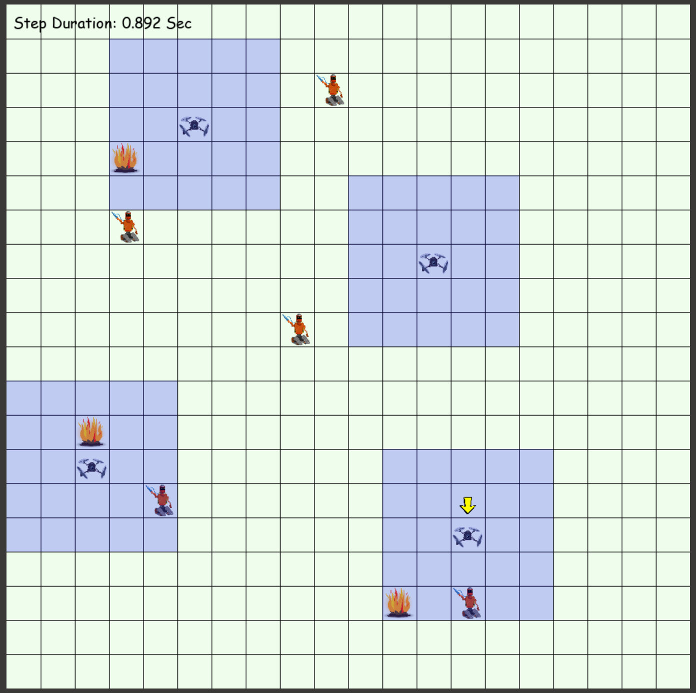
Easy scenario: 8×8 domain, 5 agents (3P, 2A), 1 initial fire



Medium scenario: 10×10 domain, 4 agents (2P, 2A), 5 initial fires



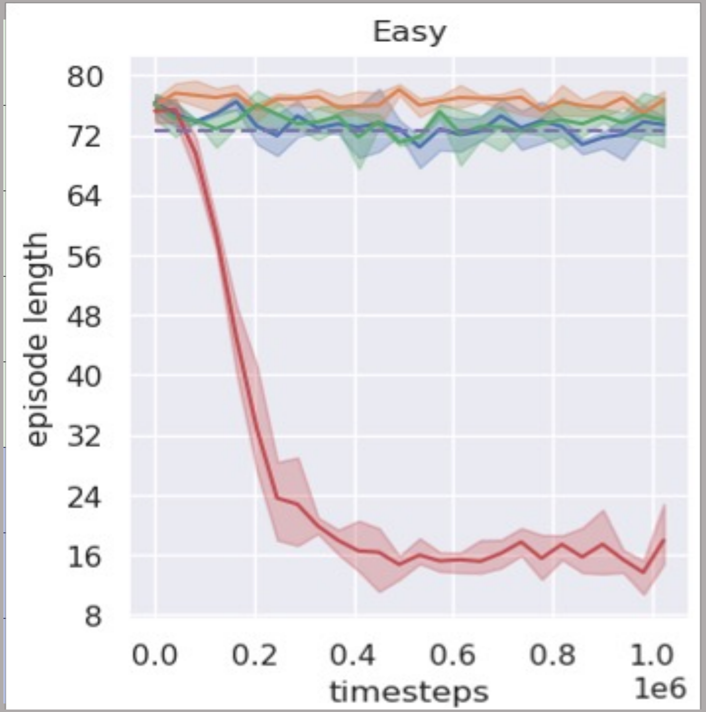
Hard scenario: 20×20 domain, 10 agents (4P, 6A), 10 initial fires



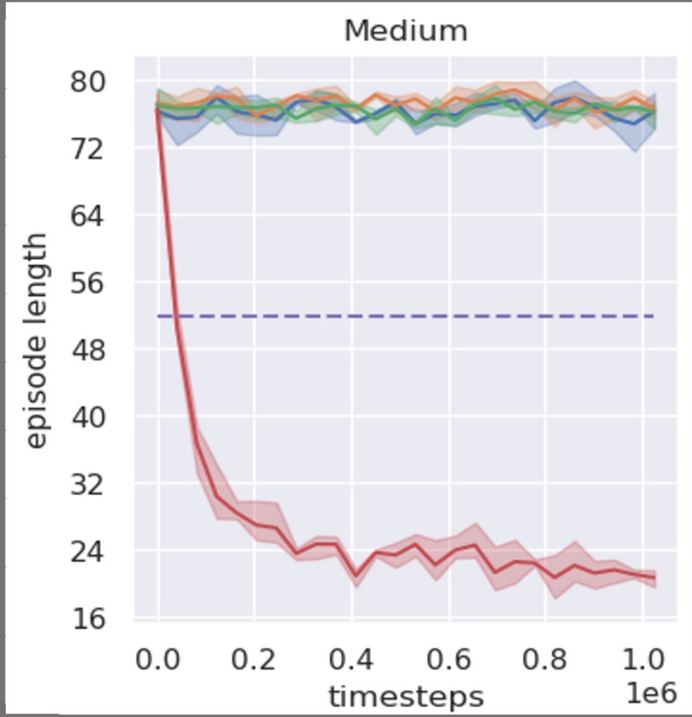
Human-Subject Dataset

- **Baseline Comparison**: Evaluate the learned policy via MixTURE and MA-LfD baselines on real human data.

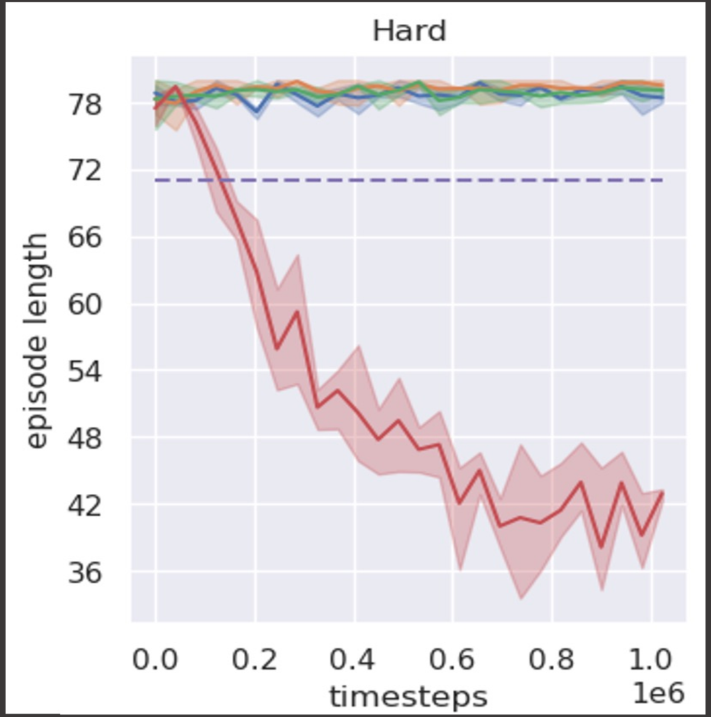
Easy scenario: 8×8 domain, 5 agents (3P, 2A), 1 initial fire



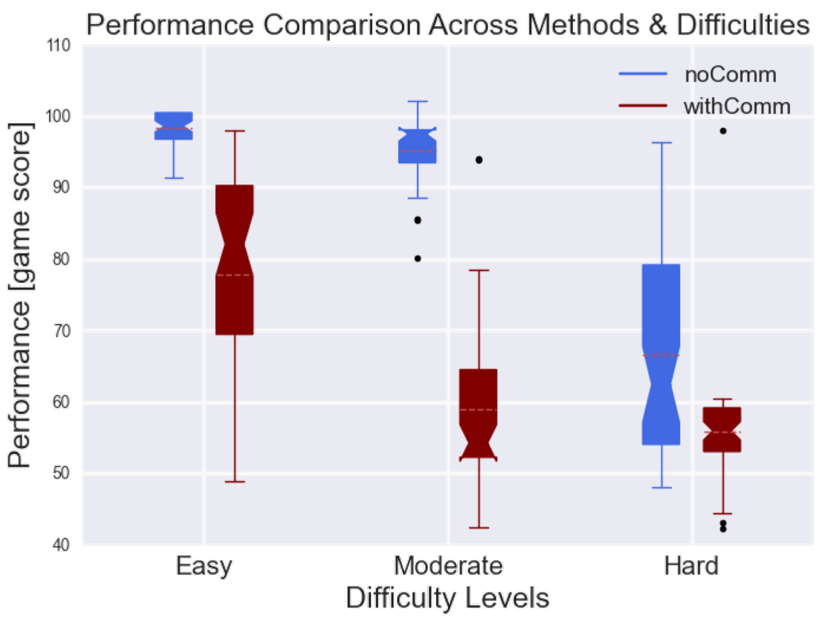
Medium scenario: 10×10 domain, 4 agents (2P, 2A), 5 initial fires

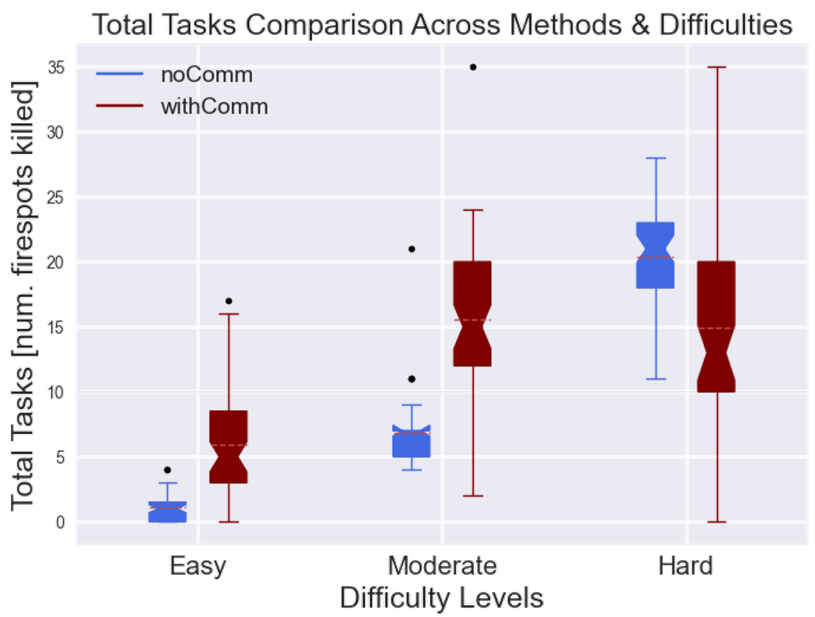


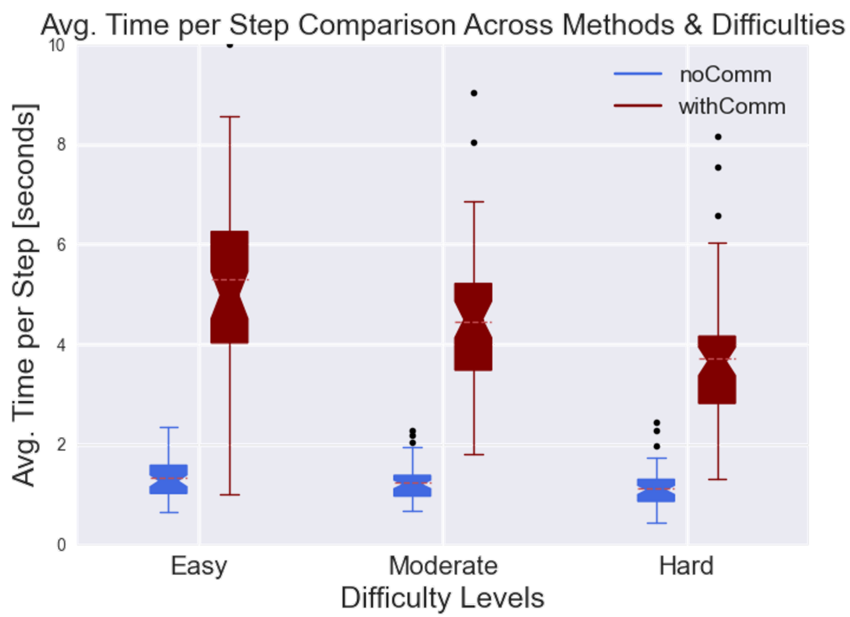
Hard scenario: 20×20 domain, 10 agents (4P, 6A), 10 initial fires



Objective Results







Summary

(1) Performance: Demonstrating communication for a multi-agent team significantly ($p < .001$) reduces the human performance in FC task.

(2) Avg. Time per Demonstration Step: Demonstrating communication for a multi-agent team significantly ($p < .001$) increases the demonstration time in FC task.

(3) Total Tasks Completed: Demonstrating communication for a multi-agent team significantly ($p < .001$) reduces the human's ability to accomplish tasks in FC.

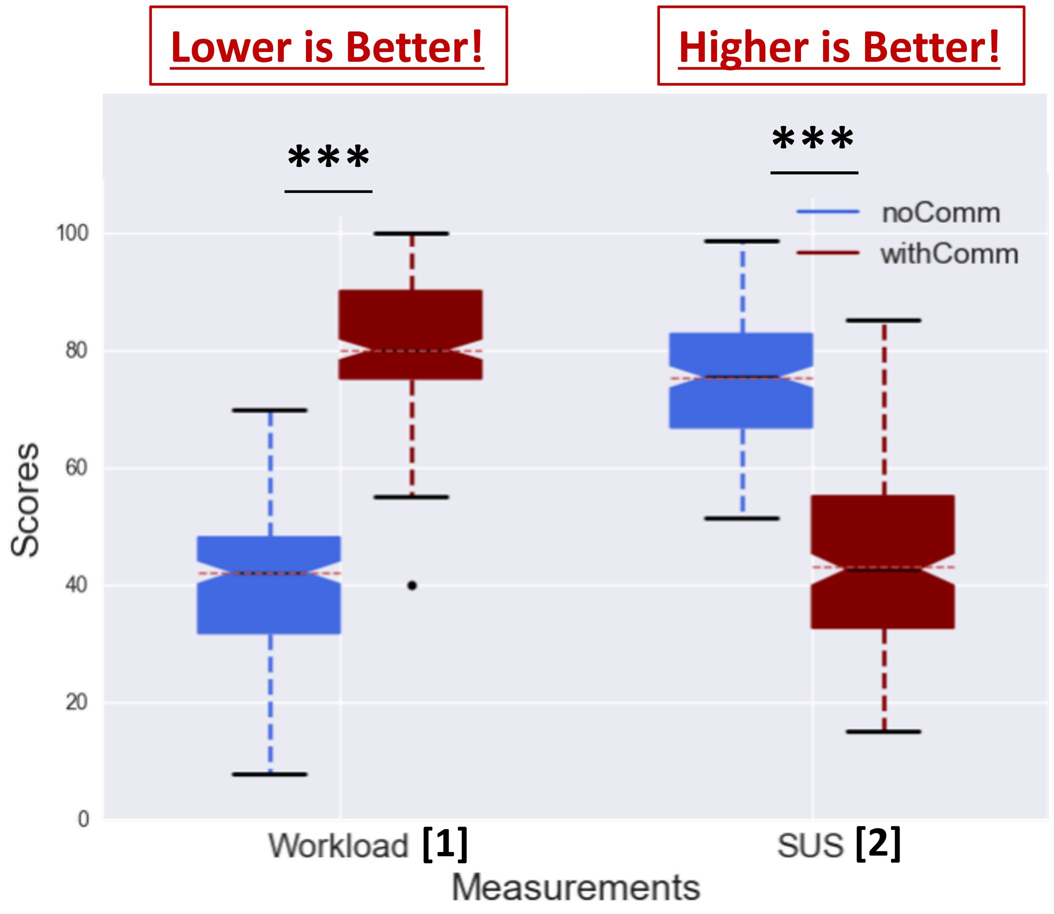
Subjective Results

Summary

(1) Workload Score – NASA TLX [1]: Demonstrating communication for a multi-agent team significantly ($p < .001$) increases the human workload in FC task (increase by 44.3%).

(2) System Useability Scale [2]: Demonstrating communication for a multi-agent team significantly ($p < .001$) reduces the system usability score for FC task (decrease by 46.7%).

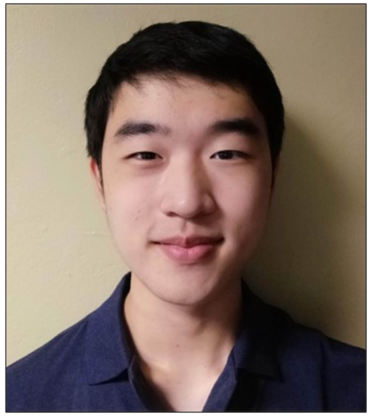
Using MixTURE bypasses the communication demonstration step and therefore leads to **lower workload** and **higher system usability score** by a human expert.



[1] Hart, Sandra G., and Lowell E. Staveland. "Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research." Advances in psychology. Vol. 52. North-Holland, 1988. 139-183.

[2] Brooke, John. "SUS-A quick and dirty usability scale." Usability evaluation in industry 189.194 (1996): 4-7.

NEURAL
INFORMATION
PROCESSING
SYSTEMS



Thank you!

SCAN ME

U.S. NAVAL
RESEARCH
LABORATORY

CORE
ROBOTICS LAB

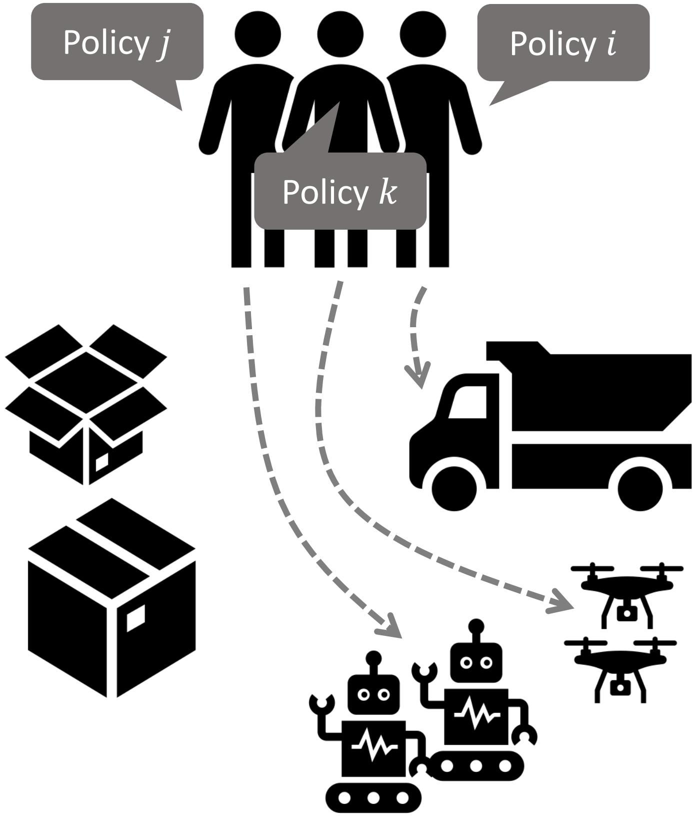
Georgia Tech

Georgia Tech | Robotics & Intelligent Machines

Appendix

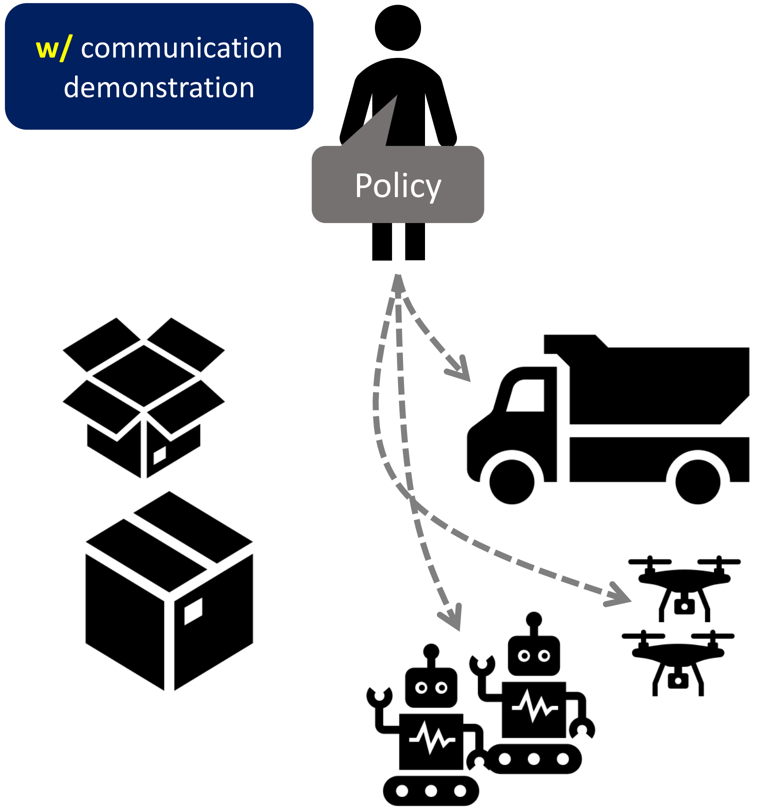
How to Incorporate Human Data for Learning Heterogeneous Multi-Agent Coordination?

Human Teams → Robot Teams



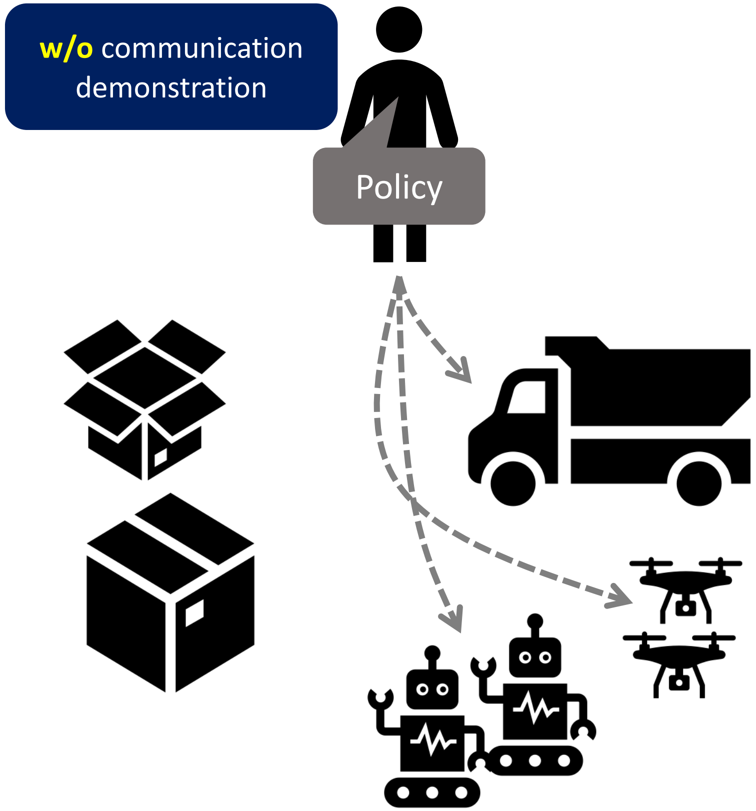
- One human expert can do the job ✗
- Need comm. & coordination among human demonstrators ✗
- Hard to translate to robot domain ✗

Single Human → Robot Teams



- One human expert can do the job ✓
- Comm. needs to be a part of the action-space ✗
- Message-space must be known ✗

Single Human → Robot Teams



- One human expert can do the job ✓
- Comm. is still a necessity & w/o it, agents cannot coordinate ✗
- Much easier to provide demonstration ✓

Empirical Evaluation: Research Questions

- **Three main research questions**:

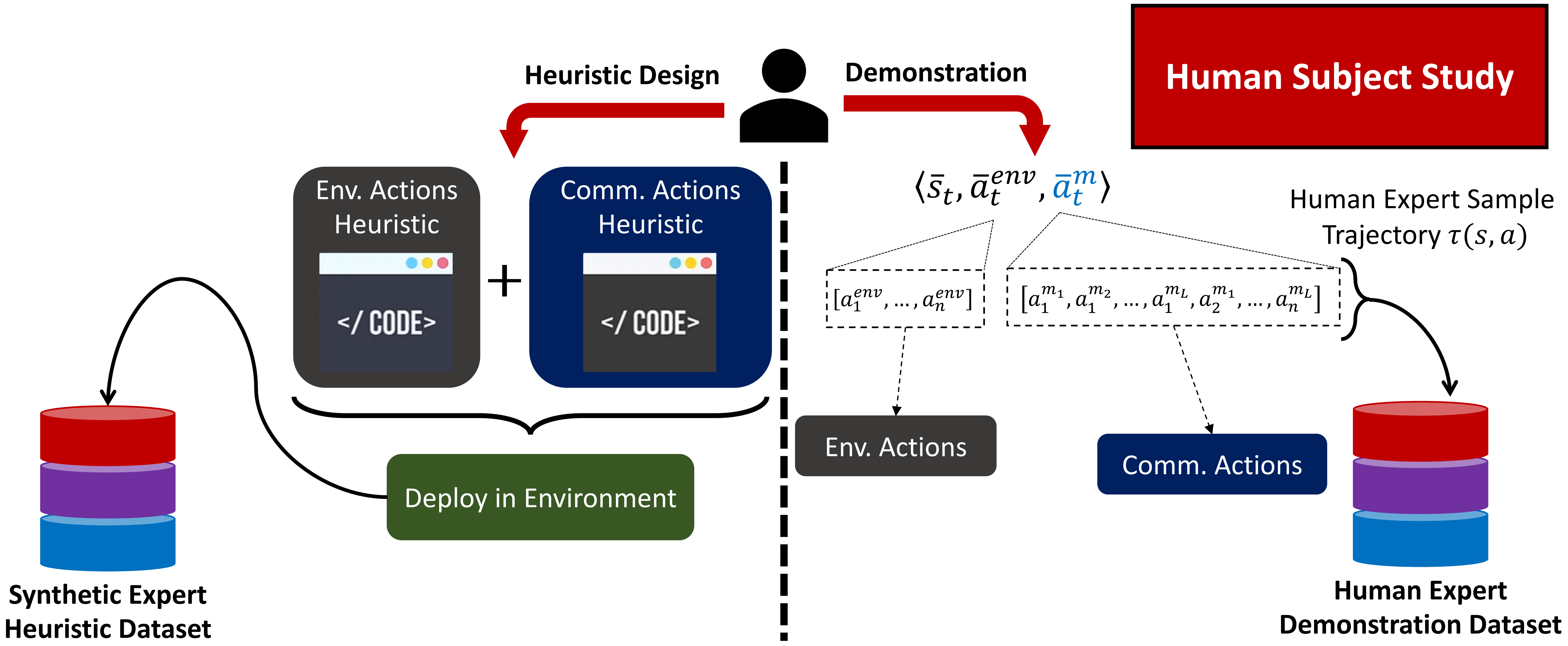
- **(RQ1) Can the MixTURE architecture learn useful coordination strategies from synthetic data (models of human experts)?**
  - Evaluate the quality of learned policies against SOTA baselines and ablations to confirm performance and sample efficiency.

- **(RQ2) Is the MixTURE architecture applicable to learning from real human data?**
  - Evaluate the performance against baseline with expert demonstrated communication.

- **(RQ3) How challenging is it for human experts to provide multi-agent demonstration and does MixTURE alleviate the challenge as compared to classic MA-LfD architectures?**
  - Compare **Workload Scores (WS)** for cases when a subject uses the MixTURE vs. a classical MA-LfD architecture.
  - Compare **System Usability Scores (SUS)** for cases when a subject uses the MixTURE vs. a classical MA-LfD architecture.

Empirical Evaluation: Evaluation Process

- **<u>Datasets</u>**: To investigate RQ1, RQ2, and RQ3:



Synthetic Expert Heuristic Dataset

- **Baseline Comparison**:

Easy scenario: 5×5 domain, 3 agents (2P, 1A), 1 prey or initial fire

Moderate scenario: 10×10 domain, 6 agents (3P, 3A), 1 prey or initial fire

Hard scenario: 20×20 domain, 10 agents (6P, 4A), 3 prey or initial fires

Summary

1- MixTURE outperforms all baselines, in all domains, and all levels of difficulty.

2- MixTURE improves sample complexity, the quality of learned policy at convergence, and can scale to various domain and robot team sizes.