# Counterfactuals and Multiple Rewards: Inducing and Explaining Good Team Behavior for Effective Agent-Human Teaming
# (FA9550-19-1-0195)


## PI: Kagan Tumer
## (Oregon State University)


**AFOSR Program Review:**
**Computational Cognition and Machine Intelligence Program**
**(October 6-8 2020, Arlington, VA)**

# Counterfactuals and Multiple Rewards: Inducing and Explaining Good Team Behavior for Effective Agent-Human Teaming

Kagan Tumer, Oregon State University

**Objective: To enable robust human-machine teaming for long-term autonomy**

- Enable complex agent-agent interactions and effective human-agent teaming
- Enable agents to prioritize "what matters when" and "what matters to human teammate"

**Approach:**

- Develop counterfactual-based rewards that incentivize globally cooperative behavior in agent-agent and human-agent teams
- Generate a Multi-Reward Learning paradigm that allows agents to prioritize rewards across time and space

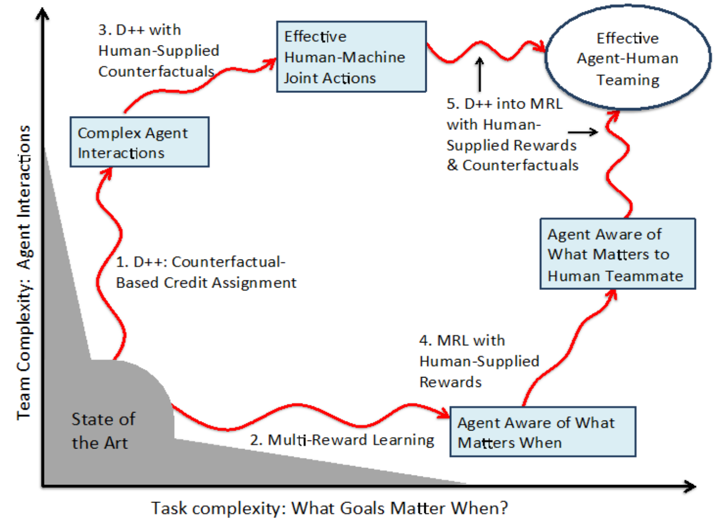
**DoD Benefits:**

- Incorporate human insight into agent decisions (enable human suggestions to agents)
- Allow agents to justify their actions (enable agent explanations to humans)

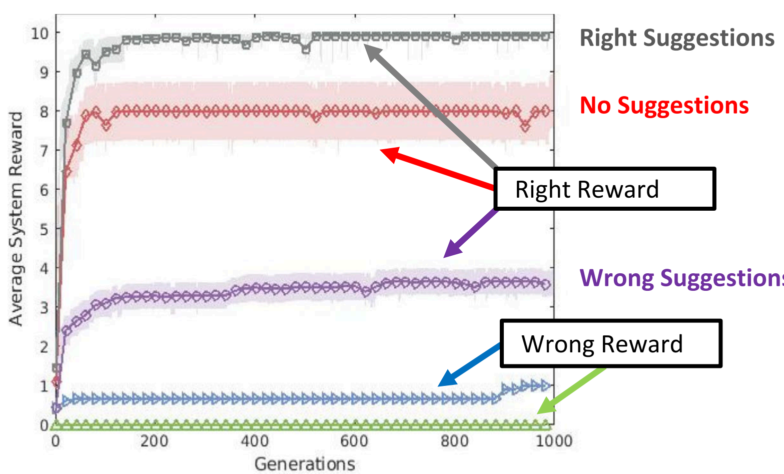
**Progress:**

- Devised counterfactual-based rewards that accept human suggestions on a tightly-coupled multi-robot coordination problem (AAMAS 2020)
- Devised Multi-Reward Learning (MRL) paradigm for agents to learn complex joint actions on a time-sensitive coordination problem (GECCO 2020)

## Project Progression Plan




## Impact of Rewards AND Suggestions



# List of Project Goals

1. Develop rewards based on counterfactual agents to incentivize globally cooperative behavior in agent-agent and human-agent teams

2. Generate a multi-reward learning paradigm that allows agents to prioritize rewards across time and space to enable long-term autonomy

3. Augment counterfactual agent rewards by enabling human team members to inject insight into agents' decision-making space

4. Augment multi-reward learning by enabling human team members to suggest local rewards/constraints for agents to include in their decision process

5. Insert counterfactual agent-based rewards into multi-reward learning paradigm to enable more complex insights and explanations to disseminate among human and agent teammates

# Progress Towards Goals

1. Develop rewards based on counterfactual agents to incentivize globally cooperative behavior in agent-agent and human-agent teams

   [AAMAS 2020: G. Dixit, S. Airiau, and K. Tumer]

   [AAMAS 2020: G. Rockefeller, S. Khadka, and K. Tumer]

2. Generate a multi-reward learning paradigm that allows agents to prioritize rewards across time and space to enable long-term autonomy

   [GECCO 2020: C. Yates, R. Christopher, and K. Tumer]

3. Augment counterfactual agent rewards by enabling human team members to inject insight into agents' decision-making space

4. Augment multi-reward learning by enabling human team members to suggest local rewards/constraints for agents to include in their decision process

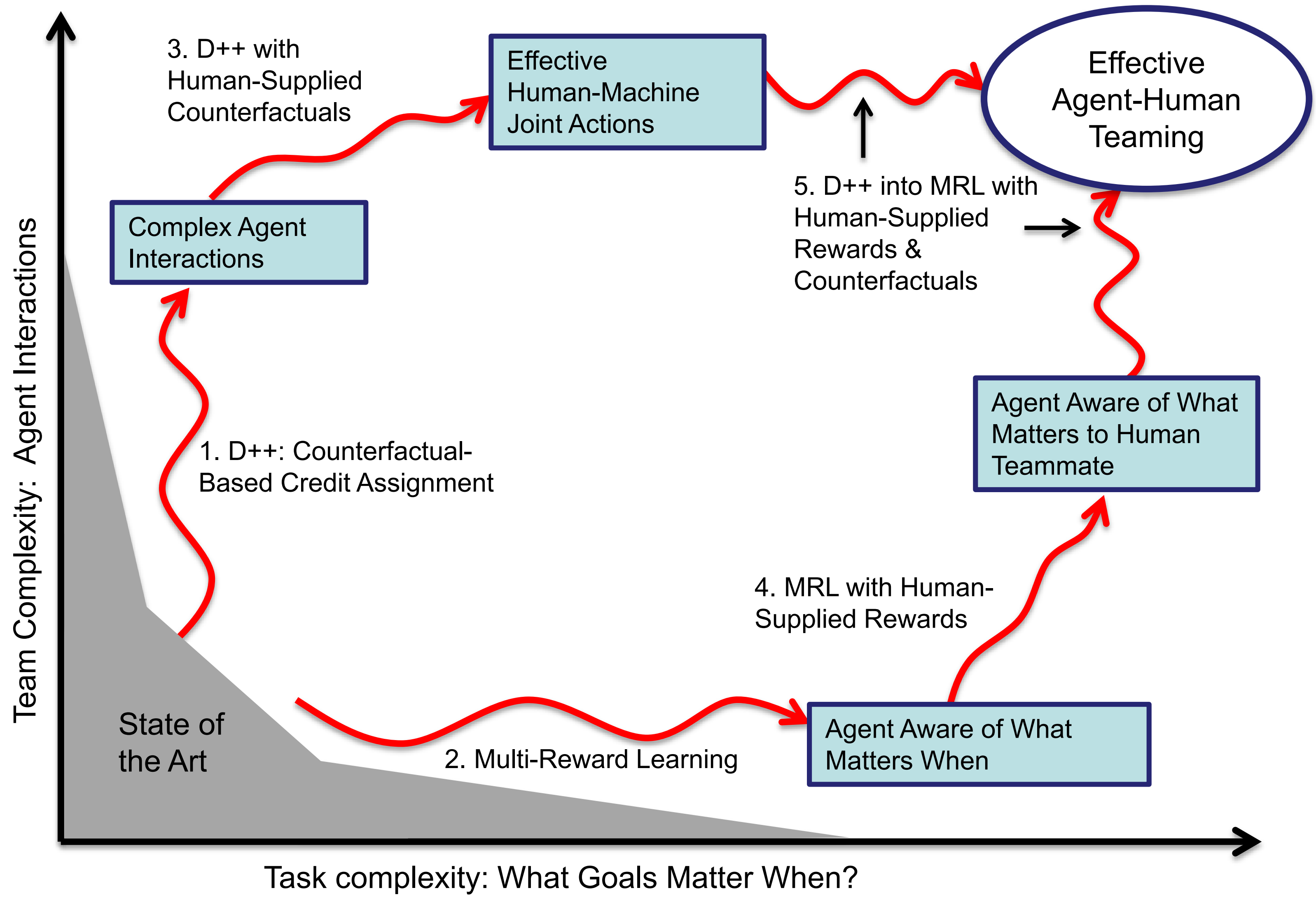
5. Insert counterfactual agent-based rewards into multi-reward learning paradigm to enable more complex insights and explanations to disseminate among human and agent teammates

# Counterfactuals & Multiple Rewards: Motivation

For effective teaming (agent-agent or human-agent) we need to address two key issues:

- How to discover good joint actions
  - Stepping stone reward for complex tasks
  - Suggestions
- How to evaluate complex sequences of actions
  - Multi-reward learning
  - Fitness Critic

# Project Roadmap



# Joint-Action Discovery

The probability of

**SUFFICIENT** agents picking the **RIGHT ACTION** at the **RIGHT TIME**

is **VANISHINGLY LOW**

Almost all learning based on one assumption:

You have to stumble upon the right action by accident

# Counterfactual Stepping Stone Rewards

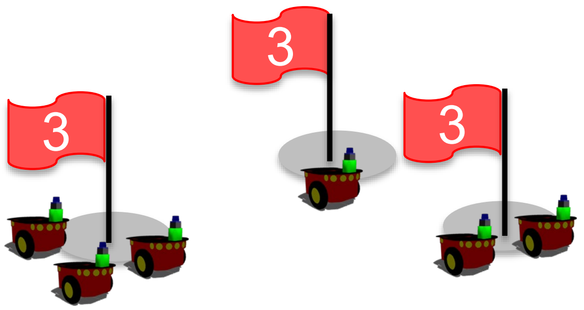
Think of “counterfactual agents” rather than counterfactual actions

$$D^n_{++}(z) = \frac{G(\vec{z}_+(\cup_{i=1,\ldots,n})i) - G(\vec{z})}{n+1}$$

Global system performance
Where “multiple copies of me” are present

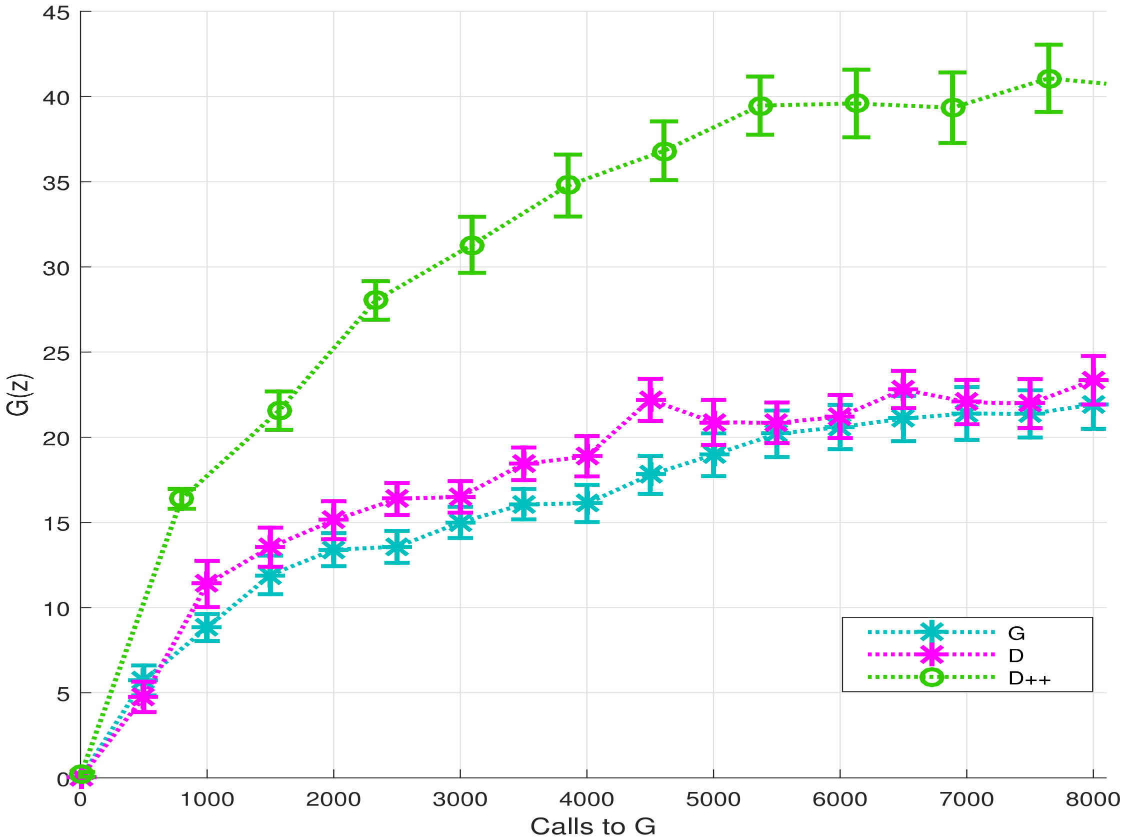
Global system performance

# D++ : Stepping Stone Rewards




| | | | |
|---|---|---|---|
| G | 1 | 1 | 1 |
| D | 1 | 0 | 0 |
| D++ | 1 | 0.33 | 0.5 |

# D++ : Stepping Stone Rewards

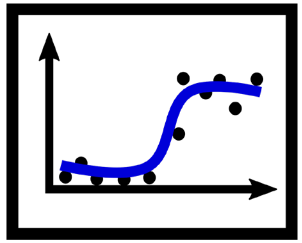



Calls to G

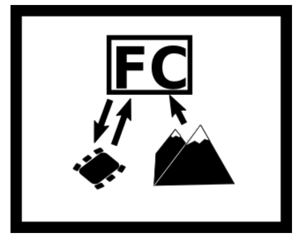
Oregon State University

# Fitness Critic

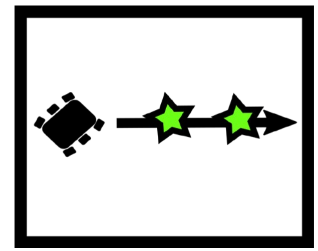
Uninformative Feedback



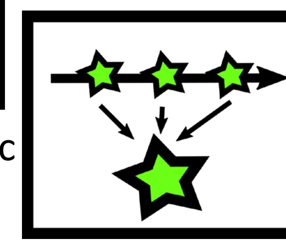
Expected Feedback



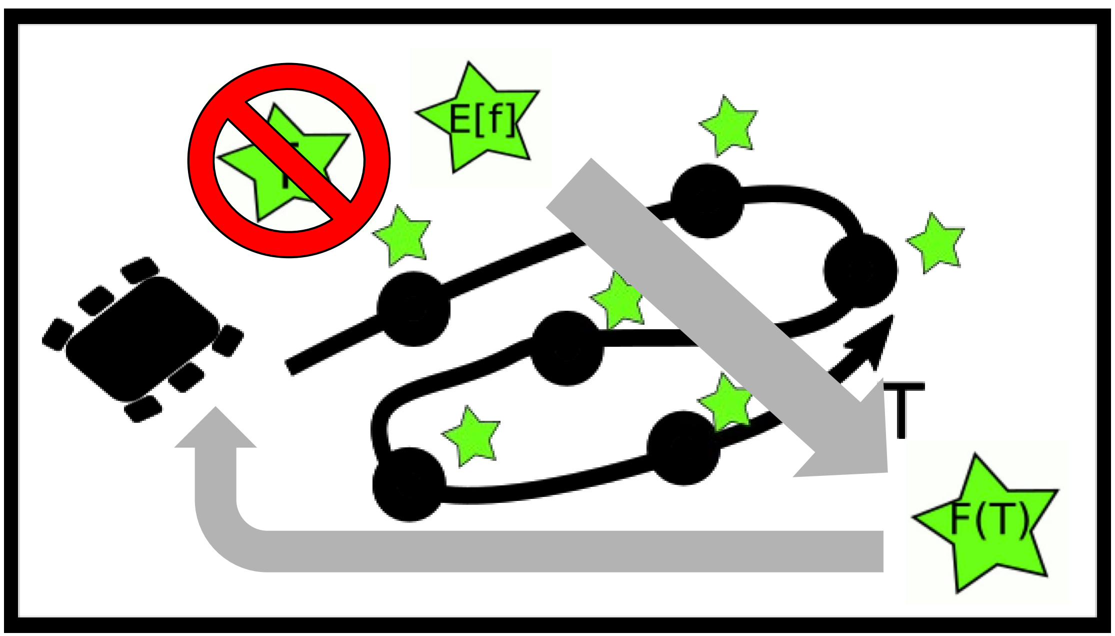

Fitness Critics



Intermediate Critic



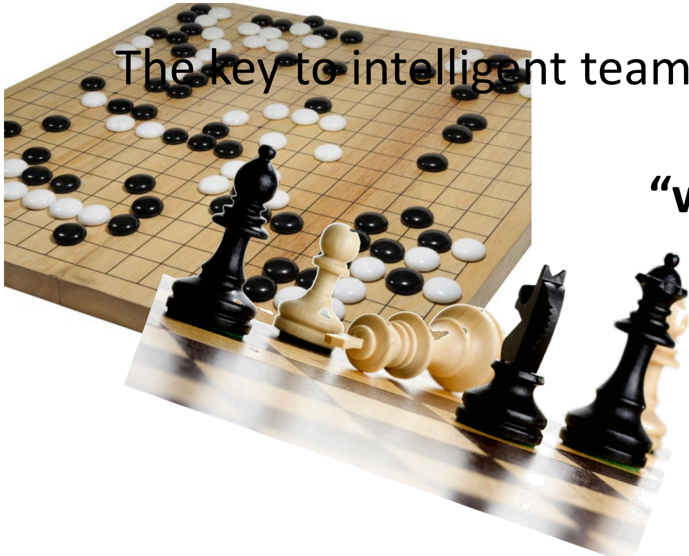
Aggregation



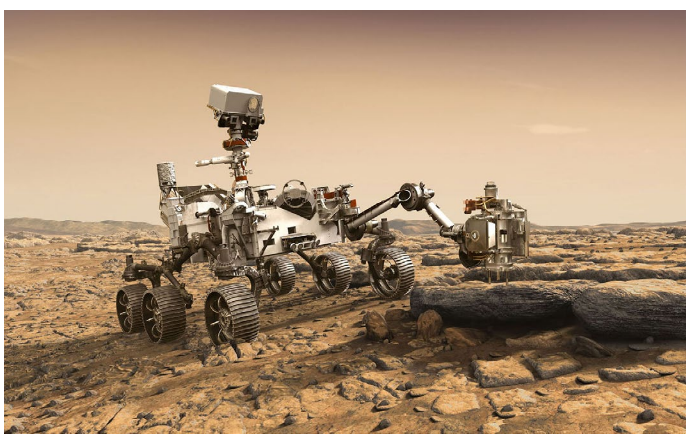
# Multi-Reward Learning

The key to intelligent team behavior in complex systems is determining

**“what matters when”**



Win / Loss



Do Experiments (#)

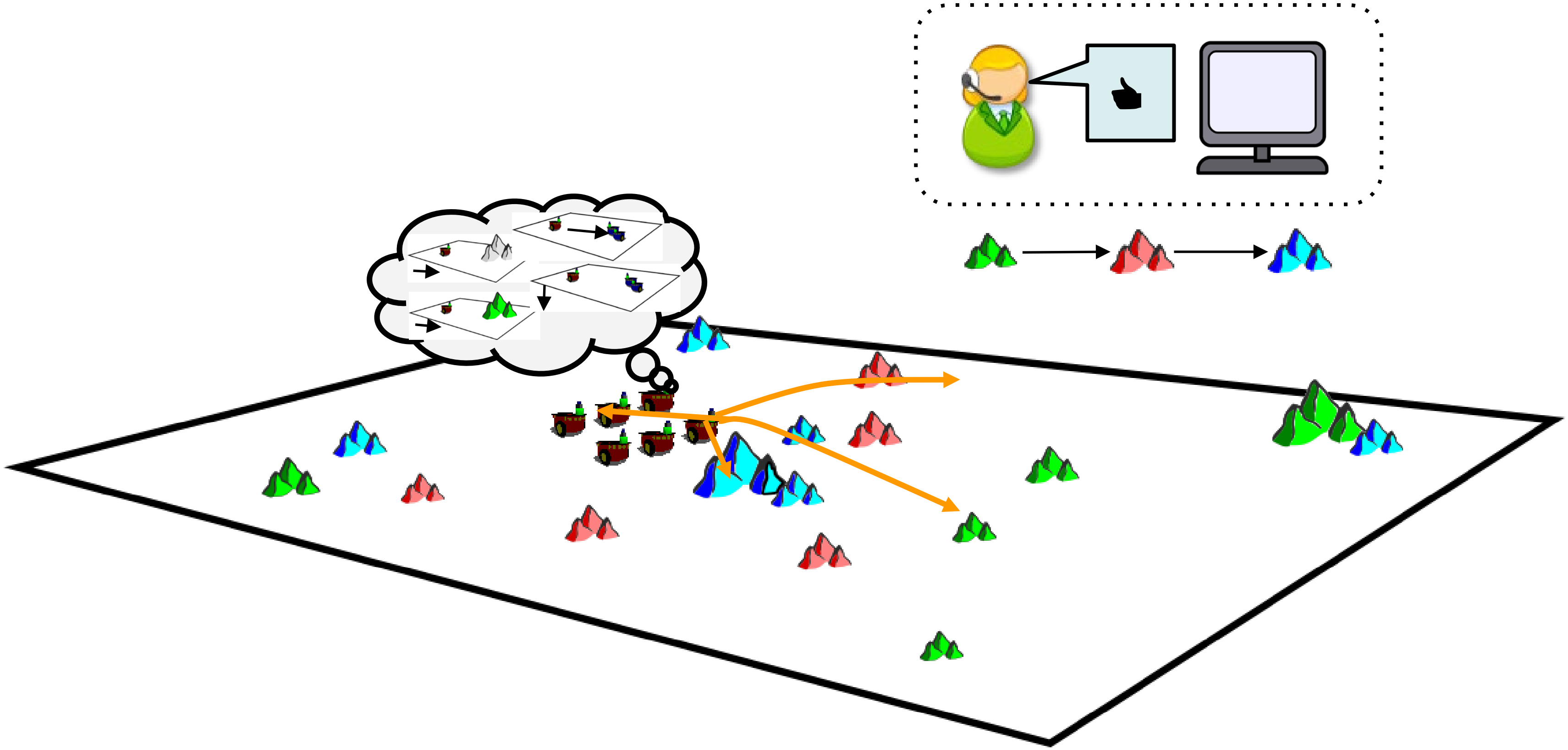
Survived (Y/N)

Area Explored ($m^2$)

Discovered Life (Y/N)

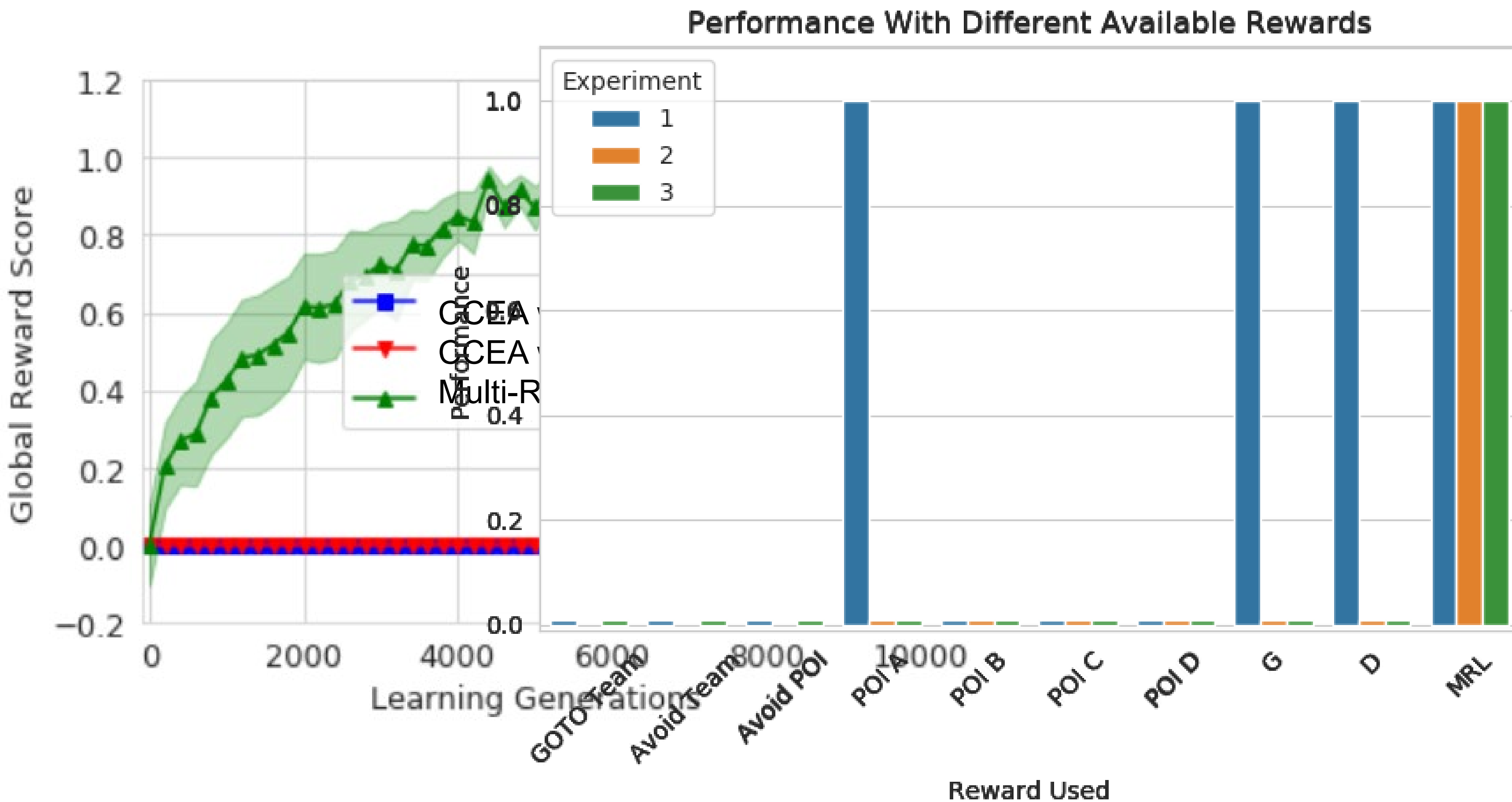
Oregon State University

# Multi-Reward Learning



Oregon State University

# Multi-Reward Learning



# List of Publications, Awards, Honors, etc. Attributed to the Grant

1. G. Dixit, S. Airiau, and K. Tumer. Gaussian Processes as Multiagent Reward Models. Proceedings of the 19th International Conference on Autonomous Agents and Multiagent Systems

   AAMAS, May 2020. (23% acceptance)

2. G. Rockefeller, S. Khadka, and K. Tumer. Multi-level Fitness Critics for Cooperative Coevolution. Proceedings of the 19th International Conference on Autonomous Agents and Multiagent Systems

   AAMAS, May 2020. (23% acceptance)

3. C. Yates, R. Christopher, and K. Tumer. Multi-Fitness Learning for Behavior-Driven Cooperation. Proceedings of the Genetic and Evolutionary Computation Conference

   GECCO, July 2020