Co-adaptive Human Machine Teaming with a Reinforcement Learning Agent

Reed Jensen, Technical Staff

MIT ILP Conference

14 November 2019

LINCOLN LABORATORY
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

This material is based upon work supported by the Under Secretary of Defense for Research and Engineering under Air Force Contract No. FA8702-15-D-0001. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Under Secretary of Defense for Research and Engineering.

Human Collaboration with Technology

Safety

Asiana Flight 214

USAir Flight 5050

Garuda Indonesia Flight 200

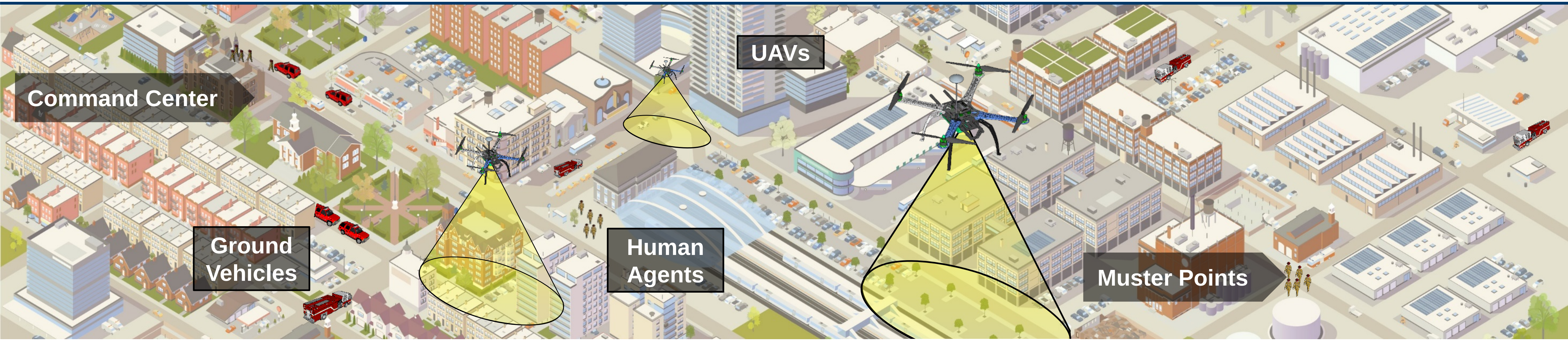
- Number of crash incidents increased when pilot and co-pilot flew together for the first time
- Due to a lack of mutual understanding / co-adaptation

Efficiency and Adoption

- Scrub nurses learn over years to anticipate surgeons' needs
- Robots require surgeon to select tools using gestures or voice commands
- This manual tasking slows down surgery, negating adoption, and potentially endangering patient

Systems (human or automated) that do not understand their human collaborator suffer from decreased safety, efficiency, and rates of adoption

Collaboration Example: Search and Rescue

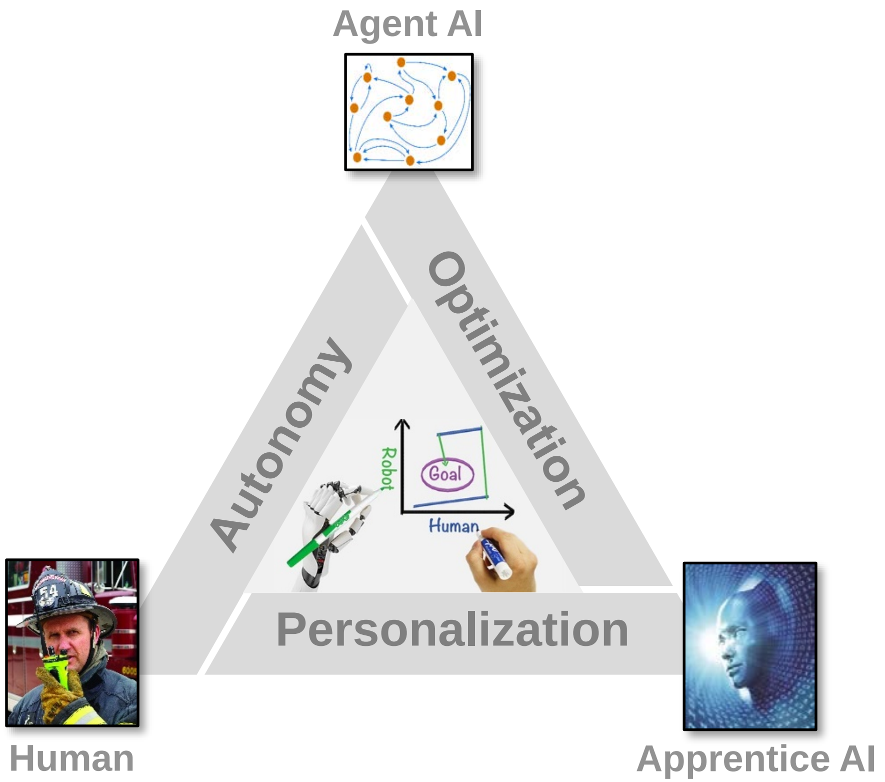


- **Humans team with autonomous vehicles to manage multiple roles and missions in complex environments**
- **Imagine introducing a swarm of smart UAVs to assist**
  - **Will the technology be used?**
  - **Will it be used effectively?**
  - **Is the human-AI system robust?**

Need to evaluate adoption and performance of human collaborators with new technology

Elements for Robust Human-AI Collaboration

- **Flexible, human-aware AI**
- **Natural interface between human and machine**
- **Low barrier of entry for human involvement**
- **Iterative feedback process that incorporates behavior and performance**



Robust Human-aware AI

Capture Human Preferences → Co-adapt → Share Tasks Over Time

For robust human-AI teams, need AI technologies that can co-adapt

Outline

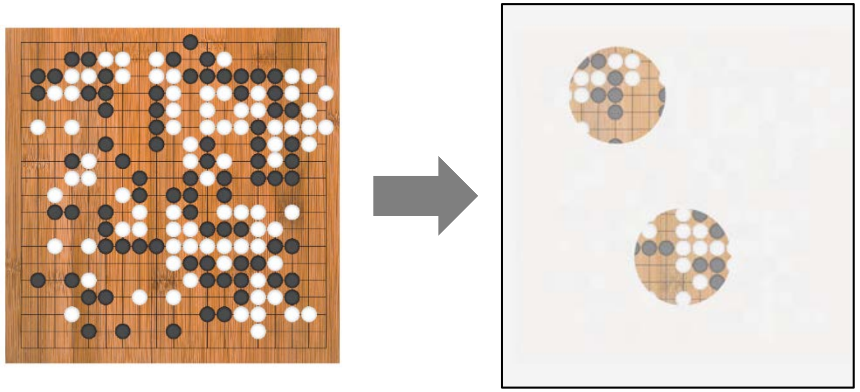
- **Motivation**
- **Development of Co-adaptive AI**
  - **Optimization under Uncertainty**
  - **Personalization**
- **Human-machine Teaming Study**
- **Summary**

Optimization under Real-world Conditions

Optimization Challenges
Environment Uncertainty
Outcome Uncertainty
Communication Limitations
Adversarial Behavior
High-dimensional Domain

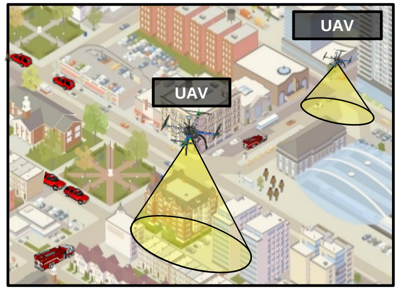
Optimization algorithms must deal with the *fog*, *friction*, and *chance* of real-world scenarios

Going Beyond Traditional Games

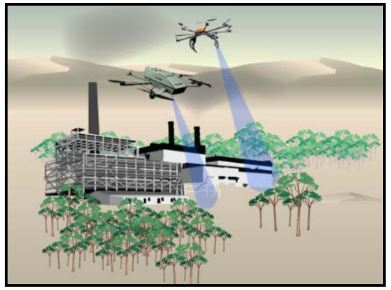


“Go” “Partially Observable Go”

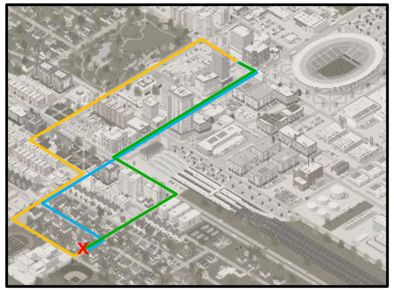
Examples of Complex Real-world Problems



Search and Rescue



Hazardous Area Assessment



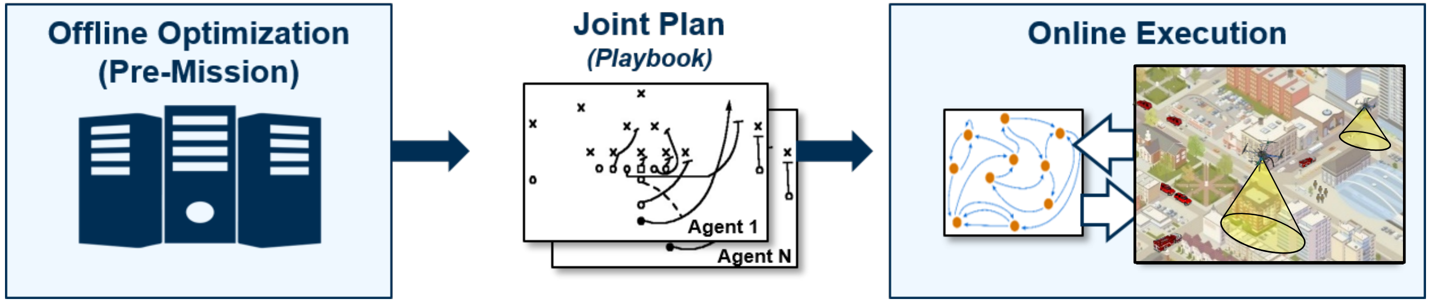
Dynamic Path Planning

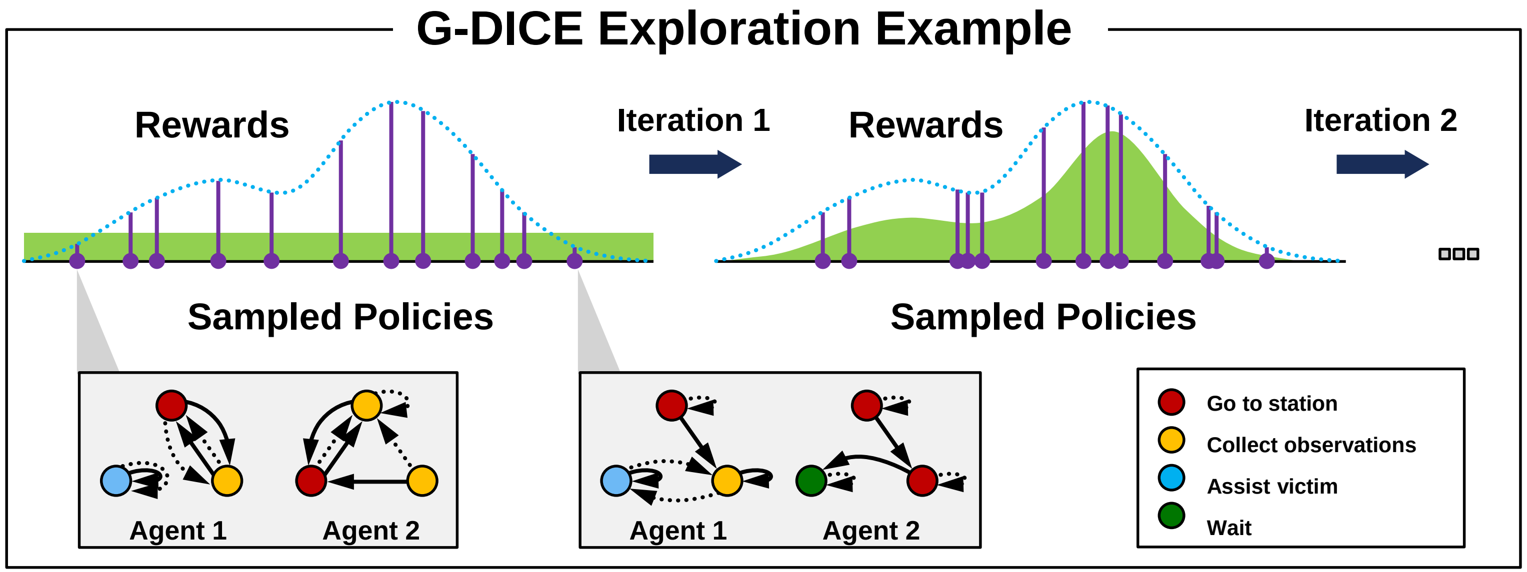
Reinforcement Learning for Autonomous Agents

Decentralized Multi-Agent Coordination with the G-DICE Algorithm

- **Joint playbook optimized using the Graph-based Directed Cross Entropy (G-DICE) algorithm**
  - General-purpose algorithm
  - Solves a Partially Observable Markov Decision Process (POMDP) with decentralized agents
  - Easily parallelized on a computing cluster

- **Playbooks / policies executed as finite state machines in real-time**
  - Require minimal computation
  - Adapt to changes and uncertainty in the environment
  - Work with or without communication

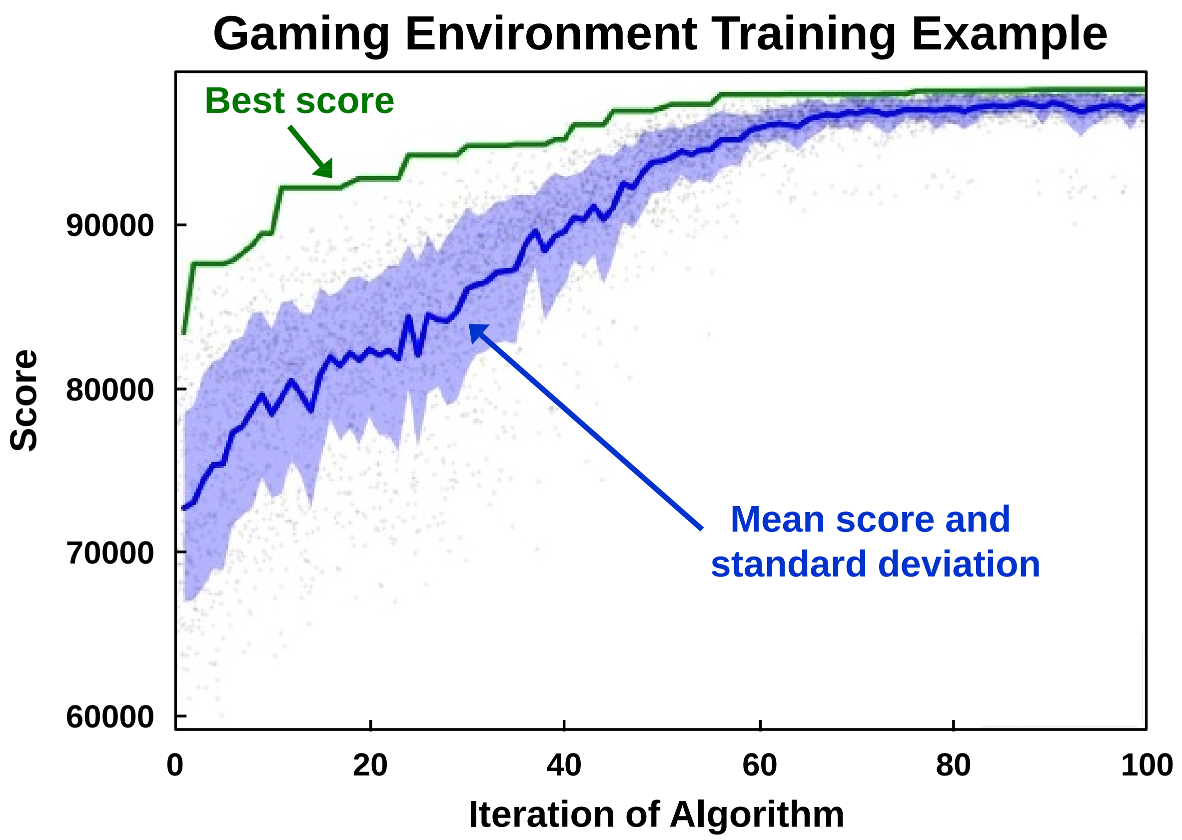


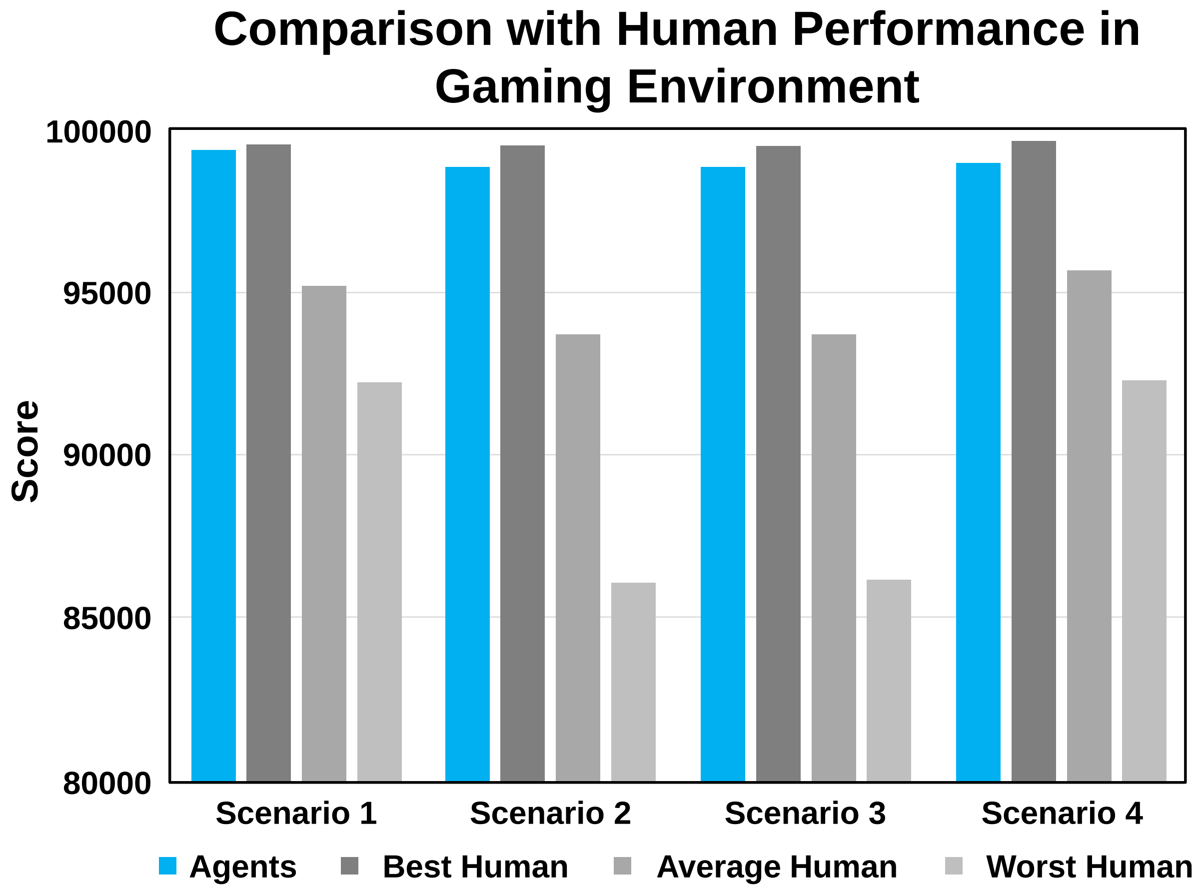


Approach can be applied to many challenging multi-agent coordination problems

Omidshafiei, Shayegan, et al. "Graph-based cross entropy method for solving multi-robot decentralized POMDPs." 2016 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2016.

G-DICE Agent Validation Results





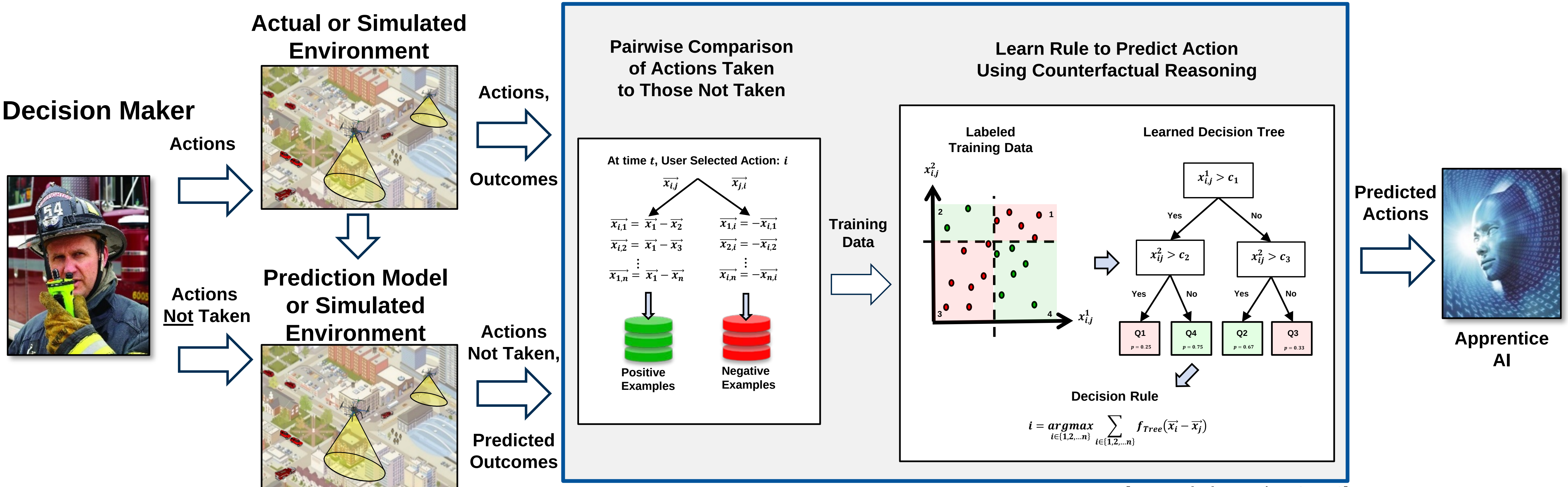
Autonomous agent performance in game compares with that of high-performing humans

Outline

- Motivation
- Development of Co-adaptive AI
  - Optimization under Uncertainty
  - Personalization
- Human-machine Teaming Study
- Summary

Modeling Human Behavior with Machine Learning: Apprentice AI

Apprenticeship Scheduler



*[**M. Gombolay** et al.; IJCAI'16]*

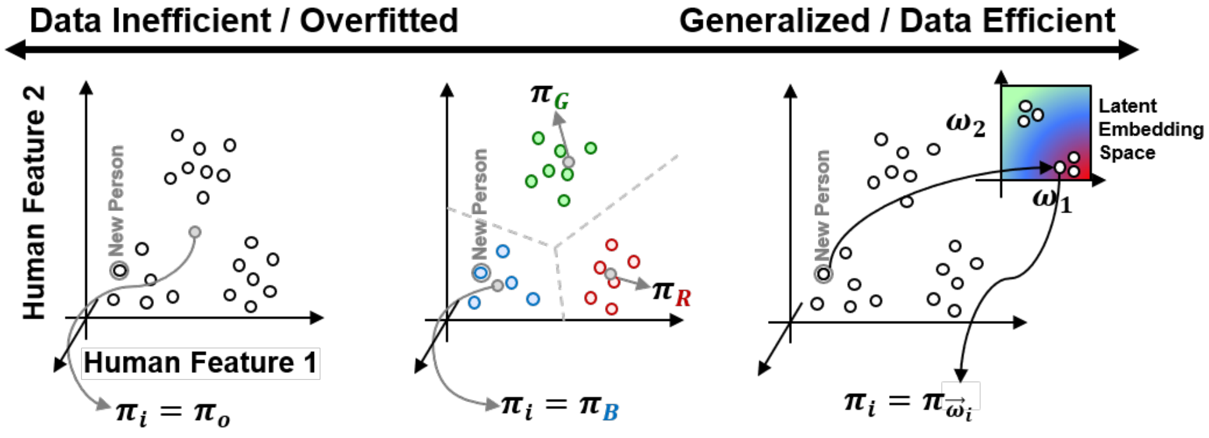
Apprentice AI predicts future user actions by forming a decision tree using pointwise comparisons between actions taken and not

Gombolay, M. C. et al. "Apprenticeship Scheduling: Learning to Schedule from Human Experts." In Proc. International Joint Conference on Artificial Intelligence (IJCAI), 2016

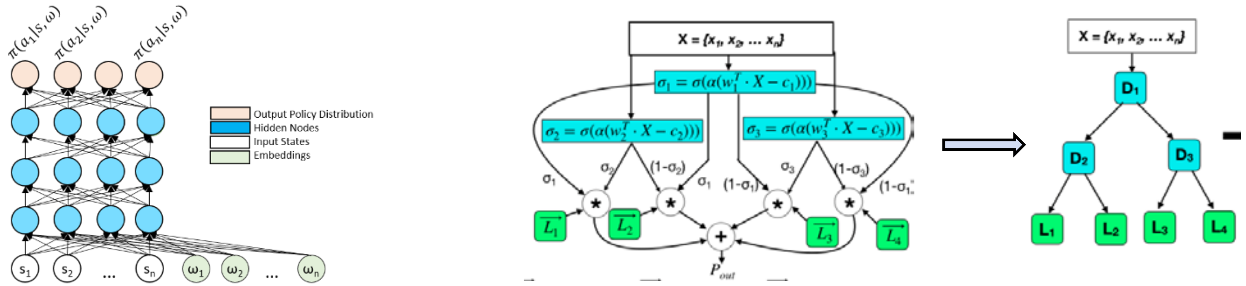
Personalization with Apprenticeship Learning

- **Learn a personalized model using demonstrations from a heterogeneous population**
  - **Actions, observations, features are determined from gameplay**
  - **Personalized model leverages data from all users**
  - **Individual customization can be learned online**
- **World / transition model learned from gameplay**
- **Model variables represent individual qualities and can be interpreted from a decision tree**

Personalization with Heterogeneous Demonstrators



Personalized Neural Networks and Decision Trees



Multi-layer Long Short-Term Memory Network with Personalized Embeddings

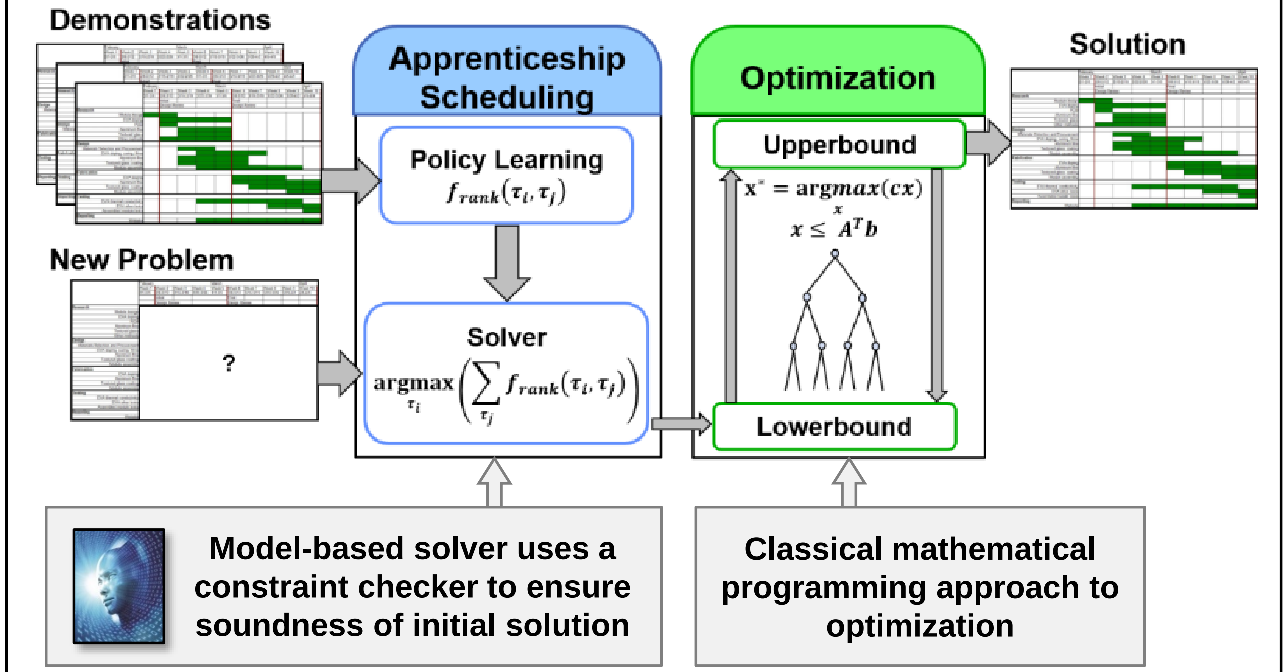
Conversion of Neural Network to Differentiable Decision Tree for Interpretability

Paleja, Rohan, and Matthew Gombolay. "Inferring Personalized Bayesian Embeddings for Learning from Heterogeneous Demonstration." arXiv preprint arXiv:1903.06047 (2019).

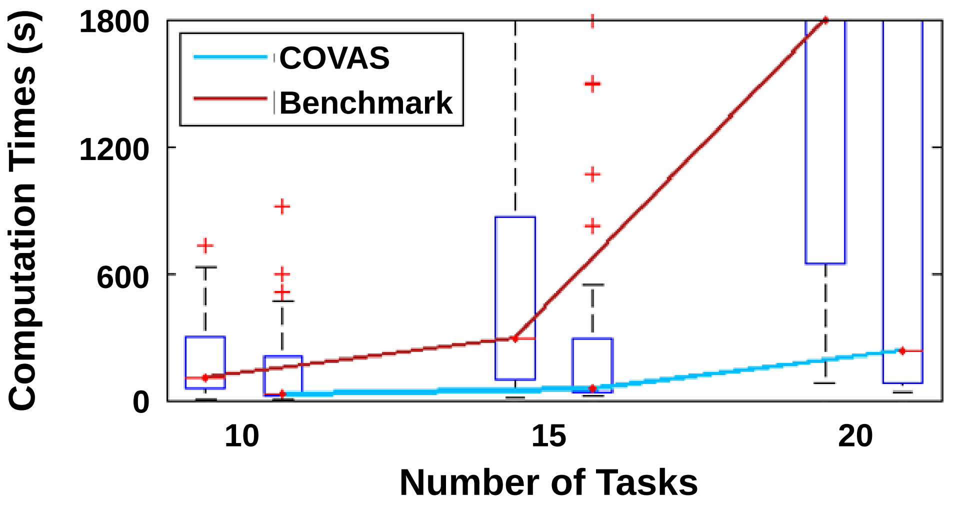
Collaborative Optimization Via Apprenticeship Scheduling (COVAS)

- Traditional optimization approaches rely on "warm starts," which provide an initial guess lower bound on the optimality
- COVAS is a human-machine optimization technique that learns from human demonstrations how to warm start

COVAS Architecture



Compute Performance for a Complex Scheduling Problem



COVAS can solve problems with more tasks an order of magnitude faster than pure optimization

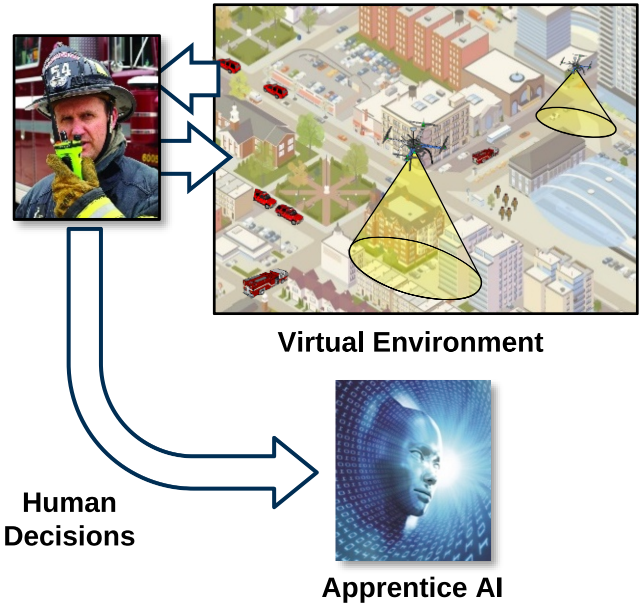
Gombolay, M. C. et al. "Human-Machine Collaborative Optimization via Apprenticeship Scheduling." Journal of Artificial Intelligence Research, 63, 2018, 1-49.

Outline

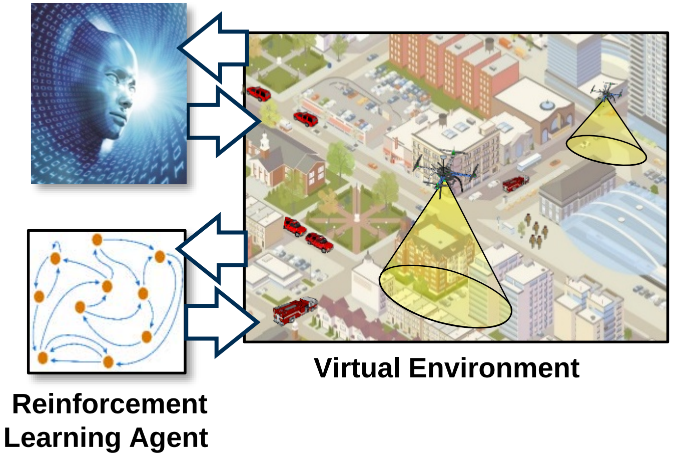
- Motivation
- Development of Co-adaptive AI
  - Optimization under Uncertainty
  - Personalization
- Human-machine Teaming Study
- Summary

Human-Machine Teaming with Autonomous Agents

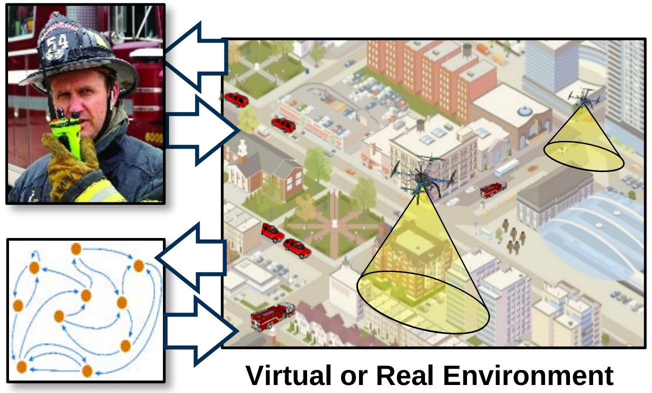
Learn model of human team member



Optimize automated agent using apprentice as surrogate for human



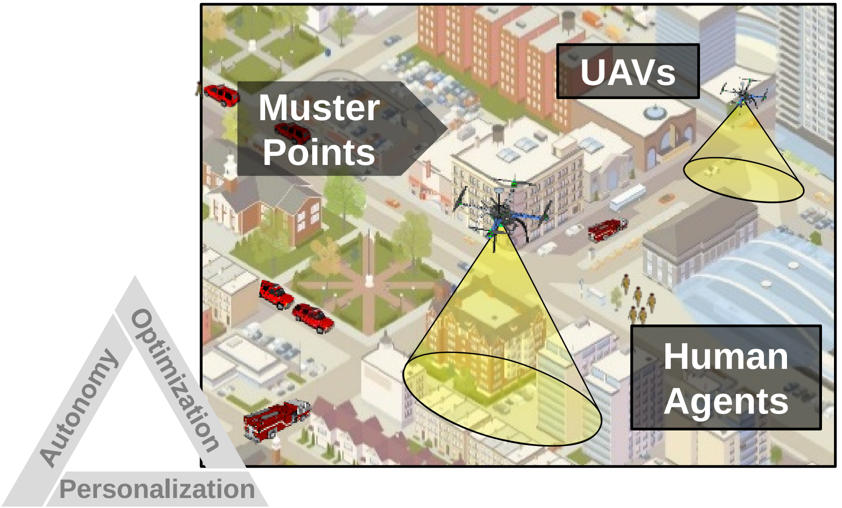
AI complements human teammates playing scenario

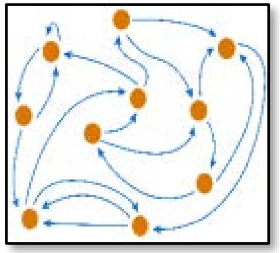
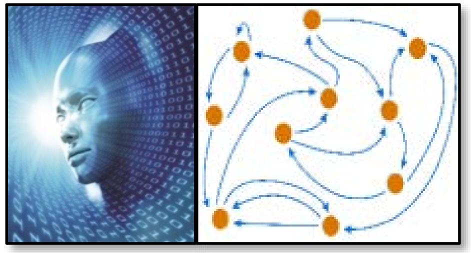


Iteration with co-adaptive AI enables robust human-AI teaming

Human-Machine Teaming Study Overview

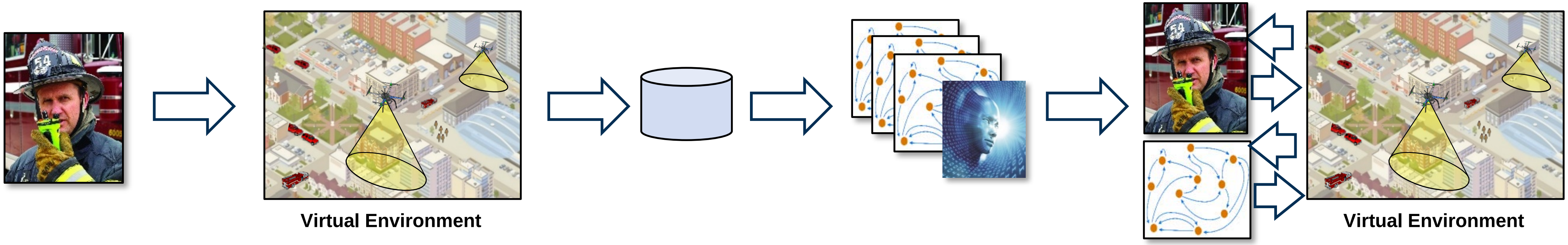
- **Hypothesis**
  - **Human team with personalized, optimized autonomous helper agent will be preferred and most efficient**
- **Scenario**
  - **Human given a task in a virtual environment to rescue victims using rescue personnel and an unmanned air vehicle (UAV) decoy**
  - **Player must use both personnel and the UAV to obtain perfect score**
  - **UAV controlled by an automated algorithm trained under four conditions:**



Manual (untrained)	Optimized Without User Input	Preference-based	Preference-based and Optimized
			
Human user controls personnel and UAV	UAV finds optimal stations ignoring user preferences	UAV finds stations based on user demonstrations	UAV finds optimal stations given user demonstrations

Experiment Implementation

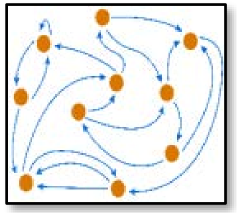
Phase 1 "Learning"	Phase 2 "Performance"	Phase 3 "Training"	Phase 4 "Testing"
Tutorials and Practice	Collecting Player Data	Training Algorithms Offline	Teaming with Algorithm



- **Participants play long to enough to settle on a strategy but not enough to get perfect score**
- **UAV agent algorithm trained offline between Performance and Testing sessions**
- **Participants team with UAV agent but are not told algorithm objectives or training conditions**

Example Human-Machine Teaming Results

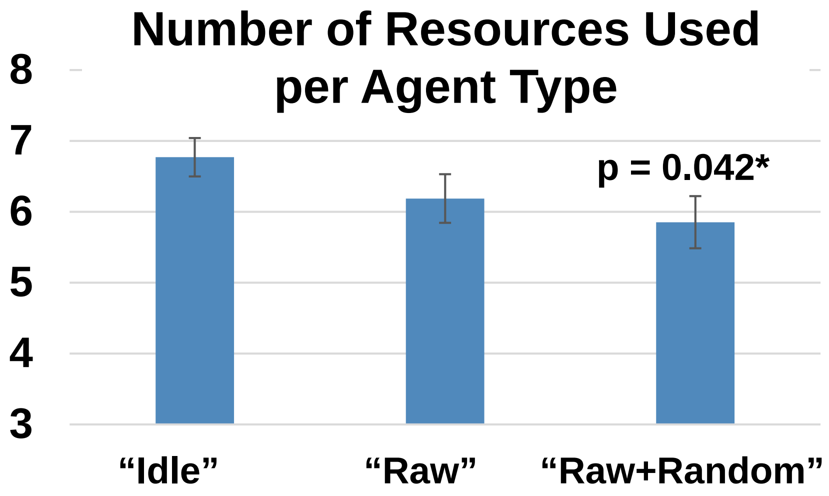
Human Agnostic (“Idle”)

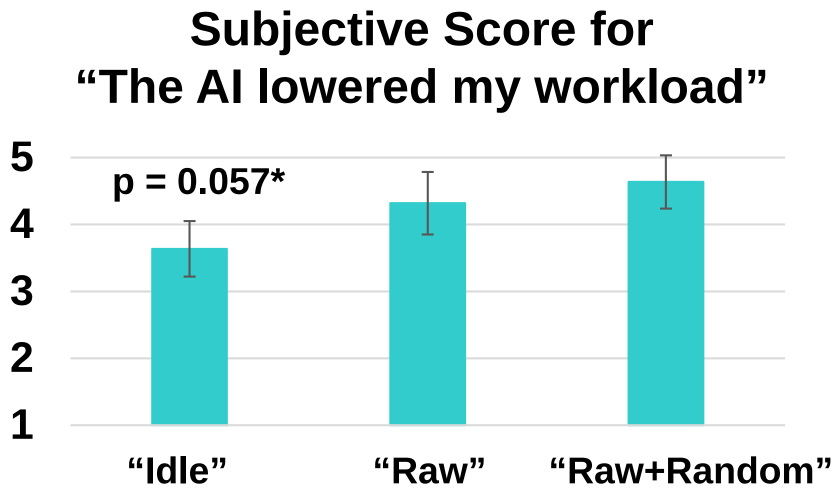


Raw Playback

Raw Playback + Randomization

Human-Trained Apprentice





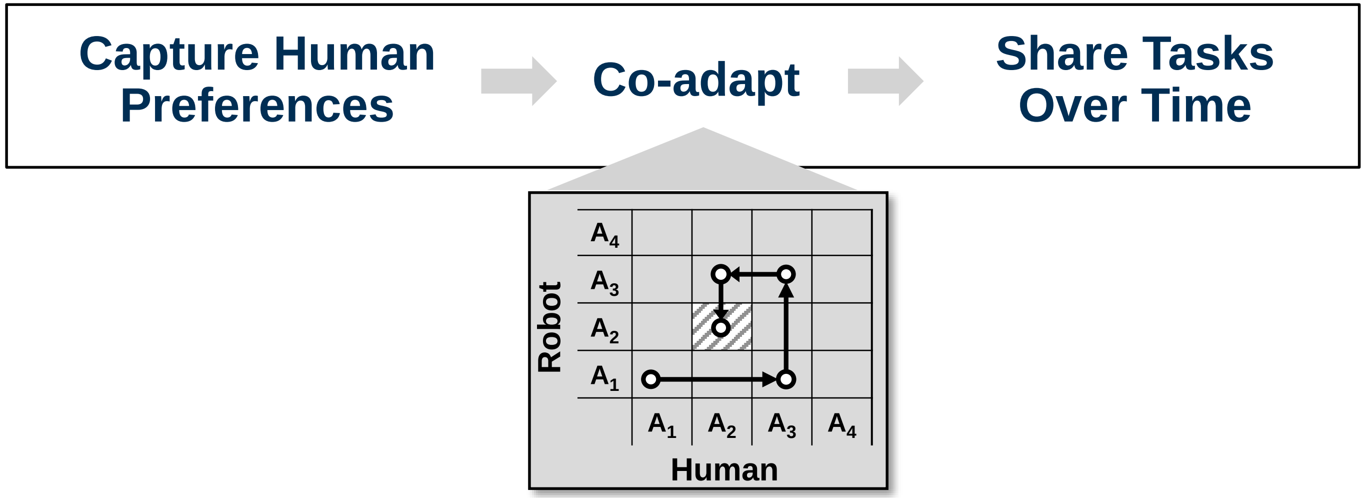
Players used fewer resources when teaming with AI trained with previous gameplay

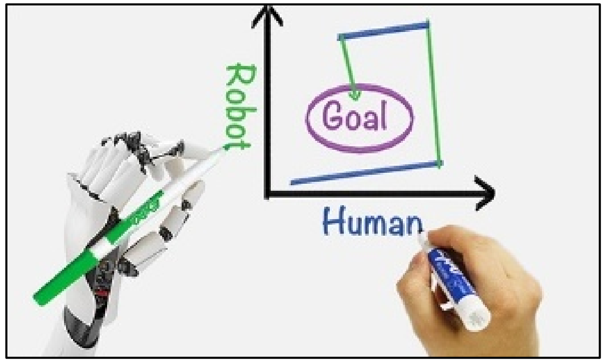
Players indicated lower workload when using AI trained with previous gameplay

N = Approximately 25 participants   * Additional collections needed to increase statistical power

Next Steps

- **Complete 2019 study analysis**
  - **Tested all four conditions, including apprentice with personalized embeddings**
  - **Over 40 participants**
- **Teach learning agents to co-adapt as human changes behavior**
  - **Agent learns changing objectives**
  - **Assess convergence or divergence of human and AI within game-theoretic framework**
- **Dynamically share roles and tasks over time**
  - **Human takes over or cedes tasks depending on cognitive load and performance**
  - **AI naturally adjusts**





Summary

- **Must consider the human element when designing AI solutions with human teammates**
- **Flexible AI solutions exist that can incorporate individual preferences and population behaviors in a natural way**
- **Preliminary experiment results suggest reinforcement learning algorithm with human demonstrations may increase efficiency and reduce workload**
- **Work to incorporate co-adaptation and task sharing ongoing**

Acknowledgements

- Georgia Tech
  - Matthew C. Gombolay
  - Rohan Paleja
- MIT Campus
  - Jonathan P. How
- MIT LL
  - J. Daniel Griffith
  - Emily P. Anesta
- Northeastern University
  - Christopher Amato