Shield Decentralization for Safe Multi-Agent Reinforcement Learning

Daniel Melcer, Christopher Amato, Stavros Tripakis
Northeastern University

NeurIPS 2022, November 28 - December 9

Intro & Related Work

- RL is great but unsafe

Image: NTSB

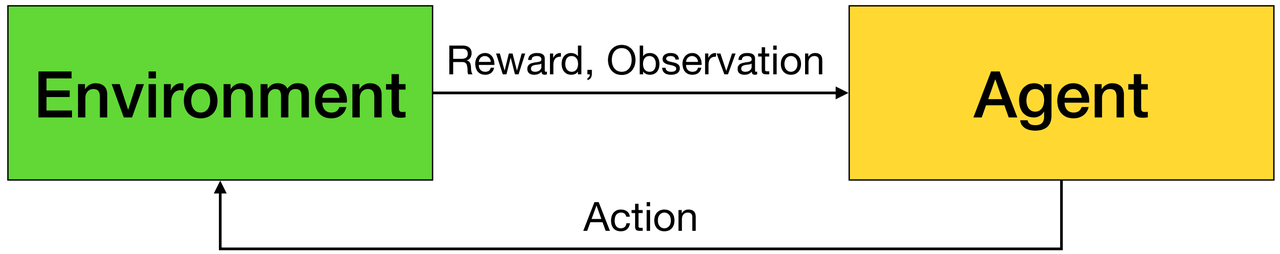
Intro & Related Work

- RL is great but unsafe
- Single-agent shielding framework: Alshiekh et al. 2017

Image: NTSB

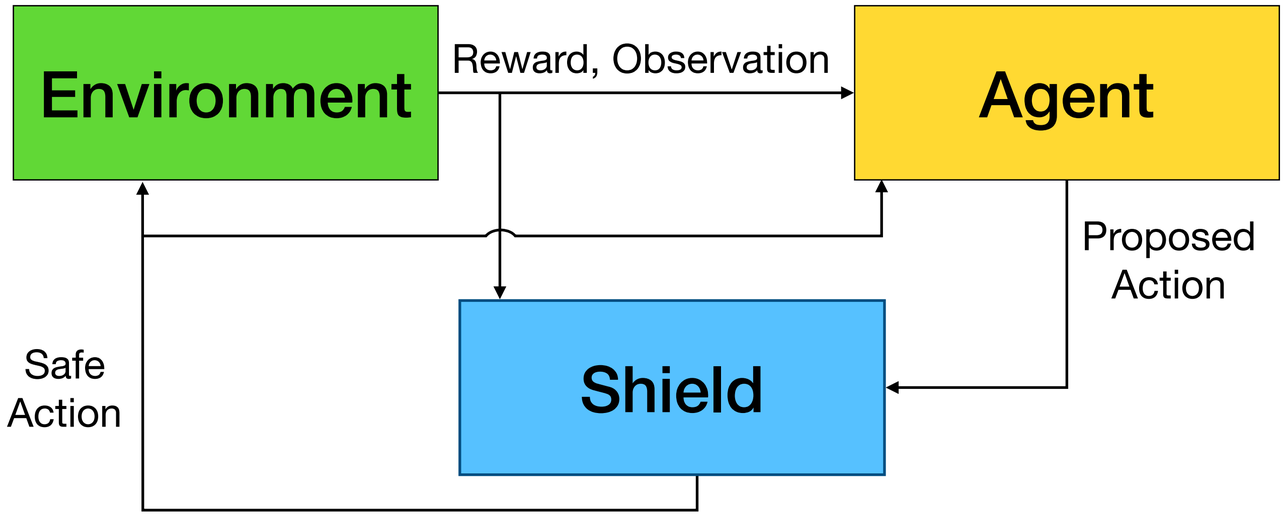
Standard RL

Agent-Environment Interaction



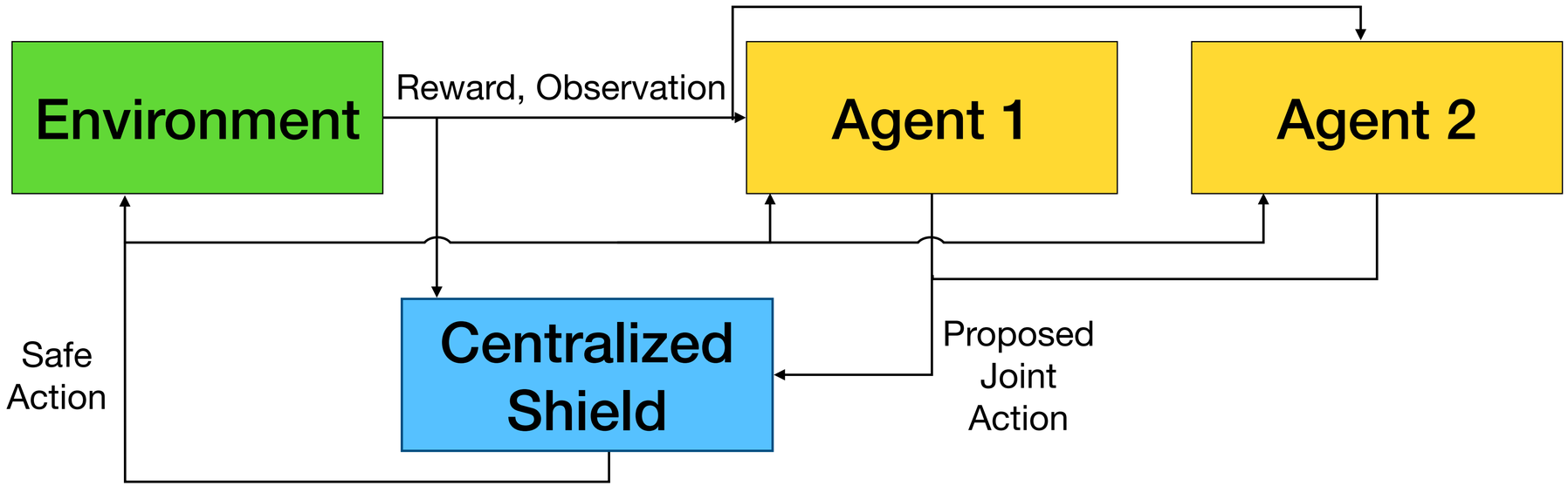
Single-Agent Shielding

Agent-Environment-Shield Interaction



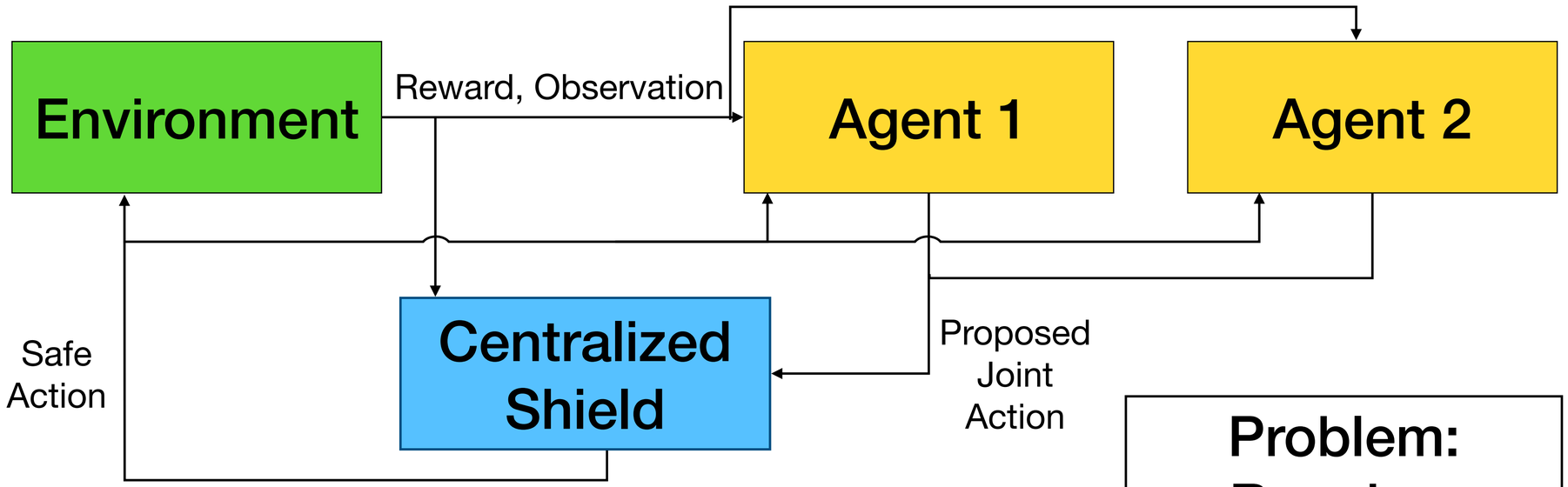
Multi-Agent Shielding

ElSayed-Aly et al. 2021



Multi-Agent Shielding

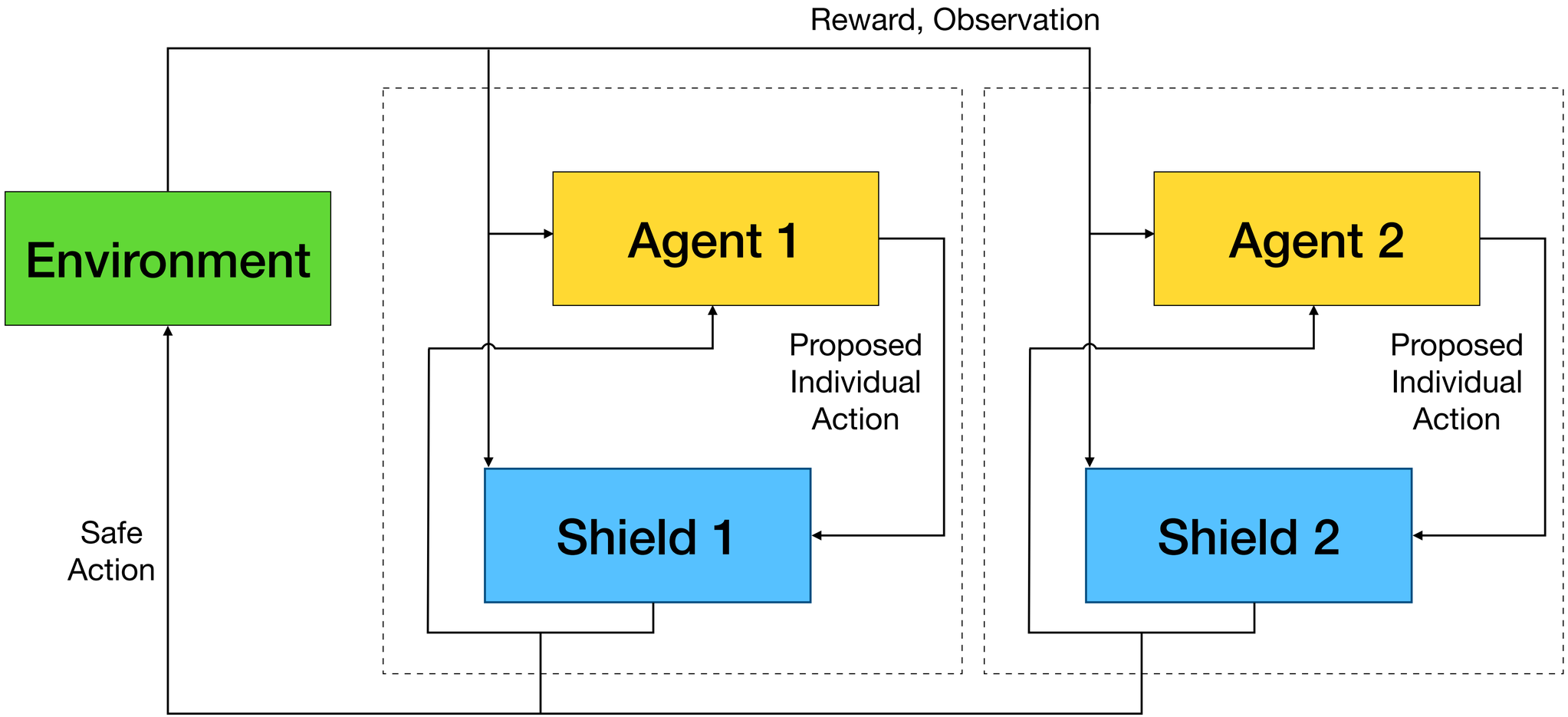
ElSayed-Aly et al. 2021



Problem: Requires communication between agents

Decentralized Shielding

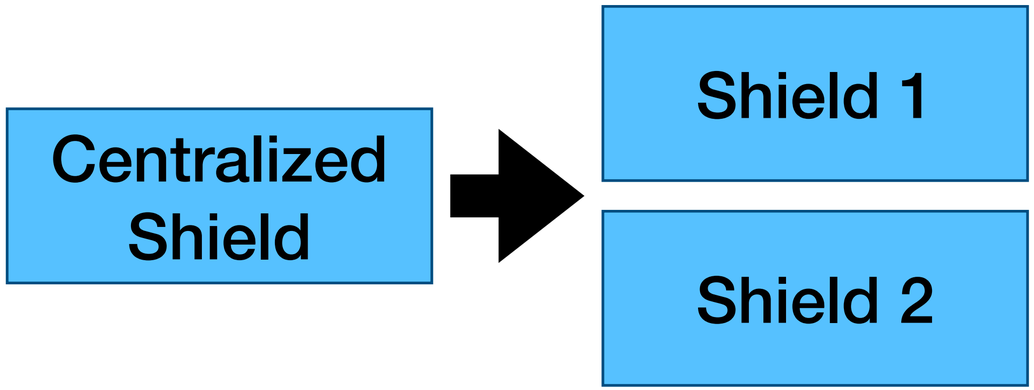
Without Communication



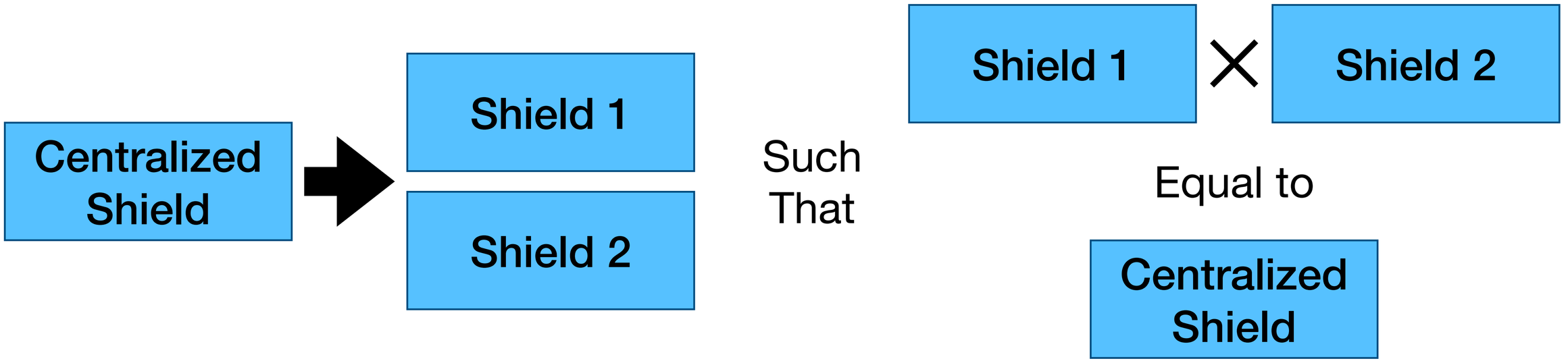
Our Proposal: Shield Decomposition

Centralized Shield

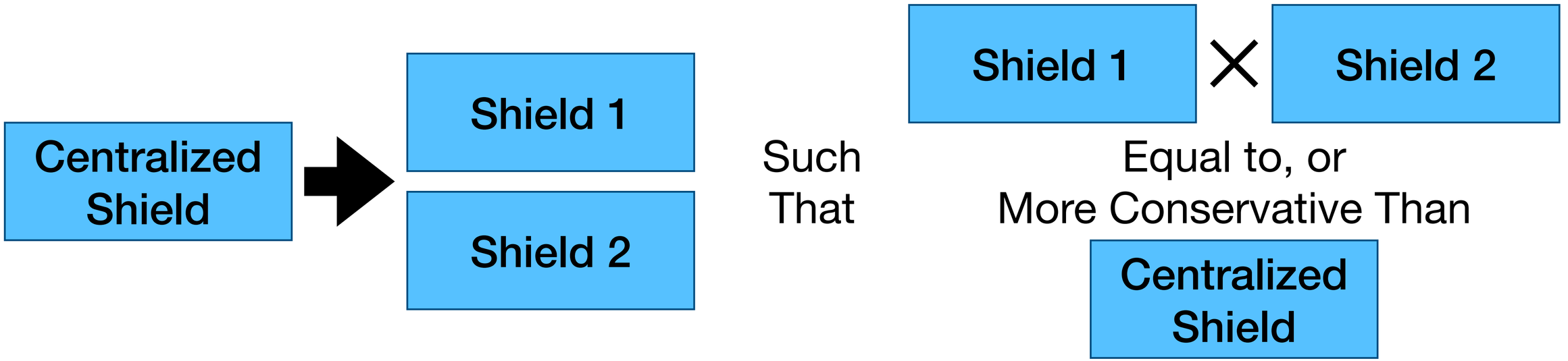
Our Proposal: Shield Decomposition



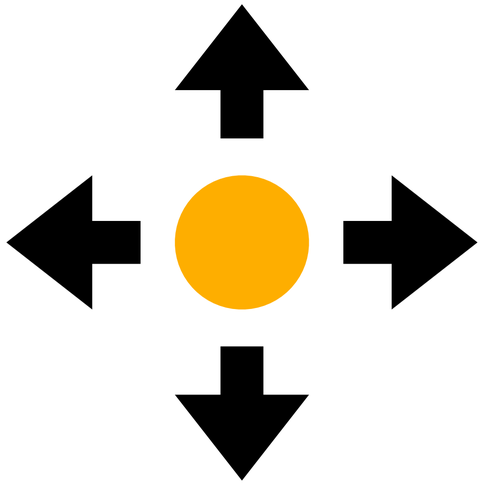
Our Proposal: Shield Decomposition

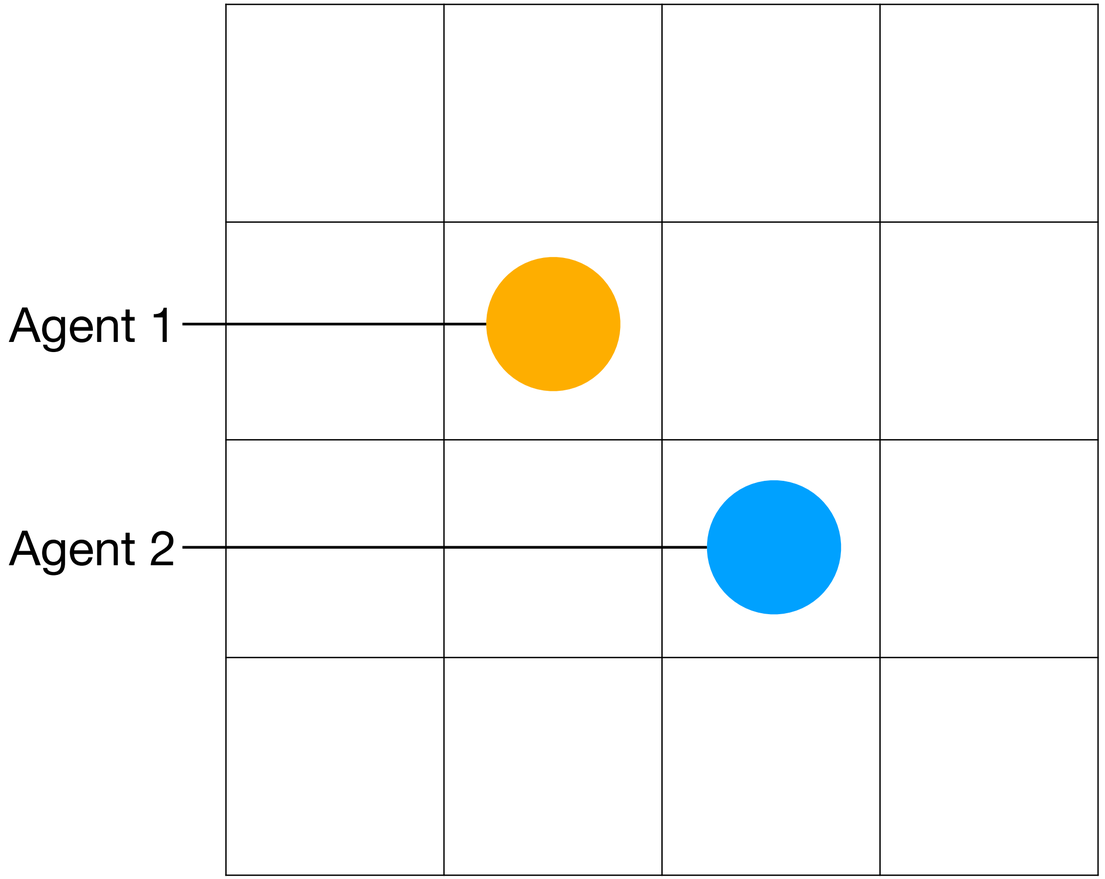


Our Proposal: Shield Decomposition

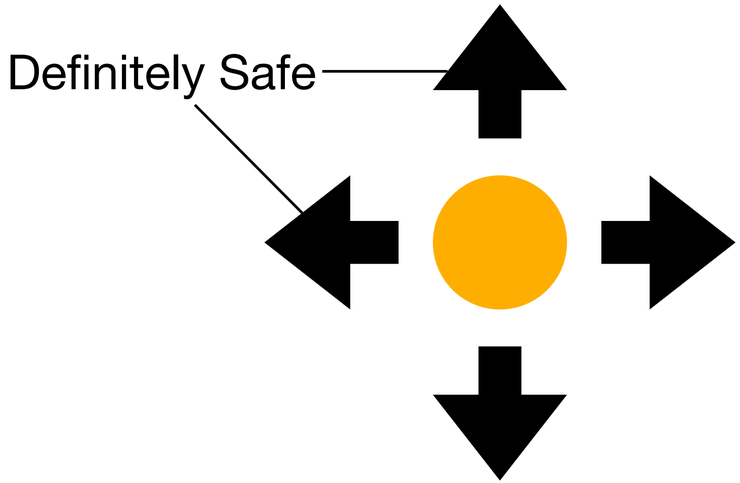


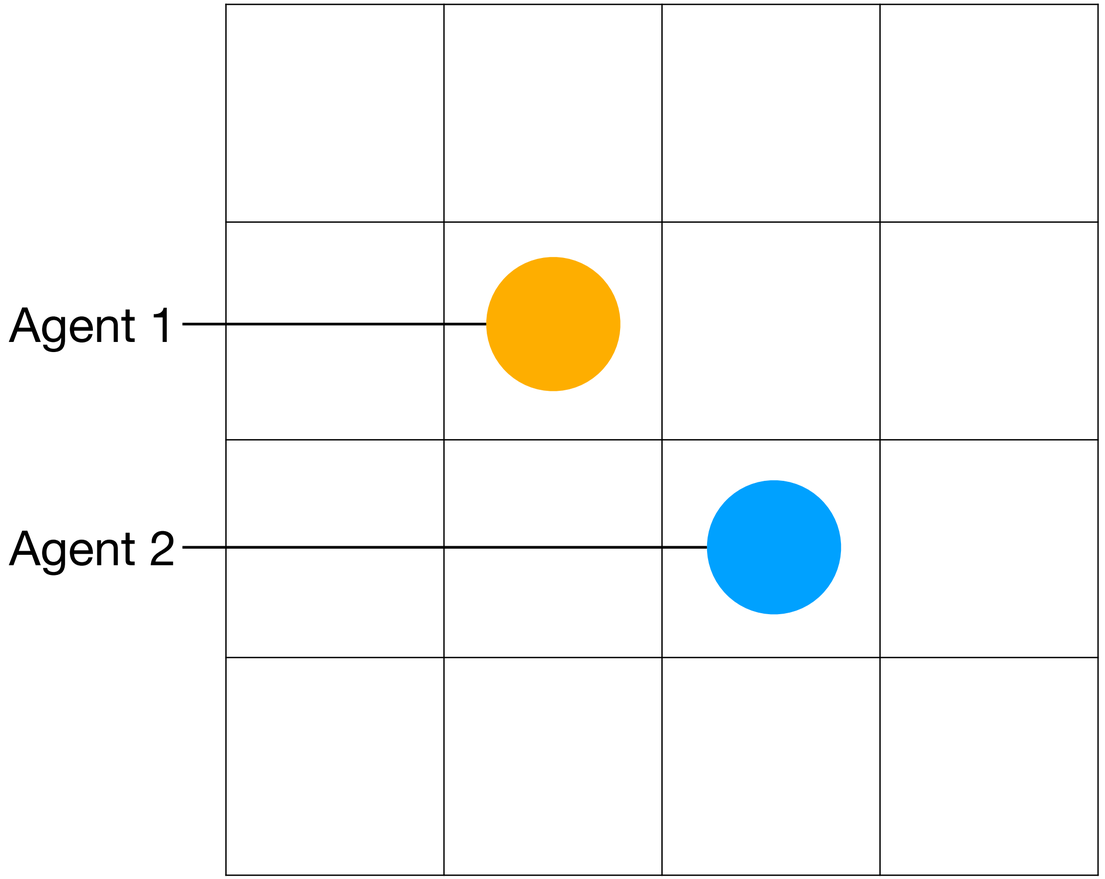
What Actions are Safe?



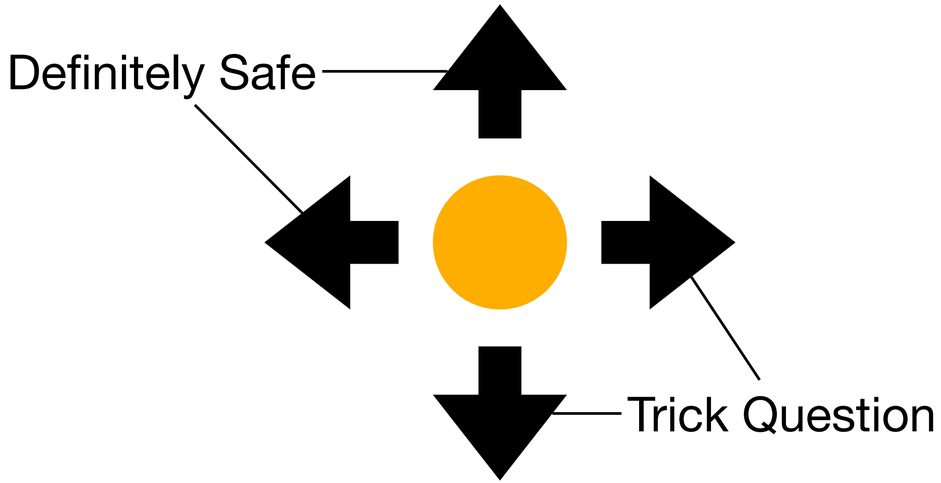


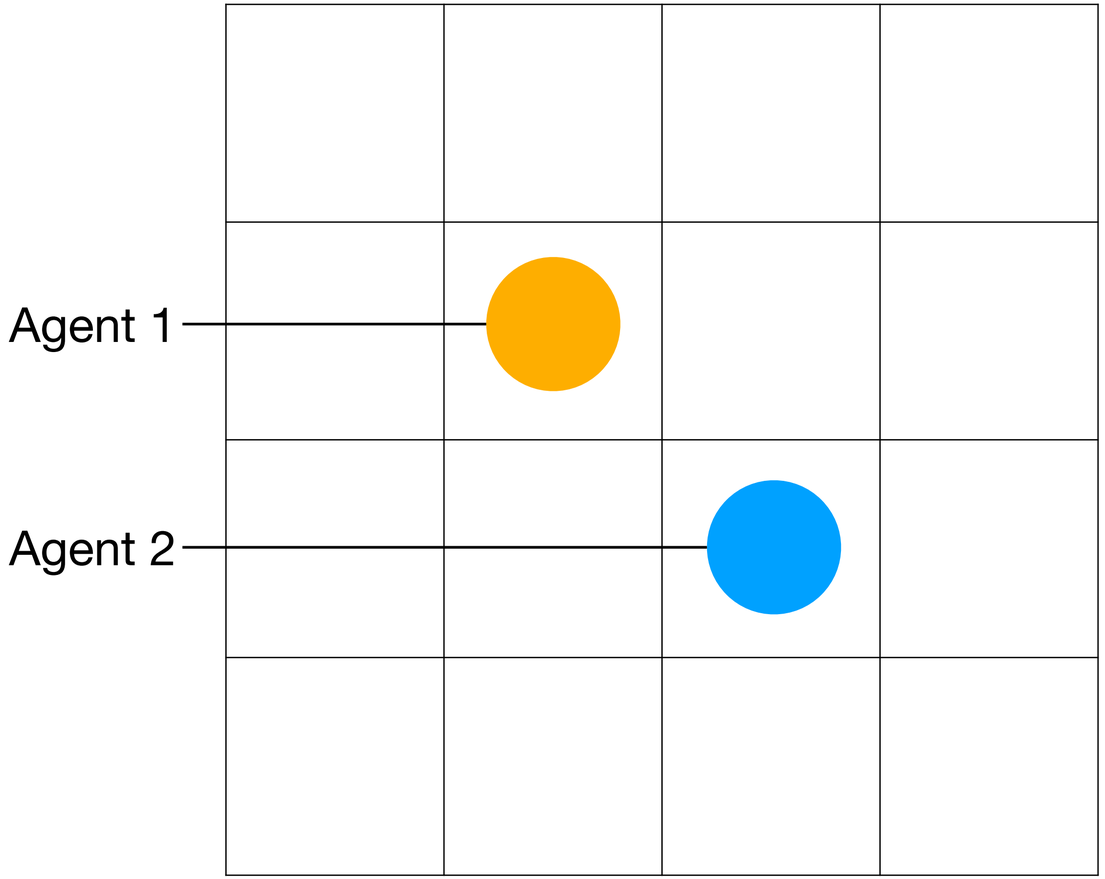
What Actions are Safe?





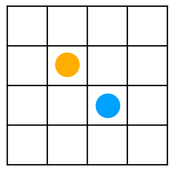
What Actions are Safe?

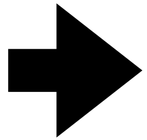




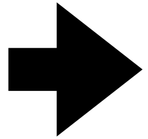
Centralized Shield: Safe Joint Actions

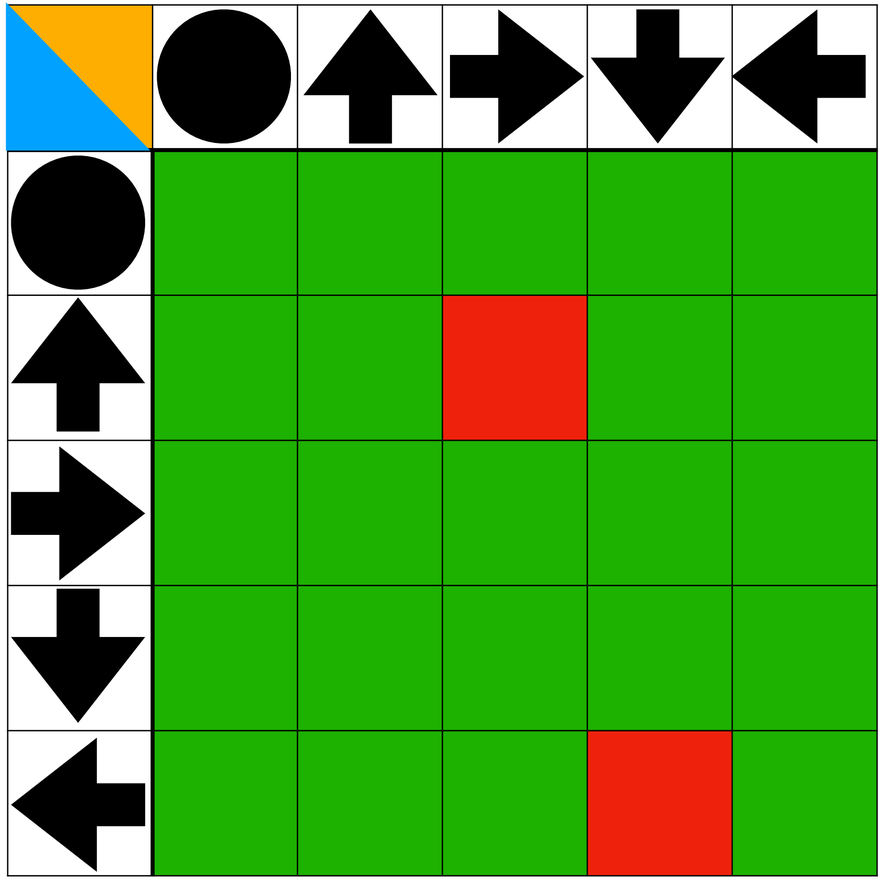
Current Observation



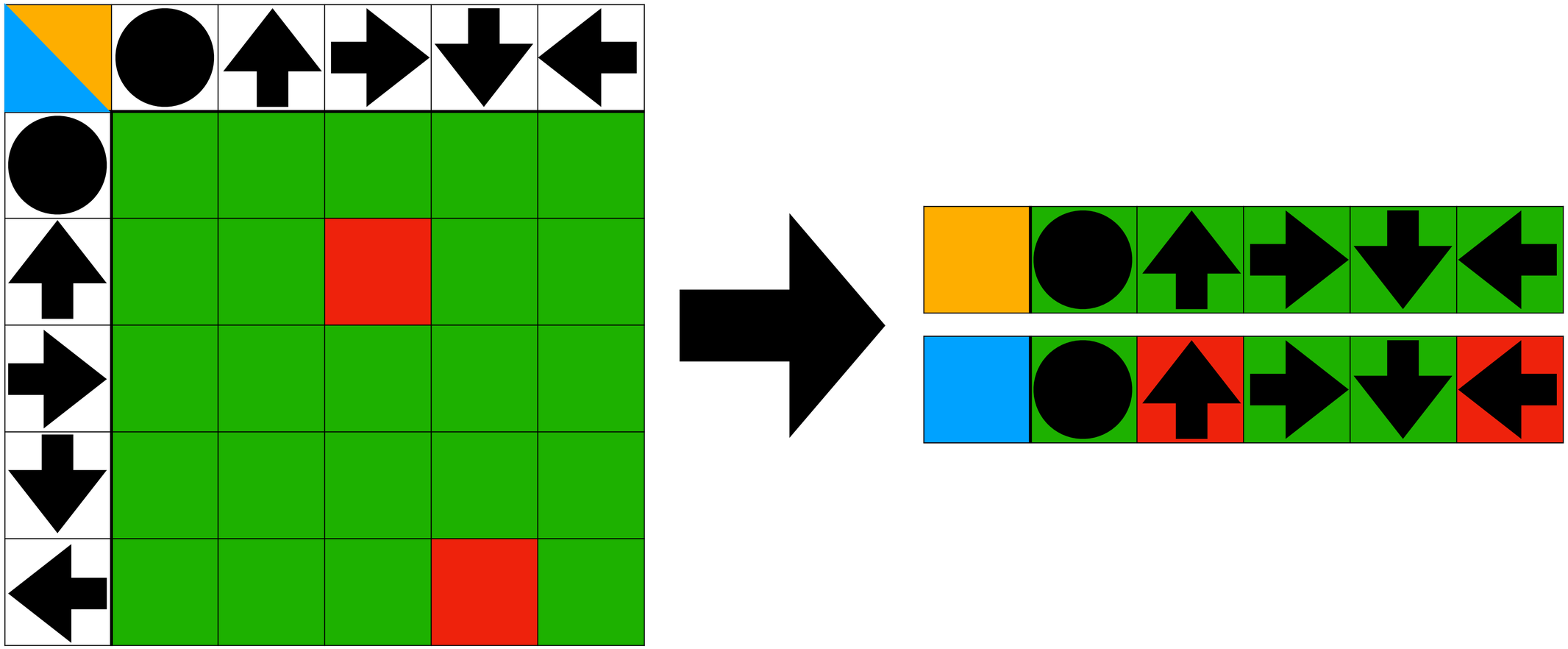


Centralized Shield

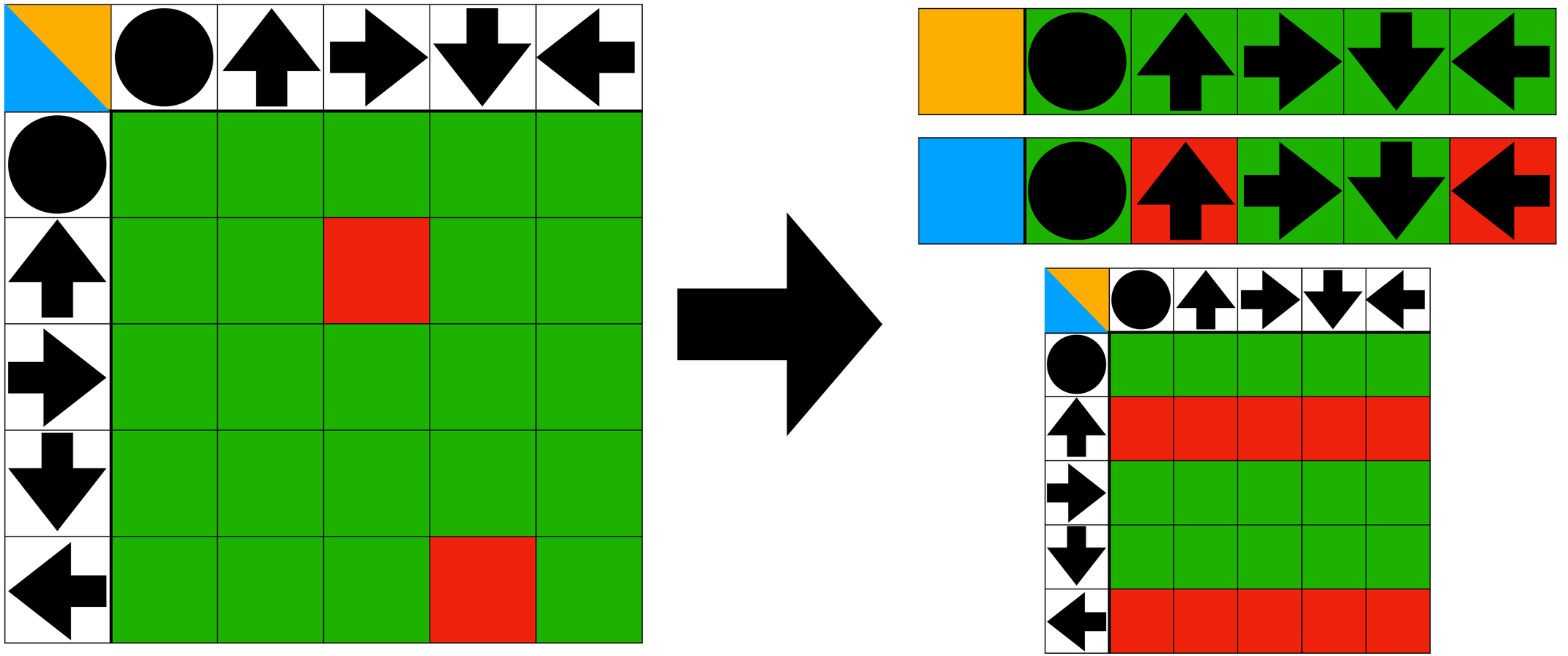




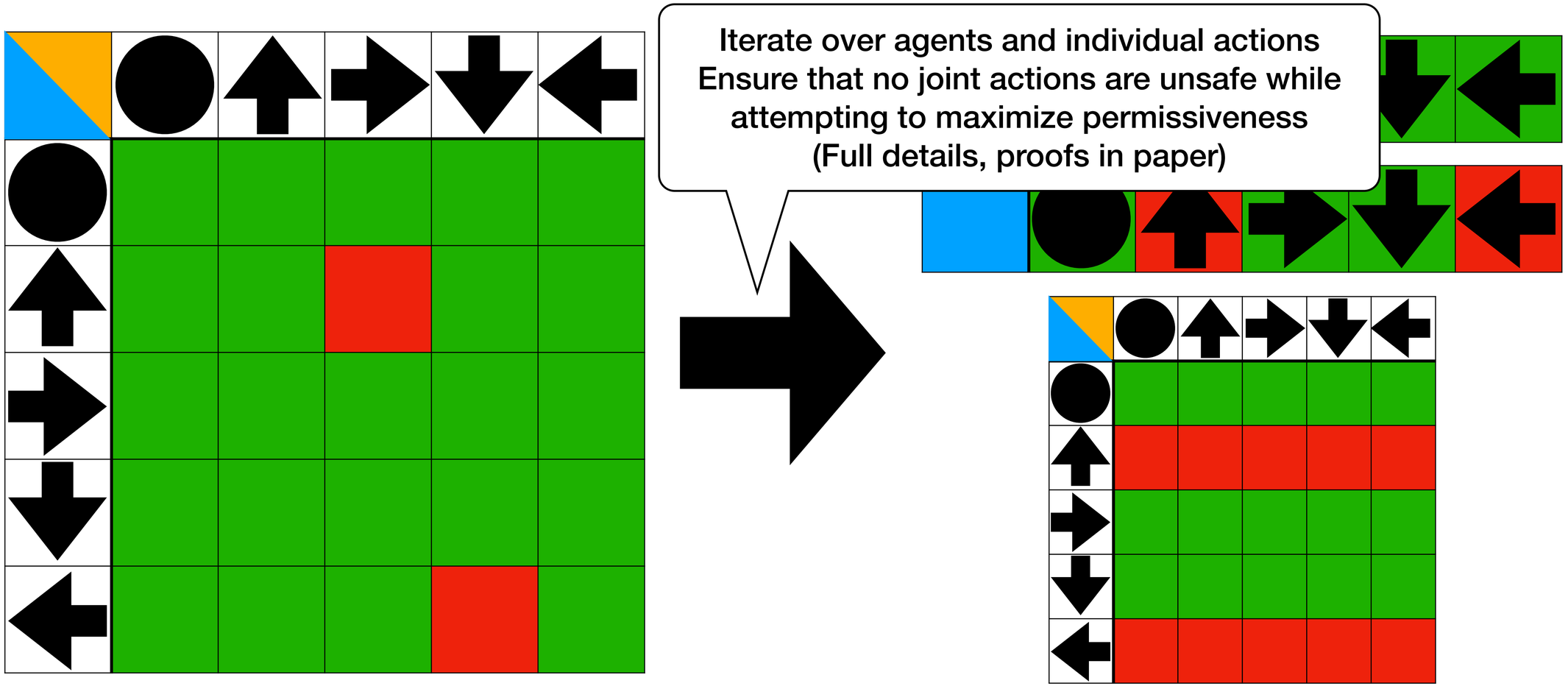
Shield Decentralization



Shield Decentralization



Shield Decentralization



Experiments

What are the goals?

Show that decentralized shields are:

Experiments

What are the goals?

Show that decentralized shields are:

1. Safe

Experiments

What are the goals?

Show that decentralized shields are:

1. Safe
2. Performant

Scenario 1: Gridworld-Collision

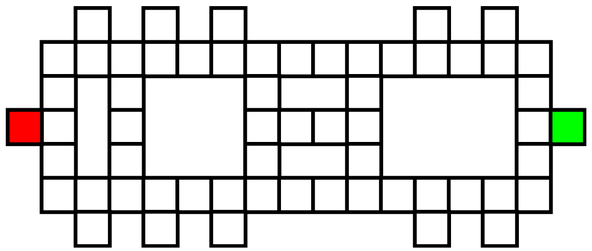
Setup

Scenario 1: Gridworld-Collision

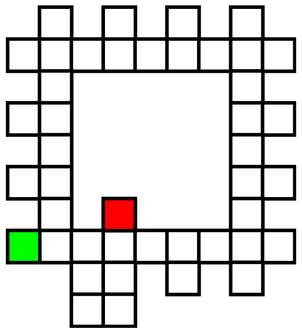
Setup

- Multi-agent gridworlds (Melo and Veloso, 2009)

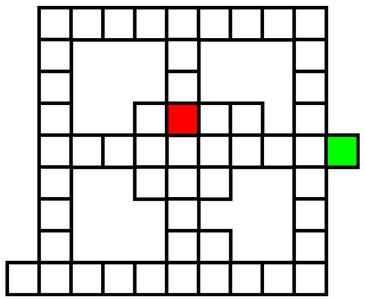
MIT



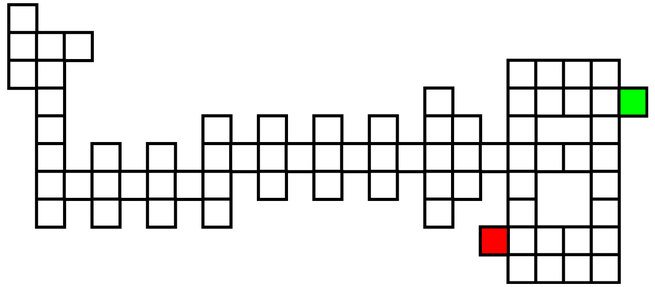
ISR



Pentagon



SUNY

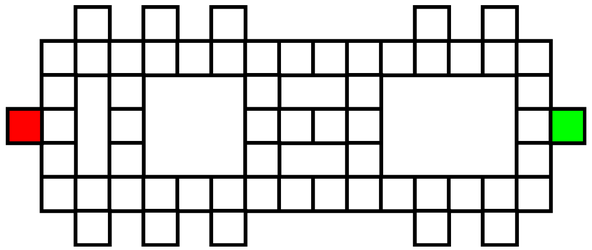


Scenario 1: Gridworld-Collision

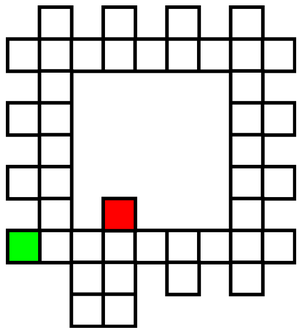
Setup

- Multi-agent gridworlds (Melo and Veloso, 2009)
- Safety Specification: Don’t collide, same shield for all maps

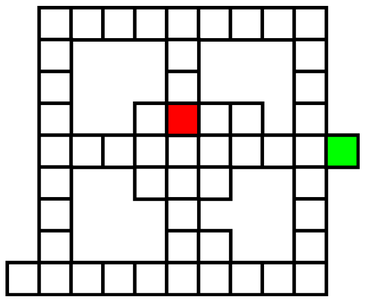
MIT



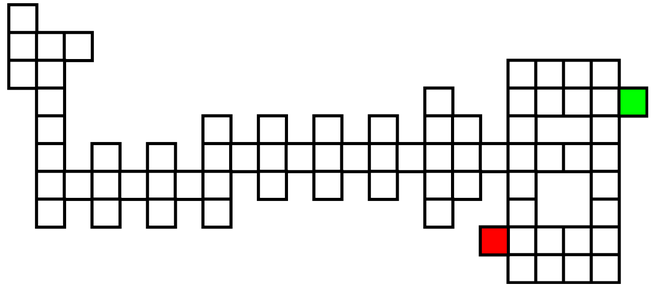
ISR



Pentagon



SUNY

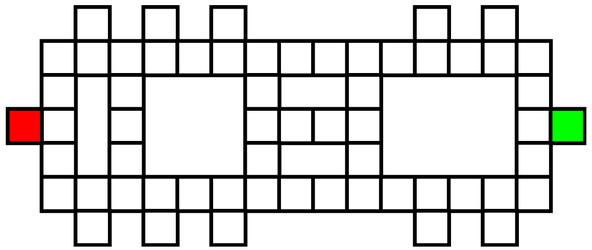


Scenario 1: Gridworld-Collision

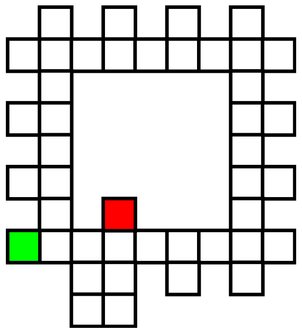
Setup

- Multi-agent gridworlds (Melo and Veloso, 2009)
- Safety Specification: Don’t collide, same shield for all maps
- Tabular Q-Learning, ϵ-greedy actions

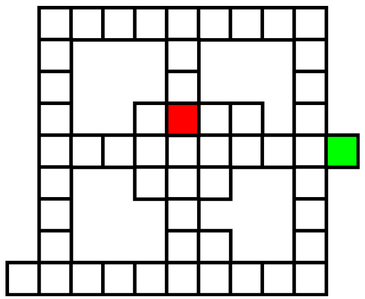
MIT



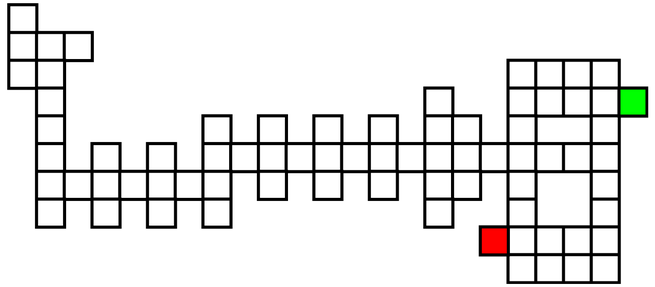
ISR



Pentagon



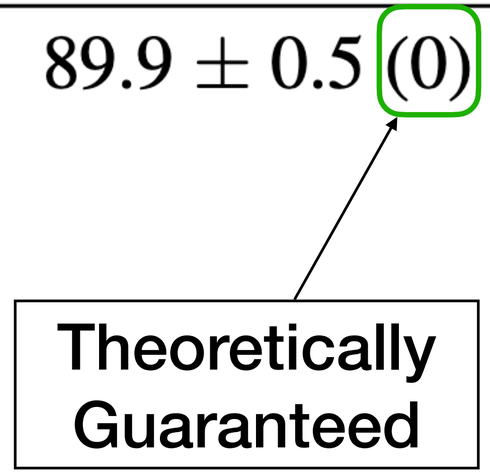
SUNY



Scenario 1: Gridworld-Collision

Results

Map Name	Centralized	Decentralized	No Shield
ISR	90.4 ± 0.6 (0)	89.9 ± 0.5 (0)	90.0 ± 0.6 (12.3)



Scenario 1: Gridworld-Collision

Results

Map Name	Centralized	Decentralized	No Shield
ISR	90.4 ± 0.6 (0)	89.9 ± 0.5 (0)	90.0 ± 0.6 (12.3)

Theoretically Guaranteed

Plus thousands of violations during training

Scenario 1: Gridworld-Collision

Results

About the same performance

Map Name	Centralized	Decentralized	No Shield
ISR	90.4 ± 0.6 (0)	89.9 ± 0.5 (0)	90.0 ± 0.6 (12.3)

Theoretically Guaranteed

Plus thousands of violations during training

More in the paper

Or visit our poster!

- Additional Experiments (momentum environments, DQN)

More in the paper

Or visit our poster!

- Additional Experiments (momentum environments, DQN)
- Unambiguous states

More in the paper

Or visit our poster!

- Additional Experiments (momentum environments, DQN)
- Unambiguous states
- Transforming lists of actions into individual shields

More in the paper

Or visit our poster!

- Additional Experiments (momentum environments, DQN)
- Unambiguous states
- Transforming lists of actions into individual shields
- More details + proof of algorithm

Conclusion
And Future Work

- Algorithm preserves safety & performance

Conclusion
And Future Work

- Algorithm preserves safety & performance
- Ongoing: Partial Observability

Thank You!

Chris Amato

Stavros Tripakis

Special Thanks To:
My Advisors (Above)
NSF (SaTC Award CNS-1801546, CAREER Award IIS-2044993)
Army Research Office (Award W911NF2010265)
Northeastern University Research Computing

Questions?

Email: melcer.d@northeastern.edu

Full paper: #3410