DRIMA: Disentangling Sources of Risk for Distributional Multi-Agent Reinforcement Learning

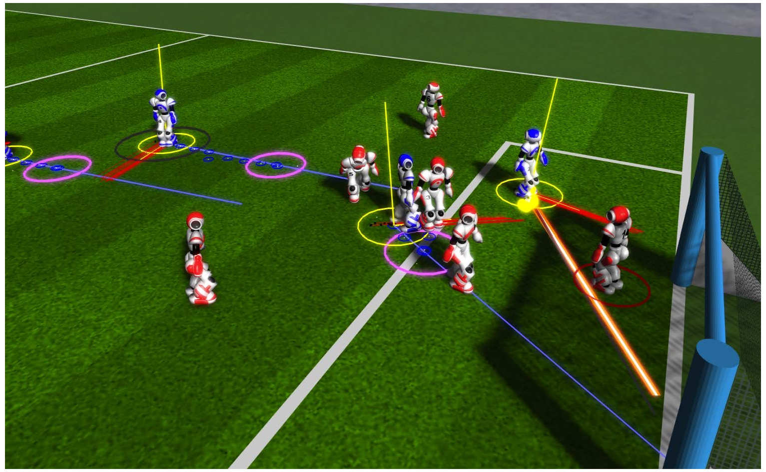
Kyunghwan Son, Junsu Kim, Sungsoo Ahn, Roben Delos Reyes,
Yung Yi, Jinwoo Shin

KAIST POSTECH

Cooperative Multi-Agent Reinforcement Learning

- Distributed multi-agent systems with a shared reward
- Each agent has an partial observation
- Centralized training with decentralized execution has shown great success
- Agents still often fail to cooperative in highly stochastic environments

Drone Swarm Control



Cooperation Game

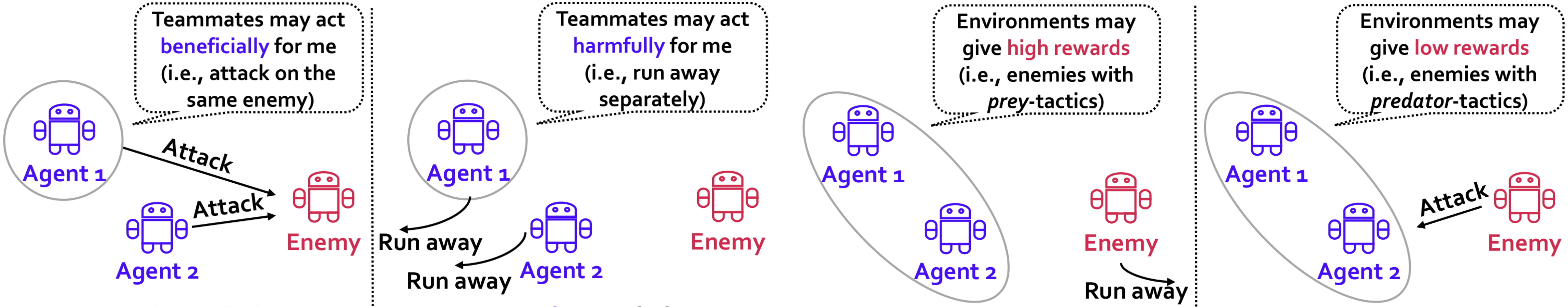
Network Optimization

Risk-Sensitive Reinforcement Learning

- Risk rather than simple expectations for return distribution caused by state transitions, rewards, and actions
- Risk-sensitive policies act with a risk measure, such as variance
- Main goal
  - Applying risk-sensitive technique to multi-agent reinforcement learning to learn more robust policies against various factors of uncertainty

Two Types of Uncertainty in MARL

- Cooperative uncertainty stems from how the agents cannot communicate with each other
- Environmental uncertainty is caused by stochastic transition and the rewarding mechanism of the environment



Cooperative optimism (i.e., $w_{\mathtt{agt}}$ is near 1)

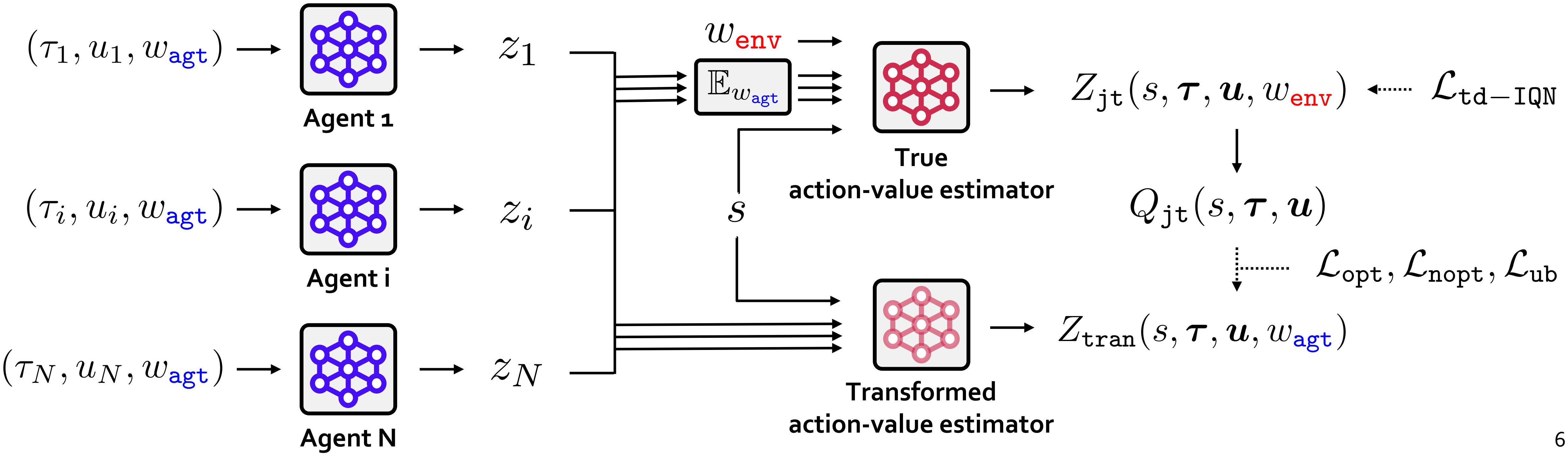
Cooperative pessimism (i.e., $w_{\mathtt{agt}}$ is near 0)

Environmental optimism (i.e., $w_{\mathtt{env}}$ is near 1)

Environmental pessimism (i.e., $w_{\mathtt{env}}$ is near 0)

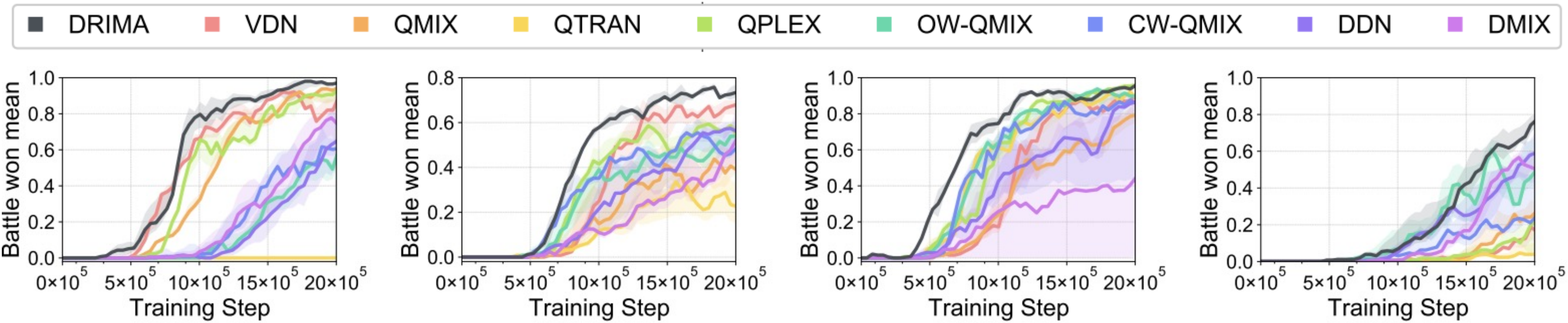
DRIMA: Disentangled Risk-Sensitive MARL

- True action-value estimator which learns the joint action-value captures environmental risk with a risk level w_{env}
- Transformed action-value estimator which learns an action-value guided by the true-action value estimator while capturing cooperative risk with a risk level w_{agt}



Results: Explorative Scenarios

- Question 1: Are there scenarios that existing methods cannot solve? Can DRIMA solve them through disentangling sources of risk?
- Explorative: Agents behave exploratory in the training phase
- DRIMA obtains significant gains, where separating cooperative risk and environmental risk is critical



3s_vs_5z

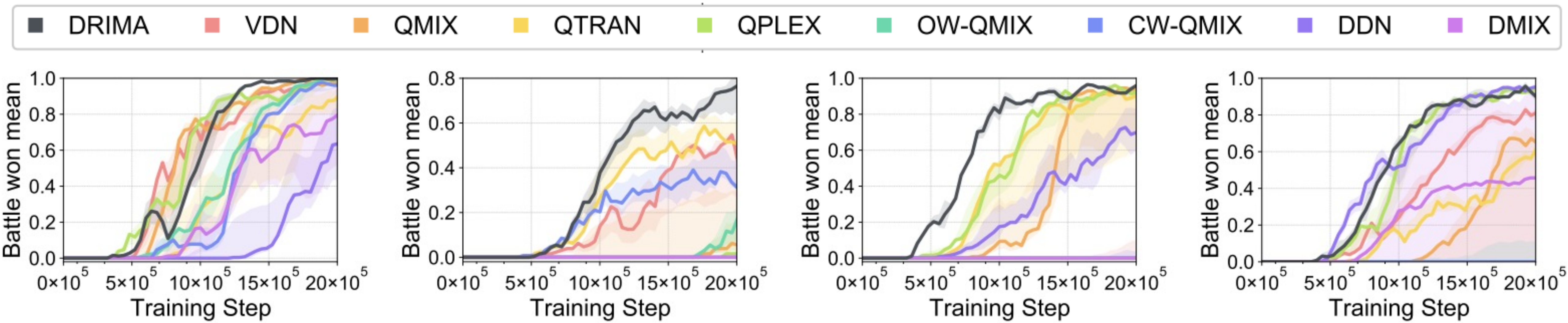
5m_vs_6m

8m_vs_9m

MMM2

Results: Dilemmatic Scenarios

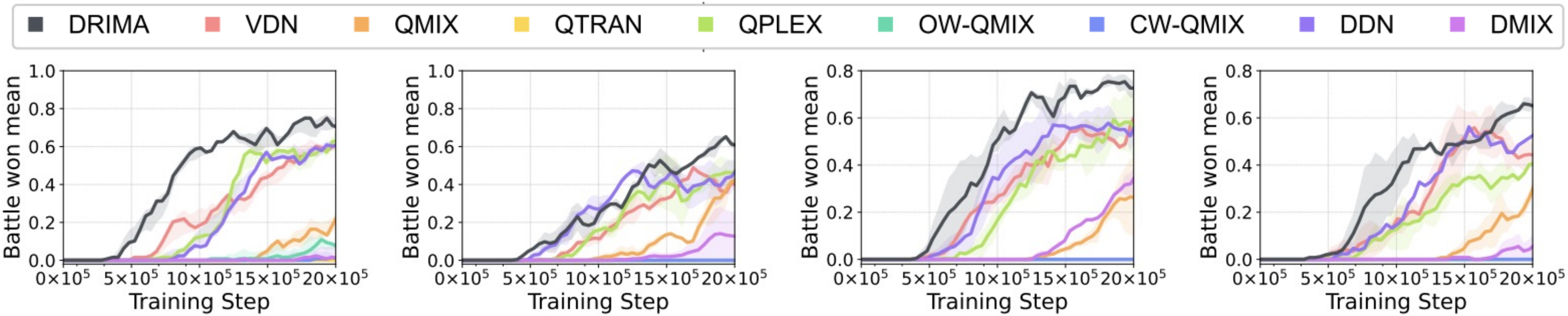
- Question 1: Are there scenarios that existing methods cannot solve? Can DRIMA solve them through disentangling sources of risk?
- Dilemmatic: A social dilemma exists in which agents can learn local optimum policies
- DRIMA also obtains significant gain



3s_vs_5z   5m_vs_6m   8m_vs_9m   MMM2

Results: Noisy Scenarios

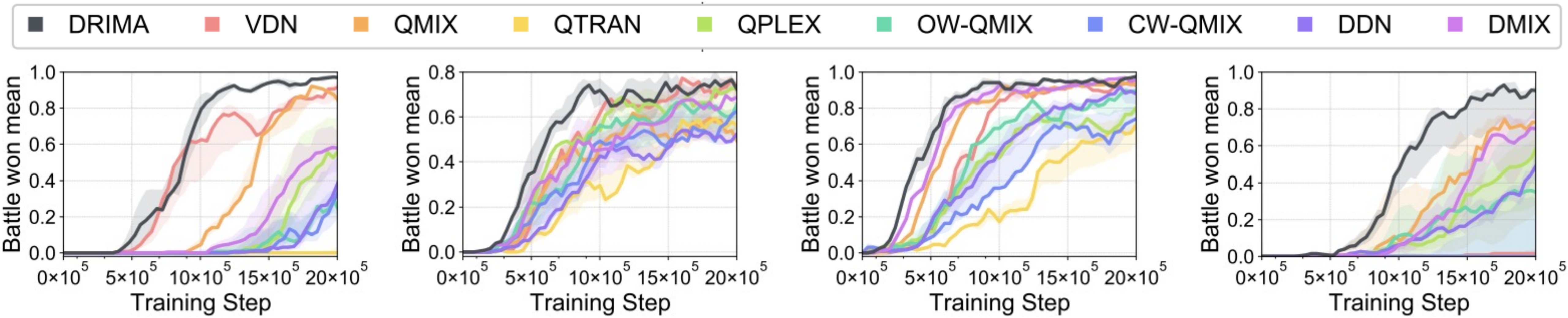
- Question 2: Can DRIMA achieve robust performance through risk sensitivity control in the presence of noisy agents?
- Noisy: During the test phase, some agents may behave incorrectly
- DRIMA achieves robust and high performance through risk sensitivity control even in the presence of noisy agents



3s_vs_5z (random) 3s_vs_5z (adversarial) 5m_vs_6m (random) 5m_vs_6m (adversarial)

Results: Basic Scenarios

- Question 3: Can DRIMA improve sample efficiency and final performance over baseline methods, even under traditional scenarios?
- Basic: Traditional multi-agent reinforcement learning environment
- DRIMA generally achieves the state-of-the-art performance both sample-efficiency and asymptotically



3s_vs_5z 5m_vs_6m 8m_vs_9m MMM2

Thank You!