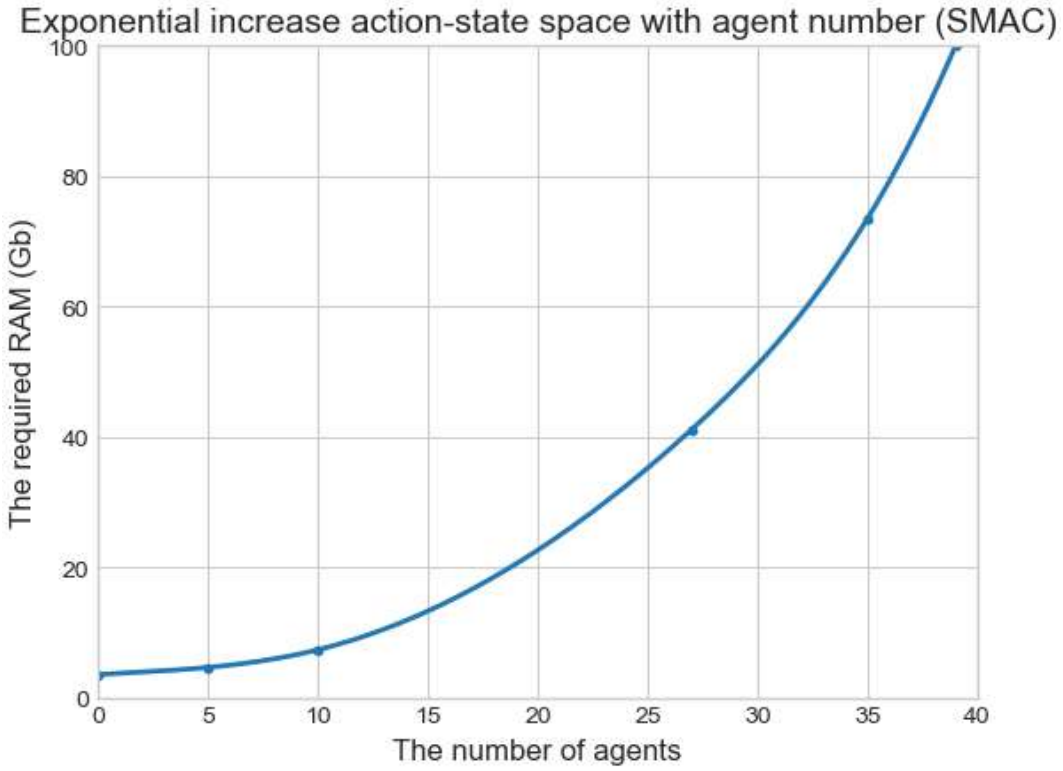
Integrating Suboptimal Human Knowledge with Hierarchical Reinforcement Learning for Large-Scale Multiagent Systems

Dingbang Liu[1,2], Shohei Kato[2], Wen Gu[3], Fenghui Ren[1], Jun Yan[1], Guoxin Su[1]
University of Wollongong[1], Nagoya Institute of Technology[2], Japan Advanced Institute of Science and Technology[3]

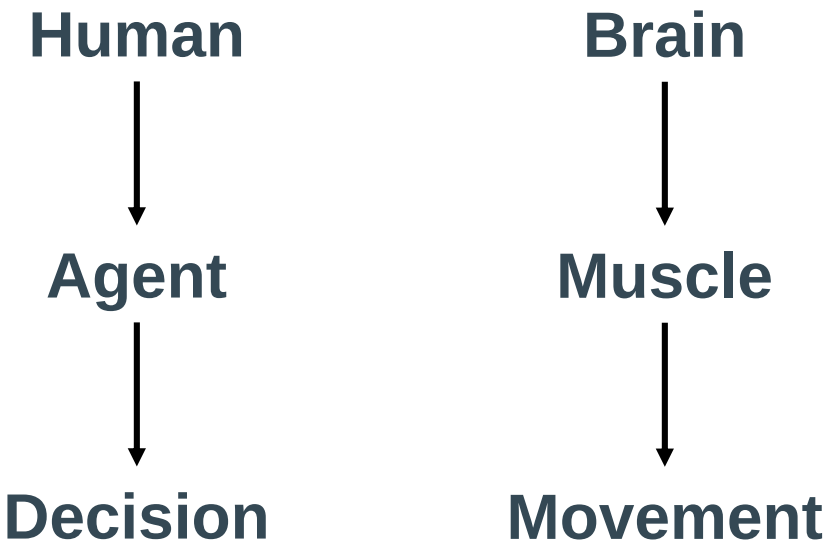
Introduction

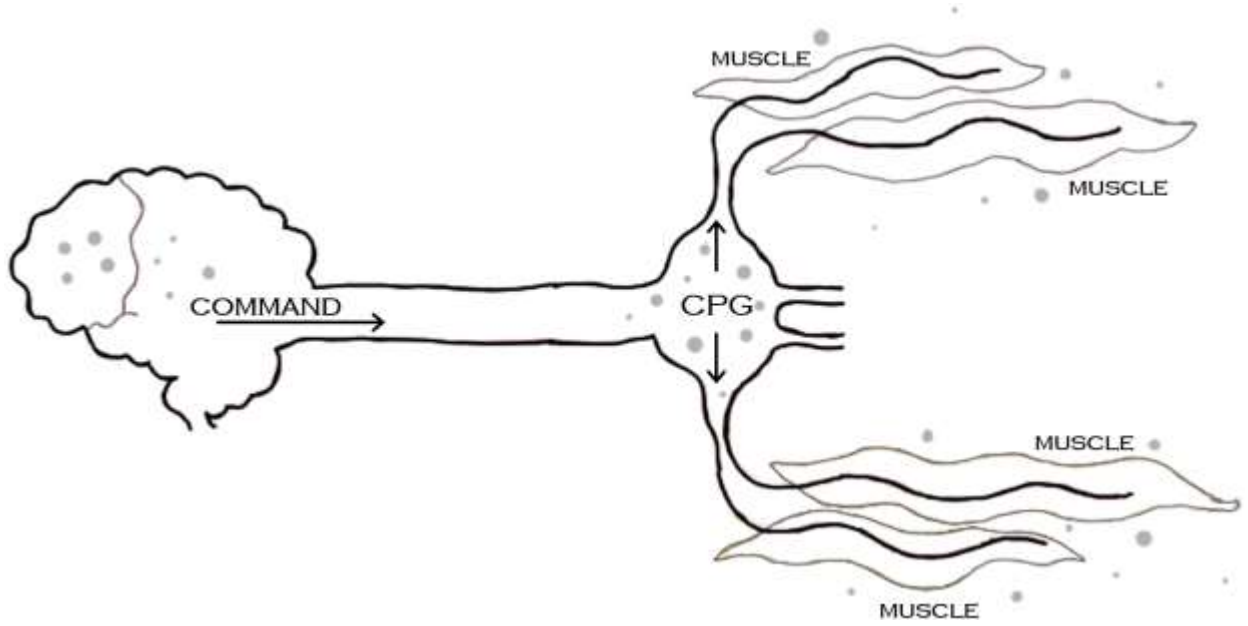
- **Learning difficulty in large-scale MAS**: Sample inefficiency and curse of dimensionality make it challenging to learn from scratch;
- **Human-On-The-Loop**: As the most intuitive and common source, humans can provide demonstrations for agents by personally executing tasks;
- **Human burden mitigation**: Providing step-by-step demonstrations in MAS can be overwhelming.



Intuition

Inspired by the hierarchical control of human daily activity, we attempt to integrate abstract human knowledge into MARL algorithms

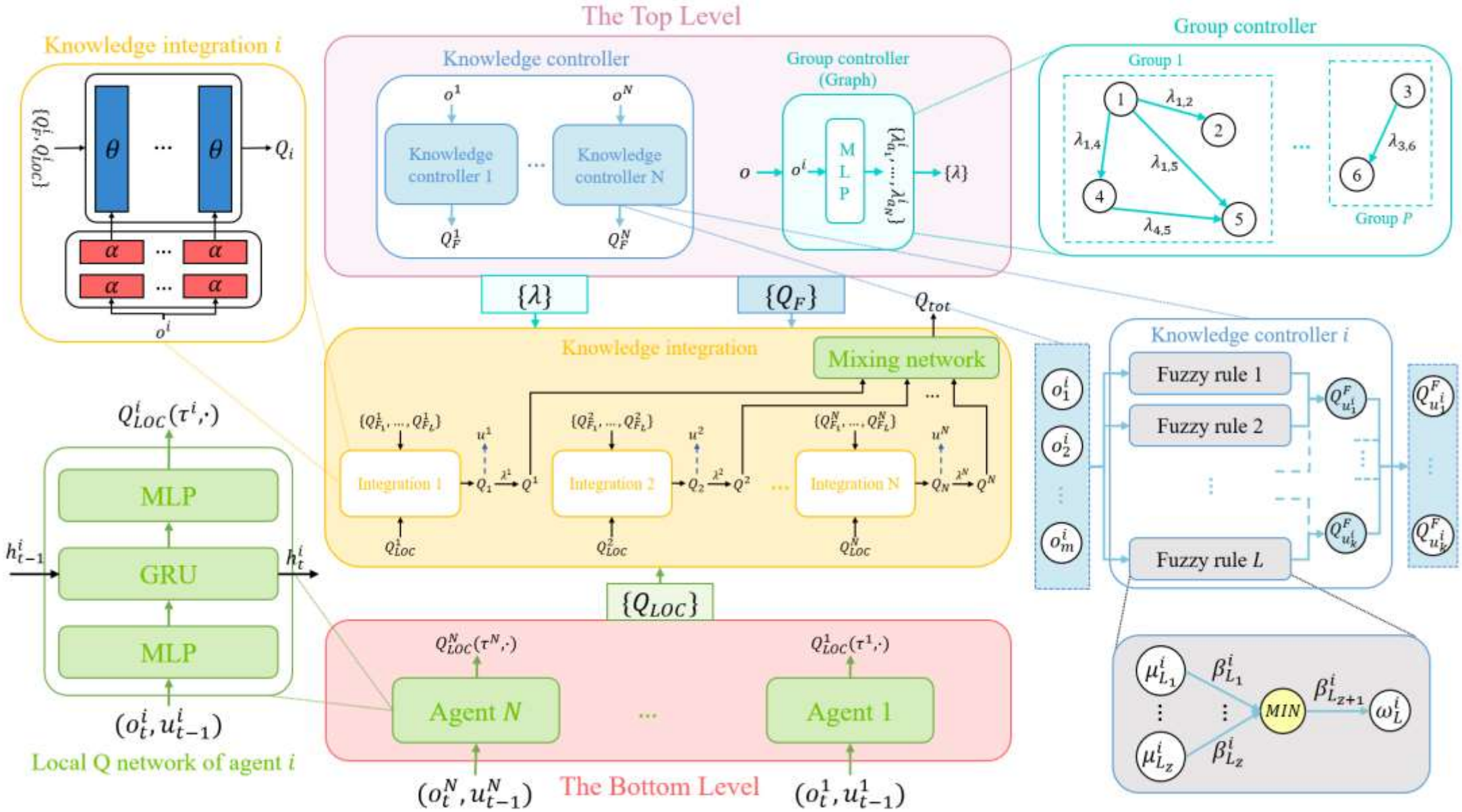




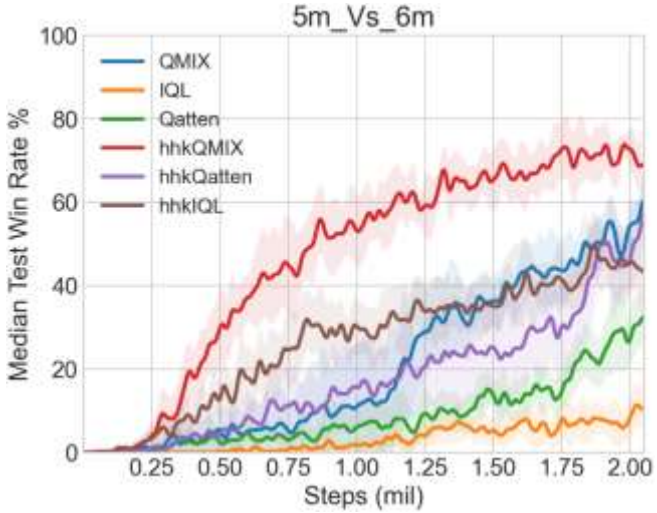
Human daily hierarchical control

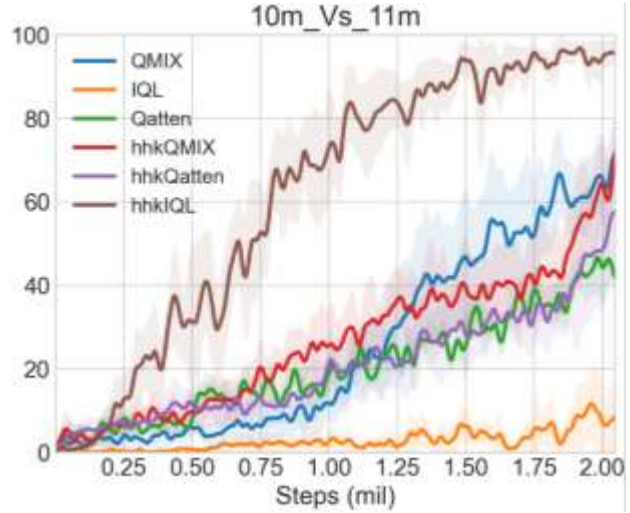
Humans are naturally adept at abstracting and providing high-level knowledge, while agents are expected to autonomously decide the utilization of the proposed knowledge.

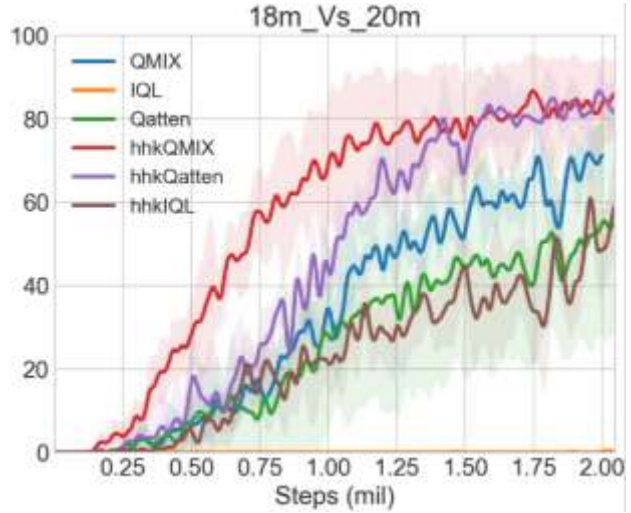
Methodology

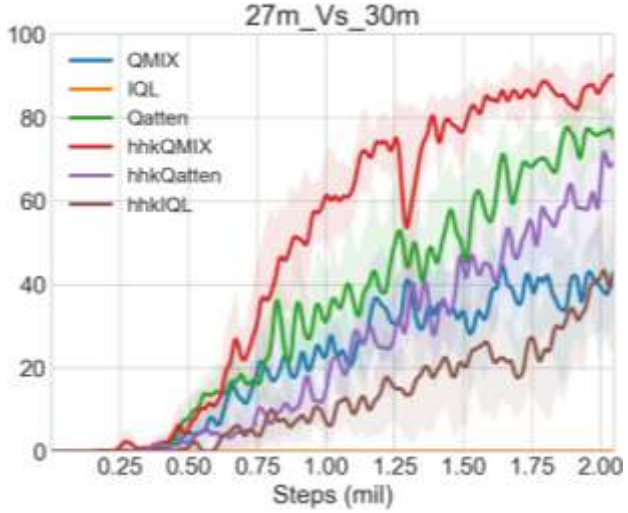


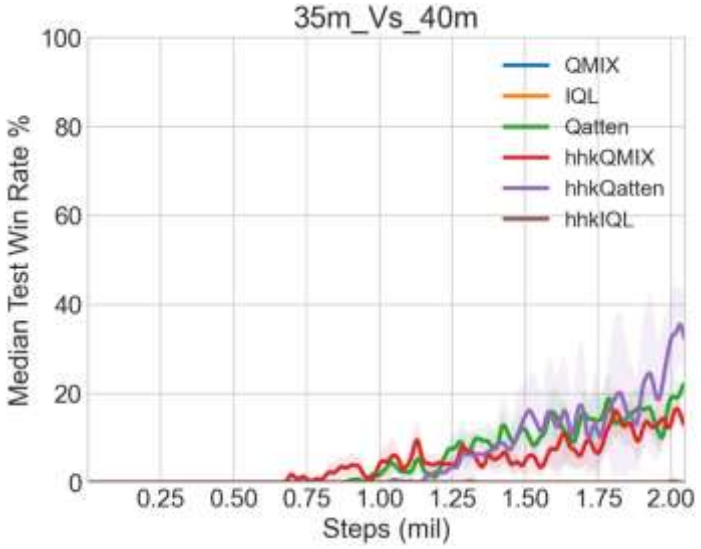
Experiment

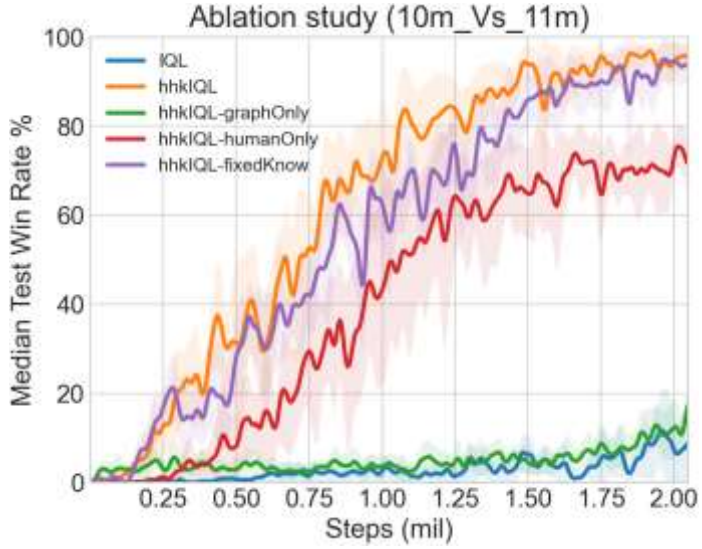

5m_Vs_6m
Median Test Win Rate %
Steps (mil)
QMIX
IQL
Qatten
hhkQMIX
hhkQatten
hhkIQL

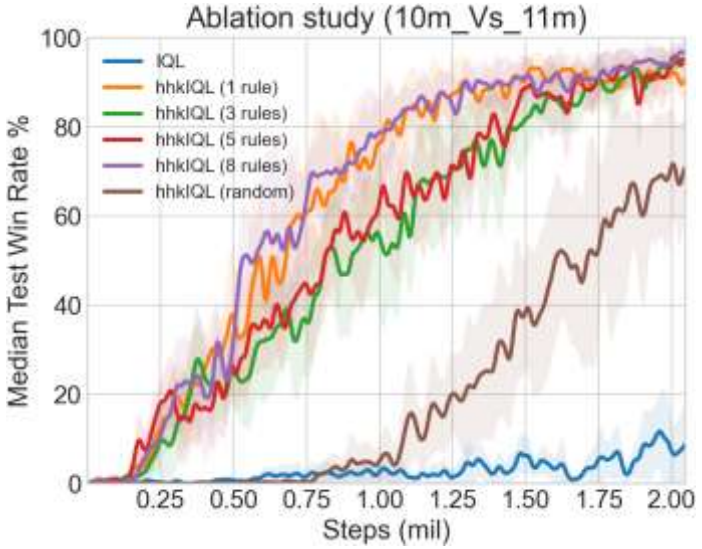

10m_Vs_11m
Steps (mil)
QMIX
IQL
Qatten
hhkQMIX
hhkQatten
hhkIQL


18m_Vs_20m
Steps (mil)
QMIX
IQL
Qatten
hhkQMIX
hhkQatten
hhkIQL


27m_Vs_30m
Steps (mil)
QMIX
IQL
Qatten
hhkQMIX
hhkQatten
hhkIQL


35m_Vs_40m
Median Test Win Rate %
Steps (mil)
QMIX
IQL
Qatten
hhkQMIX
hhkQatten
hhkIQL


Ablation study (10m_Vs_11m)
Median Test Win Rate %
Steps (mil)
IQL
hhkIQL
hhkIQL-graphOnly
hhkIQL-humanOnly
hhkIQL-fixedKnow


Ablation study (10m_Vs_11m)
Median Test Win Rate %
Steps (mil)
IQL
hhkIQL (1 rule)
hhkIQL (3 rules)
hhkIQL (5 rules)
hhkIQL (8 rules)
hhkIQL (random)

Conclusion

- **A novel hierarchical learning framework for enhancing coordination in large-scale MAS by leveraging suboptimal human knowledge**;
- **Allowing humans to provide knowledge at the top level while agents develop their own policies at the bottom**;
- **This end-to-end method can help improve learning speed and final performance, even when integrating low-performance human knowledge.**

Thank you !