LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

mobile and distributed systems group

VAST: Value Function Factorization with Variable Agent Sub-Teams

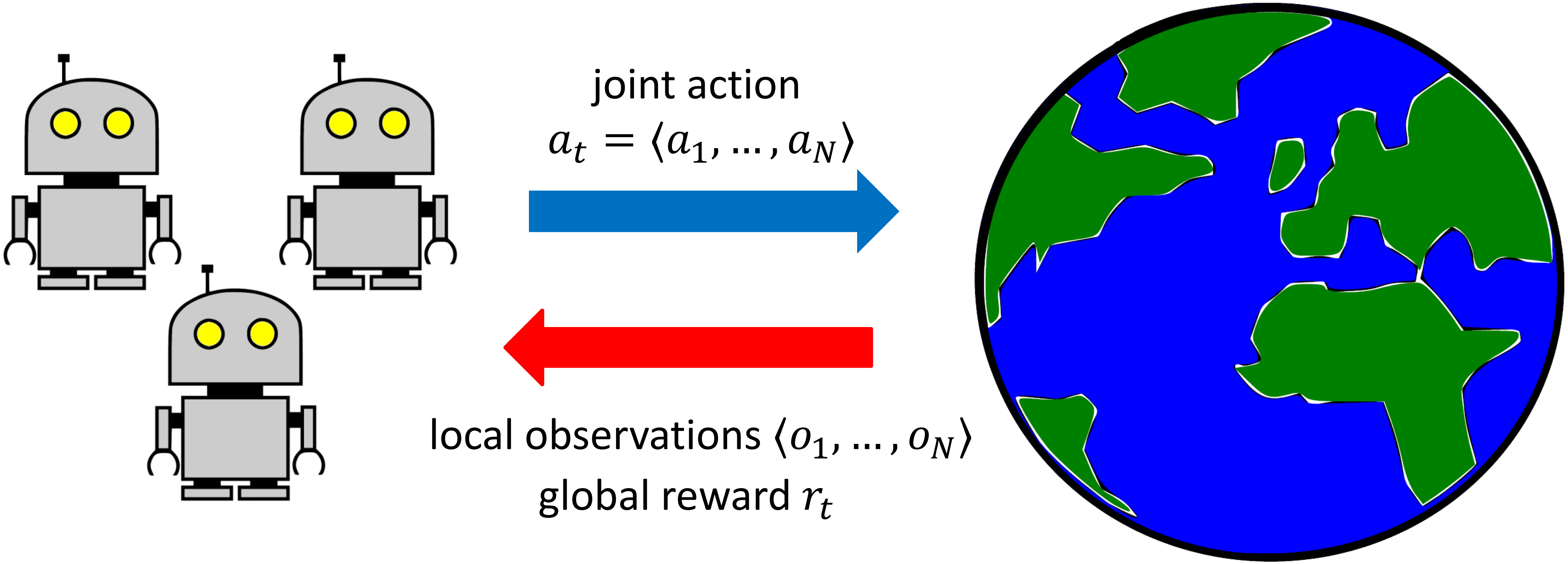
NeurIPS 2021

Thomy Phan[1], Fabian Ritz[1], Lenz Belzner[2],
Philipp Altmann[1], Thomas Gabor[1], Claudia Linnhoff-Popien[1]

[1]LMU Munich, [2]Technische Hochschule Ingolstadt

Preliminaries

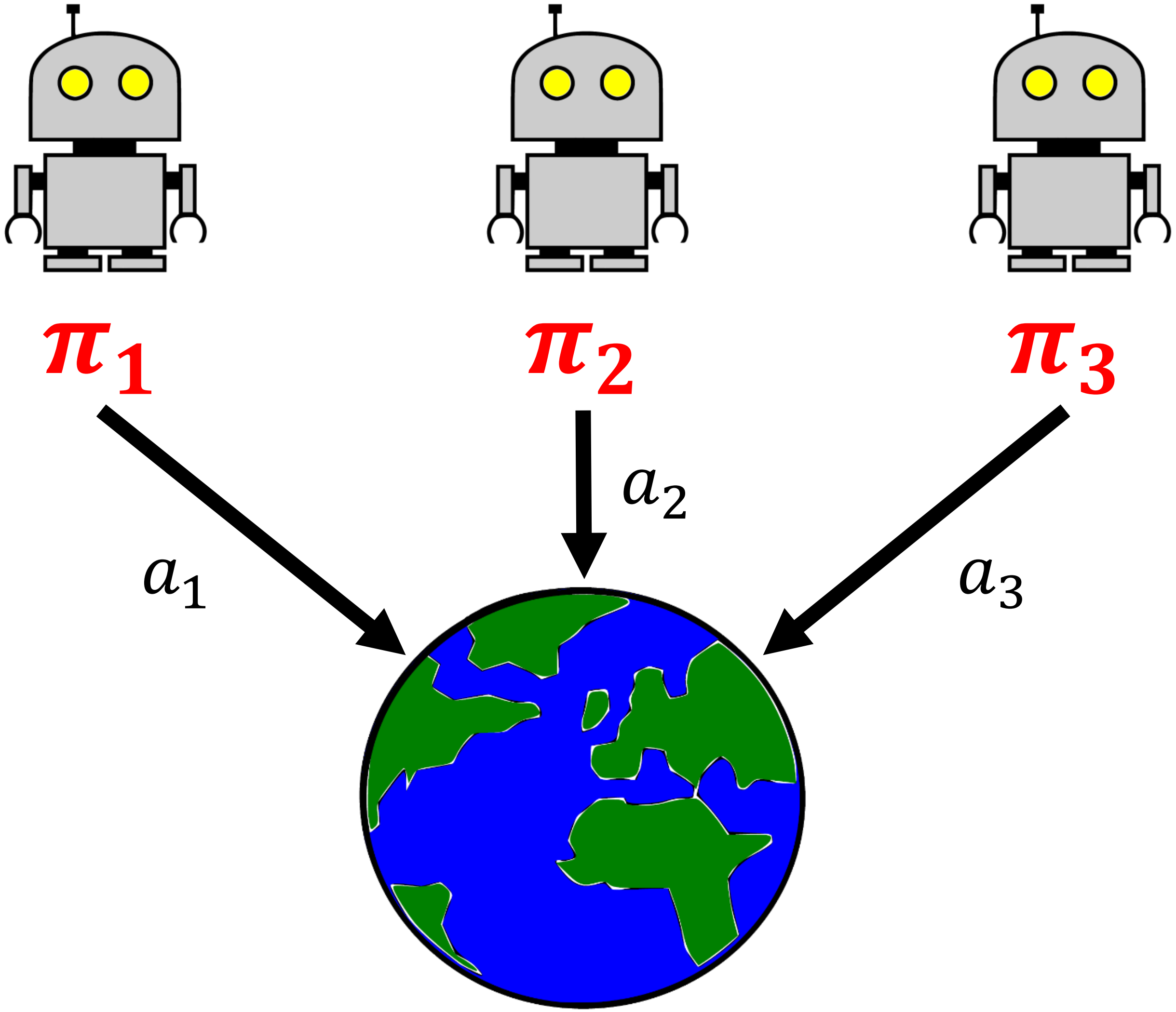
Cooperative Multi-Agent Systems (MAS)



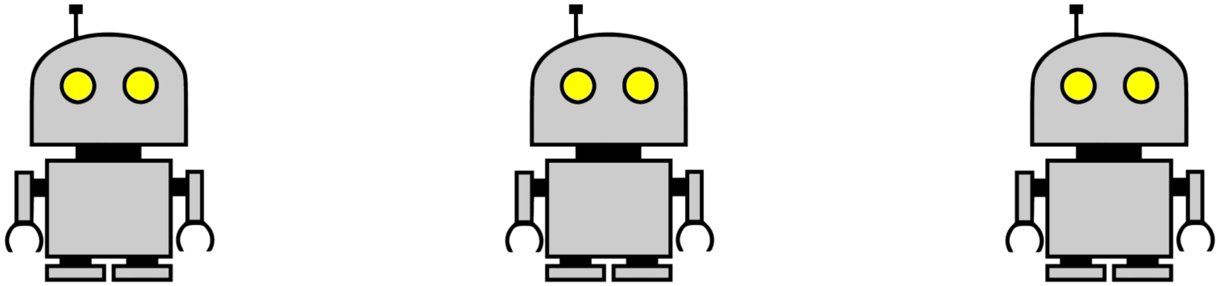
Goal: Maximize expectation of the return $\sum_{k=1}^{\infty} \gamma^k r_{t+k}$

Multi-Agent Reinforcement Learning

Policies



Value-based Multi-Agent Reinforcement Learning



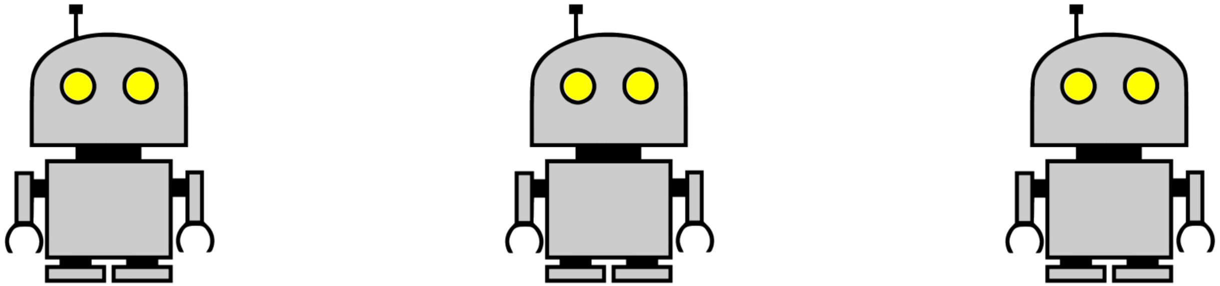
Policies π_1 π_2 π_3

↑ ↑ ↑

Value Functions Q_1 Q_2 Q_3

$$Q_i(\tau_i, a_i) = \mathbb{E}[\sum_{k=1}^{\infty} \gamma^k r_{t+k}]$$

Value Function Factorization



Local Value Functions

Q_1 Q_2 Q_3

Ψ

Factorization Operator

VDN
QMIX
QTRAN
W-QMIX
Qatten
QPLEX
...

Centralized Value Function

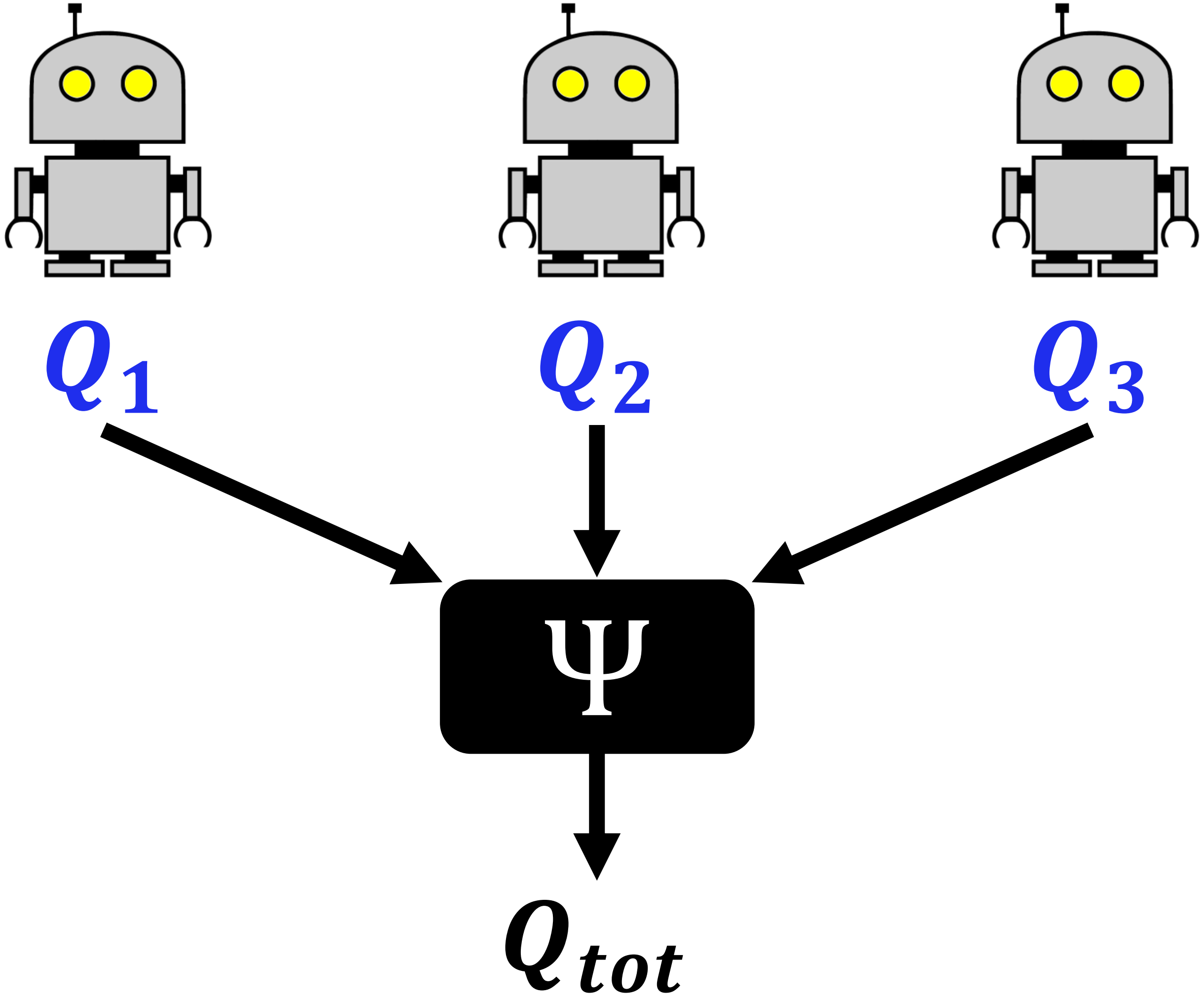
$$Q_{tot} = \mathbb{E}[\sum_{k=1}^{\infty} \gamma^k r_{t+k}]$$

Individual-Global-Max (IGM) Consistency

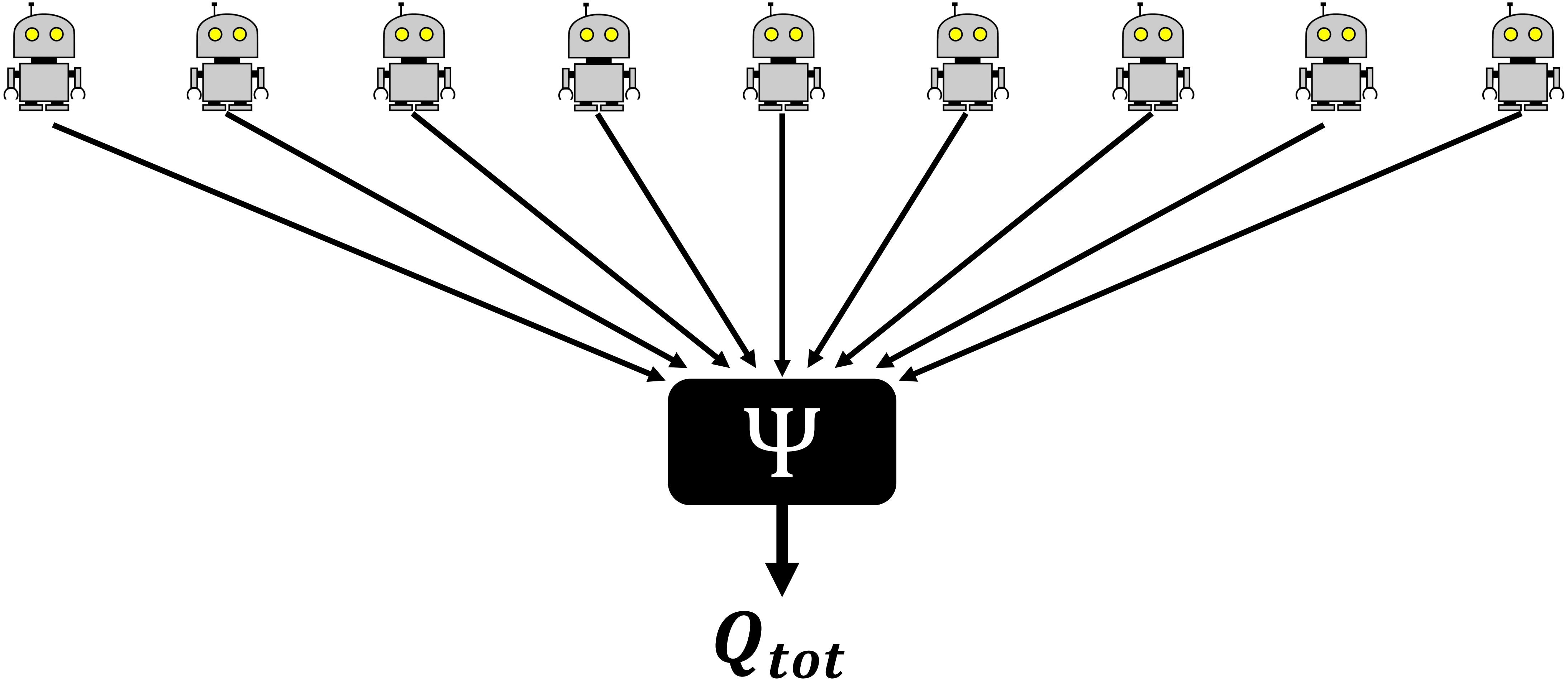
$$argmax_{\boldsymbol{a}_t} \boldsymbol{Q}_{tot}(\boldsymbol{\tau}_t, \boldsymbol{a}_t) = \begin{pmatrix} argmax_{a_{t,1}} Q_1(\tau_{t,1}, a_{t,1}) \\ \dots \\ argmax_{a_{t,N}} Q_N(\tau_{t,N}, a_{t,N}) \end{pmatrix}$$

Factorization operators must ensure IGM consistency

Scalability of Value Function Factorization

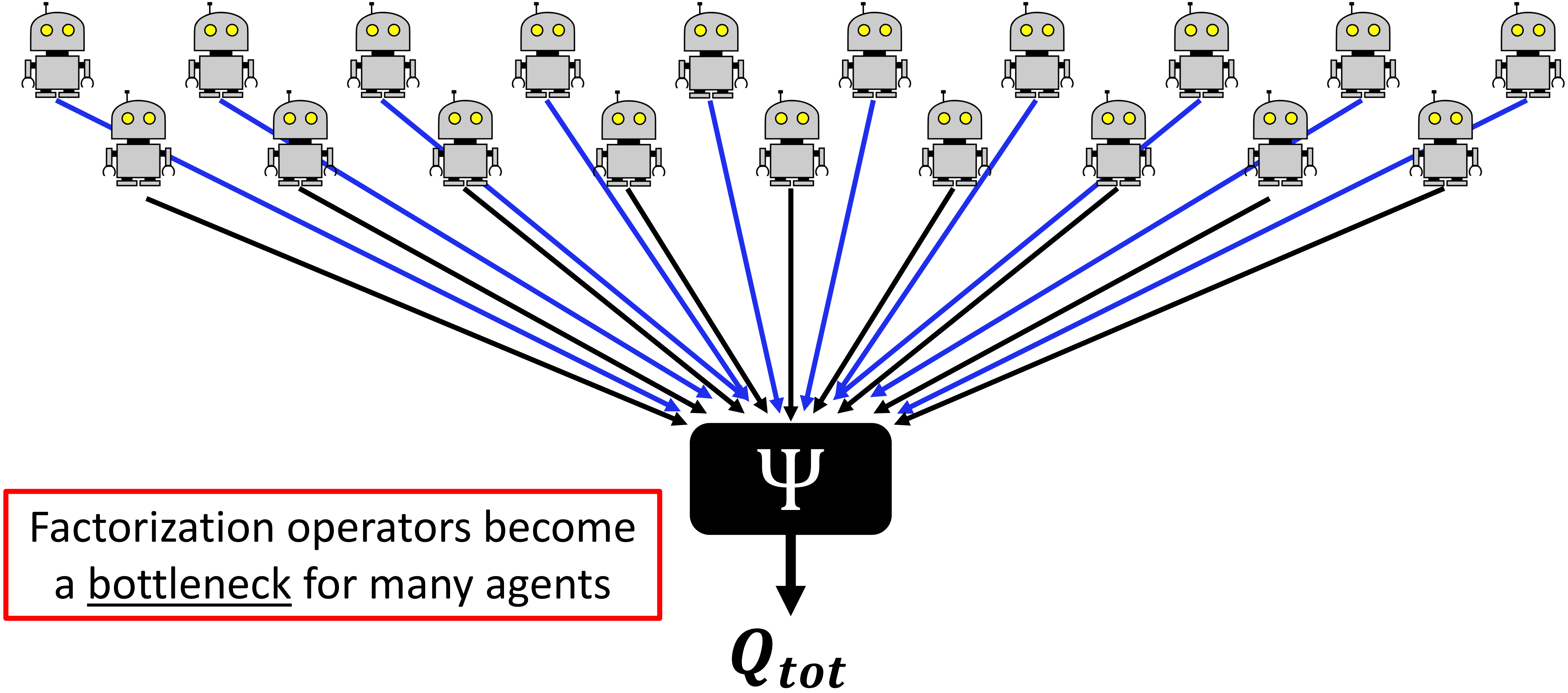


Scalability of Value Function Factorization



Thomy Phan, thomy.phan@ifi.lmu.de

Scalability of Value Function Factorization

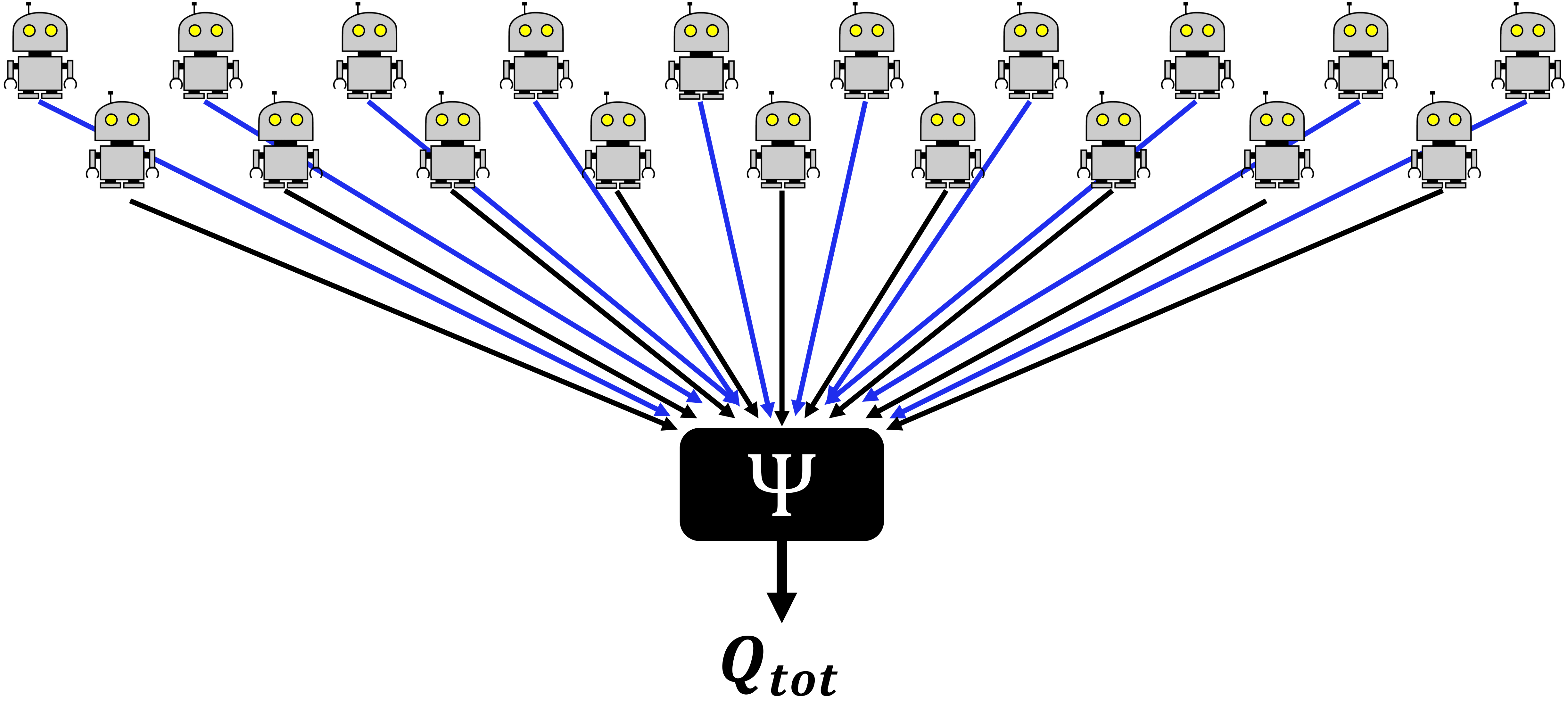


Factorization operators become a bottleneck for many agents

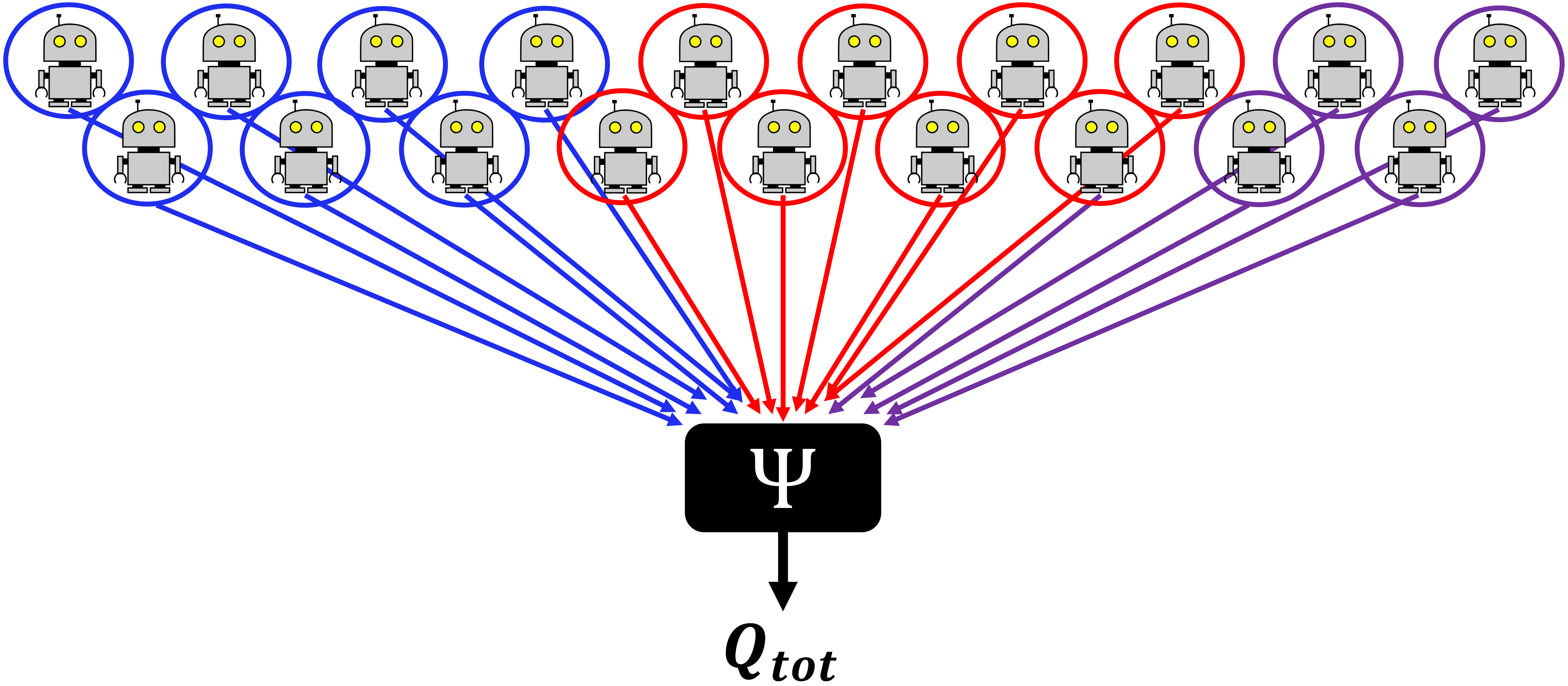
LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Value Function Factorization with Variable Agent Sub-Teams (VAST)

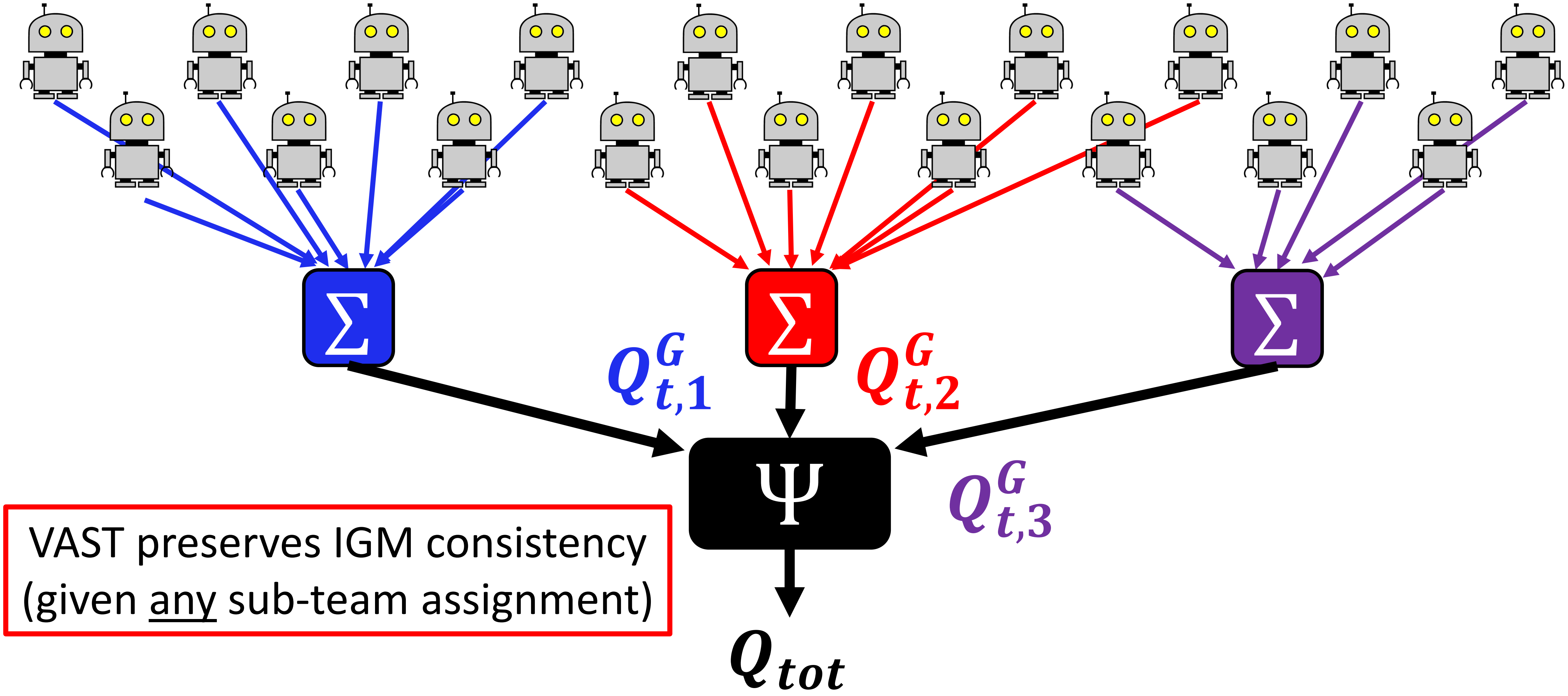
Idea of VAST



Idea of VAST: Sub-Team Assignment



Idea of VAST: Factorization on Sub-Team Values



VAST preserves IGM consistency (given any sub-team assignment)

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Meta-Gradient Learning for Sub-Team Assignment

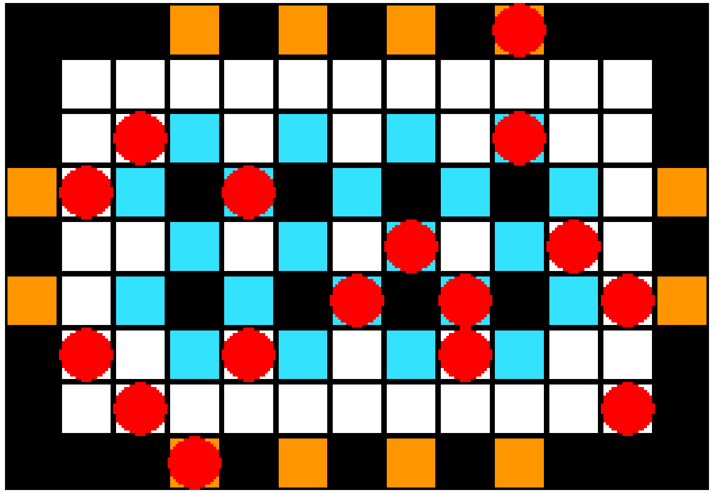
- **Idea:** Optimize sub-team assignments using a high-level objective J
  - Decide at each state s_t which sub-team k agent i should be assigned to
  - Learn meta-policy $\mathcal{X}(k|s_t, i, \tau_{t,i})$ via gradient ascent w.r.t. objective J

$$\hat{A}(k, i, s_t, \tau_{t,i}) \nabla log \mathcal{X}(k|s_t, i, \tau_{t,i})$$

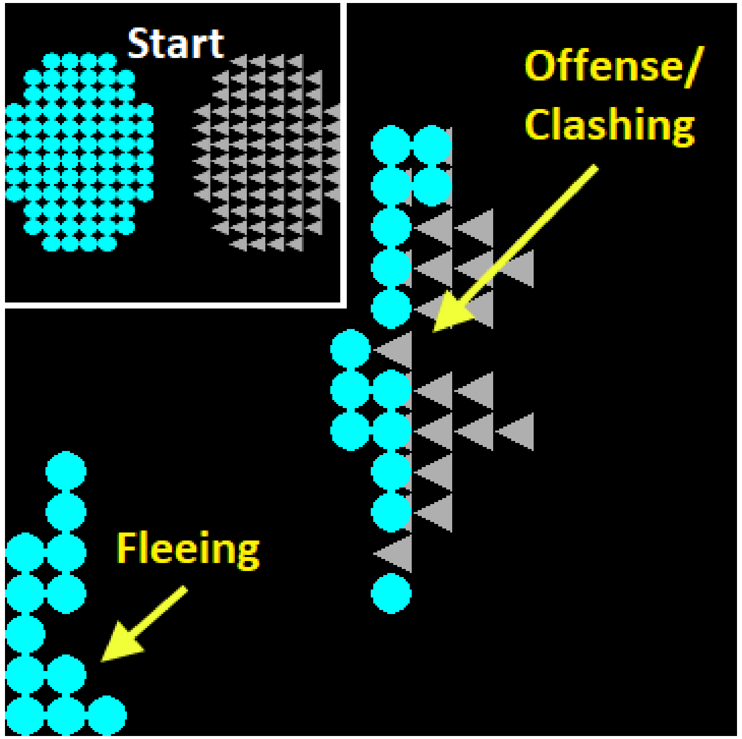
  - Advantage function $\hat{A}$ can be defined using domain knowledge or reward-based metrics (e.g., return, TD-error, …)

Results

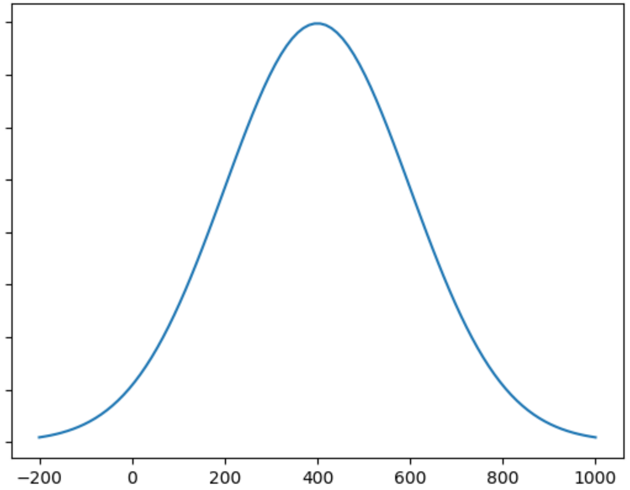
Evaluation Domains



Warehouse
(4 – 16 agents)

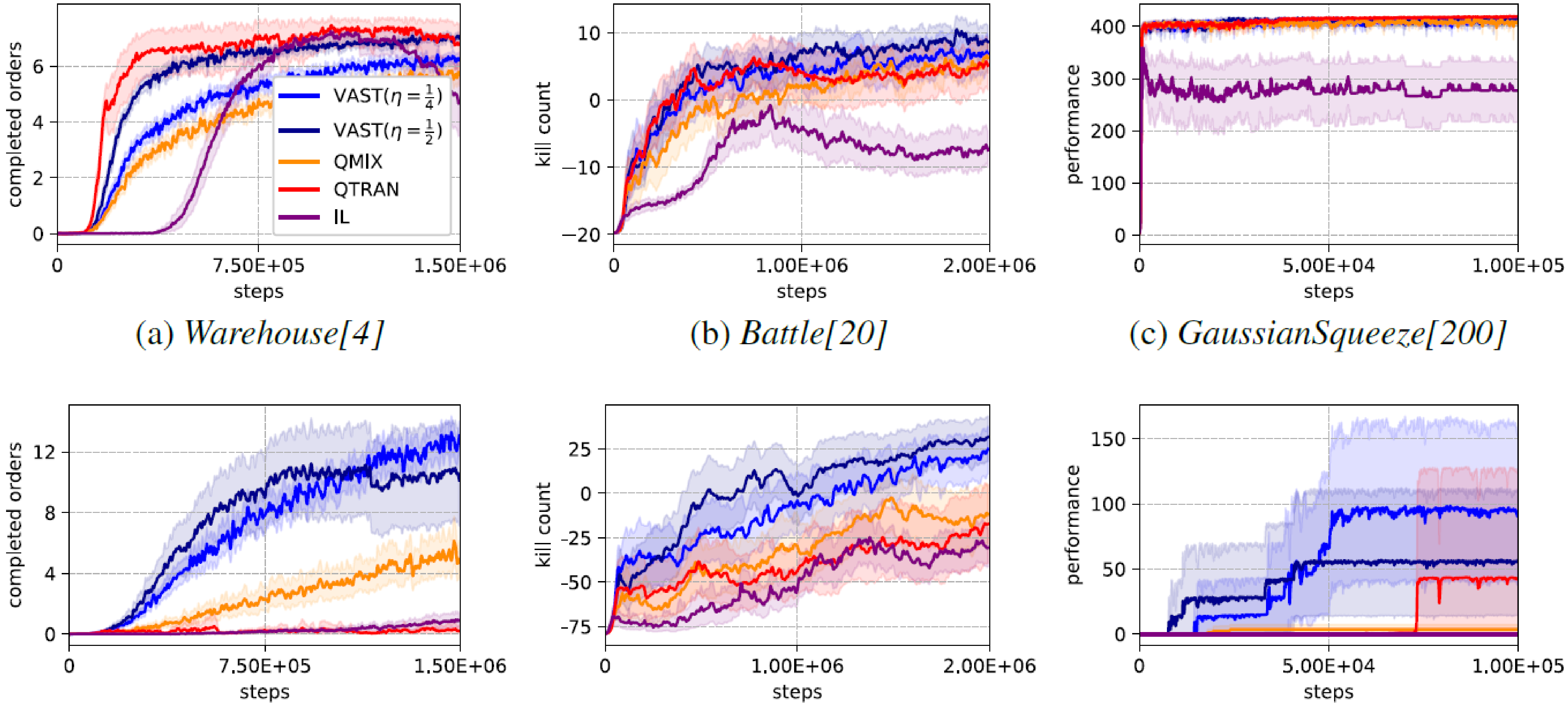


Battle
(20 – 80 agents)



Gaussian Squeeze
(200 – 800 agents)

State-of-the-Art Comparison

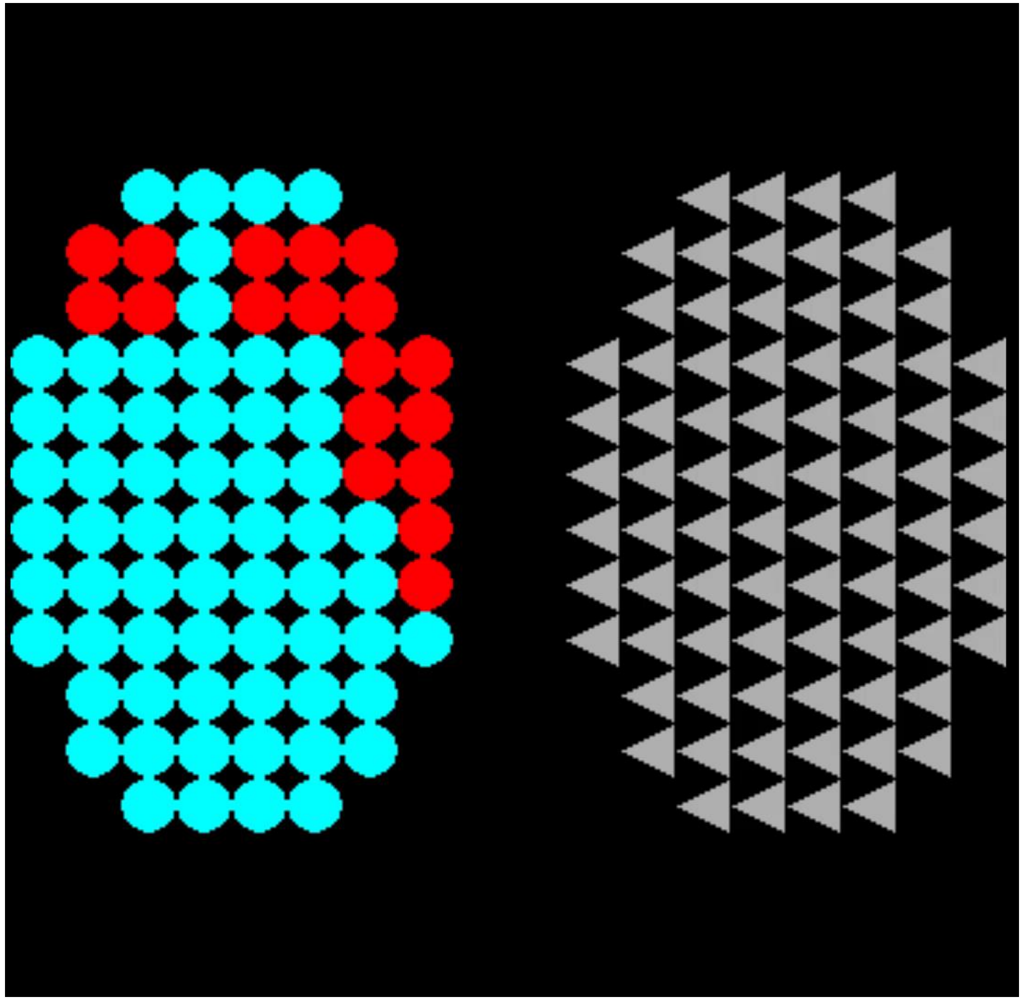


(a) *Warehouse[4]* (b) *Battle[20]* (c) *GaussianSqueeze[200]*

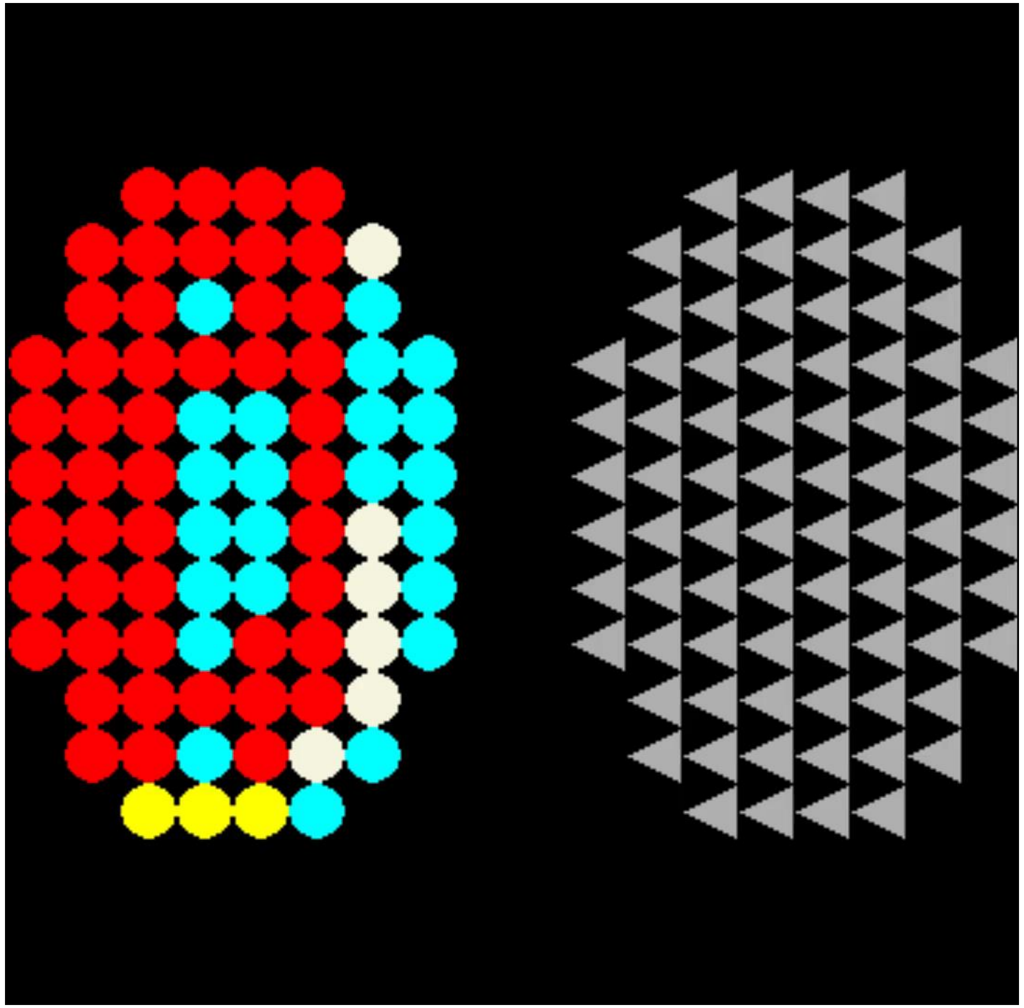
(d) *Warehouse[16]* (e) *Battle[80]* (f) *GaussianSqueeze[800]*

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Meta-Gradient Generated Sub-Teams in Battle[80]



Meta-Gradient Generated Sub-Teams in Battle[80]



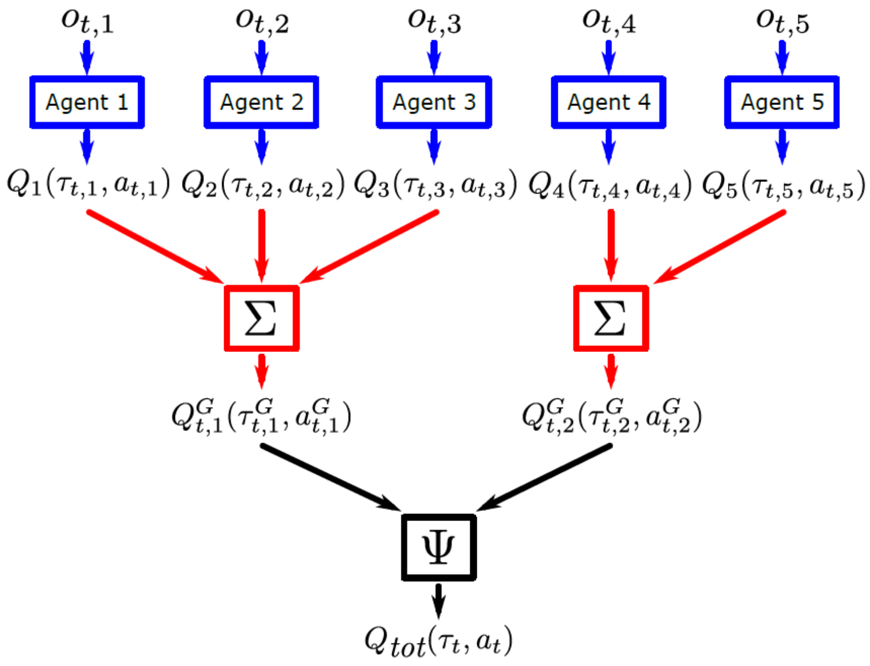
Conclusion

Conclusion and Future Work

- VAST can improve scalability of value factorization w.r.t. many agents
- IGM consistency is preserved by VAST
- Meta-gradient based sub-teams can improve performance of VAST

Future Work

- Deeper hierarchies of sub-teams
- Non-linear factorization of sub-team values



Code available at →

VAST: Value Function Factorization with Variable Agent Sub-Teams

NeurIPS 2021

Thomy Phan[1], Fabian Ritz[1], Lenz Belzner[2],
Philipp Altmann[1], Thomas Gabor[1], Claudia Linnhoff-Popien[1]

[1]LMU Munich, [2]Technische Hochschule Ingolstadt