Composition and Zero-Shot Transfer with Lattice Structures in Reinforcement Learning

Geraud Nangue Tasse*, Steven James, Benjamin Rosman

University of the Witwatersrand, Johannesburg, South Africa

RLC 2025

Provably!!!

UNIVERSITY OF THE WITWATERSRAND JOHANNESBURG

MIND Machine Intelligence and Neural Discovery Institute

RAIL LAB

JAIR Journal of Artificial Intelligence Research

Motivation

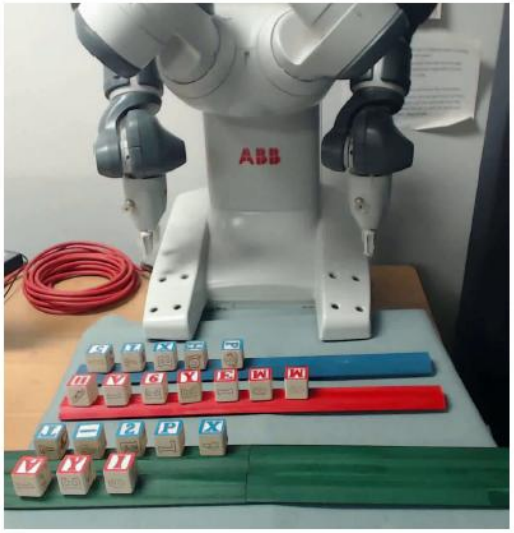
We want **instructable agents** (e.g. via language) that can **solve tasks beyond Boolean rewards** [1,2], and **generalise compositionally** to new tasks.

- "**Serve breakfast** with **plain toast** *and* **ketchup**…"

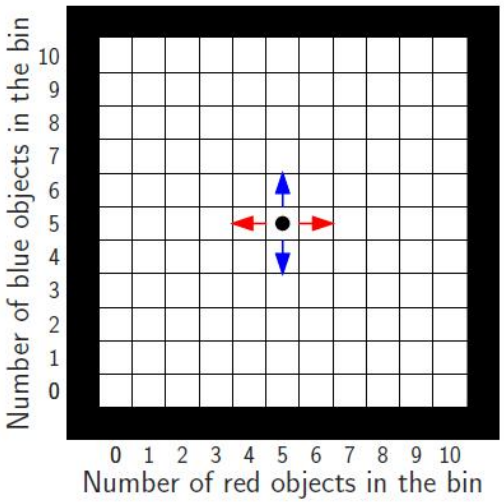
- Neural networks **struggle to generalize compositionally**[3]

[1] G. Nangue Tasse et al., "A Boolean task algebra for reinforcement learning" NeurIPS, 2020.
[2] R. T. Icarte et al., "Using reward machines for high-level task specification and decomposition in reinforcement learning," ICML, 2018
[3] B. M. Lake et al., "Generalization without systematicity: On the compositional skills of sequence-to-sequence recurrent networks. PMLR 2018.

Motivation

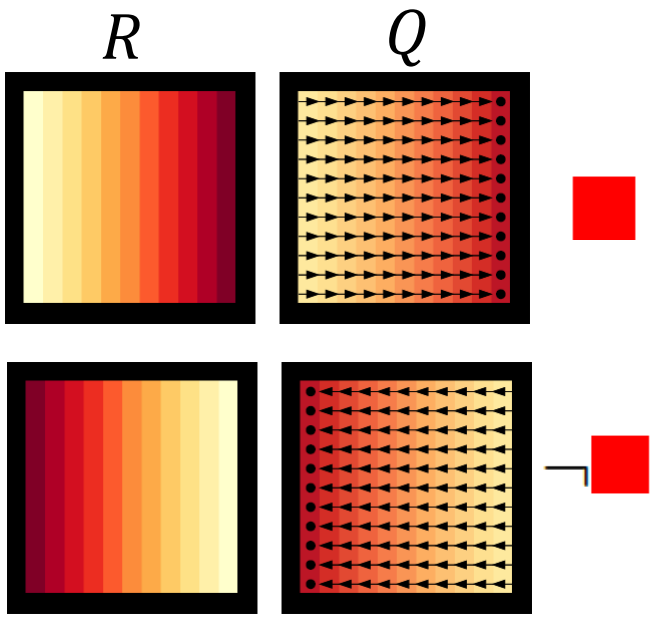


Bin packing domain



Gridworld representation

Pack **all** **red**



Unpack **all** **red**

W1: Task Composition

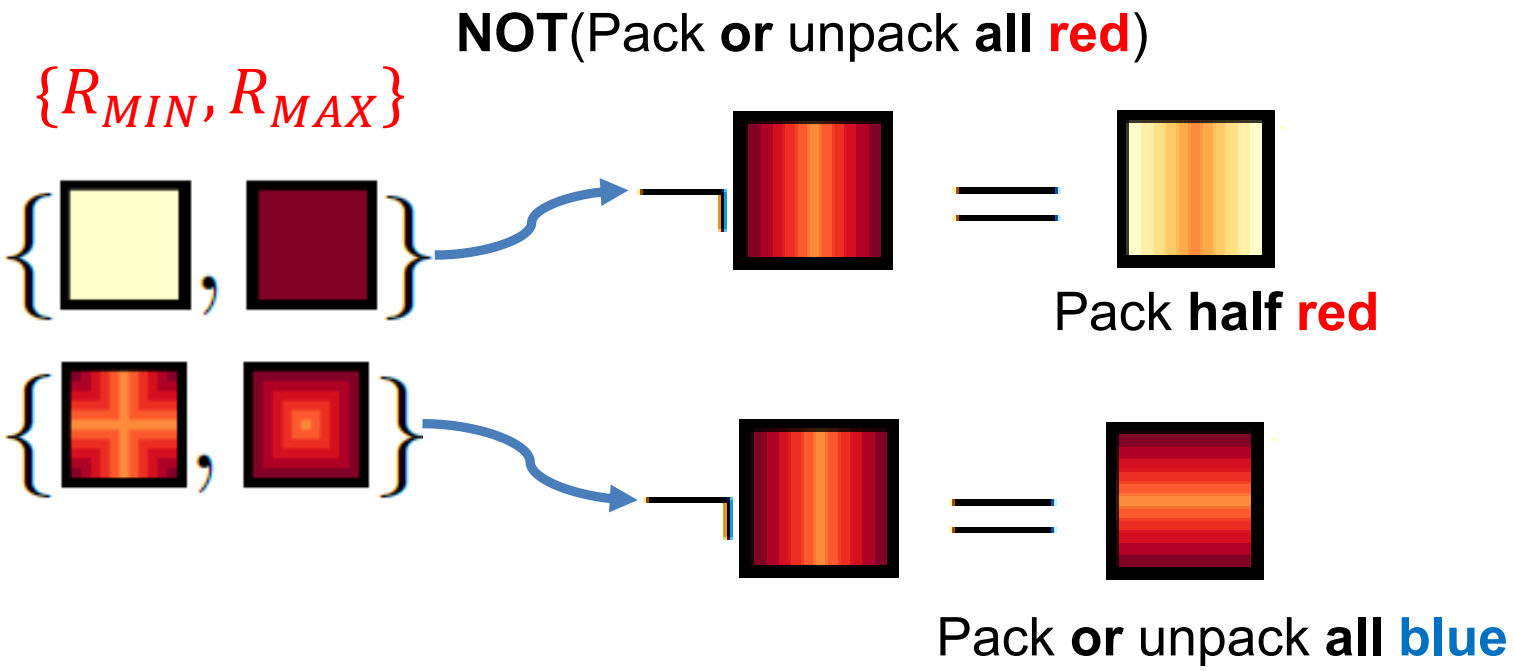
Logic Operators

- ***OR:*** $A \vee B \coloneqq \mathbf{max}\{R_A(s,a,s'), R_B(s,a,s')\}$
- ***AND:*** $A \wedge B \coloneqq \mathbf{min}\{R_A(s,a,s'), R_B(s,a,s')\}$
- ***NOT:*** $\neg A \coloneqq \left(\boldsymbol{R_{MAX}(s,a,s')} + \boldsymbol{R_{MIN}(s,a,s')}\right) - R_A(s,a,s')$

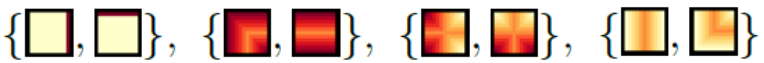
Reward bounds

W2: Different Composition Semantics

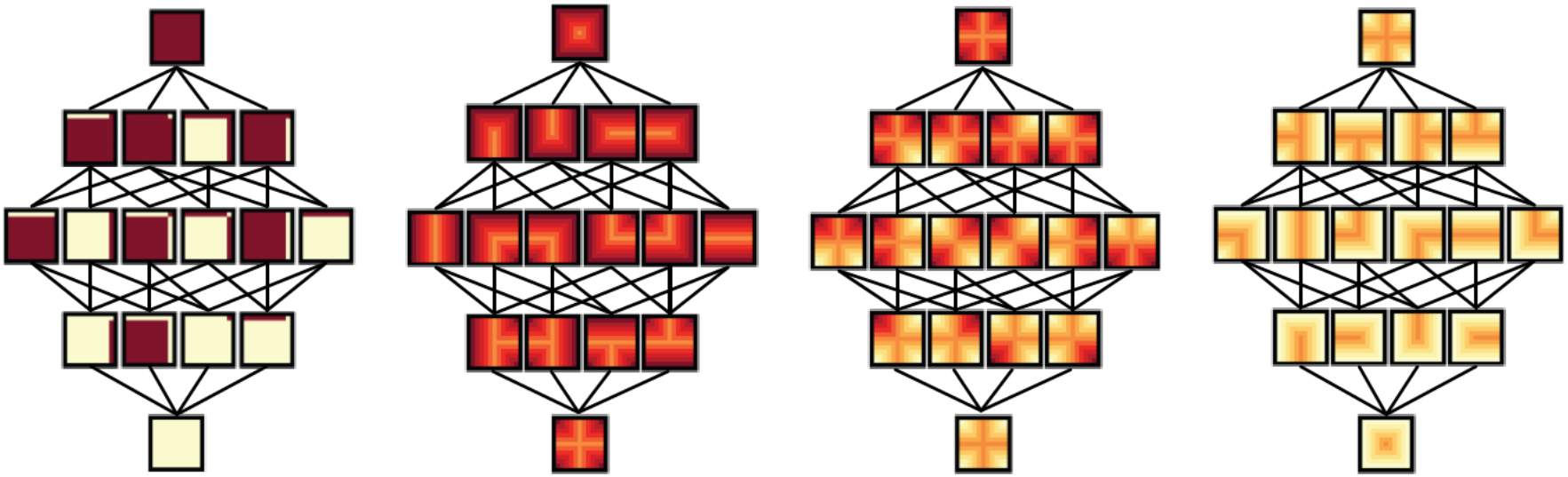
Bounds change semantics:



W3: Boolean Tasks with Dense Rewards

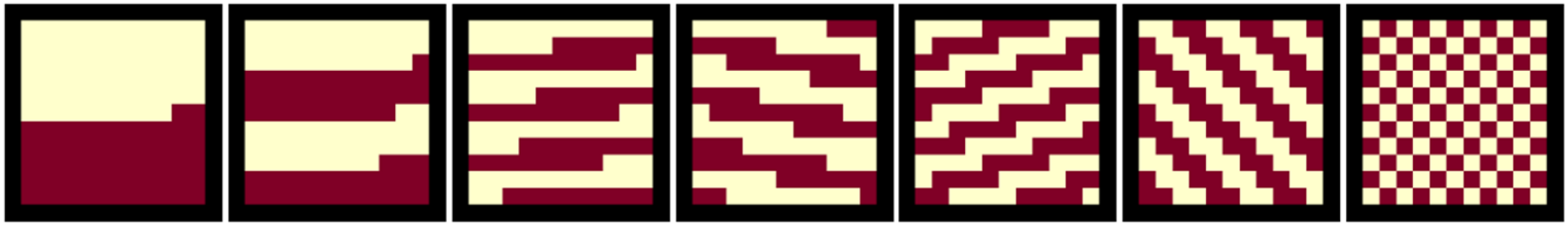
Given a basis: 

Sub spaces



W4: Construct a Basis

Bounds: {  }



W5: World Value Function

Intuition: while solving one task, we should learn about other tasks that we may need to solve in the future

$$\bar{r}(s, g, a) = \begin{cases} \bar{r}_{MIN} & if\ g \neq s \in G \\ r(s, a) & otherwise \end{cases}$$

$$\bar{\pi}(s, a, g) \rightarrow [0,1]$$

$$Q^{\bar{\pi}}(s, g, a) = \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t \bar{r}(s_t, g, a_t)\right]$$

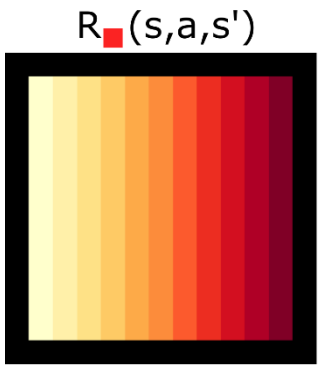
Agent-chosen goals

Mastery!!!

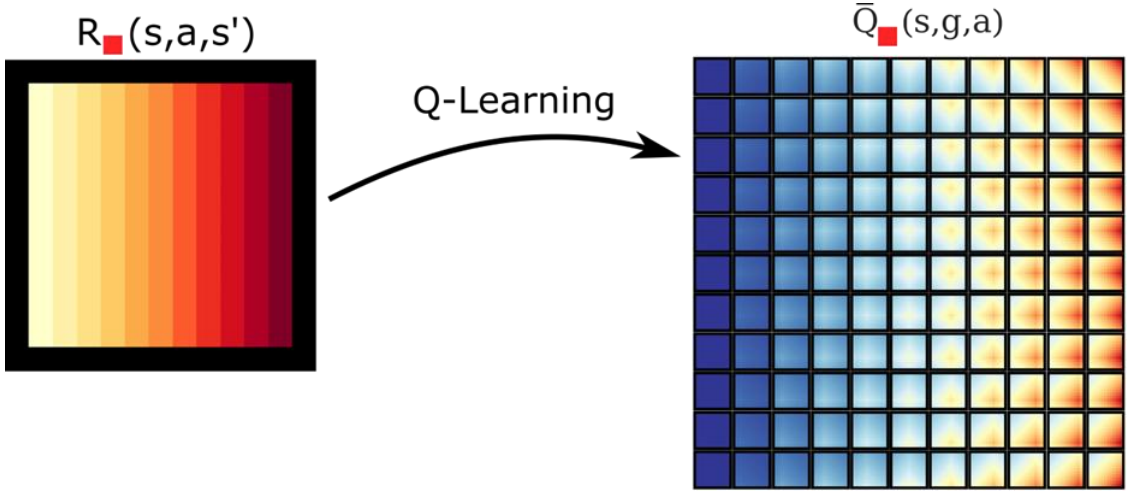
Purely driven by reward maximixation

No spoon-feeeding

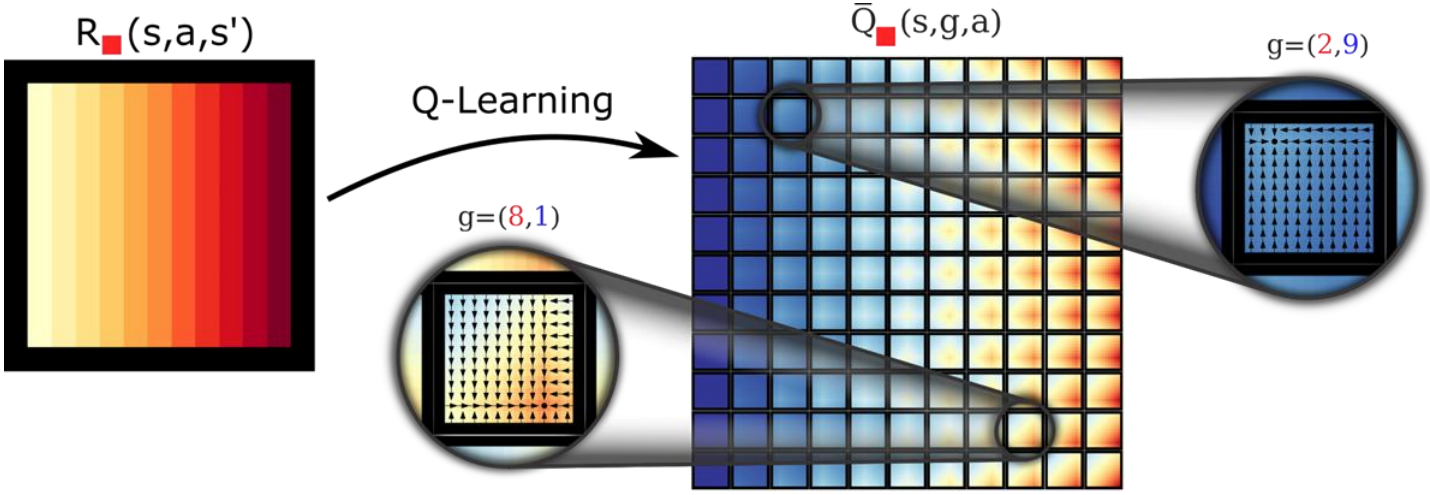
W6: Mastery



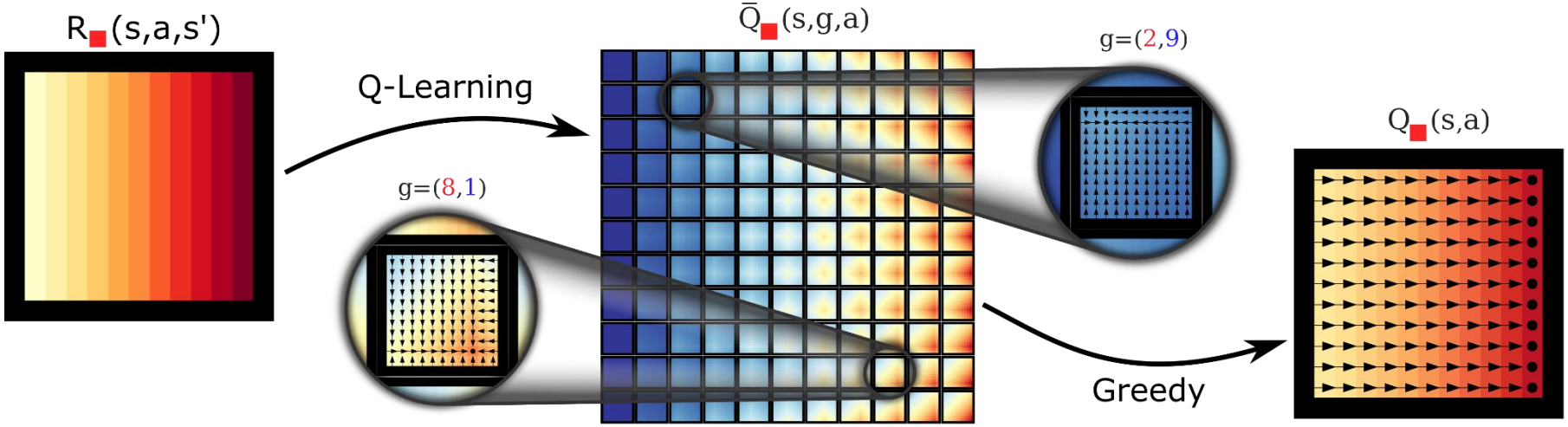
#6 Mastery



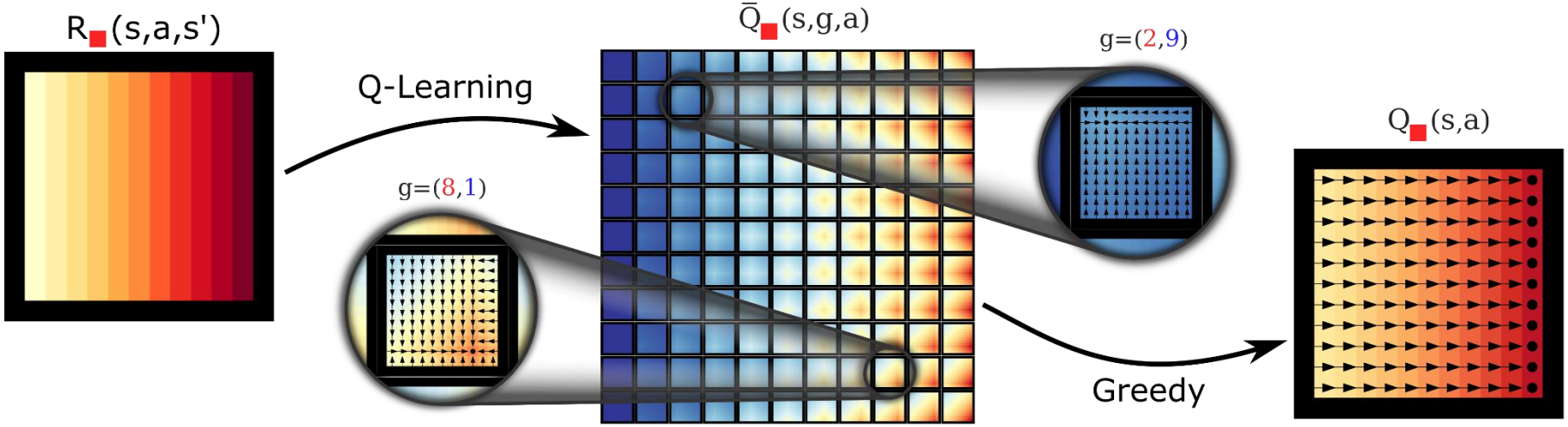
W6: Mastery



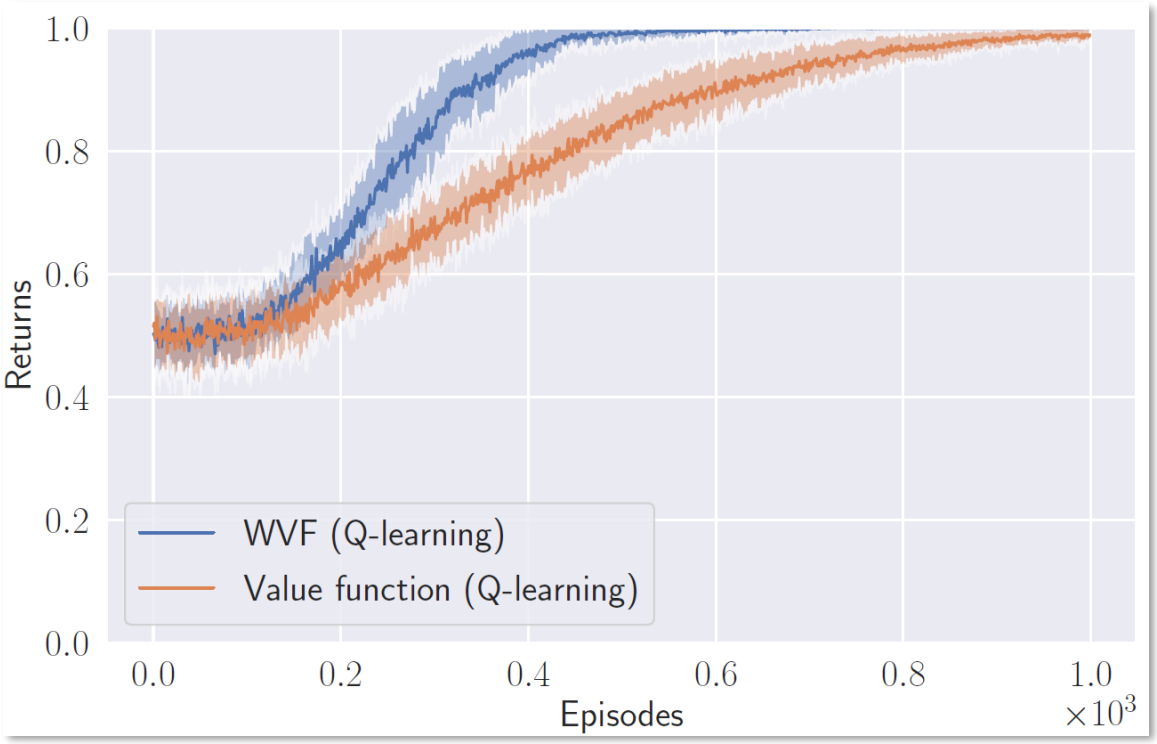
W6: Mastery



W6: Mastery



W7: Better sample-efficiency



W8: Skill Composition

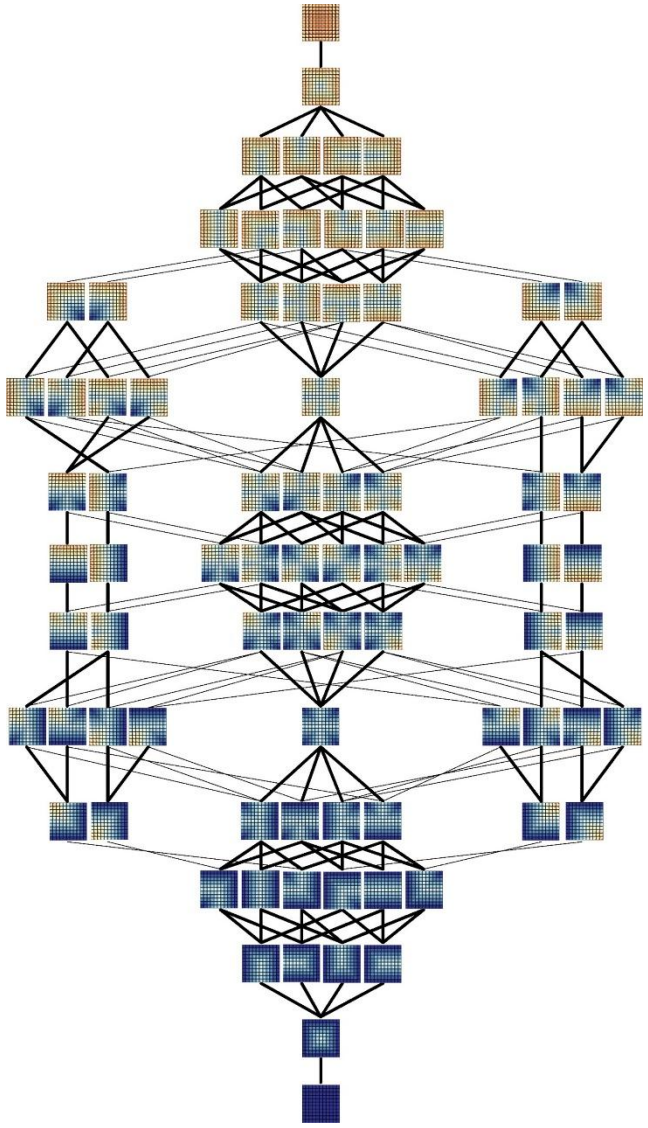
Logic Operators

- ***OR:***
  - $Q_A \vee Q_B \coloneqq \mathbf{max}\{Q_A(s,a,g), Q_B(s,a,g)\}$
- ***AND:***
  - $\boldsymbol{Q_A \wedge Q_B} \coloneqq \mathbf{min}\{Q_A(s,a,g), Q_B(s,a,g)\}$
- ***NOT:***
  - $\neg Q_A \coloneqq \left(\boldsymbol{Q_{MAX}(s,a,g)} + \boldsymbol{Q_{MIN}(s,a,g)}\right) - Q_A(s,a,g)$

W9: Homomorphism over tasks and skills

Logic Operators

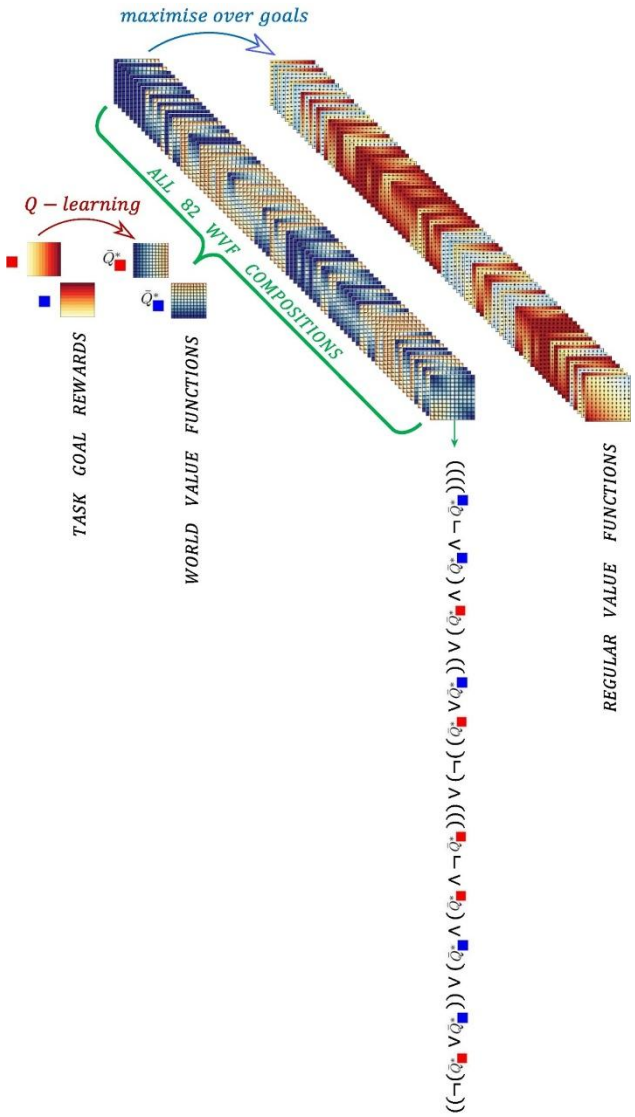
- ***OR:***
  - $Q_A \vee Q_B \coloneqq \mathbf{max}\{Q_A(s,a,g), Q_B(s,a,g)\}$
- ***AND:***
  - $\boldsymbol{Q_A \wedge Q_B} \coloneqq \mathbf{min}\{Q_A(s,a,g), Q_B(s,a,g)\}$
- ***NOT:***
  - $\neg Q_A \coloneqq \left(\boldsymbol{Q_{MAX}(s,a,g)} + \boldsymbol{Q_{MIN}(s,a,g)}\right) - Q_A(s,a,g)$



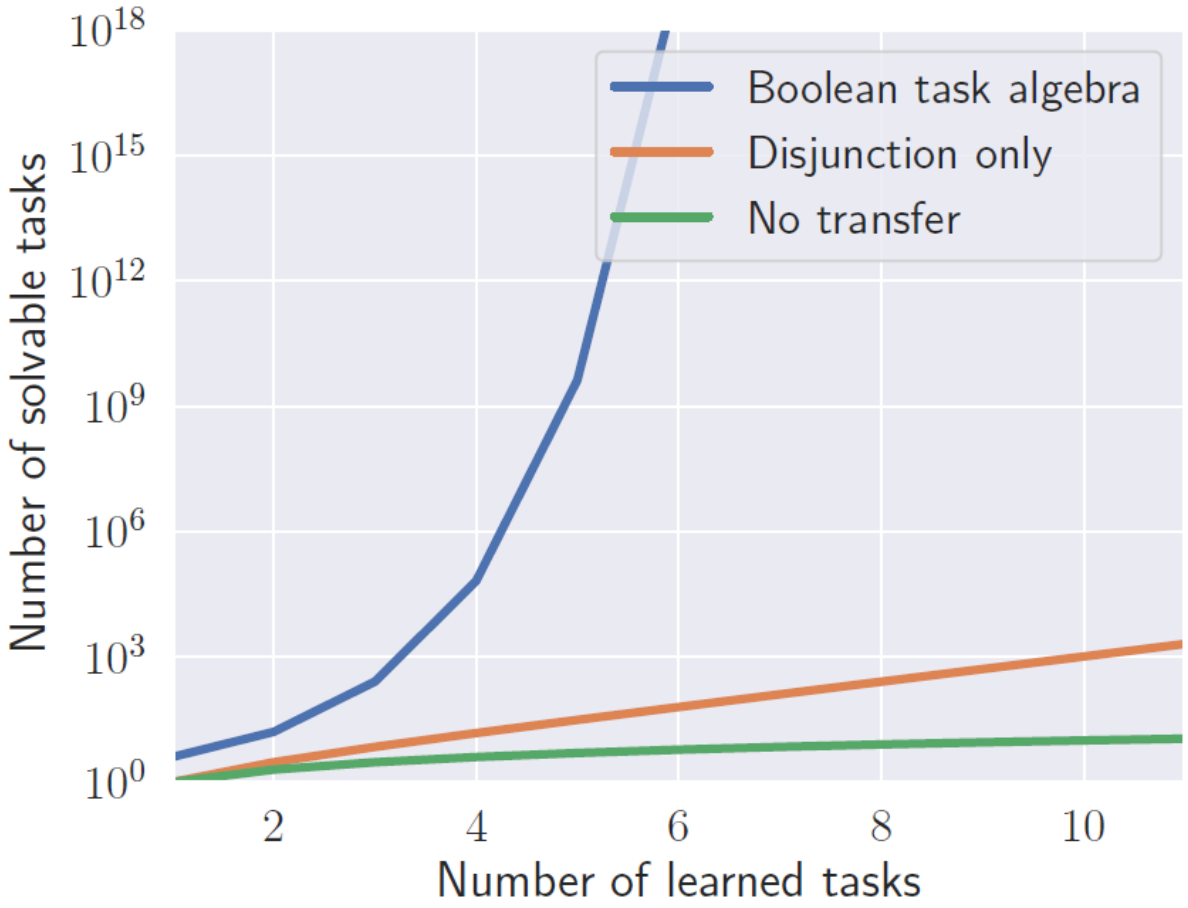
W10: Combinatorial explosion of skills

Logic Operators

- ***OR:***
  - $Q_A \vee Q_B \coloneqq \mathbf{max}\{Q_A(s,a,g), Q_B(s,a,g)\}$
- ***AND:***
  - $\boldsymbol{Q_A \wedge Q_B} \coloneqq \mathbf{min}\{Q_A(s,a,g), Q_B(s,a,g)\}$
- ***NOT:***
  - $\neg Q_A \coloneqq \left(\boldsymbol{Q_{MAX}(s,a,g)} + \boldsymbol{Q_{MIN}(s,a,g)}\right) - Q_A(s,a,g)$

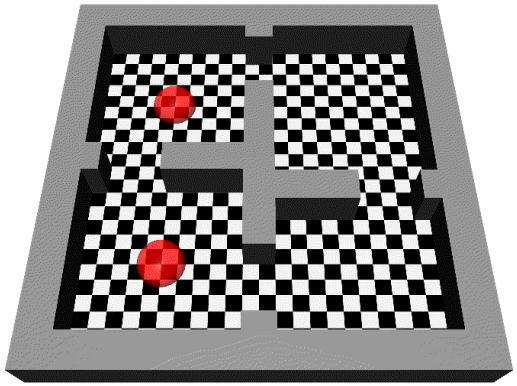


W11: **Super-Exponential** explosion of skills

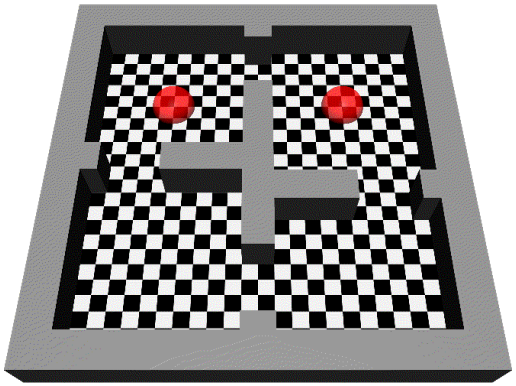


W12: Any RL Algorithm and Environment

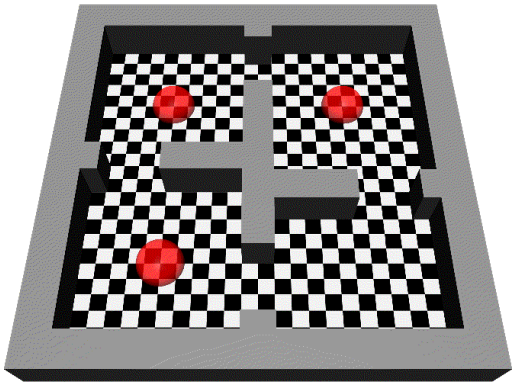
DQN/TD3/**SAC**/PPO/etc
Tabular/Discrete/**continuous**/etc



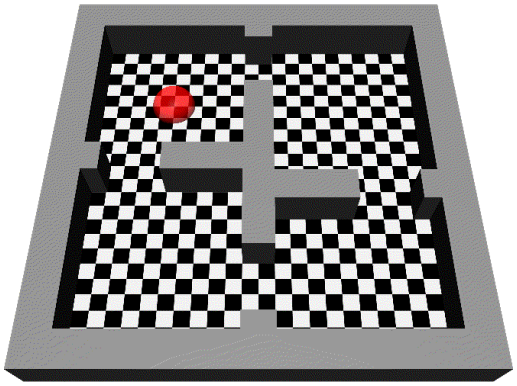
Skill: LEFT



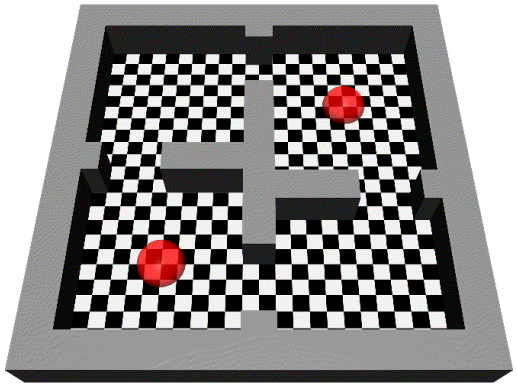
Skill: TOP



TOP **OR** LEFT



TOP **AND** LEFT



TOP **XOR** LEFT

And much more Ws!!!

Come chat at our poster