Generalisation in Lifelong RL via Logical Composition

Geraud Nangue Tasse, Steven James, Benjamin Rosman

University of the Witwatersrand

ICLR 2022

Motivation

- In general:
  - **Lifelong** agents that **reuse** past skills

Motivation

- In general:
  - **Lifelong** agents that **reuse** past skills

- In particular:
  - Are new tasks **expressible** in terms of learned ones?
    - If yes, **zero-shot learning**?
    - If no, **few-shot learning**?

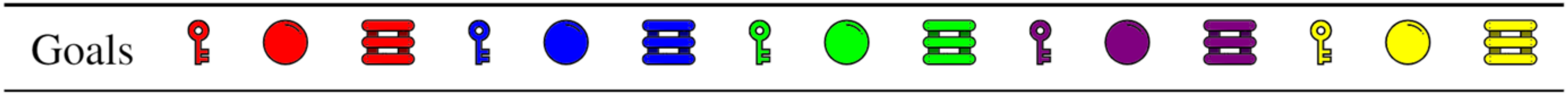
Motivation

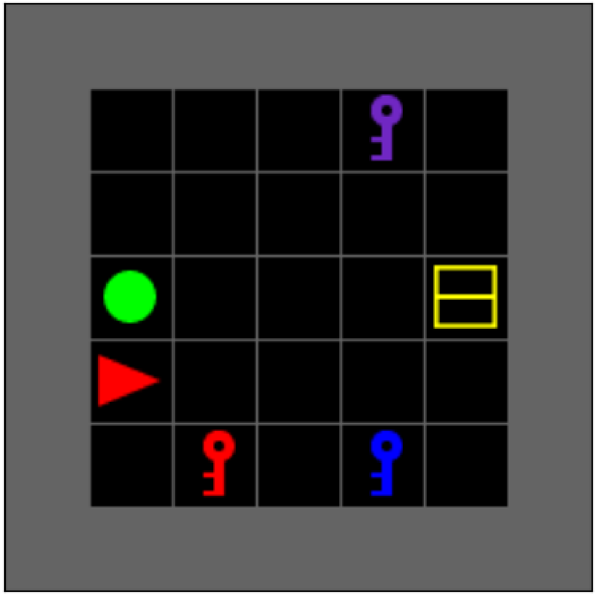
- In general:
  - **Lifelong** agents that **reuse** past skills

- In particular:
  - Are new tasks **expressible** in terms of learned ones?
    - If yes, **zero-shot learning**?
    - If no, **few-shot learning**?
  - How about lifelong **generalisation**?

Problem Setting

- Pickup-object domain





Lifelong RL with Composition

Goals	🔑	🔴	🟥	🔑	🔵	🟦	🔑	🟢	🟩	🔑	🟣	🟪	🔑	🟡	🟨
🟩	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0
🟪	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0
🟨	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1
🔑	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0

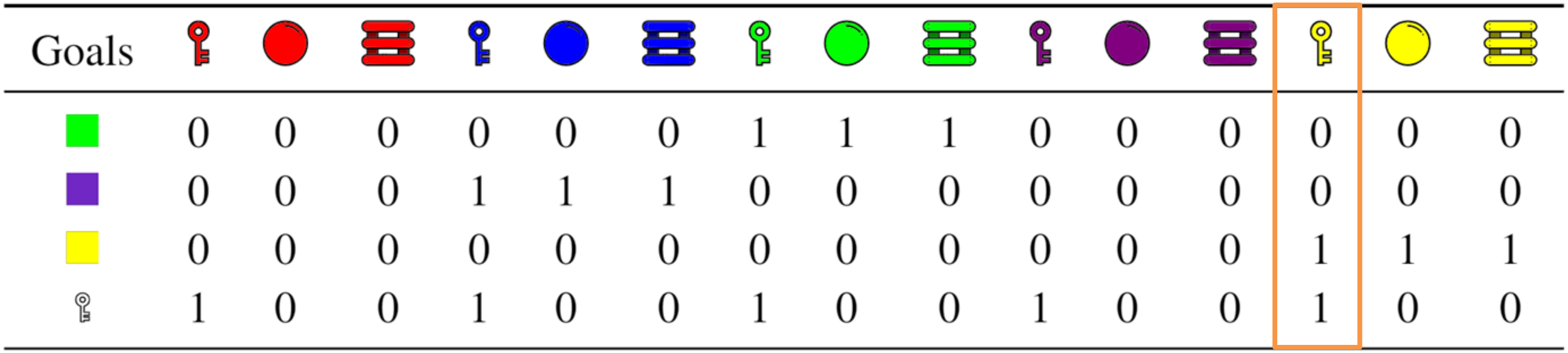
Learned

Lifelong RL with Composition

Goals	🔑	🔴	🟥	🔑	🔵	🟦	🔑	🟢	🟩	🔑	🟣	🟪	🔑	🟡	🟨
🟩	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0
🟪	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0
🟨	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1
🔑	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0

Pick up
yellow keys

Lifelong RL with Composition



Goals	🔑	🔴	🟥	🔑	🔵	🟦	🔑	🟢	🟩	🔑	🟣	🟪	🔑	🟡	🟨
🟩	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0
🟪	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0
🟨	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1
🔑	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0

SOPGOL

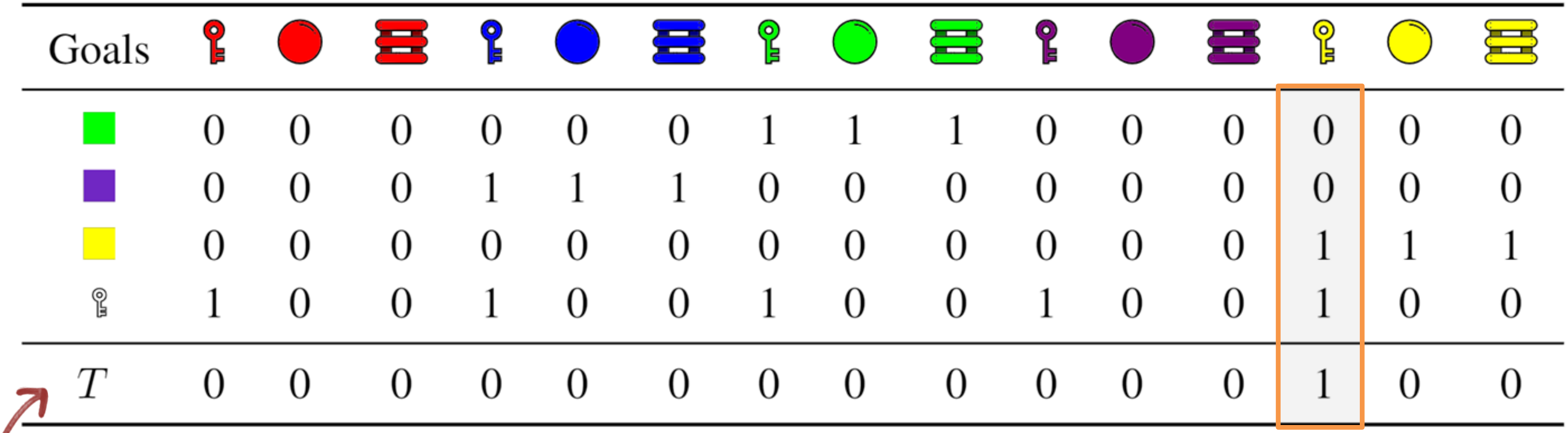
Lifelong RL with Composition

Goals	🔑	🔴	🟥	🔑	🔵	🟦	🔑	🟢	🟩	🔑	🟣	🟪	🔑	🟡	🟨
🟩	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0
🟪	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0
🟨	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1
🔑	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0
T	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0

Learned

SOPGOL

Lifelong RL with Composition



Goals	🔑	🔴	🟥	🔑	🔵	🟦	🔑	🟢	🟩	🔑	🟣	🟪	🔑	🟡	🟨
🟩	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0
🟪	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0
🟨	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1
🔑	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0
T	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0

Learned

SOPGOL

- $T' :=$ ¬🟩 $\wedge$ ¬🟪 $\wedge$ 🟨 $\wedge$ 🔑

Not green and not blue and yellow and key

Lifelong RL with Composition

Goals	🔑	🔴	🟥	🔑	🔵	🟦	🔑	🟢	🟩	🔑	🟣	🟪	🔑	🟡	🟨
🟩	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0
🟪	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0
🟨	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1
🔑	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0
T	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0

Learned

SOPGOL

Not green and not blue and yellow and key

- ➢ $T' :=$ ¬🟩 ∧ ¬🟪 ∧ 🟨 ∧ 🔑
- ➢ $Q_{SOP} = \neg Q^*($🟩$) \wedge \neg Q^*($🟪$) \wedge Q^*($🟨$) \wedge Q^*($🔑$)$

Lifelong RL with Composition

Goals	🔑	🔴	🟥	🔑	🔵	🟦	🔑	🟢	🟩	🔑	🟣	🟪	🔑	🟡	🟨
🟩	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0
🟪	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0
🟨	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1
🔑	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0
T	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0

Learned

SOPGOL

Not green and not blue and yellow and key

- $T' \coloneqq$ $\neg$🟩 $\wedge$ $\neg$🟪 $\wedge$ 🟨 $\wedge$ 🔑
- $Q_{SOP} = \neg Q^{*}($🟩$) \wedge \neg Q^{*}($🟪$) \wedge Q^{*}($🟨$) \wedge Q^{*}($🔑$)$
- $T = T'$? **Yes!**

Lifelong RL with Composition

Goals	🔑	🔴	🟥	🔑	🔵	🟦	🔑	🟢	🟩	🔑	🟣	🟪	🔑	🟡	🟨
🟩	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0
🟪	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0
🟨	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1
🔑	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0
T	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0

Learned

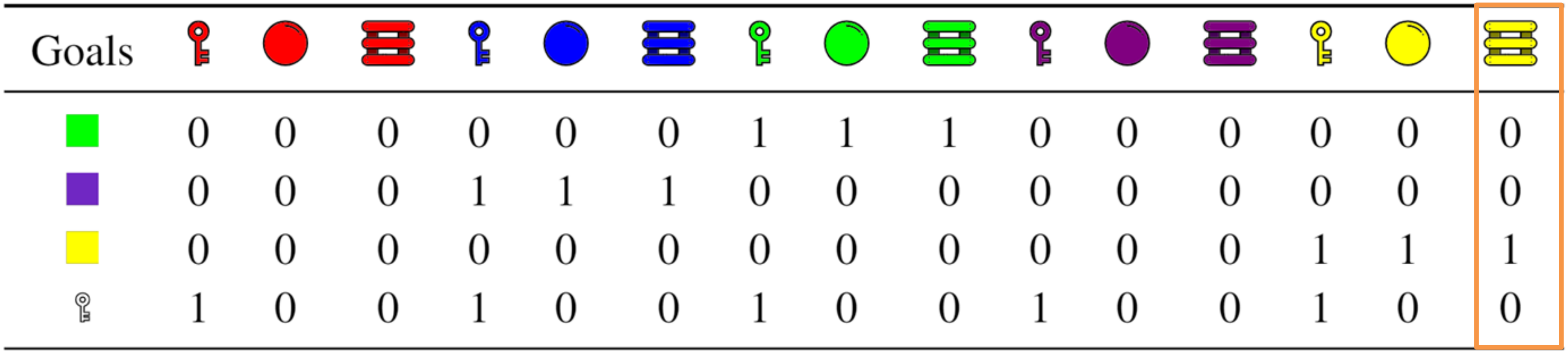
SOPGOL

Not green and not blue and yellow and key

- $T' :=$ ¬🟩 ∧ ¬🟪 ∧ 🟨 ∧ 🔑
- $Q_{SOP} = \neg Q^*($🟩$) \wedge \neg Q^*($🟪$) \wedge Q^*($🟨$) \wedge Q^*($🔑$)$
- $T = T'$? **Yes!**

Reduces the RL problem to just Supervised learning

Lifelong RL with Composition



Goals	🔑	🔴	▤	🔑	🔵	▤	🔑	🟢	▤	🔑	🟣	▤	🔑	🟡	▤
🟩	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0
🟪	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0
🟨	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1
🔑	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0

Pick up
Yellow Boxes

Lifelong RL with Composition

Goals	🔑	🔴	🟥	🔑	🔵	🟦	🔑	🟢	🟩	🔑	🟣	🟪	🔑	🟡	🟨
🟩	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0
🟪	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0
🟨	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1
🔑	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0
T	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1

Learned

SOPGOL

Lifelong RL with Composition

Goals	🔑	🔴	🟥	🔑	🔵	🟦	🔑	🟢	🟩	🔑	🟣	🟪	🔑	🟡	🟨
🟩	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0
🟪	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0
🟨	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1
🔑	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0
T	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1

Learned

SOPGOL

Not green and not blue and yellow and not key

- ➢ $T' := \neg$🟩 $\wedge \neg$🟪 $\wedge$ 🟨 $\wedge \neg$🔑
- ➢ $Q_{SOP} = \neg Q^*($🟩$) \wedge \neg Q^*($🟪$) \wedge Q^*($🟨$) \wedge \neg Q^*($🔑$)$

Lifelong RL with Composition

Goals	🔑	🔴	▤	🔑	🔵	▤	🔑	🟢	▤	🔑	🟣	▤	🔑	🟡	▤
🟩	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0
🟪	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0
🟨	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1
🔑	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0
T	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1

Learned

SOPGOL

Not green and not blue and yellow and not key

- $T' :=$ ¬🟩 ∧ ¬🟪 ∧ 🟨 ∧ ¬🔑
- $Q_{SOP} =$ ¬Q^*(🟩) ∧ ¬Q^*(🟪) ∧ Q^*(🟨) ∧ ¬Q^*(🔑)
- $T = T'$? **No!**

Lifelong RL with Composition

Goals	🔑	🔴	🟥	🔑	🔵	🟦	🔑	🟢	🟩	🔑	🟣	🟪	🔑	🟡	🟨
🟩	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0
🟪	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0
🟨	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1
🔑	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0
T	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1

Learned

SOPGOL

Not green and not blue and yellow and not key

- $T' \coloneqq$ ¬🟩 ∧ ¬🟪 ∧ 🟨 ∧ ¬🔑
- $Q_{SOP} = \neg Q^*($🟩$) \wedge \neg Q^*($🟪$) \wedge Q^*($🟨$) \wedge \neg Q^*($🔑$)$
- $T = T'$? **No!**
- Learn new Q with goal-oriented learning **(using Q_{SOP}, to speed up training)**, then add to library

Transfer After Pretraining

- Pretrained: 🟩, 🟪, 🟨, 🔑

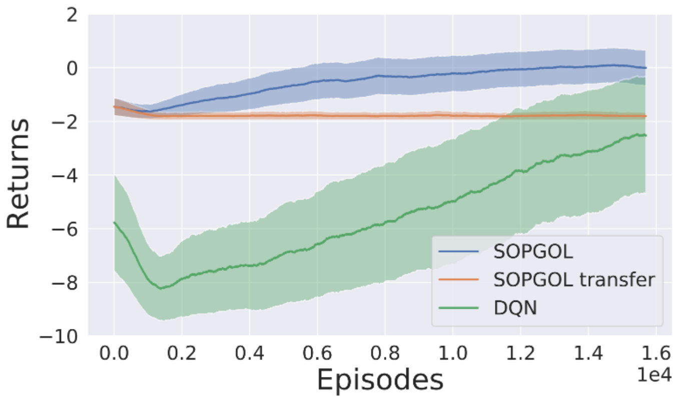
Goals	🔑	🔴	🟥	🔑	🔵	🟦	🔑	🟢	🟩	🔑	🟣	🟪	🔑	🟡	🟨
T_2	0	0	1	0	1	1	1	0	1	1	0	0	1	0	0

Pick up red/blue/green boxes, or
green/purple/yellow keys, or
blue balls

Transfer After Pretraining

- Pretrained: 🟩, 🟪, 🟨, 🔑

Goals	🔑	🔴	▤	🔑	🔵	▤	🔑	🟢	▤	🔑	🟣	▤	🔑	🟡	▤
T_2	0	0	1	0	1	1	1	0	1	1	0	0	1	0	0

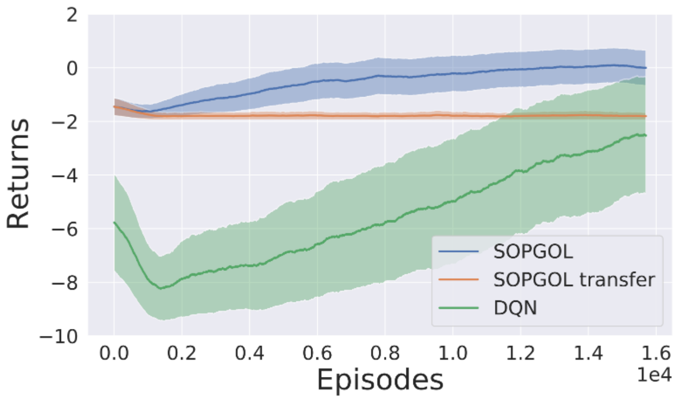


SOPGOL > DQN

Transfer After Pretraining

- Pretrained: 🟩, 🟪, 🟨, 🔑

Goals	🔑	🔴	🟥	🔑	🔵	🟦	🔑	🟢	🟩	🔑	🟣	🟪	🔑	🟡	🟨
T_2	0	0	1	0	1	1	1	0	1	1	0	0	1	0	0



SOPGOL > DQN

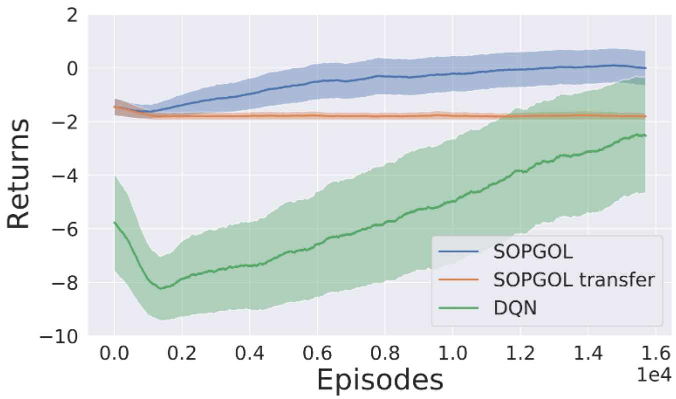
Inferred behaviour

(🟩 ∧ ¬🟪 ∧ ¬🟨) ∨ (¬🟩 ∧ 🟪 ∧ ¬🟨 ∧ 🔑) ∨ (¬🟩 ∧ ¬🟪 ∧ 🟨 ∧ 🔑) ∨ (¬🟪 ∧ ¬🟨 ∧ ¬🔑)

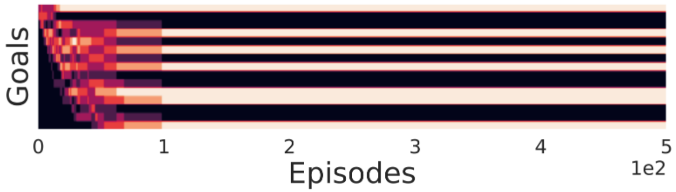
Transfer After Pretraining

- Pretrained: 🟩, 🟪, 🟨, 🔑

Goals	🔑	🔴	▤	🔑	🔵	▤	🔑	🟢	▤	🔑	🟣	▤	🔑	🟡	▤
T_2	0	0	1	0	1	1	1	0	1	1	0	0	1	0	0



SOPGOL > DQN



Learned representation of new task

Inferred behaviour { (🟩 ∧ ¬🟪 ∧ ¬🟨) ∨ (¬🟩 ∧ 🟪 ∧ ¬🟨 ∧ 🔑) ∨ (¬🟩 ∧ ¬🟪 ∧ 🟨 ∧ 🔑) ∨ (¬🟪 ∧ ¬🟨 ∧ ¬🔑)

Lifelong RL with Composition

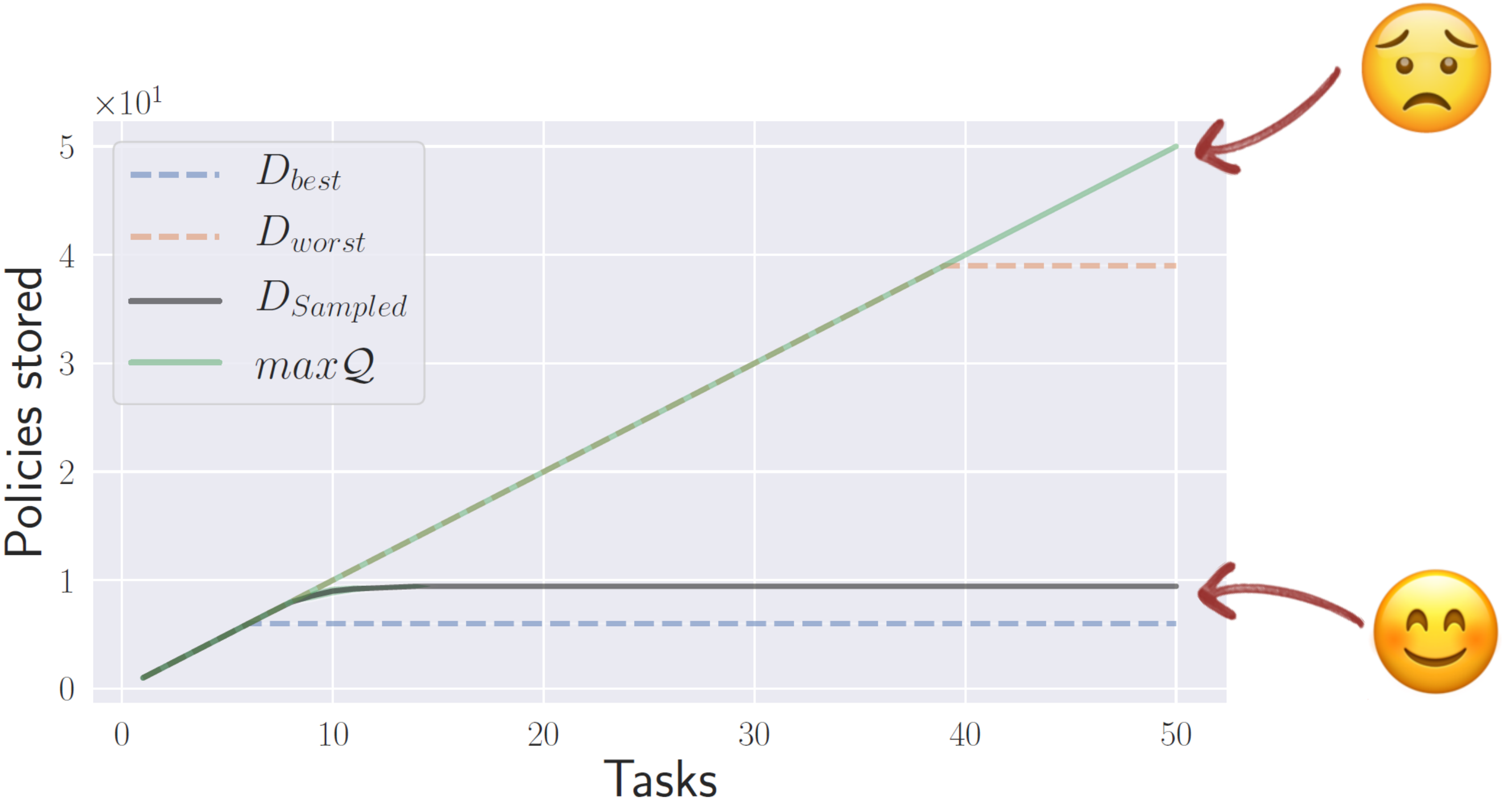
$$SOPGOL$$

Theorem: $\log(|goals|) \leq \lim_{t \to \infty}(|\boldsymbol{skills}|) \leq |goals|$

Note: $|\text{tasks}| = 2^{|goals|}$

Generalisation with Lifelong Transfer

- Four-rooms domain: 40 goals, 2^{40} ~ 1 trillion tasks



Conclusion

- We leverage logical composition for **fast transfer** between tasks.

- Leads to **quick generalisation** over any task distribution.
  - we have a **logarithmic upper bound** on the number of tasks that needs to be learned and stored.

- Leads to both **interpretable** and **sample-efficient** lifelong RL.