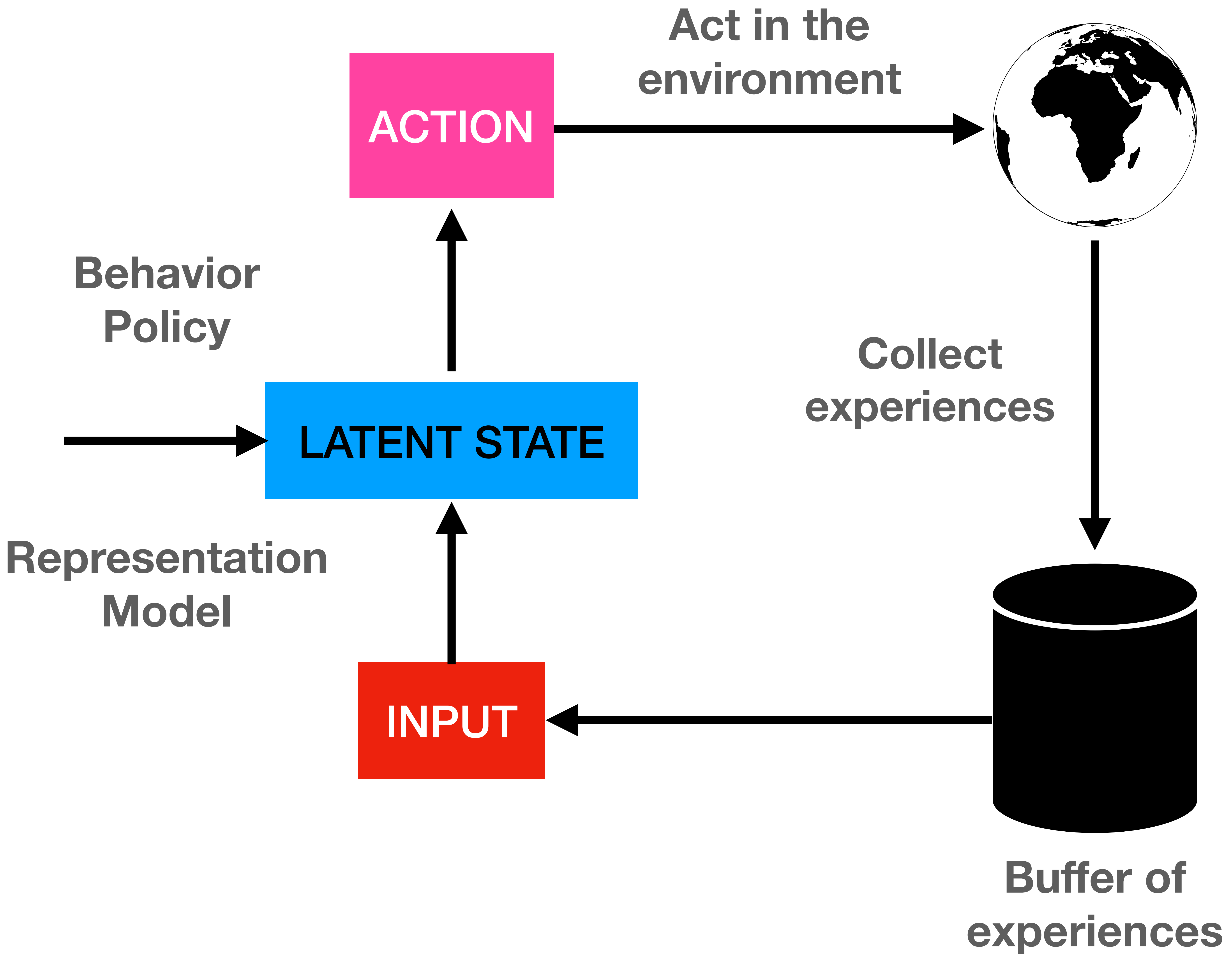
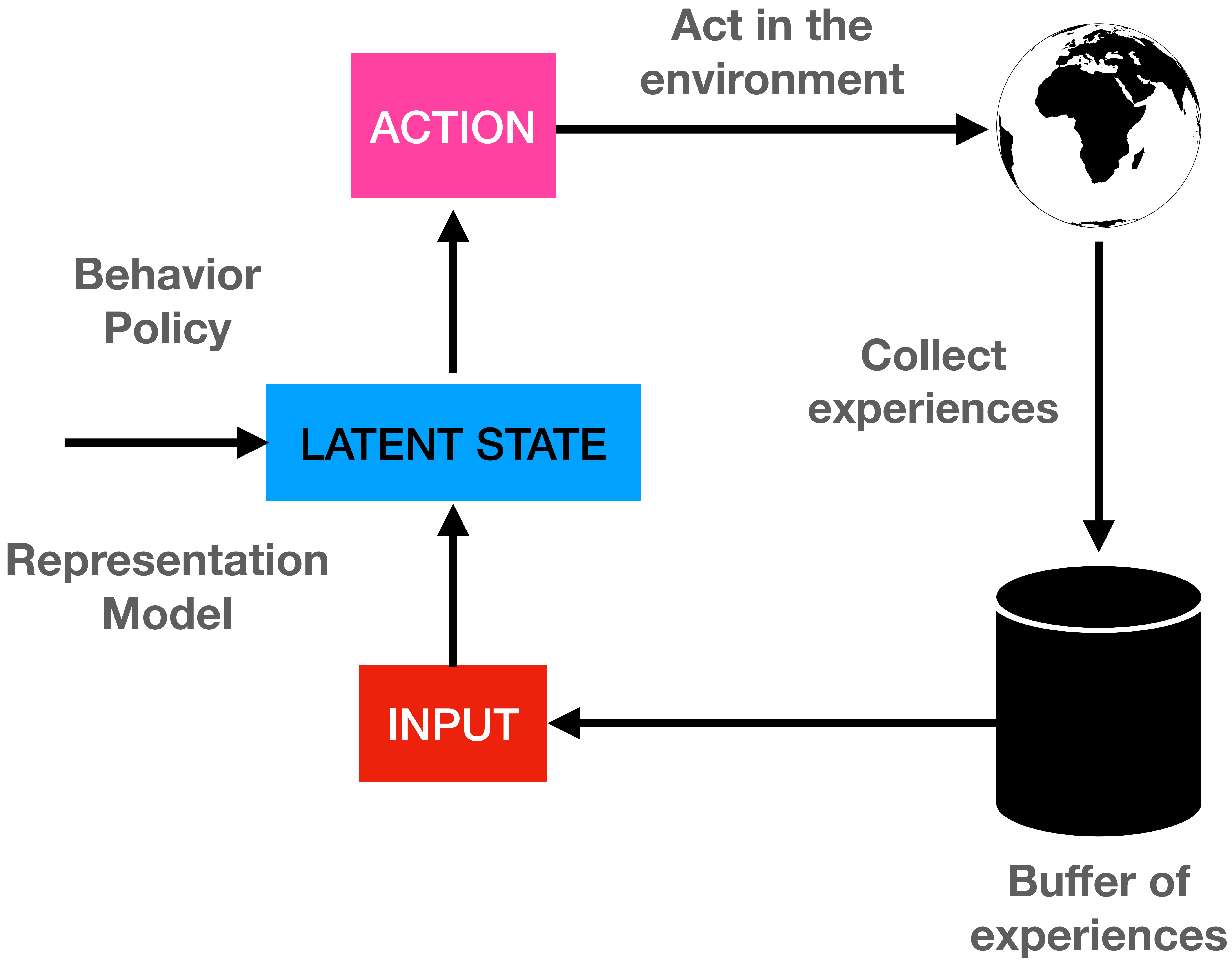
Retrieval Augmented RL: Amortized Inference over Entire Dataset (ICML'22)

Abram L. Friesen*, Andrea Banino*, Theophane Weber*, Nan Rosemary Ke*, Adria Puigdomenech Badia, Arthur Guez, Mehdi Mirza, Peter C. Humphreys, Ksenia Konyushkova, Laurent Sifre, Michal Valko, Simon Osindero, Timothy Lillicrap, Nicolas Heess, Charles Blundell

Perception-Action Loop



Perception-Action Loop

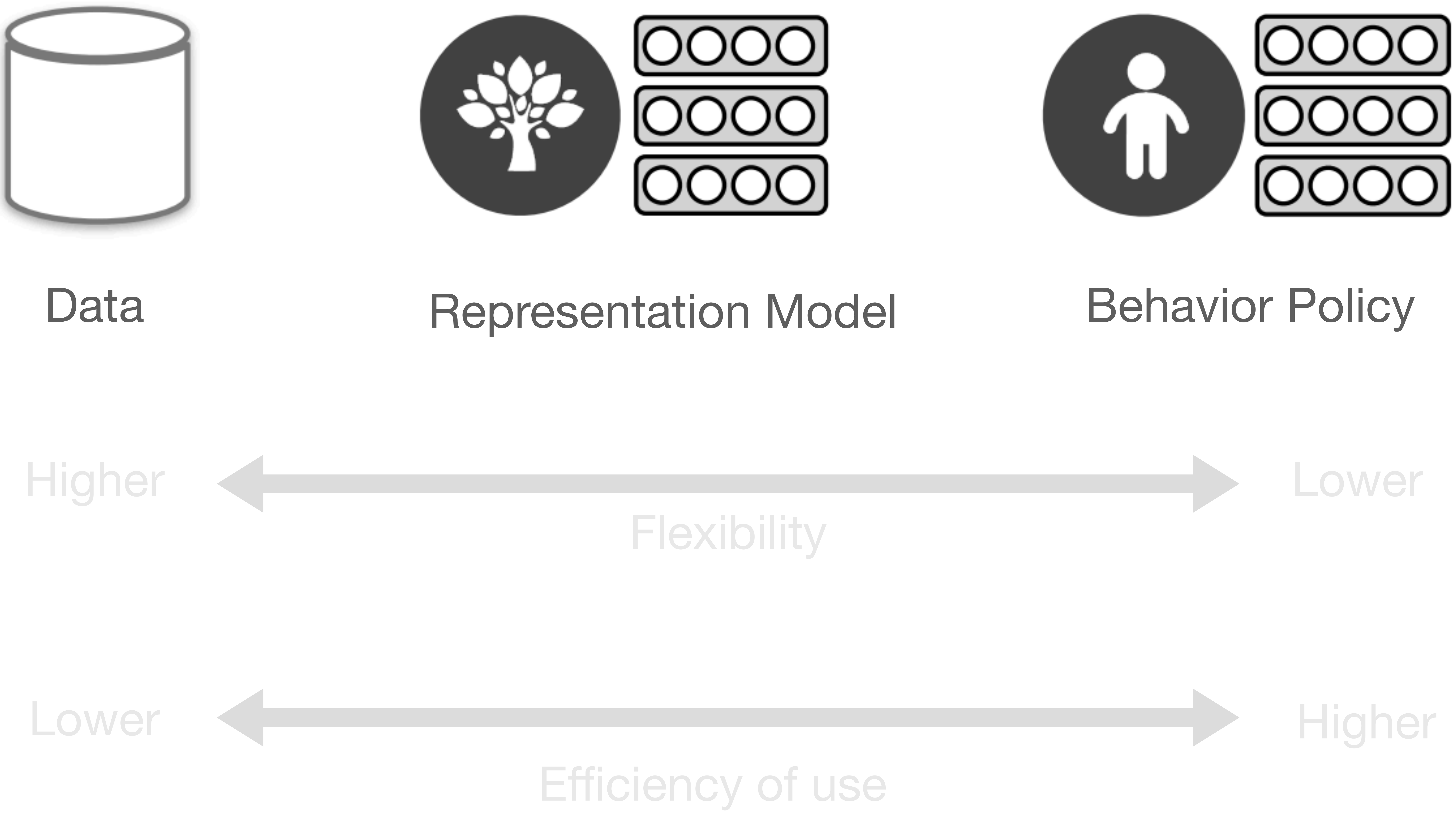


Two Limitations

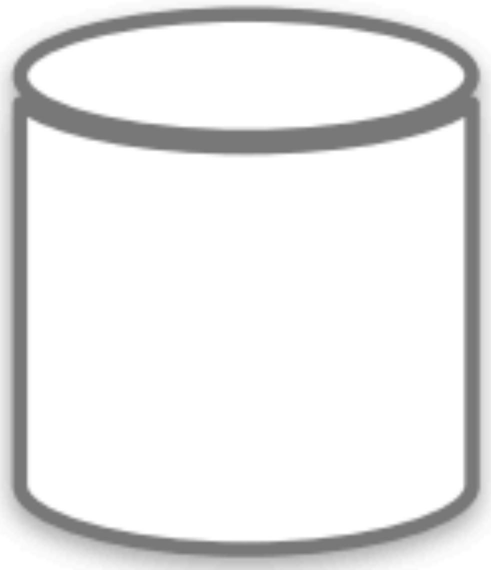
Agent's past experiences no longer play a direct role in the agent's behavior.

Not exploit specific guidance that a handful of past experiences may provide

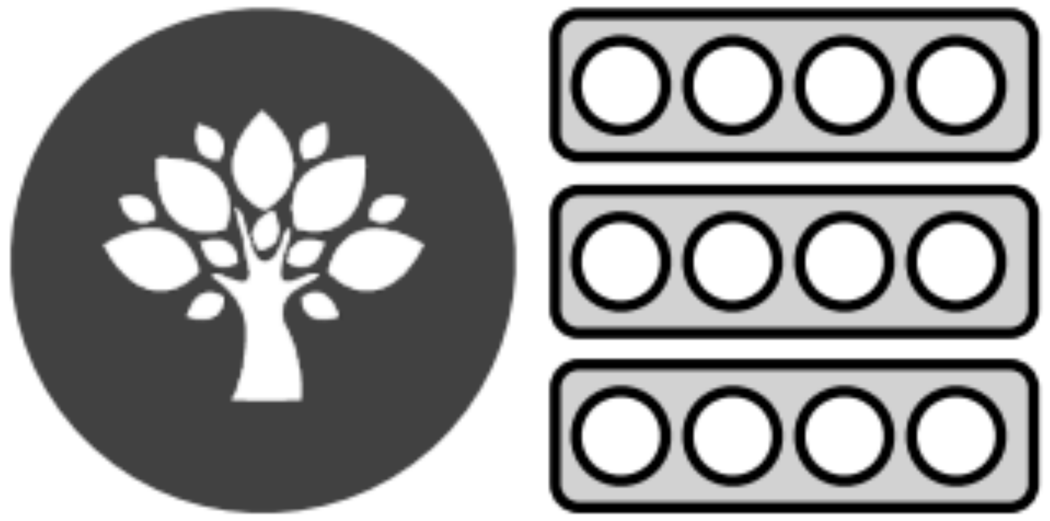
Transfer of Knowledge Across Tasks



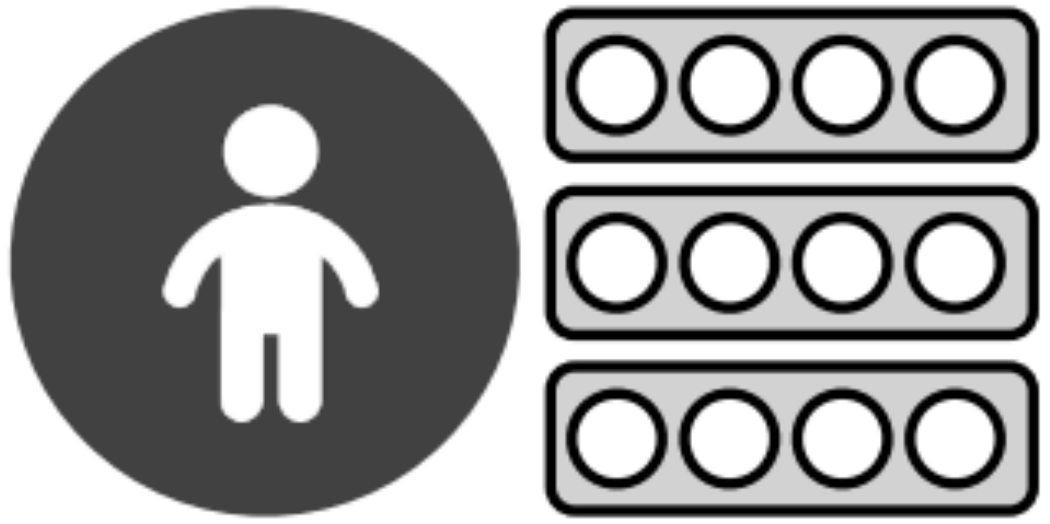
Transfer of Knowledge Across Tasks



Data



Representation Model



Behavior Policy

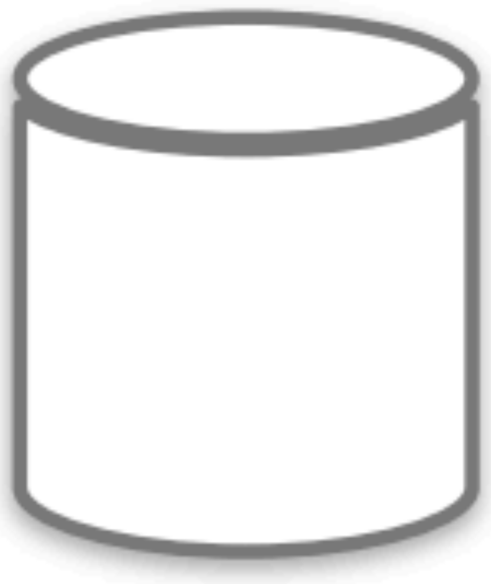
Higher ⟷ Lower

Flexibility

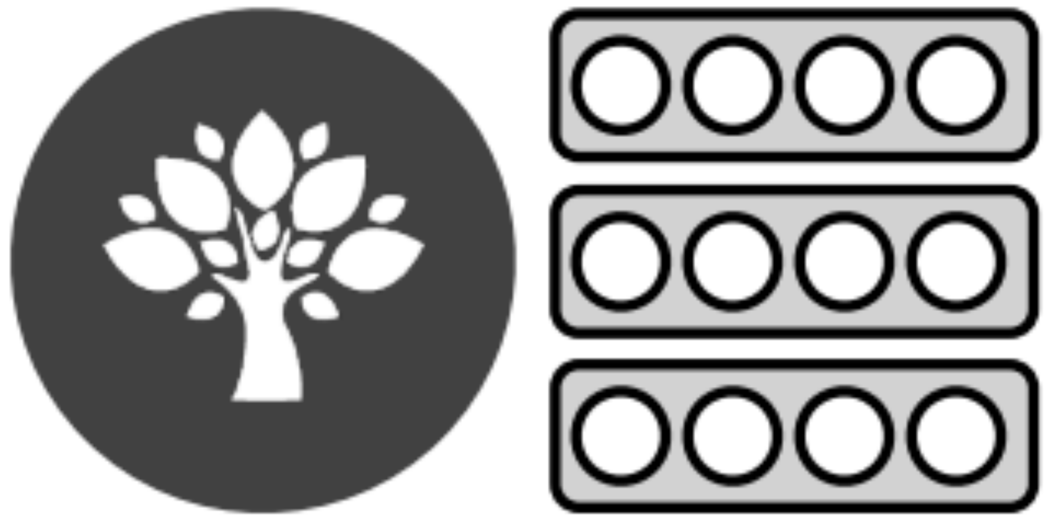
Lower ⟷ Higher

Efficiency of use

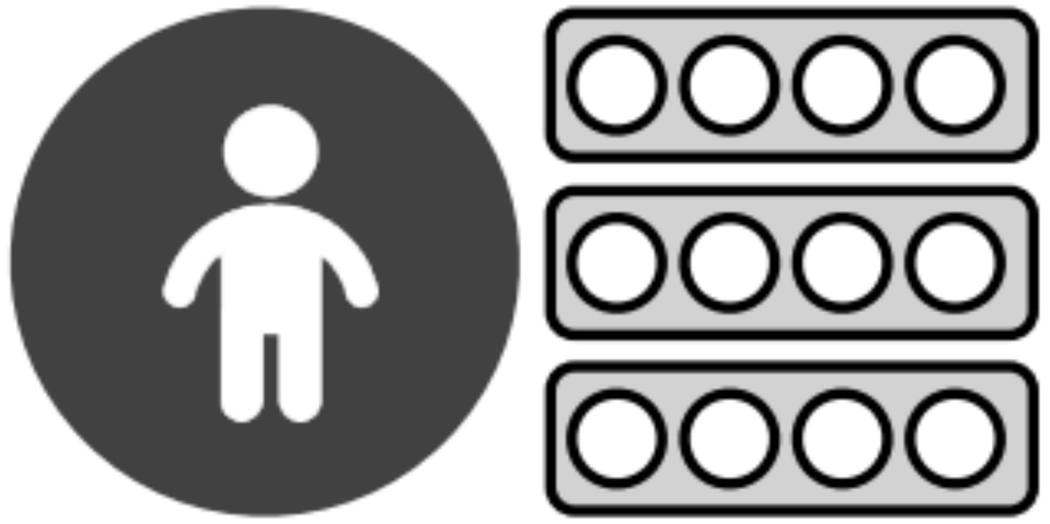
Transfer of Knowledge Across Tasks



Data



Representation Model



Behavior Policy

Higher ⟷ Lower

Flexibility

Lower ⟷ Higher

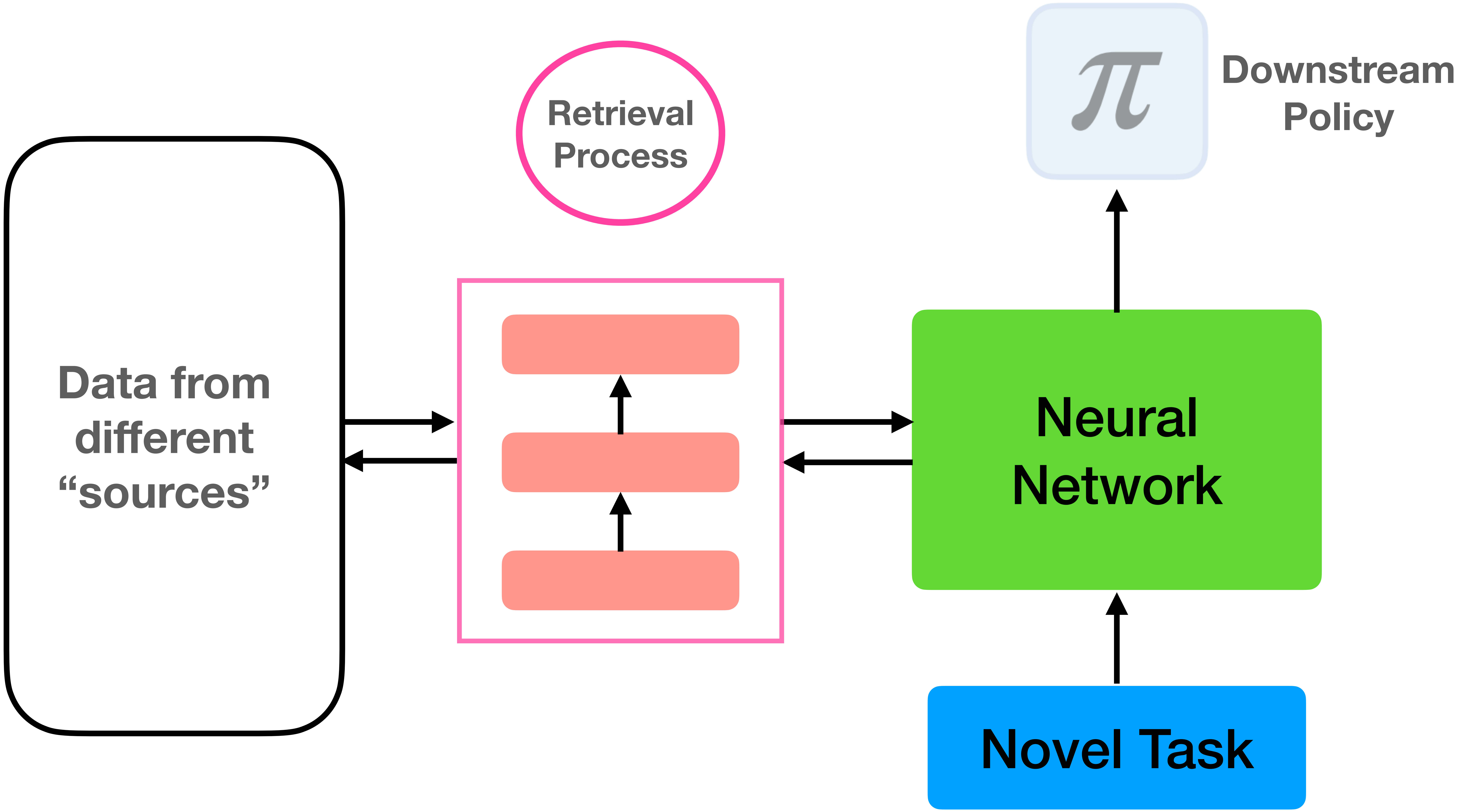
Efficiency of use

Different Sources of Data for Training RL Algorithms

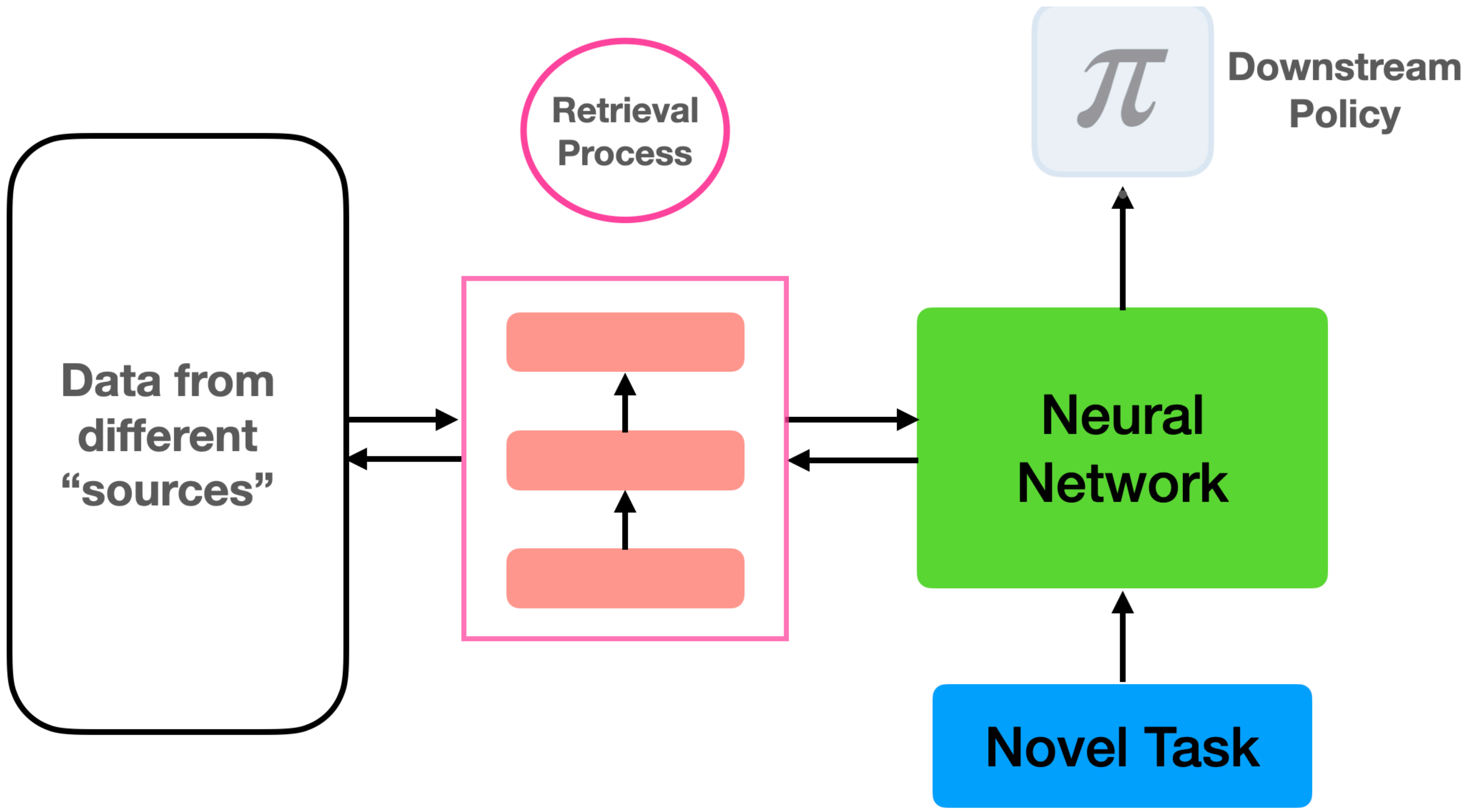
We want to create new RL algorithms that:

1. Harness data from a ***variety of sources***:
   - Past experiences of the agent,
   - Past experiences of the other agents.
   - Even imagined rollouts.
2. ***Predict*** & ***plan*** with a jumpy models that ***stitches together information*** across all of these experience types.

Retrieval Augmented Reinforcement Learning (RARL)



Retrieval Augmented Reinforcement Learning (RARL)



- Learned function encodes the data in the replay dataset.
- Retrieval process queries for data relevant to the agent in it's current context.
- Agent process uses the retrieved information to shape the value function.
- At test time, agent can ``generalize" to novel behaviors.

Retrieval Process and Agent Process

Dataset of experiences ($\mathscr{B}$)
- Contains raw data in the form of trajectories.

Retrieval Process
- Parameterized as a neural network.
- Retrieves information from dataset.

Agent Process
- Parameterized as a neural network.

Contextual Information from the agent process.

Updated state of retrieval process

$$f^{\text{retr}}_{\theta,\mathscr{B}} : m_{t-1}, s_t \mapsto m_t, u_t$$

state of retrieval process

Information retrieved.

Information provided by retrieval process.

$$f^{\text{agent}}_{\theta} : s_t, u_t \mapsto \underbrace{\pi_\theta(s_t, u_t, a)}_{\text{Executes action in the real world.}}$$

state of agent process

Retrieval Process and Agent Process

Dataset of experiences ($\mathscr{B}$)
- Contains raw data in the form of trajectories.

Contextual Information from the agent process.

Updated state of retrieval process

$$f^{\text{retr}}_{\theta,\mathscr{B}} : m_{t-1}, s_t \mapsto m_t, u_t$$

state of retrieval process

Information retrieved.

Retrieval Process
- Parameterized as a neural network.
- Retrieves information from dataset.

Information provided by retrieval process.

Agent Process
- Parameterized as a neural network.

$$f^{\text{agent}}_{\theta} : s_t, u_t \mapsto \underbrace{\pi_\theta(s_t, u_t, a)}_{\text{Executes action in the real world.}}$$

state of agent process

Retrieval Process and Agent Process

Dataset of experiences ($\mathscr{B}$)

- Contains raw data in the form of trajectories.

Contextual Information from the agent process.

Updated state of retrieval process

$$f^{\text{retr}}_{\theta,\mathscr{B}} : m_{t-1}, s_t \mapsto m_t, u_t$$

state of retrieval process

Information retrieved.

Retrieval Process

- Parameterized as a neural network.
- Retrieves information from dataset.

Information provided by retrieval process.

Agent Process

- Parameterized as a neural network.

$$f^{\text{agent}}_{\theta} : s_t, u_t \mapsto \underbrace{\pi_\theta(s_t, u_t, a)}_{\text{Executes action in the real world.}}$$

state of agent process

Experiments

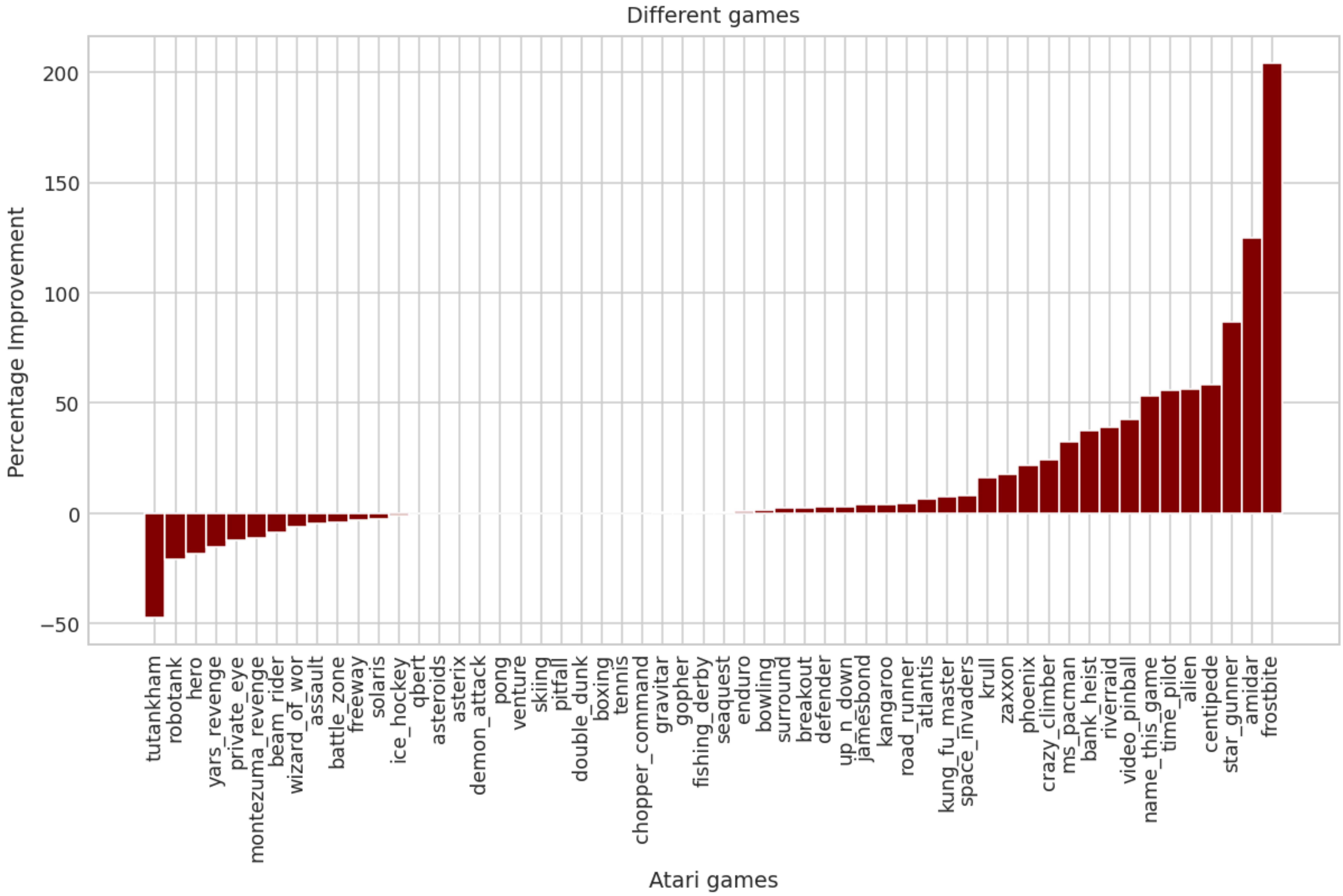
- Past experiences in the dataset [off policy RL/on-policy RL]

- (Multi-task) Data from other policies [offline RL]

- Online RL but access to (offline) dataset

- Test time generalization to new “dataset”

Atari Single Task: Retrieval Augmented Reinforcement Learning

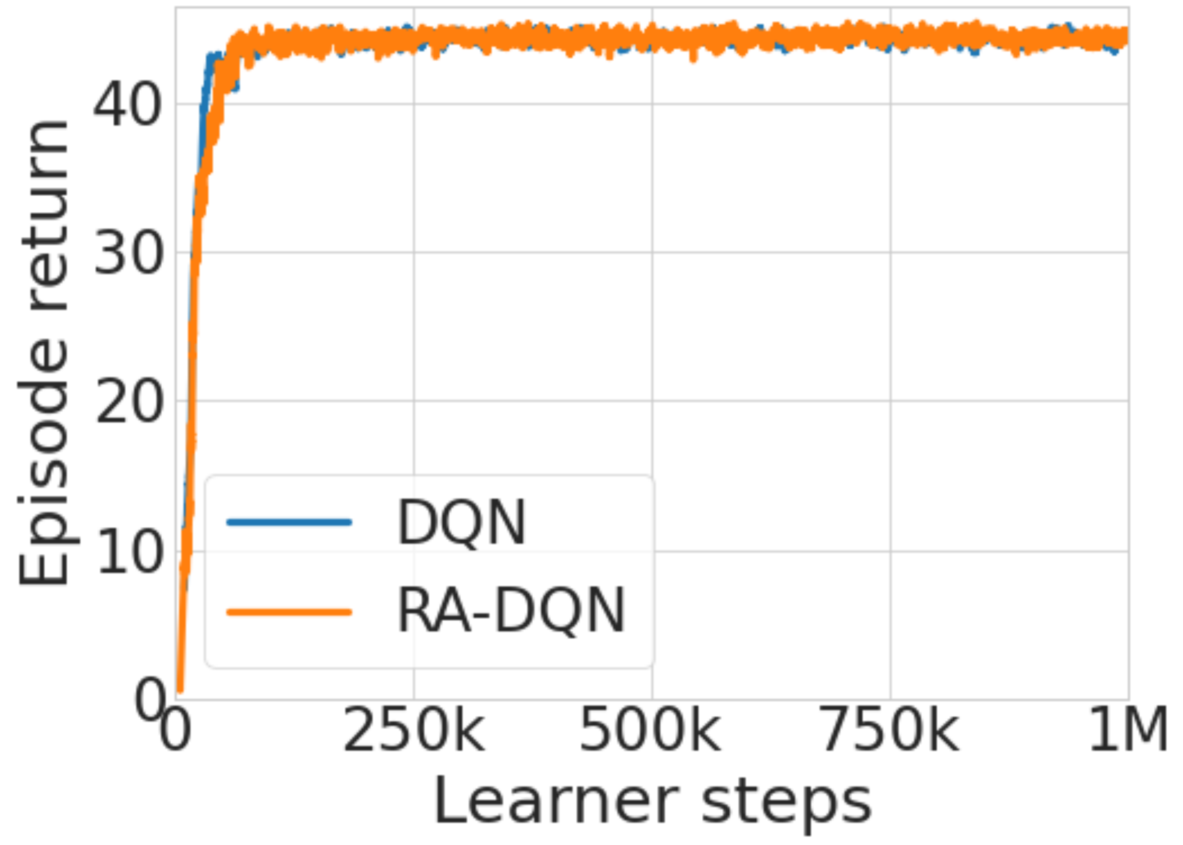


Atari: Relative percentage improvement in mean human normalized score of RA-R2D2 vs vanilla R2D2.

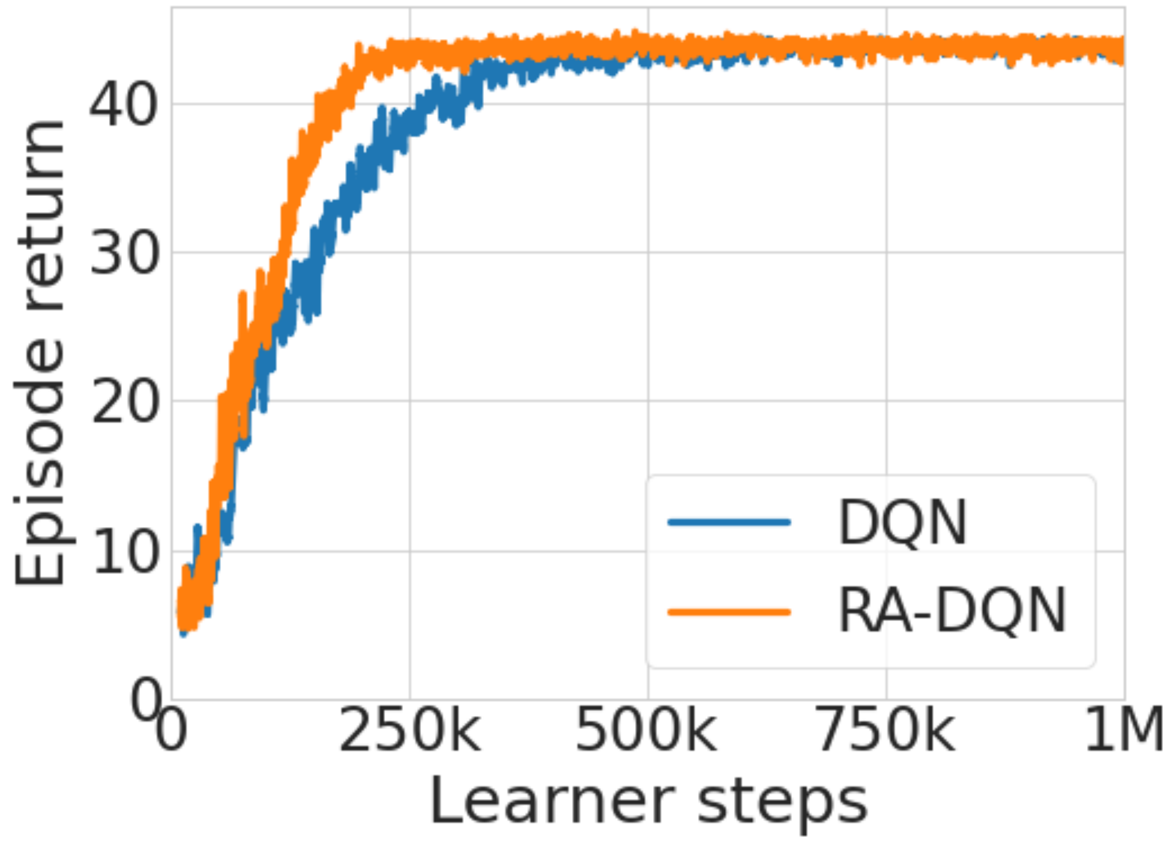
Multi-tasking Offline RL: Retrieval Augmented Reinforcement Learning

Increasing number of tasks (10, 20 and 30 tasks)

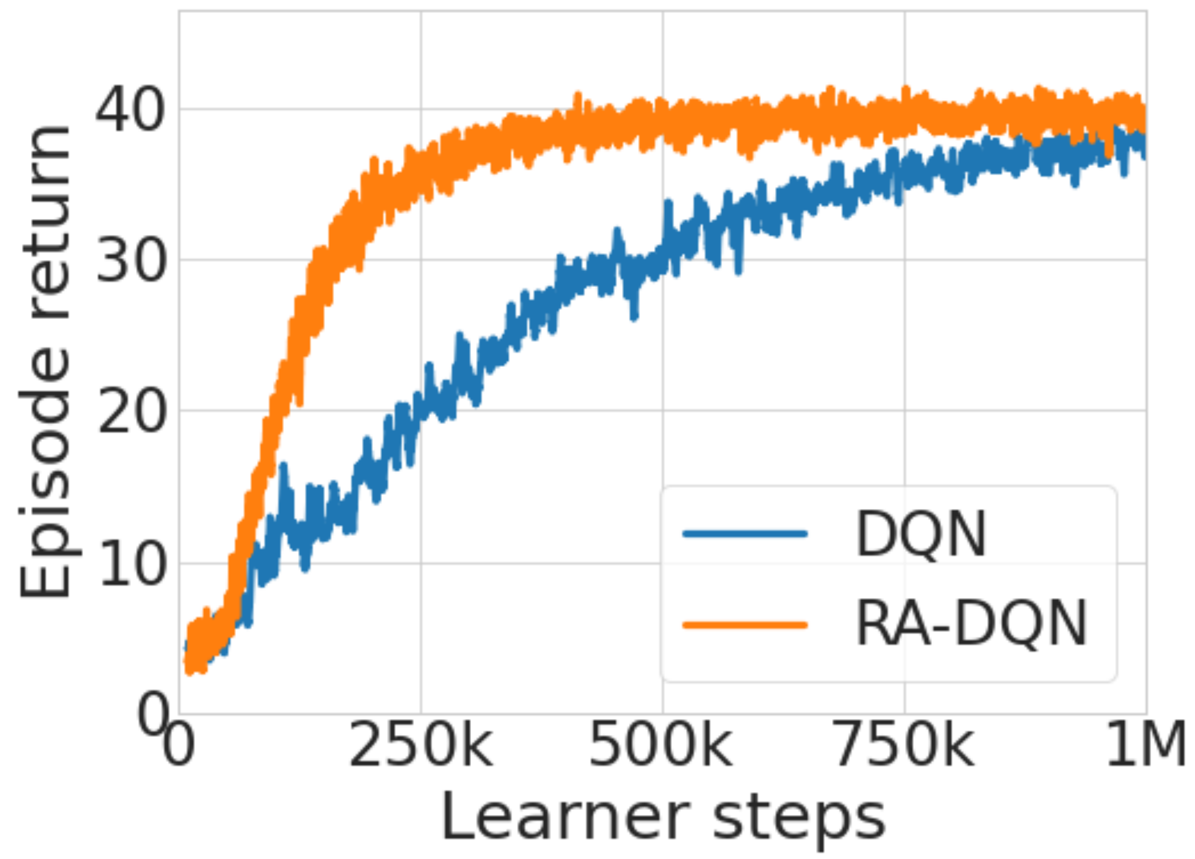
DQN (blue)
RA-DQN (Proposed, Orange)



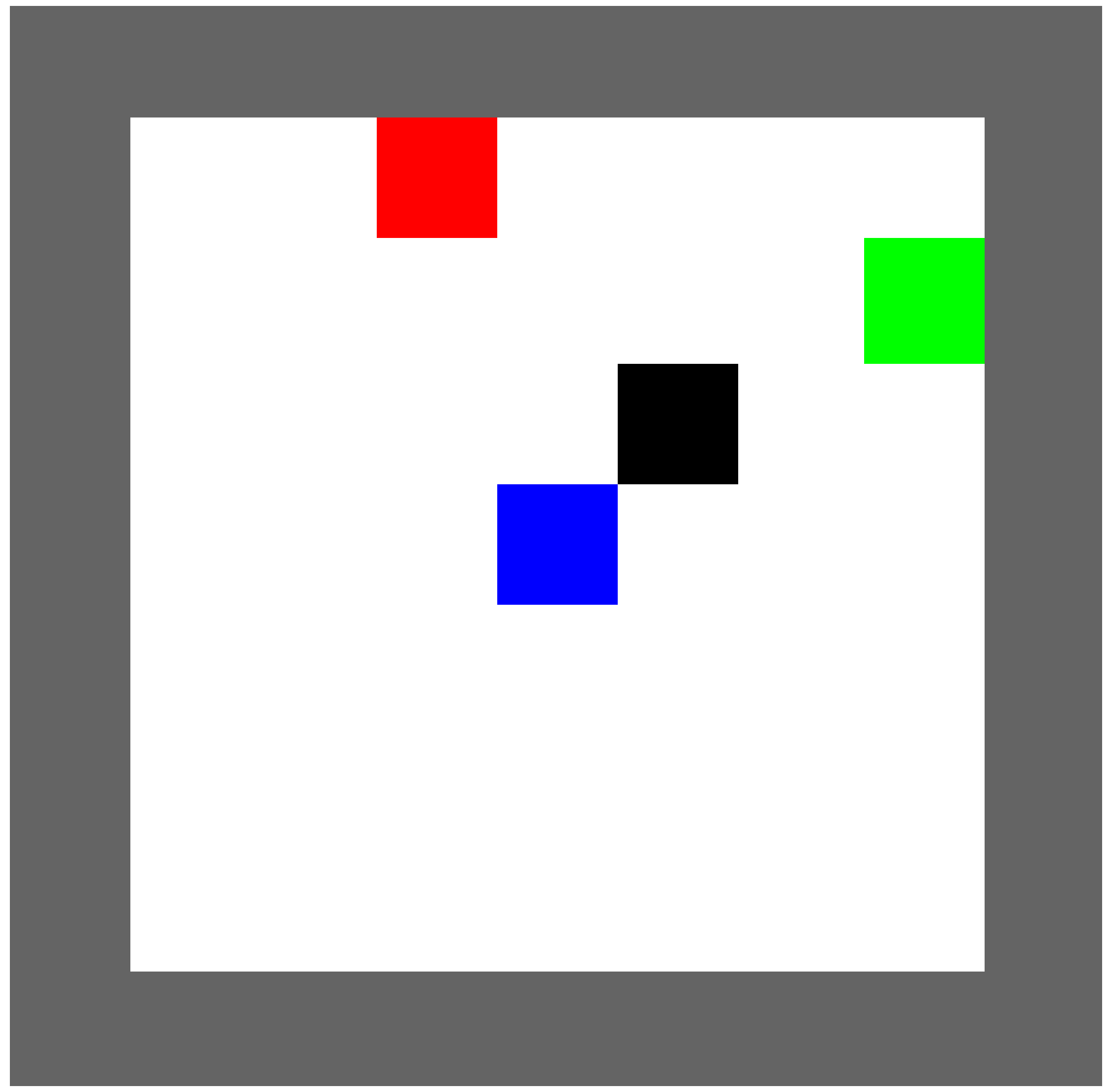
10 training tasks



20 training tasks



30 training tasks



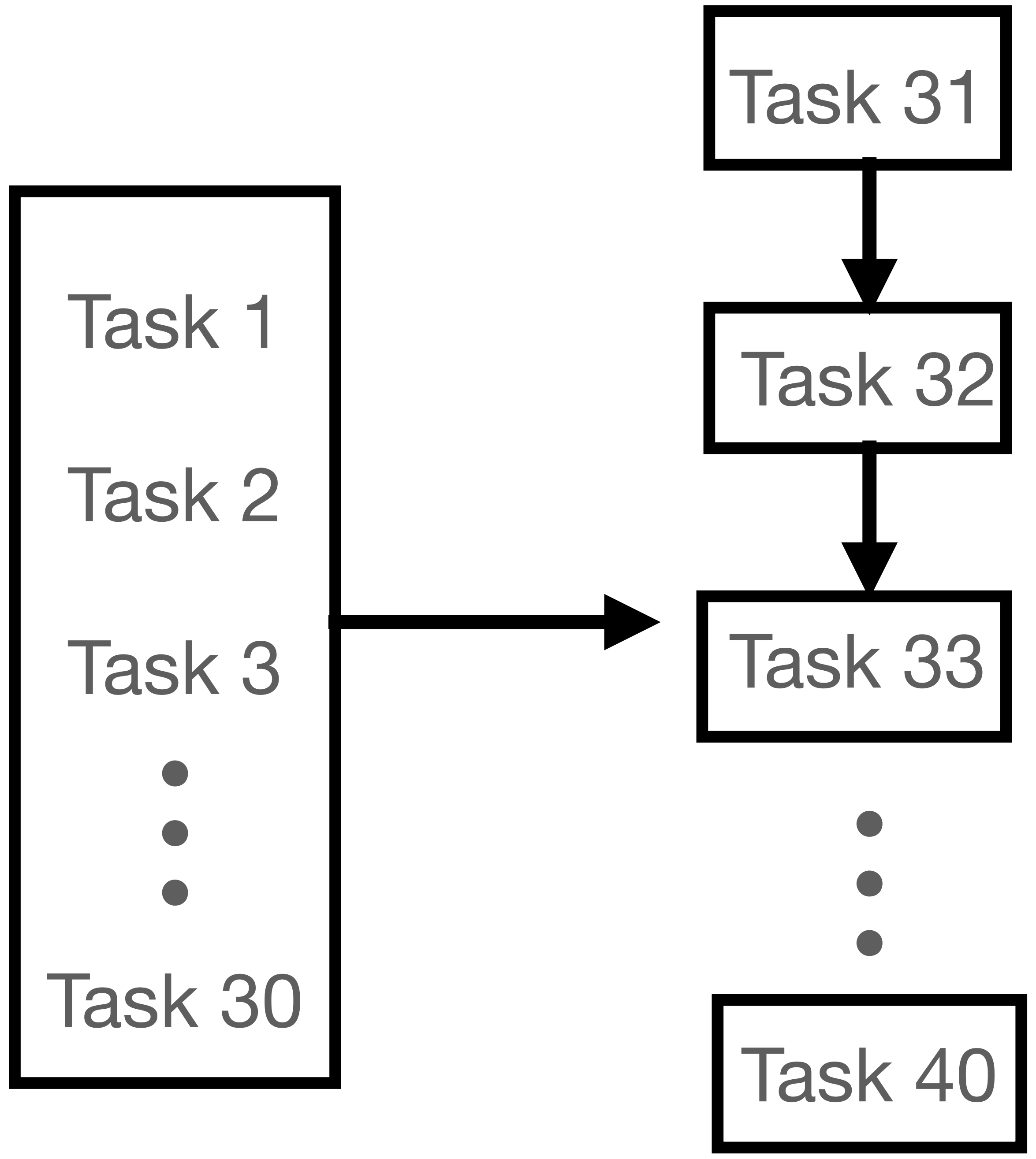
GridRoboman Env:

Task: On the board there are three colored objects and the robot is represented by a black block. The robot can move itself and move the objects.

On increasing number of tasks, the retrieval-augmented agent learns much more effectively than the baseline DQN agent.

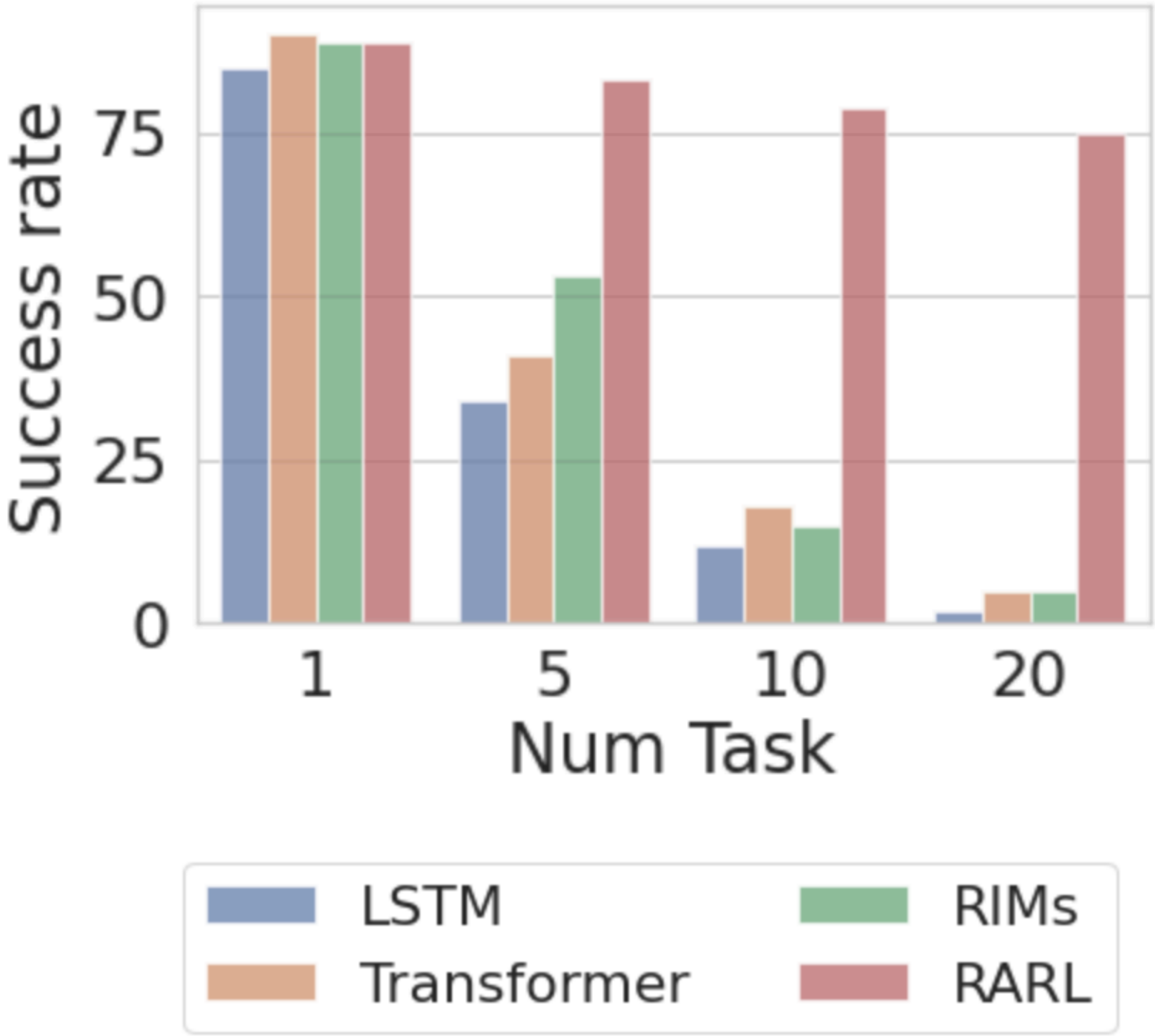
Continual Learning Results: Instruction Following

First training on training tasks, and then *sequential* "adaptation" on transfer tasks.



Training tasks

***Sequentially* train on transfer tasks**



Takeaway Lesson