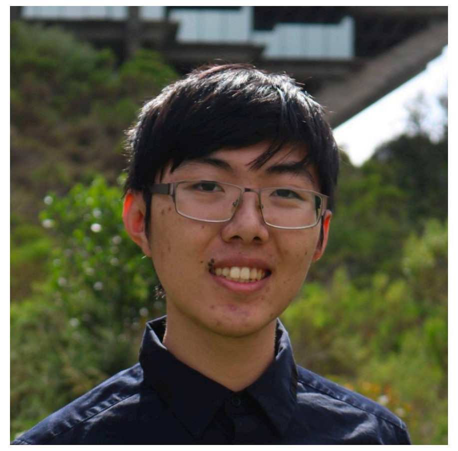
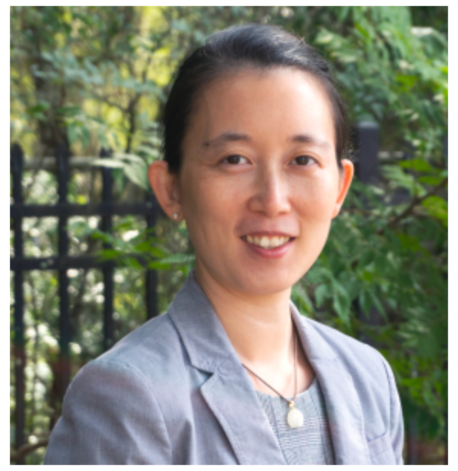
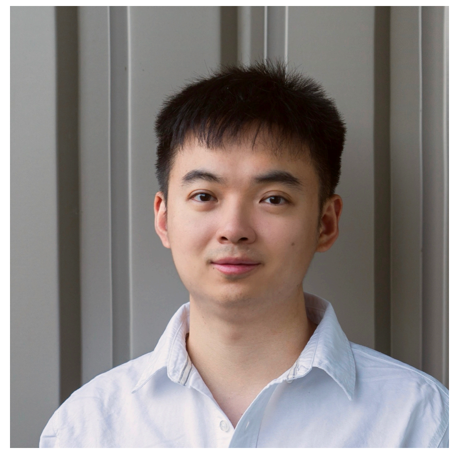
EMNLP 2023

MIX LAB

Reasoning with Language Model is Planning with World Model



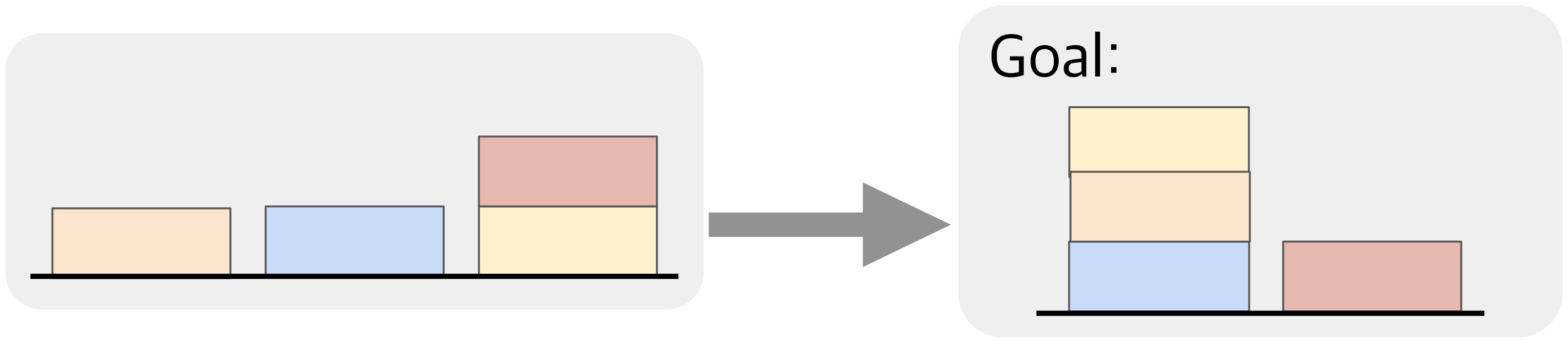



Shibo Hao* Yi Gu* Haodi Ma Joshua Hong Zhen Wang Daisy Wang Zhiting Hu

UC San Diego University of Florida MBZUAI

Reasoning with LLM vs Human

Blocksworld: How to move the blocks to the goal state?



A: Chain-of-Thoughts Prompting (CoT) with LLM

- Autoregressive decoding

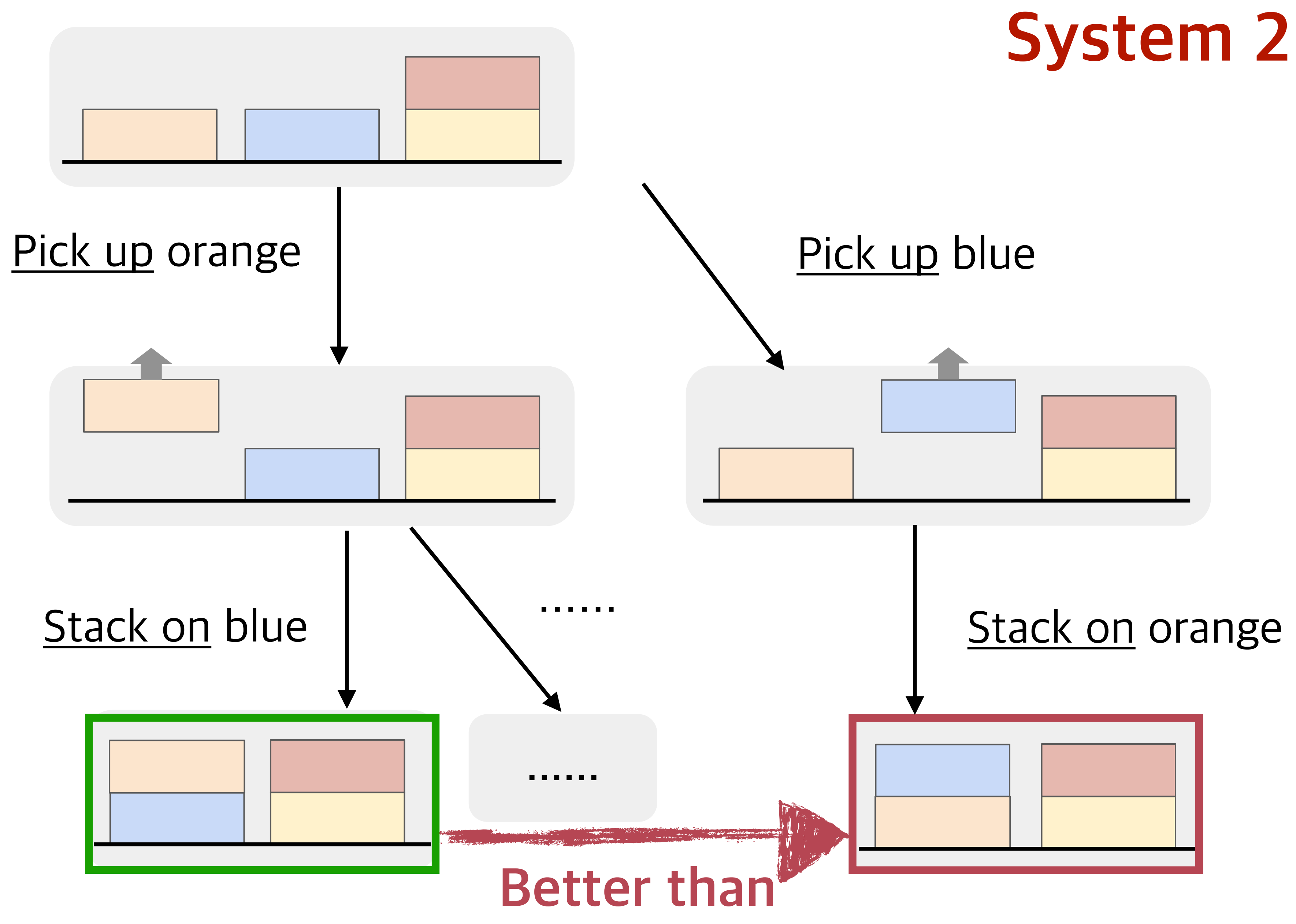
System 1

1. Pick up the orange block.
2. Stack it on the blue block.
3. Pick up the yellow block. ✗
4. Stack it on the orange block.
5. Pick up the red block.
6. Put it on the table.

Invalid Action!
The yellow block is still under the red one.

B: Human Reasoning

- Internal **world model** to track **states**
- **Explore** alternative reasoning paths
- **Assess outcomes** by looking ahead



On the planning abilities of large language models (a critical investigation with a proposed benchmark) [Valmeekam et al, 2023]
Chain-of-thought prompting elicits reasoning in large language models [Wei et al., 2022]
Mental models: Towards a cognitive science of language, inference, and consciousness [Johnson-Laird, 1983]
From System 1 Deep Learning to System 2 Deep Learning [Bengio, 2019]

Reasoning-via-Planning (RAP 🎶)

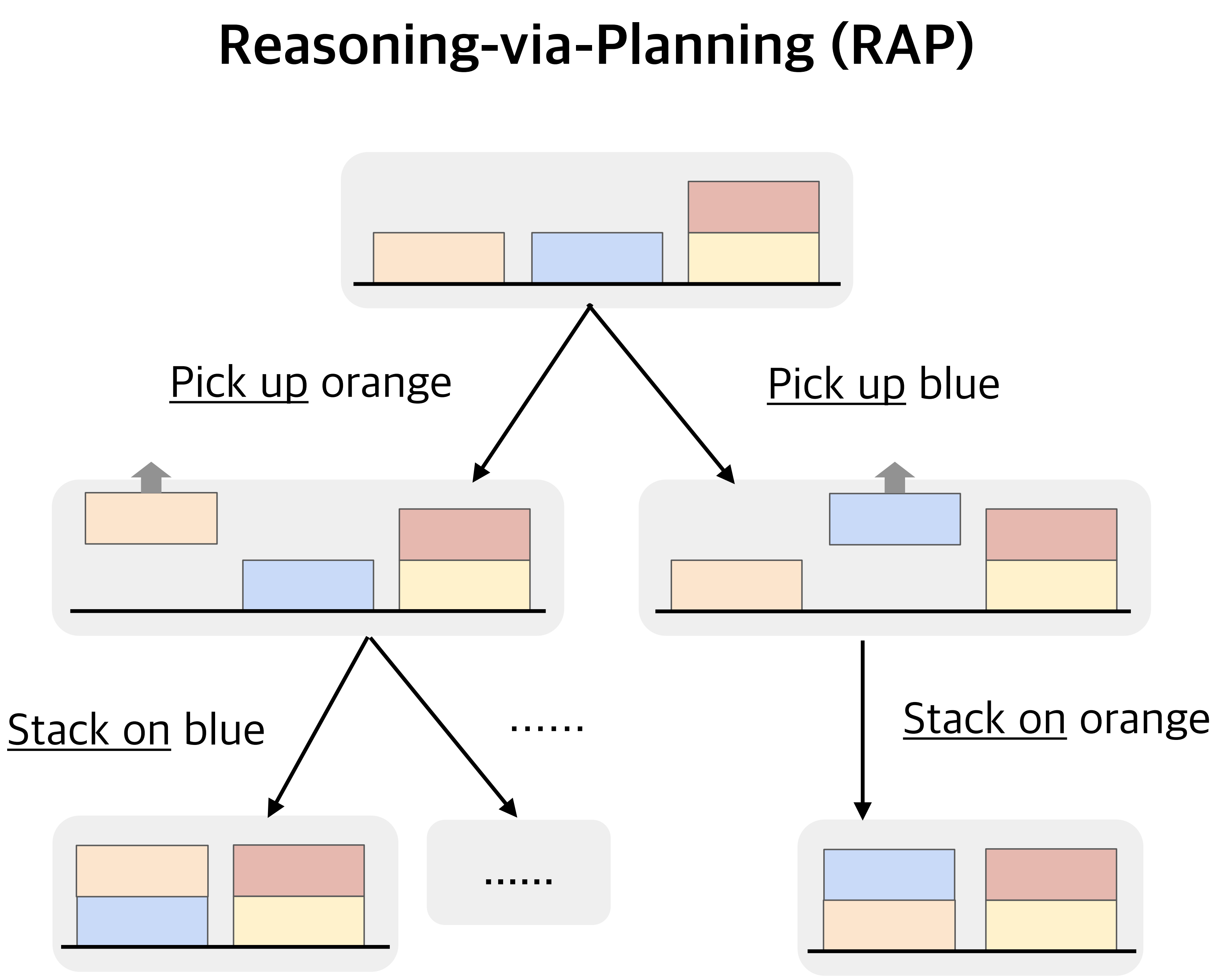
Human Reasoning

- Internal **world model** to track **states**
- **Explore** alternative reasoning paths
- **Assess outcomes** by looking ahead

How to enable LLMs to reason close to humans?

Reasoning-via-Planning: RAP 🎶

- Repurpose LLM as **world model**
- Principled **planning** algorithm
- **Rewards** to estimate outcomes

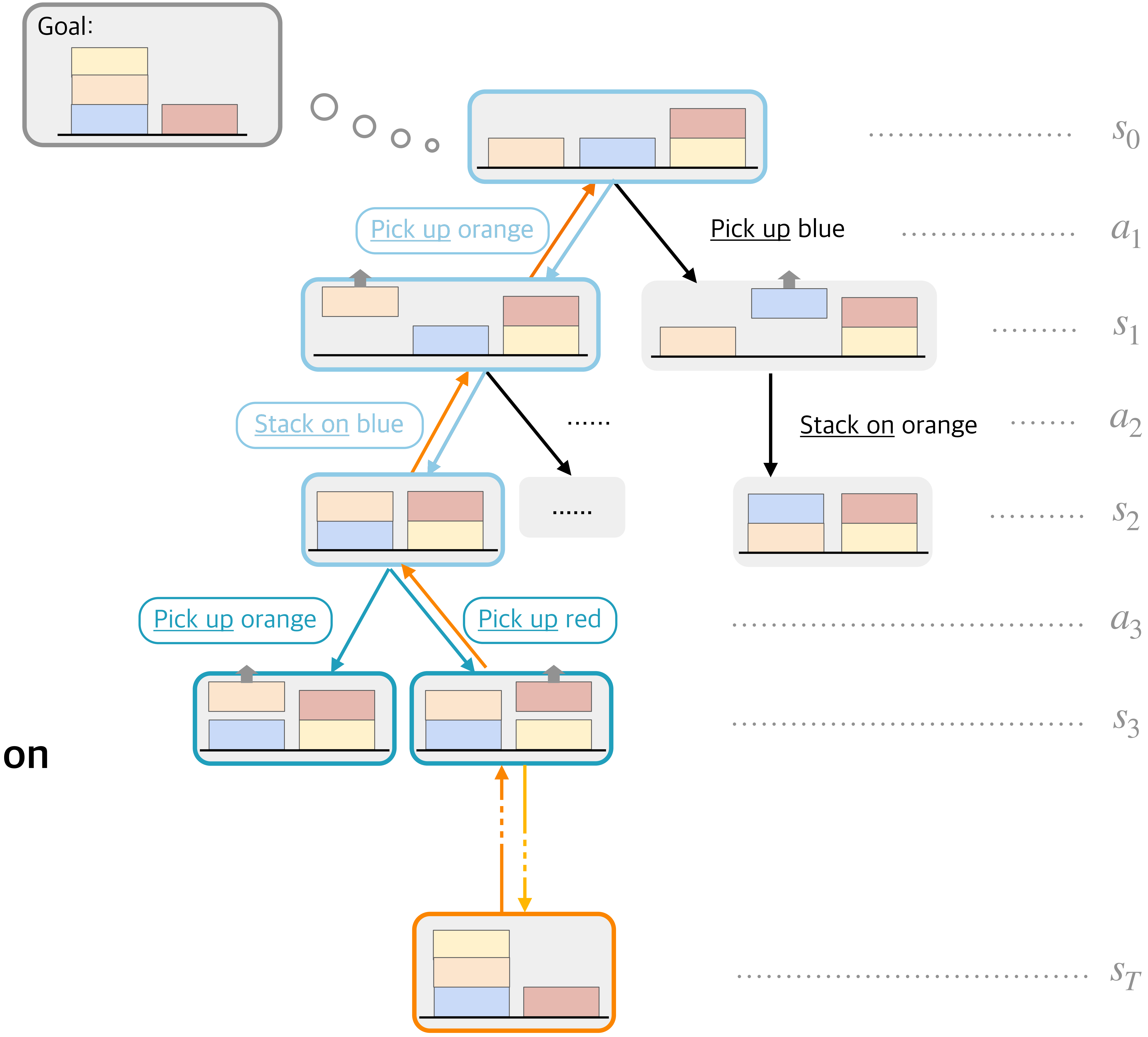


Planning Algorithm

Monte Carlo Tree Search (MCTS):
Iteratively build reasoning tree

1. Selection
2. Expansion
3. Simulation
4. Back-propagation

Balanced exploration and exploitation



Rewards in RAP

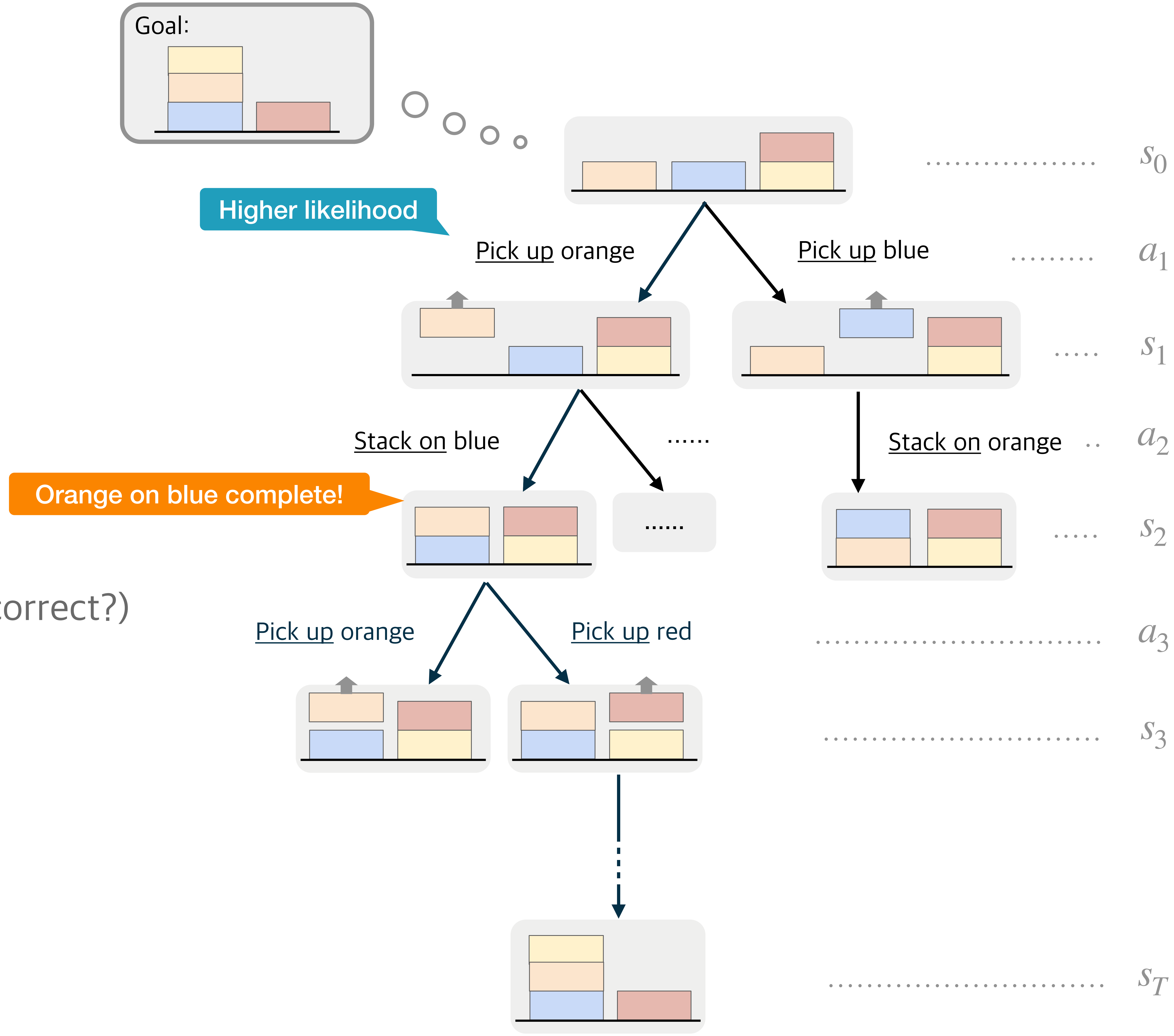
Reward design is **flexible**

In Blocksworld:

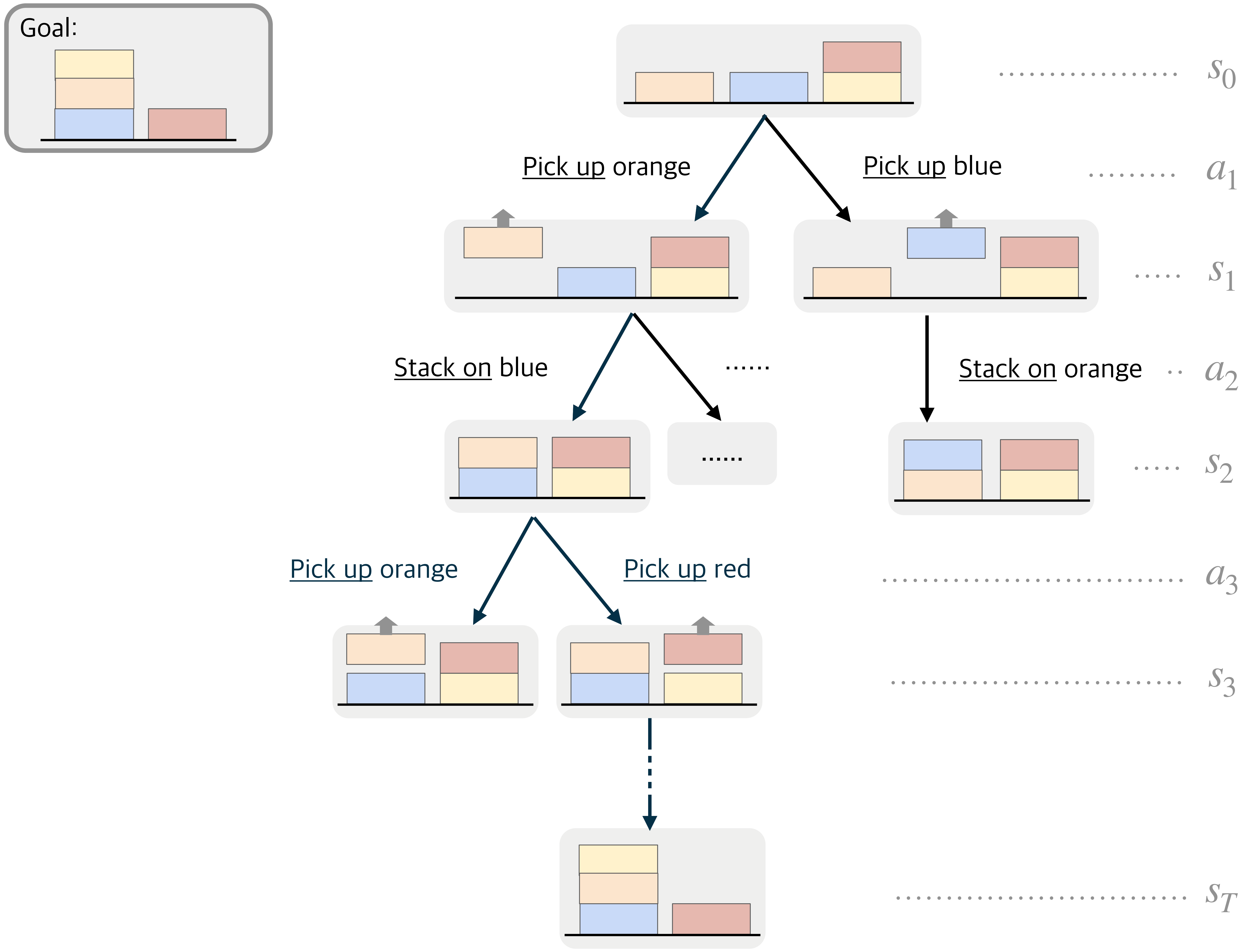
- **Likelihood** of actions
- **Task-heuristic** (# of subgoals)

Other possible rewards:

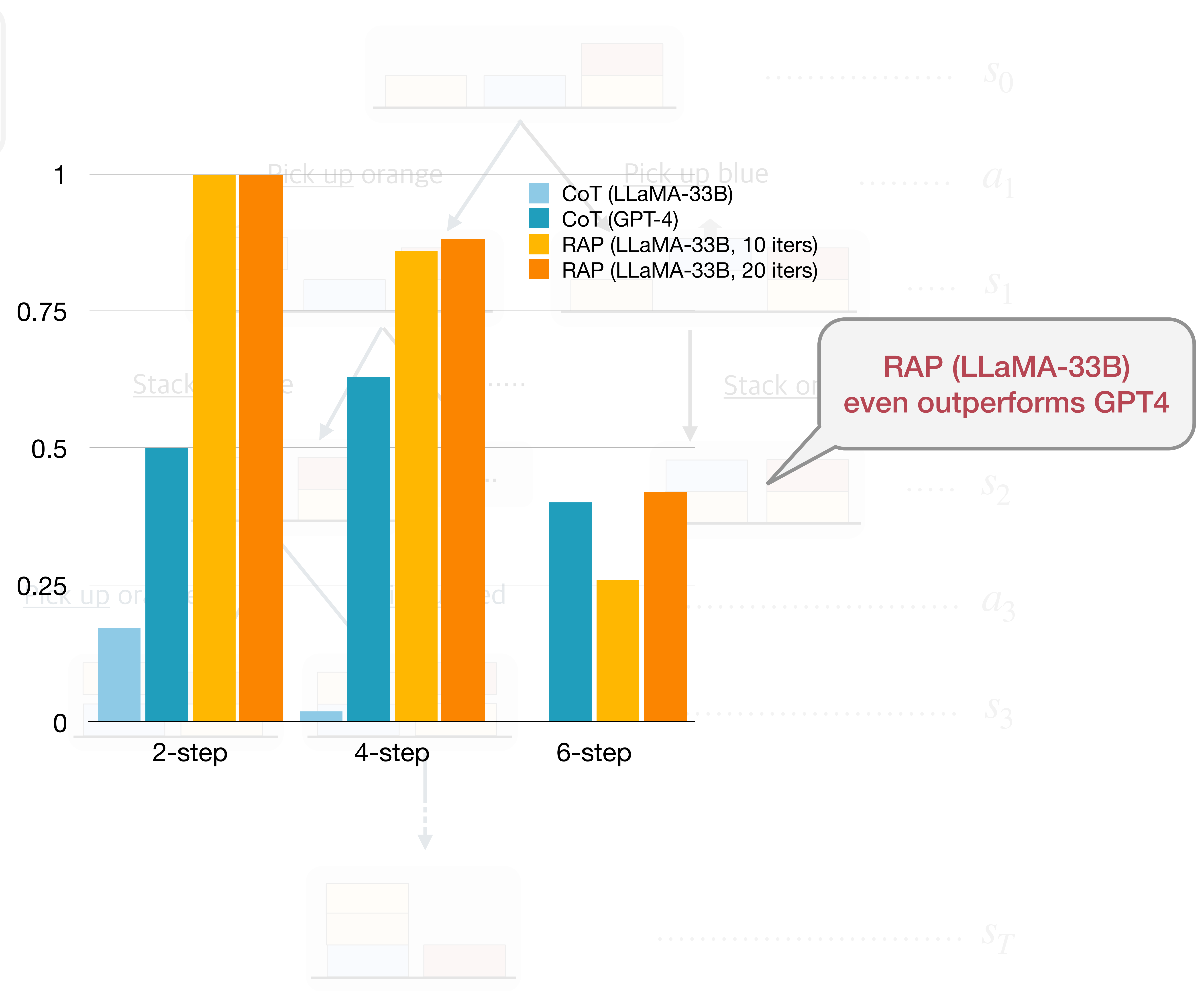
- **Self-evaluation** by LLM (e.g. useful? correct?)
- **Confidence** of next state
- ……



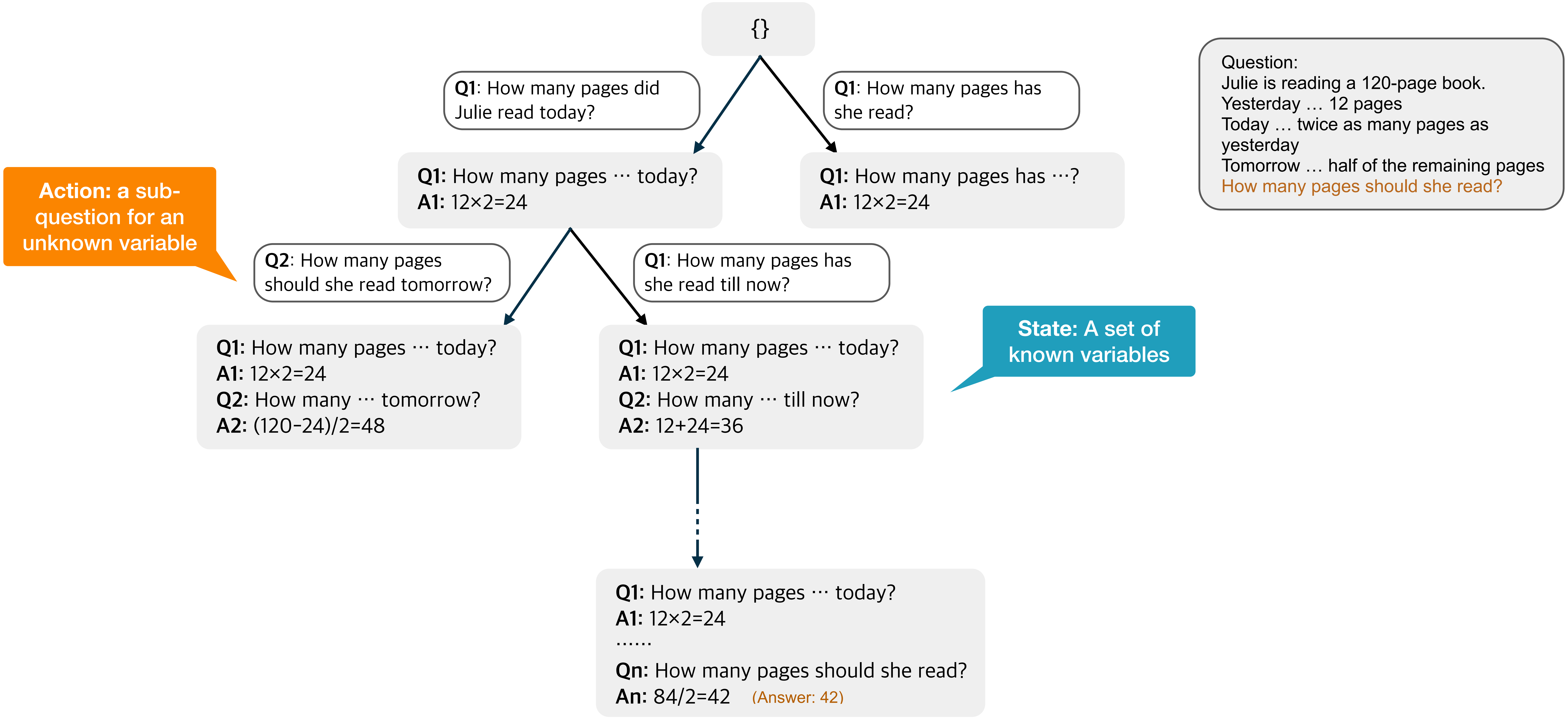
RAP on Plan Generation (Blocksworld)



RAP on Plan Generation (Blocksworld)

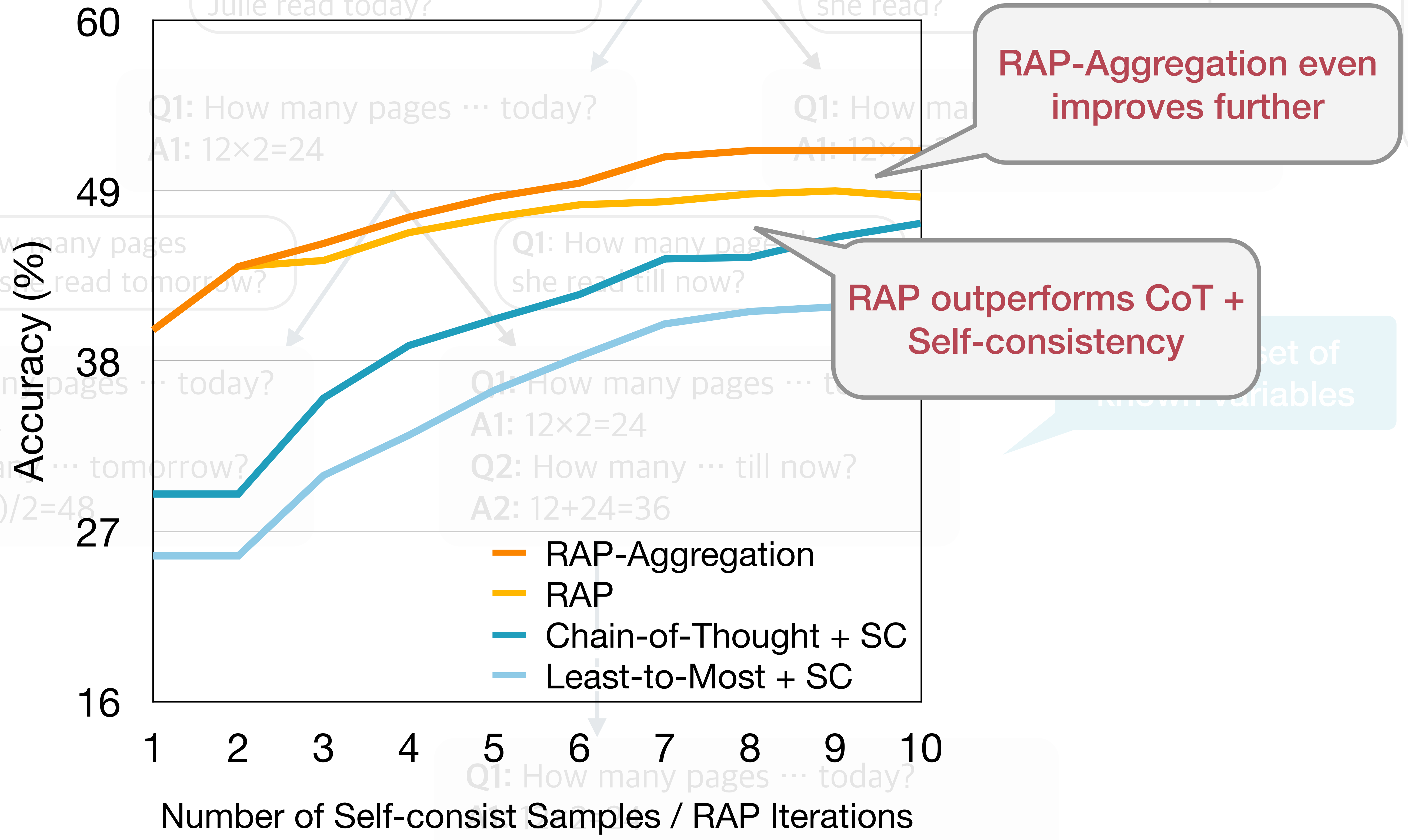


RAP on Mathematical Reasoning (GSM8k)



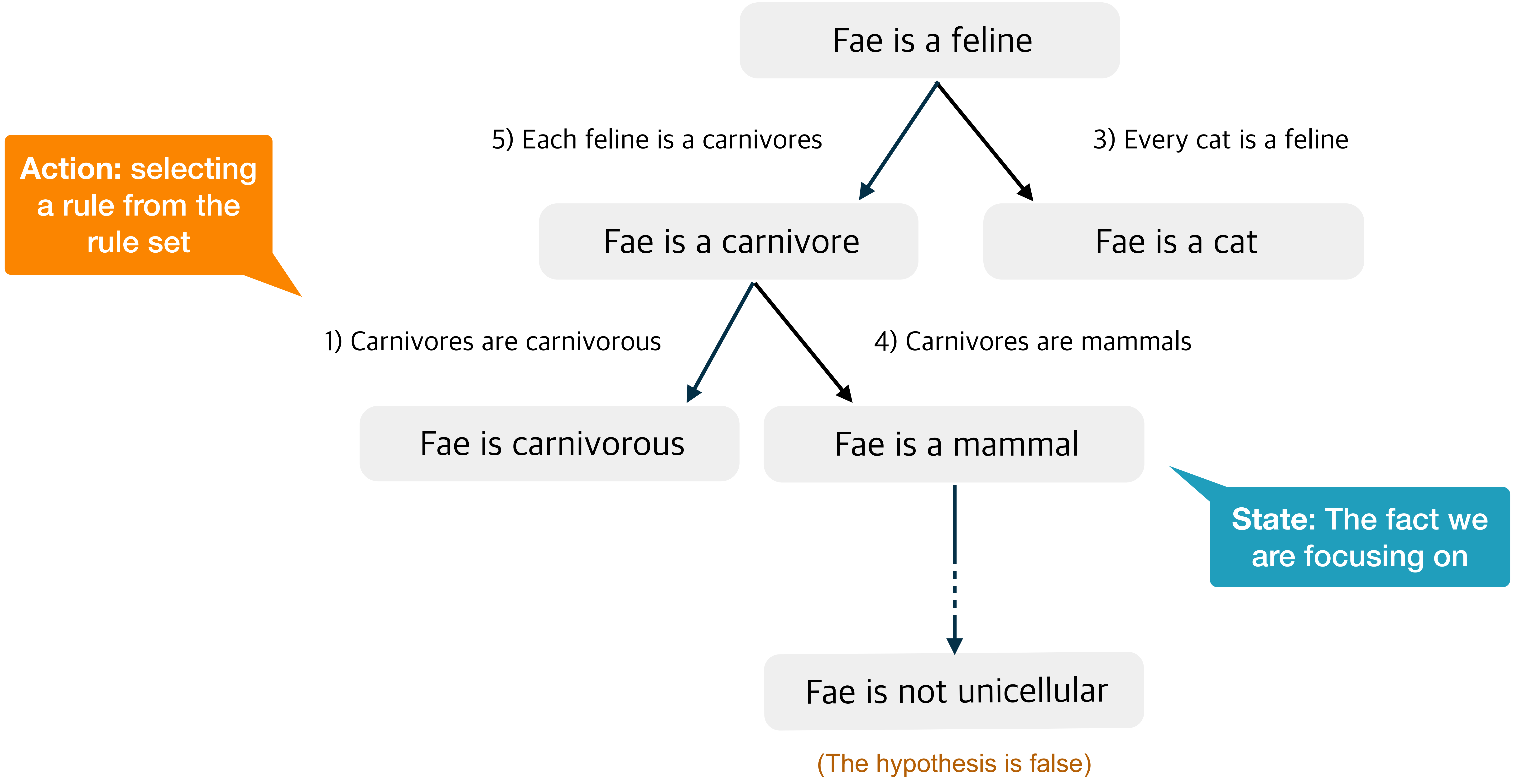
Training verifiers to solve math word problems. [Cobbe et al., 2021]

RAP on Mathematical Reasoning (GSM8k)



Training verifiers to solve math word problems. [Cobbe et al., 2021]

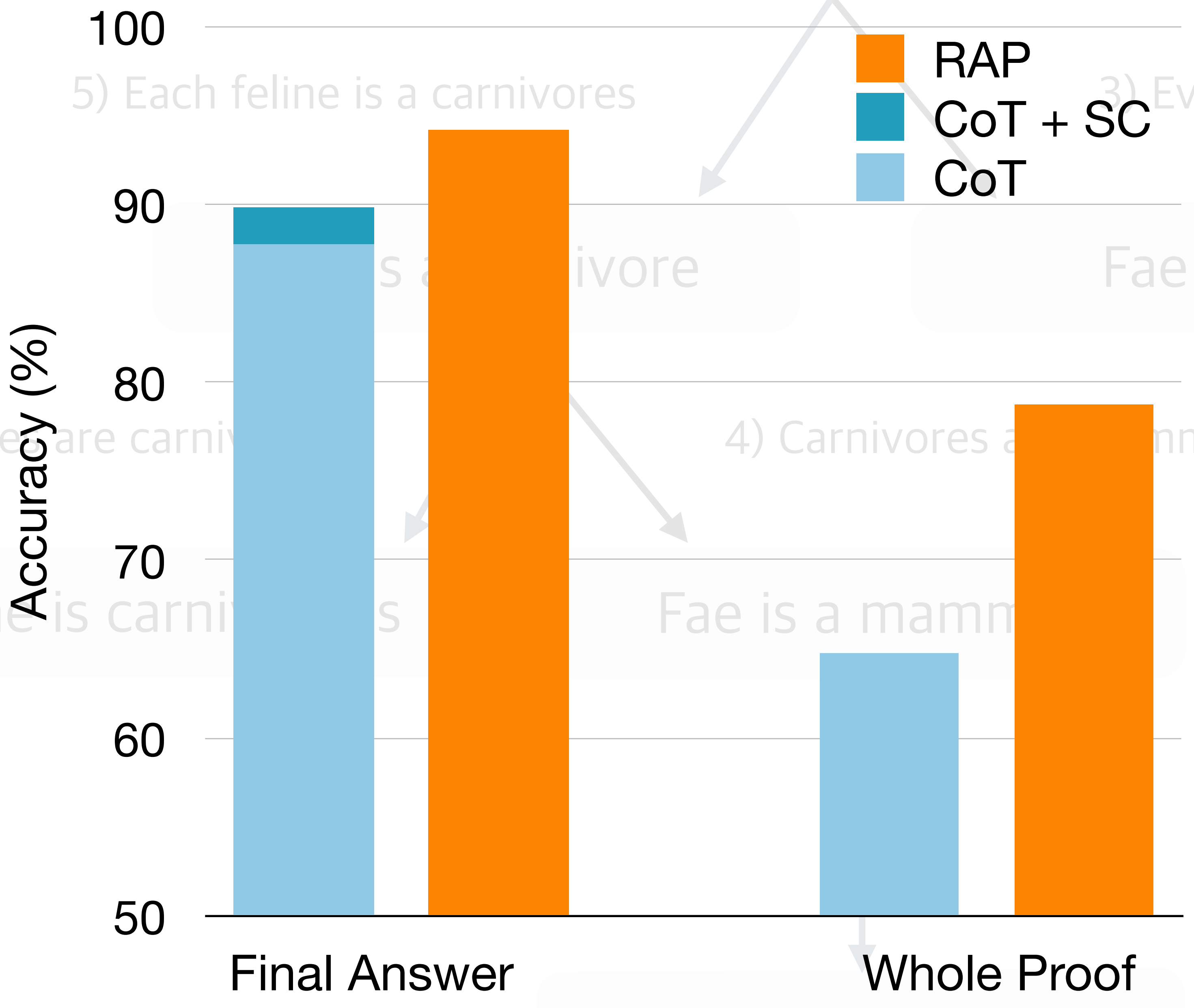
RAP on Logical Reasoning (PrOntoQA)



Rules:
(1) Carnivores are carnivorous
(2) Animals are not unicellular
(3) Every cat is a feline
(4)…
Fact: Fae is a feline
Hypothesis: Fae is unicellular?

Language models are greedy reasoners: A systematic formal analysis of chain-of-thought. [Saparov and He, 2022]

RAP on Logical Reasoning (PrOntoQA)



RAP outperforms CoT much in proof accuracy

Language models are greedy reasoners: A systematic formal analysis of chain-of-thought. [Saparov and He, 2022]

Discussion and Future Works

- We are developing LLM Reasoners - a library to conduct complex reasoning with advanced algorithms
  - Covering multiple algorithms (**BeamSearch**, **ToT**, **RAP** 🎶, etc.)
  - Flexible LLM interfaces and intuitive visualization of reasoning trees
- Possible future works
  - **Fine-tuning LLMs** to better reason and **serve as a world model**
  - Building **multi-modal world model**
  - Combining LLM with **external tools** as a more **powerful** world model

LLM Reasoners

llm-reasoners.net

Self-Evaluation Guided Beam Search for Reasoning [Xie et al., 2023]
Grace: Discriminator-Guided Chain-of-Thoughts Reasoning [Khalifa et al., 2023]
Tree of thoughts: Deliberate problem solving with large language models. [Yao et al., 2023]

Takeaways

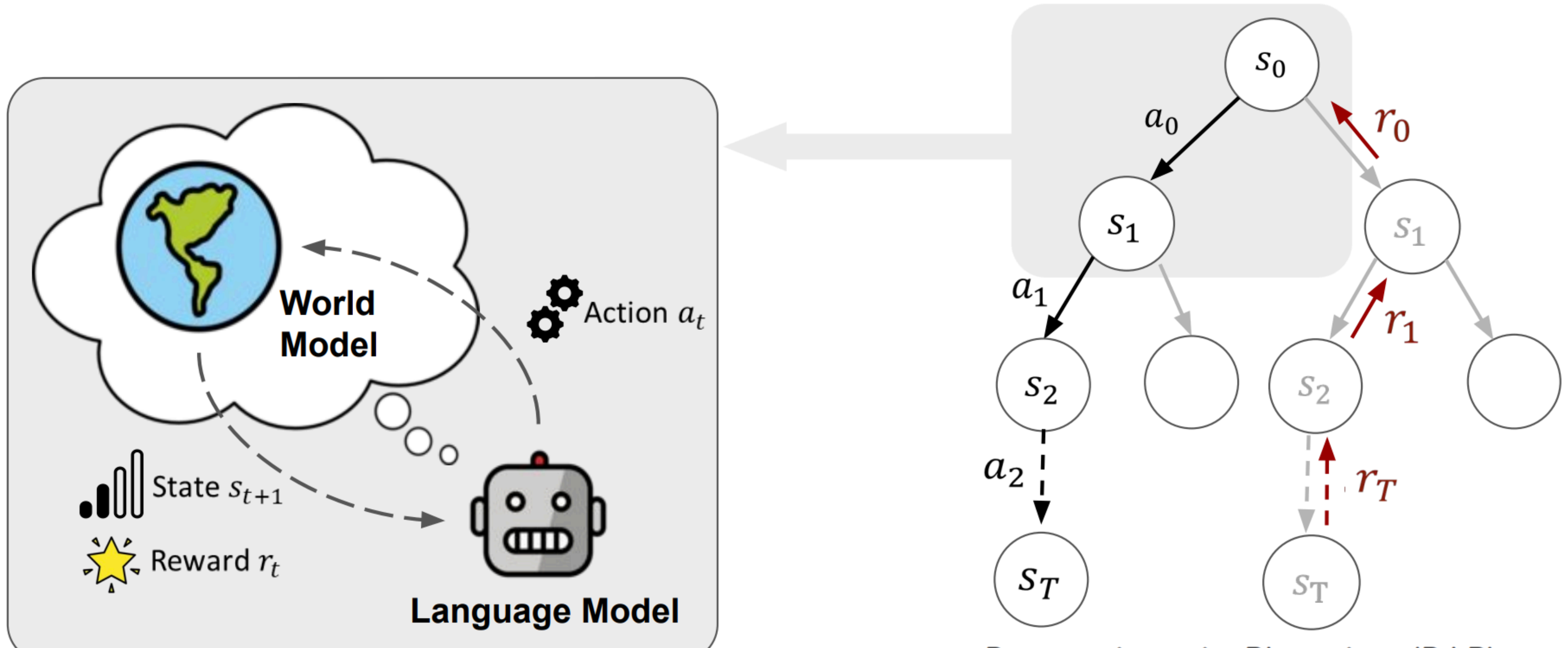
RAP 🎶: LLM reasoning as human-like strategic planning

- Repurpose LLM as **world model** to explicitly simulate future states
- Principled **planning** algorithm to balance exploration and exploitation
- Flexible **rewards** to estimate outcomes
- **Superior results in diverse domains**

arXiv GitHub

LLM Reasoners

llm-reasoners.net



Reasoning via Planning (RAP)