ELLA: Exploration through Learned Language Abstraction

Suvir Mirchandani Siddharth Karamcheti Dorsa Sadigh

iliad

Image: Szot et al. 2021

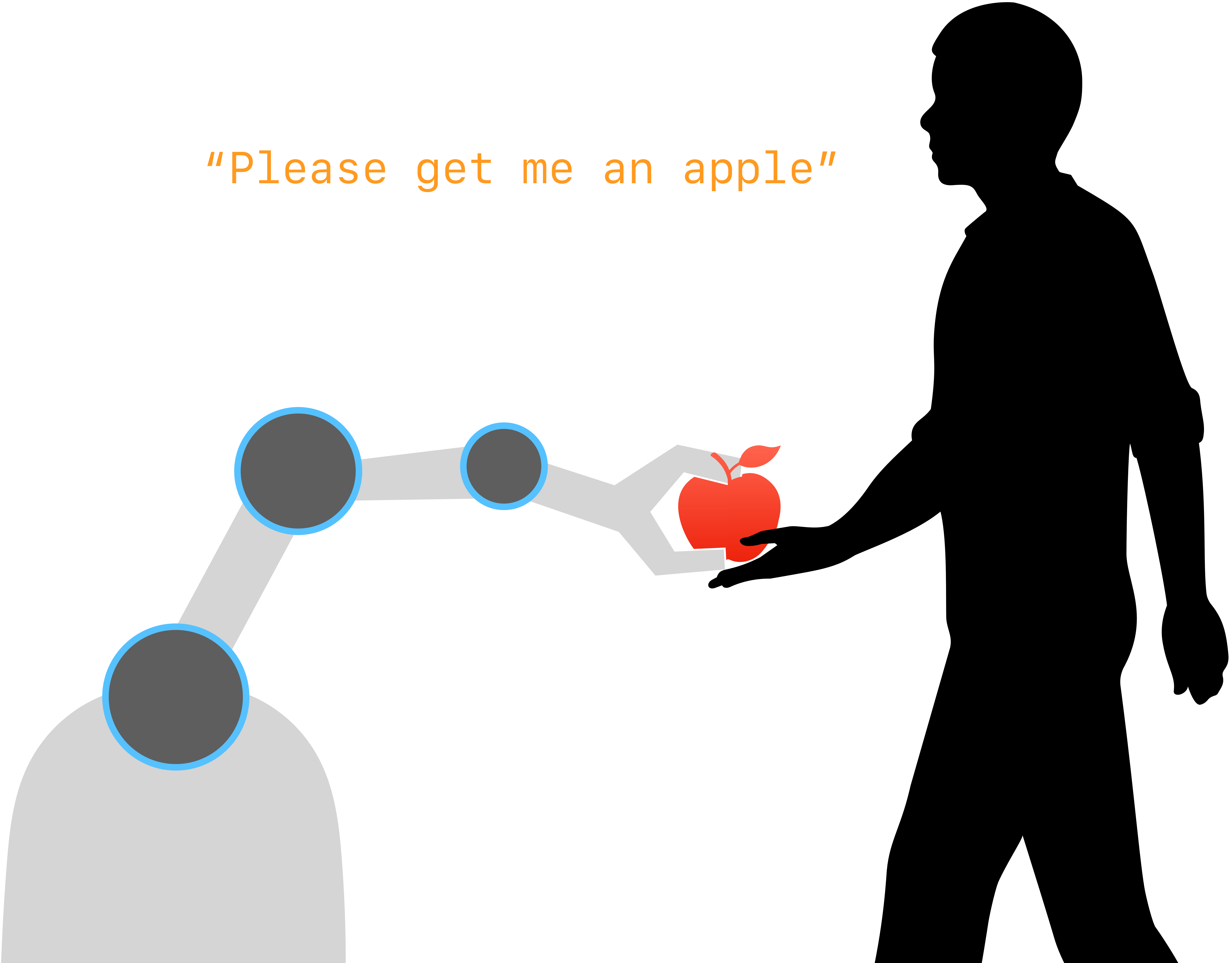
"Please get me an apple"

Image: Szot et al. 2021

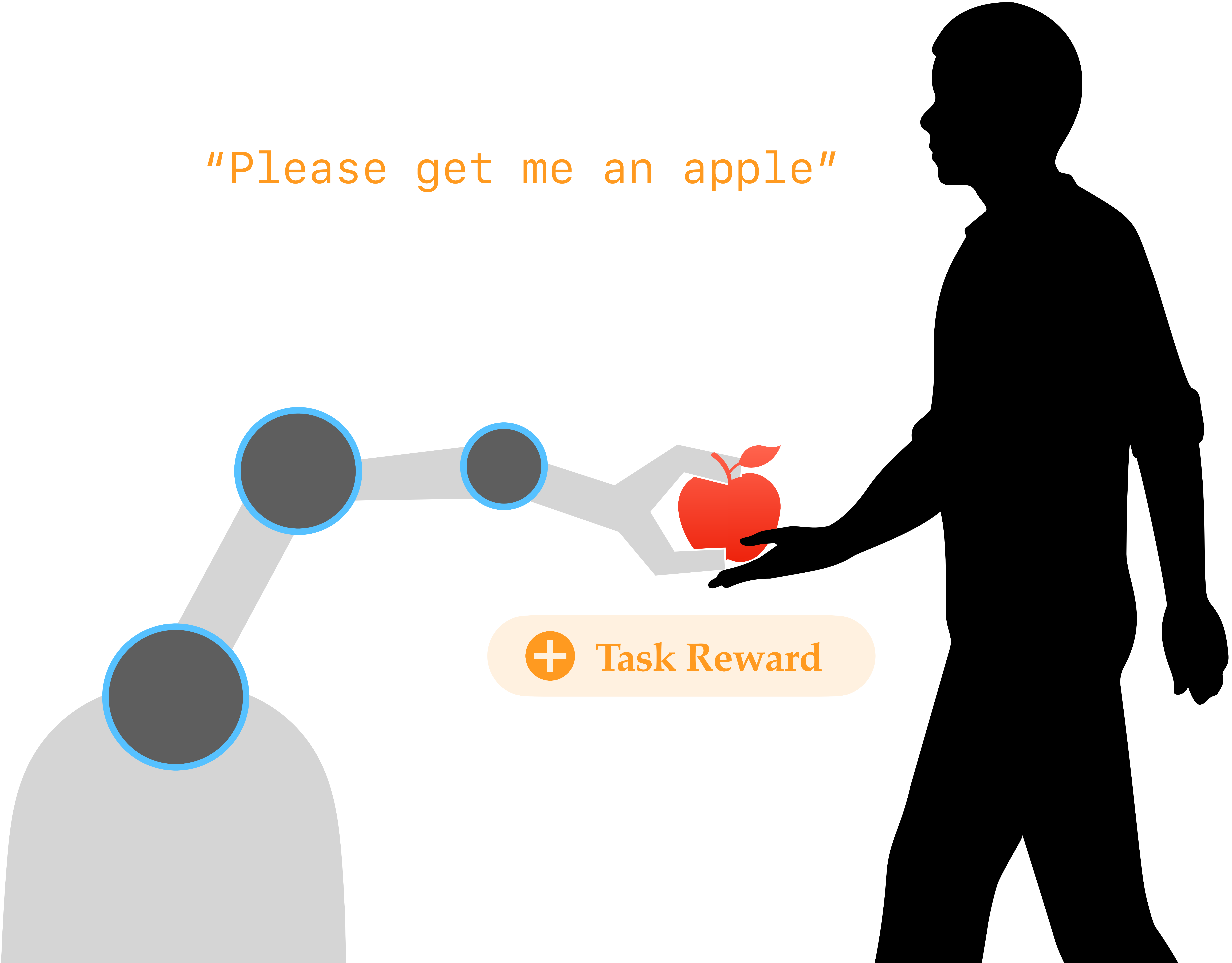
"Go to the table"
"Pick up an apple"
"Drive to me"

Image: Szot et al. 2021

“Please get me an apple”

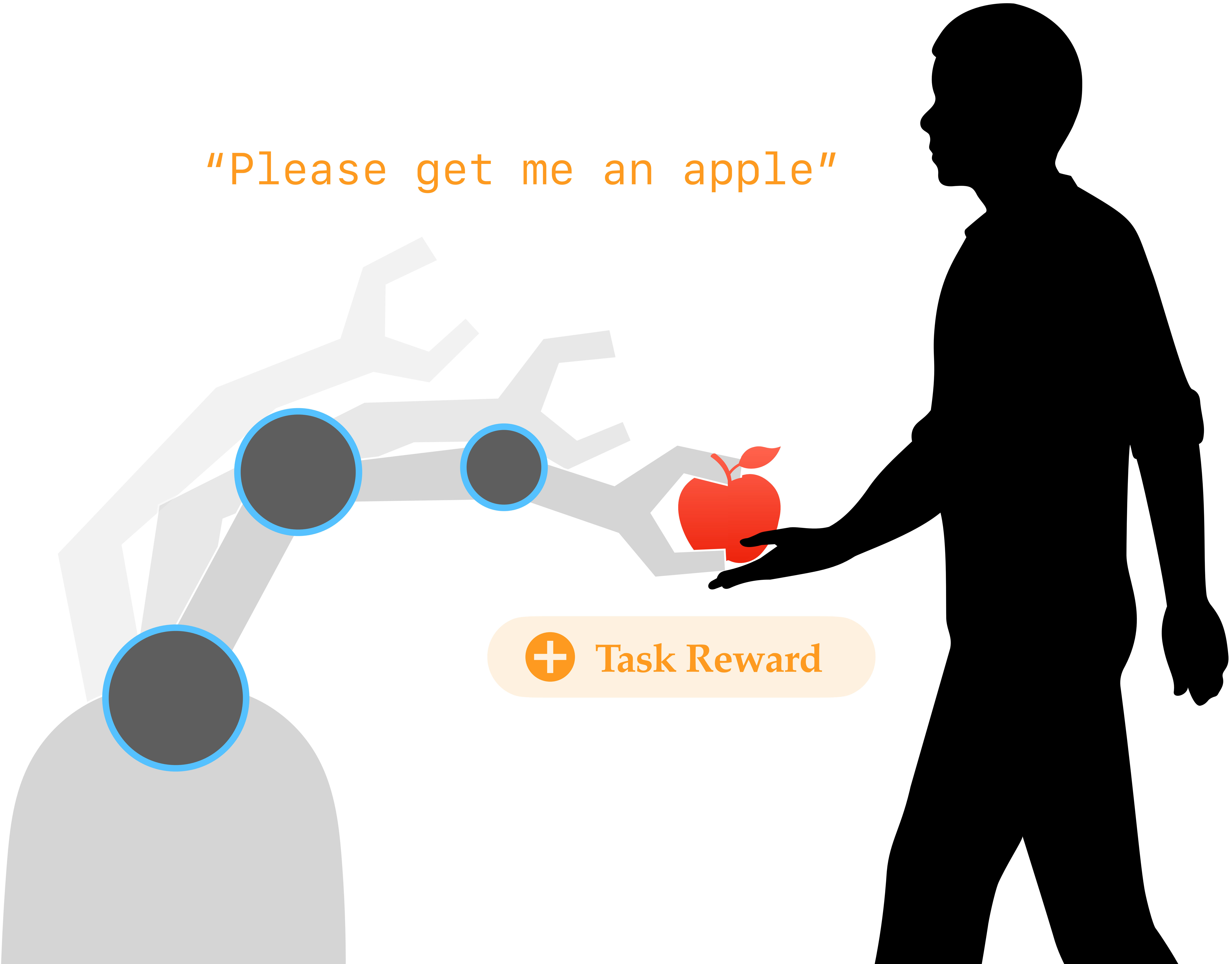


“Please get me an apple”



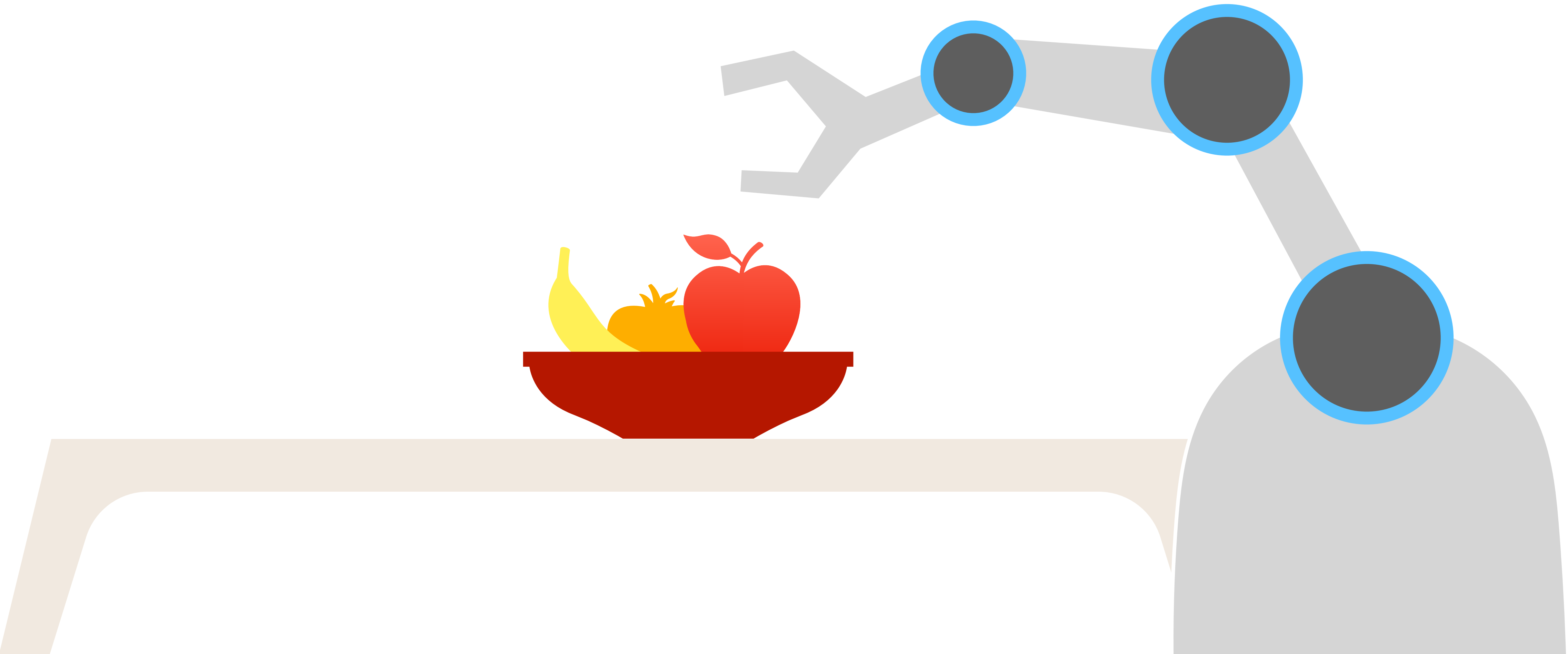
⊕ Task Reward

“Please get me an apple”

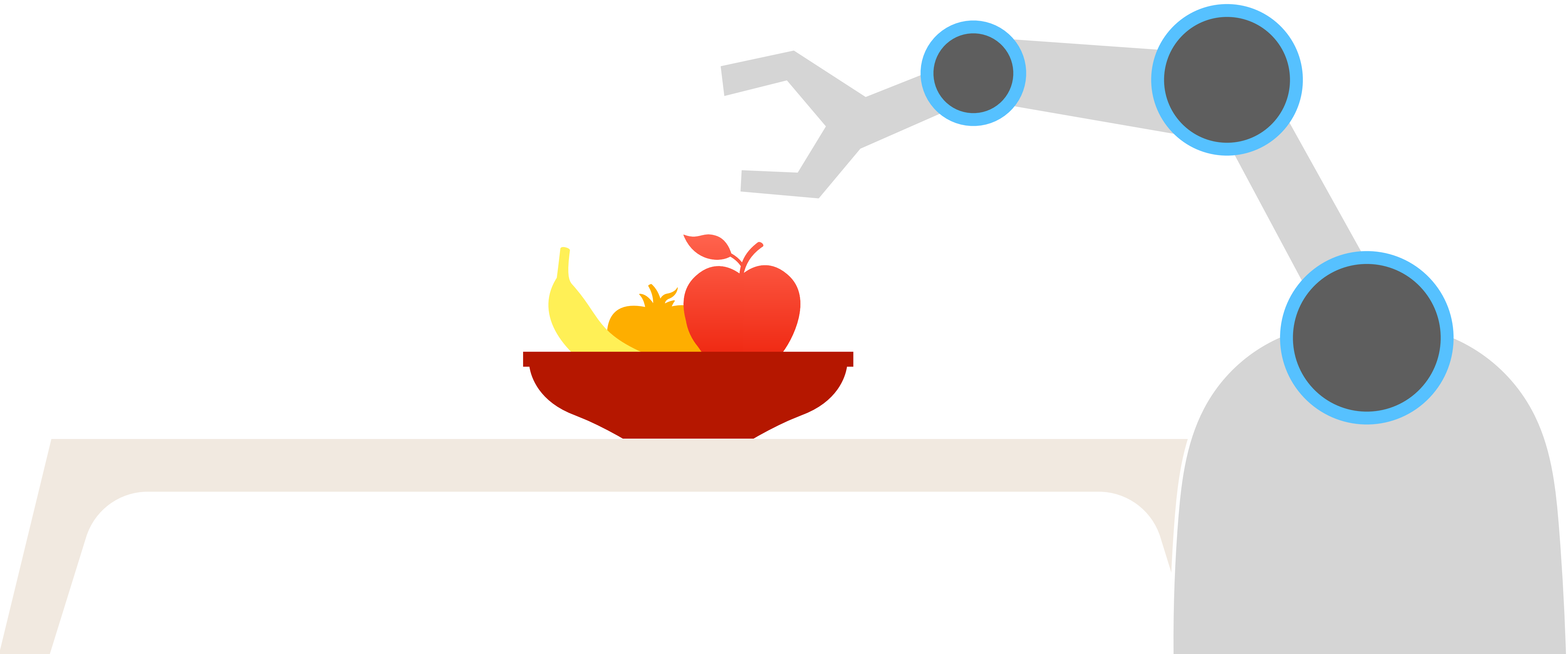


⊕ Task Reward

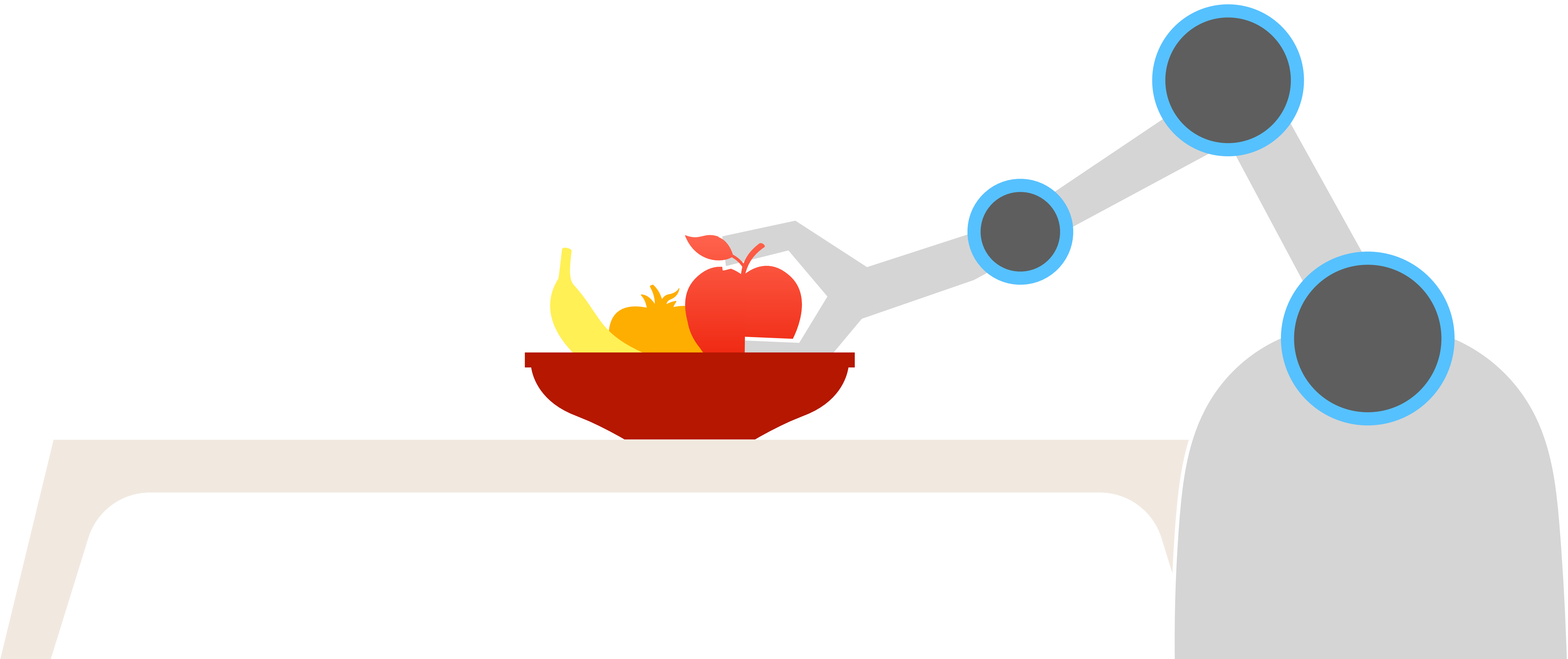
“Go to the table”



“Go to the table”

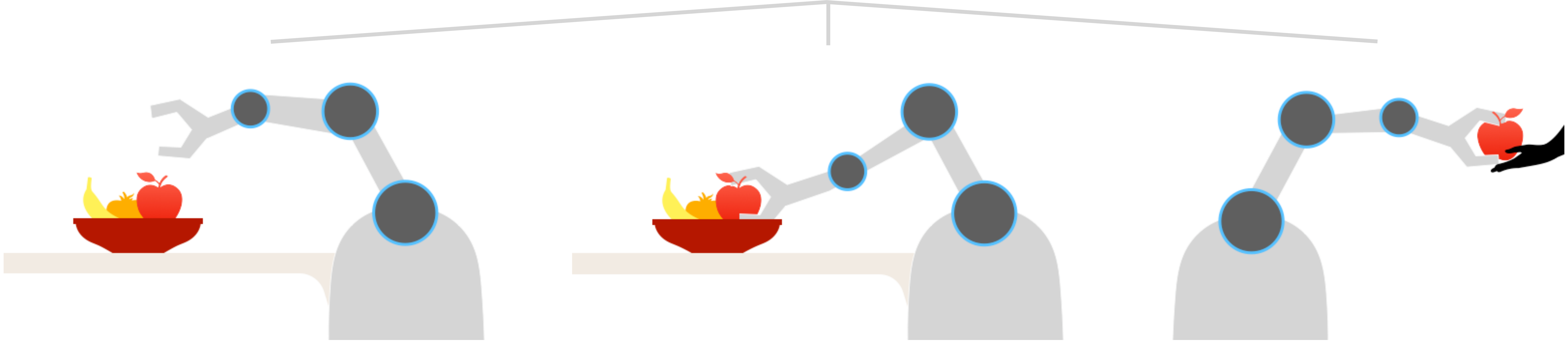


“Pick up an apple”



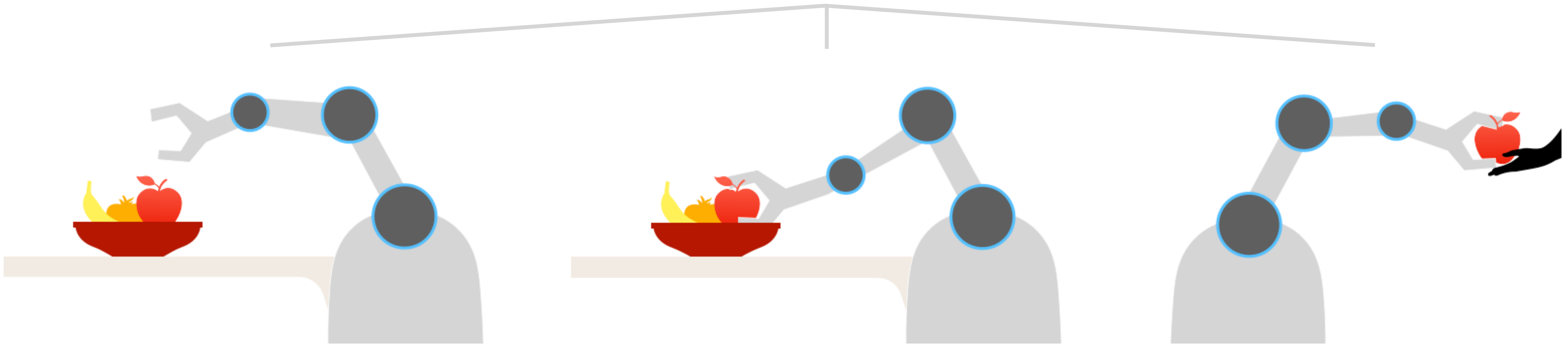
Learning from sparse rewards for complex tasks is not sample-efficient.

“Please get me an apple”



“Go to the table” “Pick up an apple” “Bring it to me”

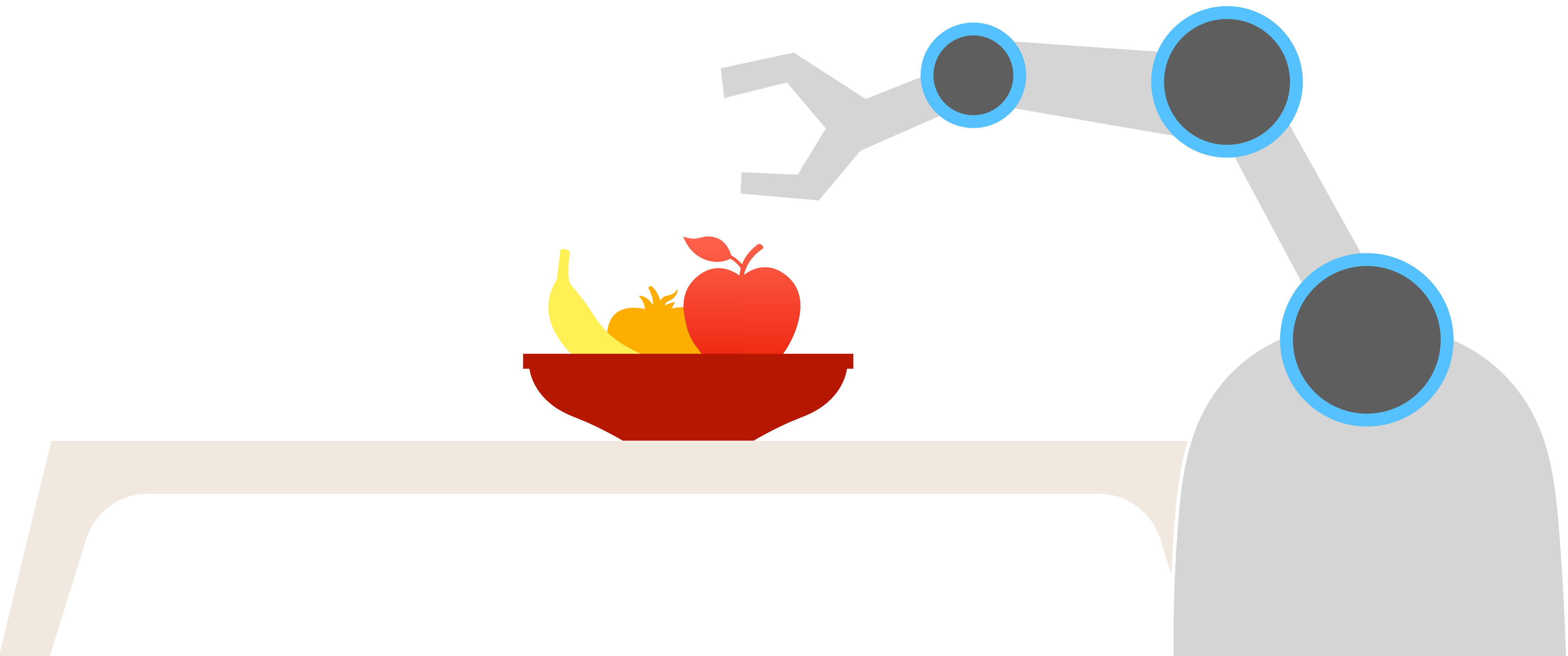
“Please get me an apple”



“Go to the table” “Pick up an apple” “Bring it to me”

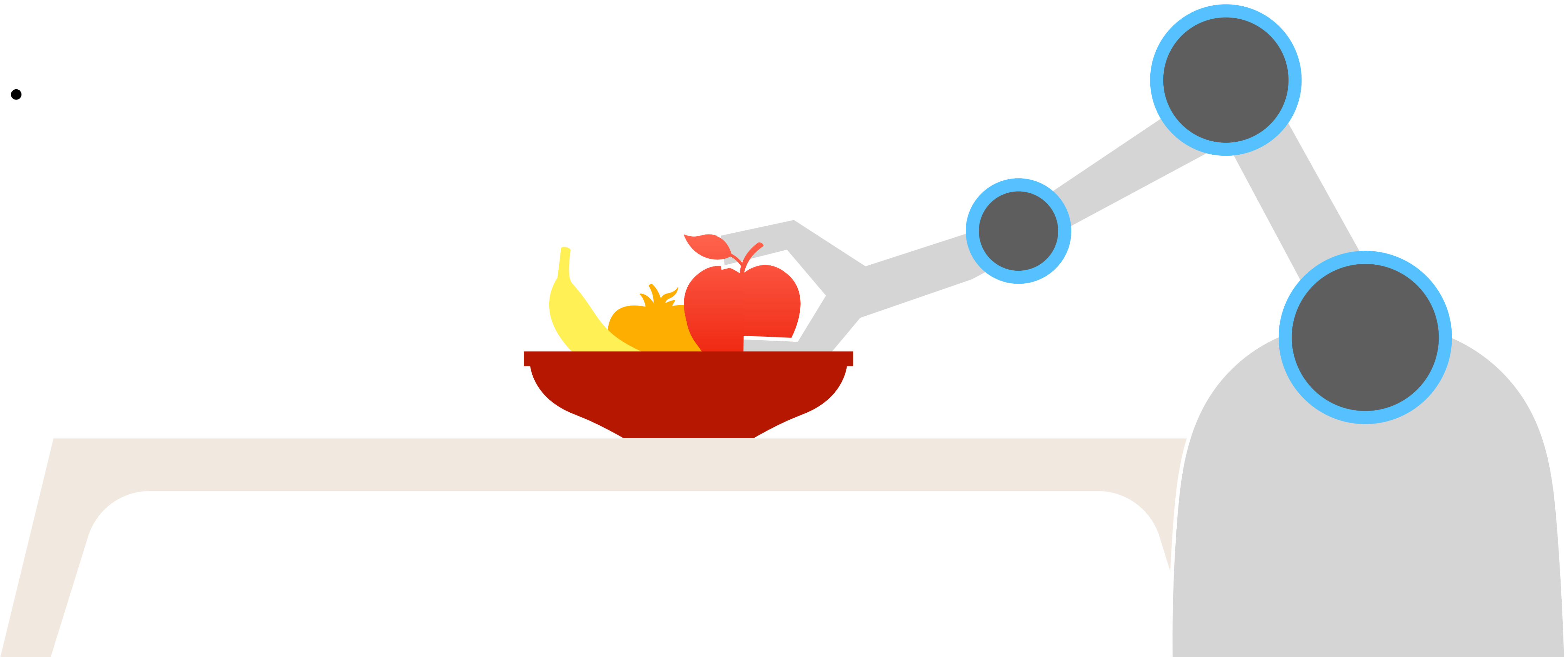
Can we use the principle of abstraction in language to guide exploration?

“Go to the table”



…

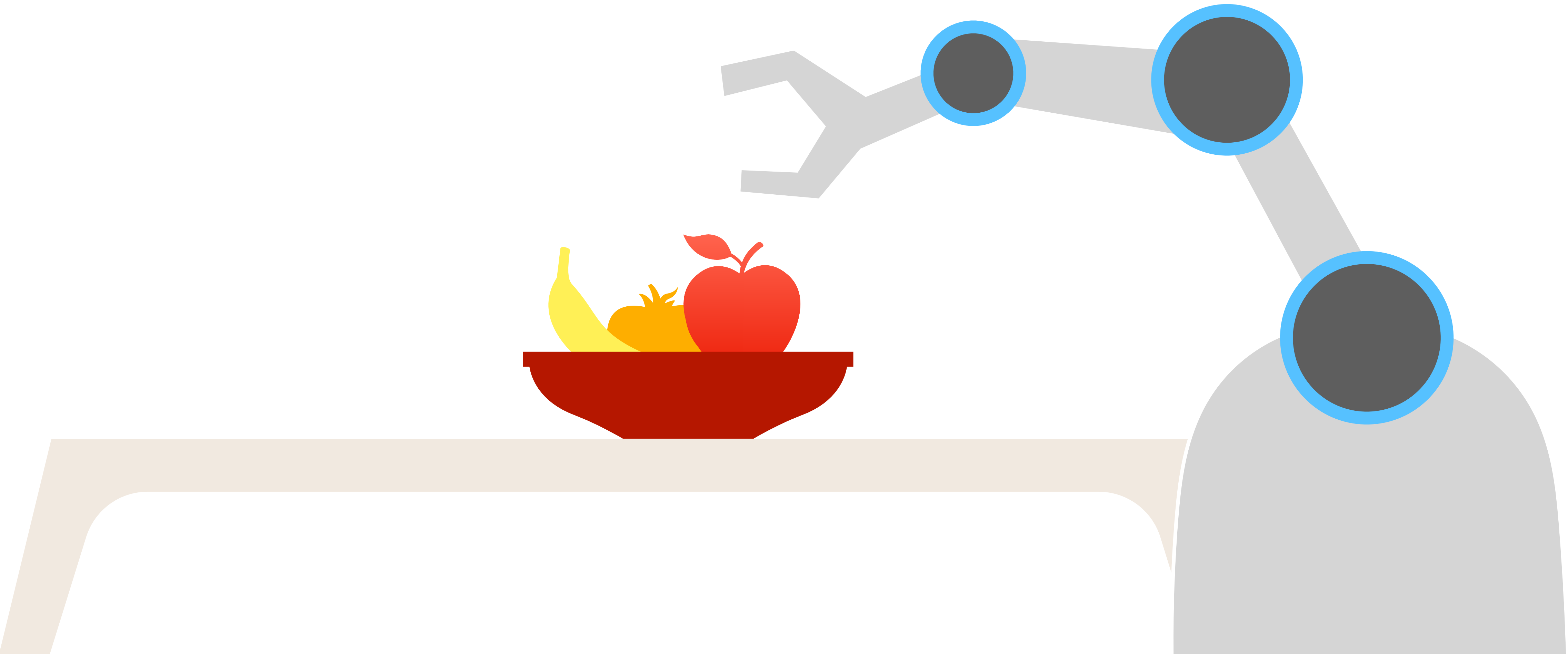
“Pick up an apple”



…

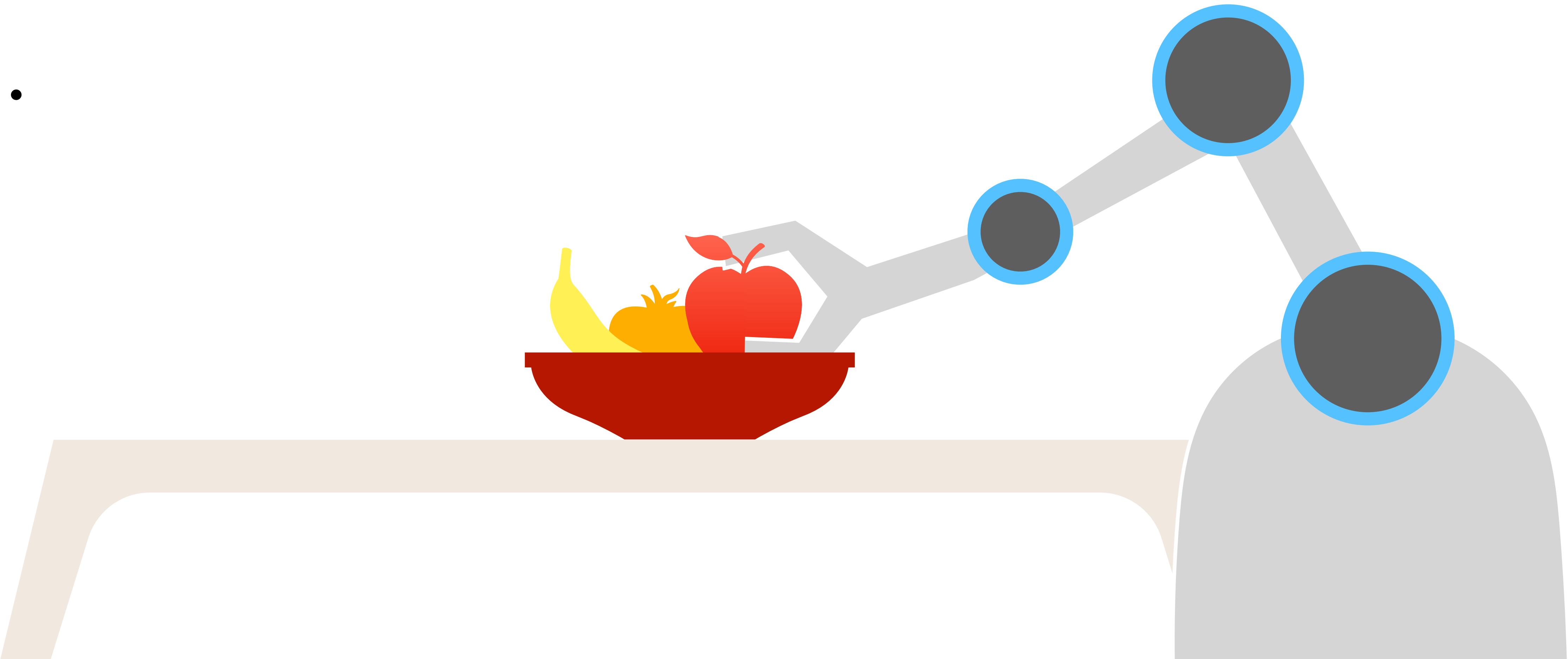
“Go to the table”

⊕ Subtask Reward



…

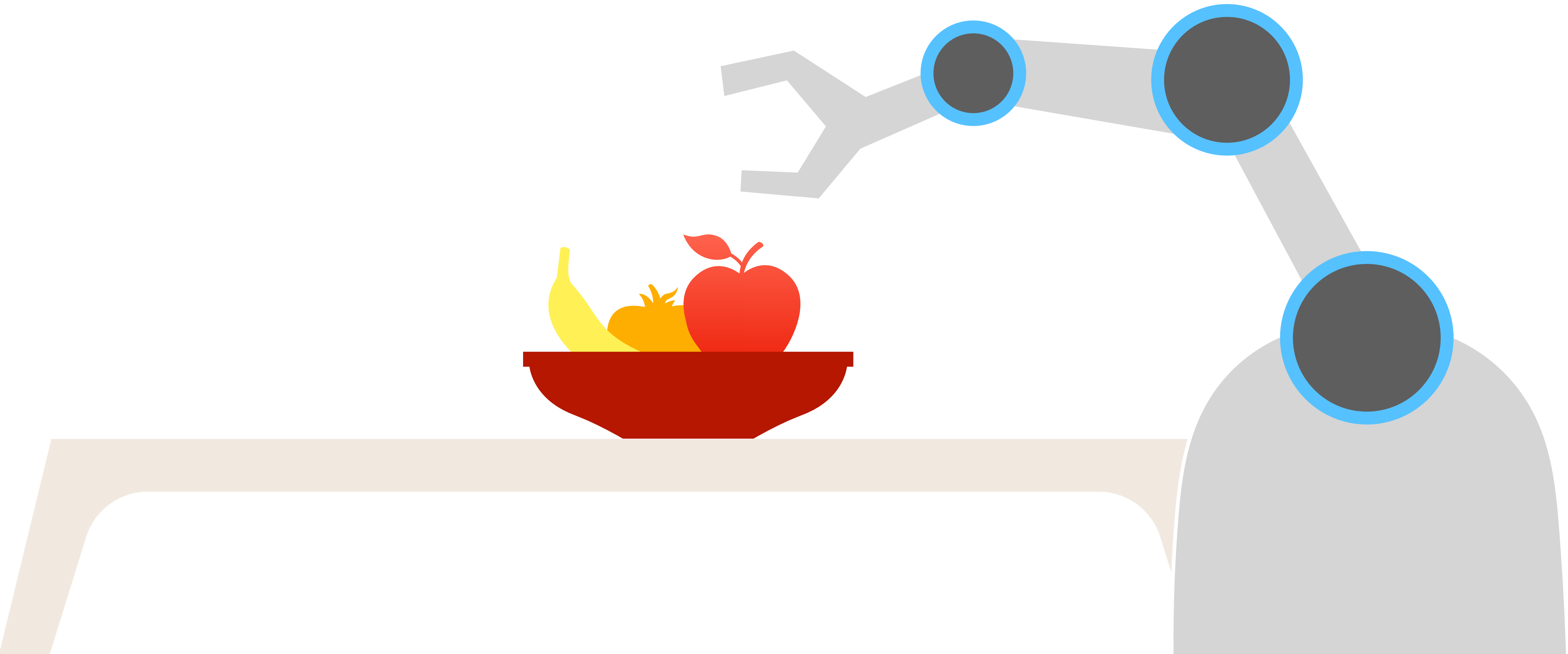
“Pick up an apple”



…

“Go to the table”

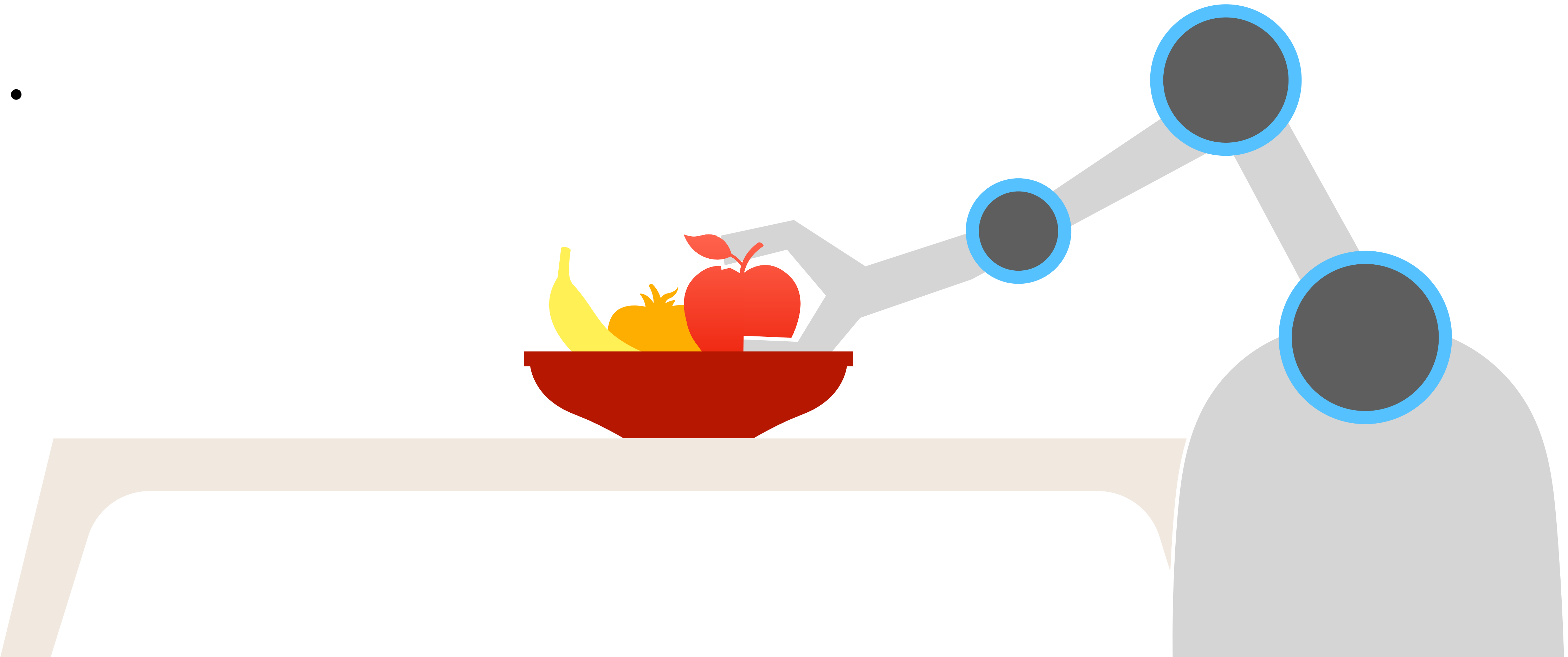
⊕ Subtask Reward



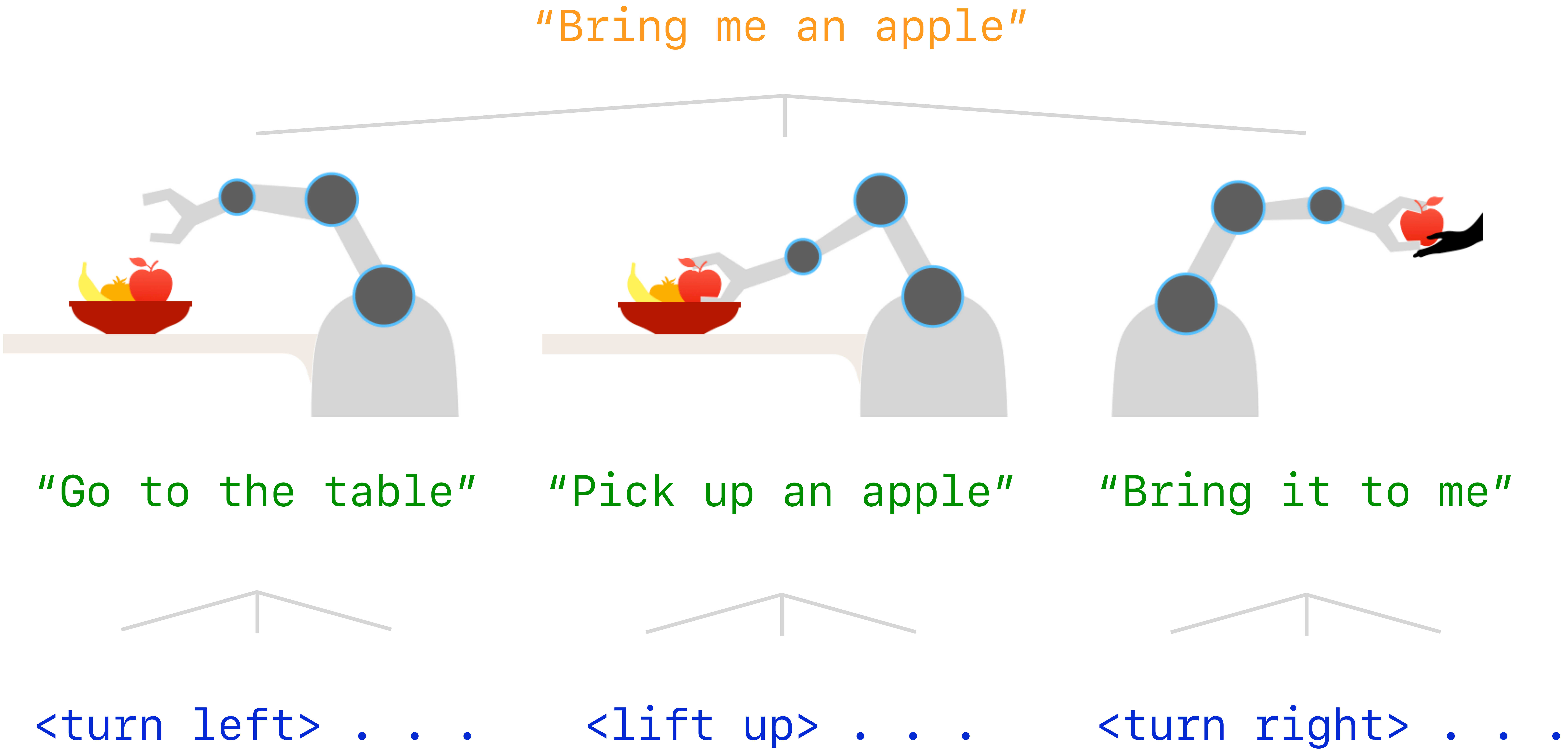
…

“Pick up an apple”

⊕ Subtask Reward



…



"Bring me an apple"

"Go to the table"

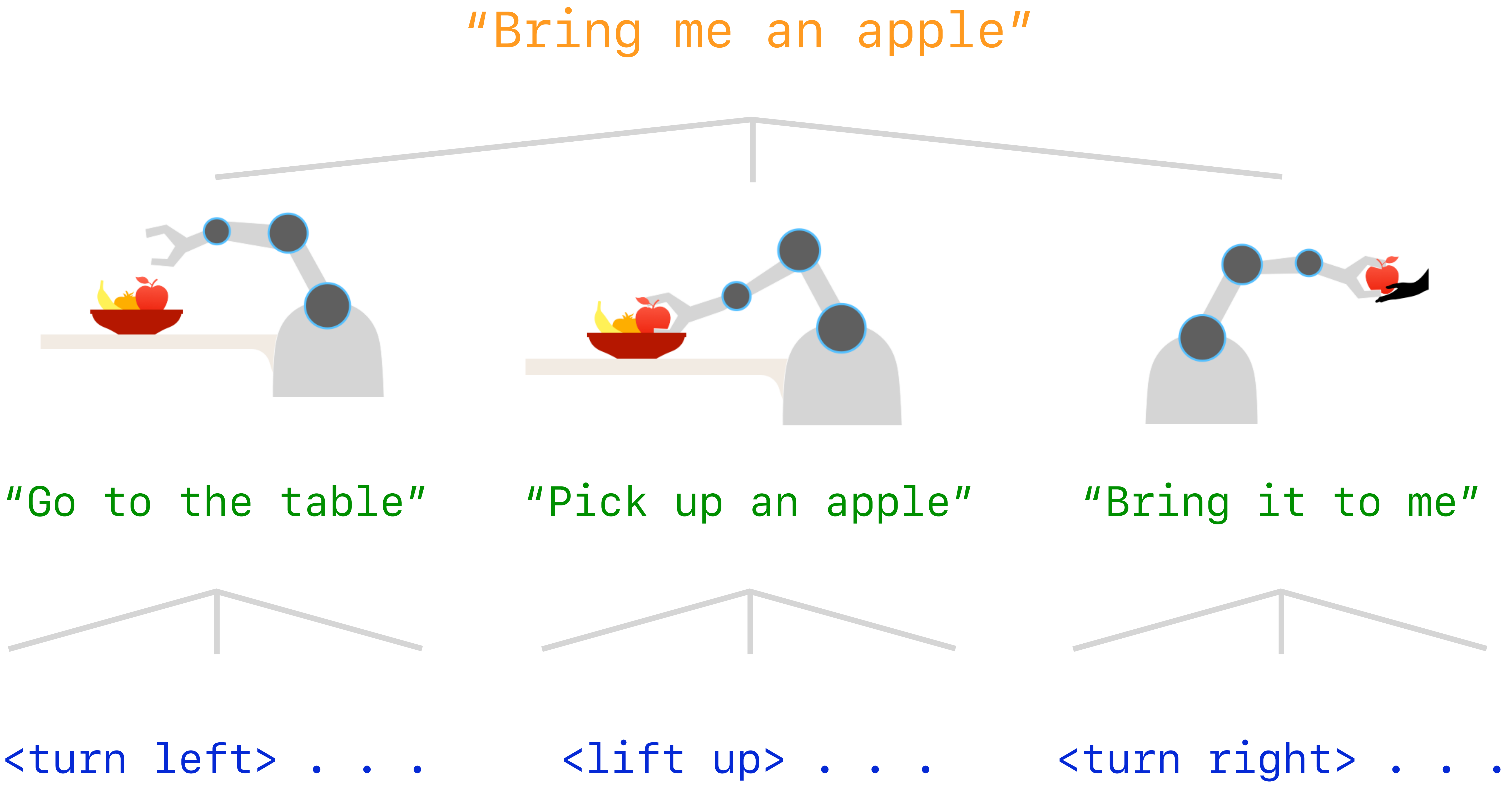
"Pick up an apple"

"Bring it to me"

<turn left> . . .

<lift up> . . .

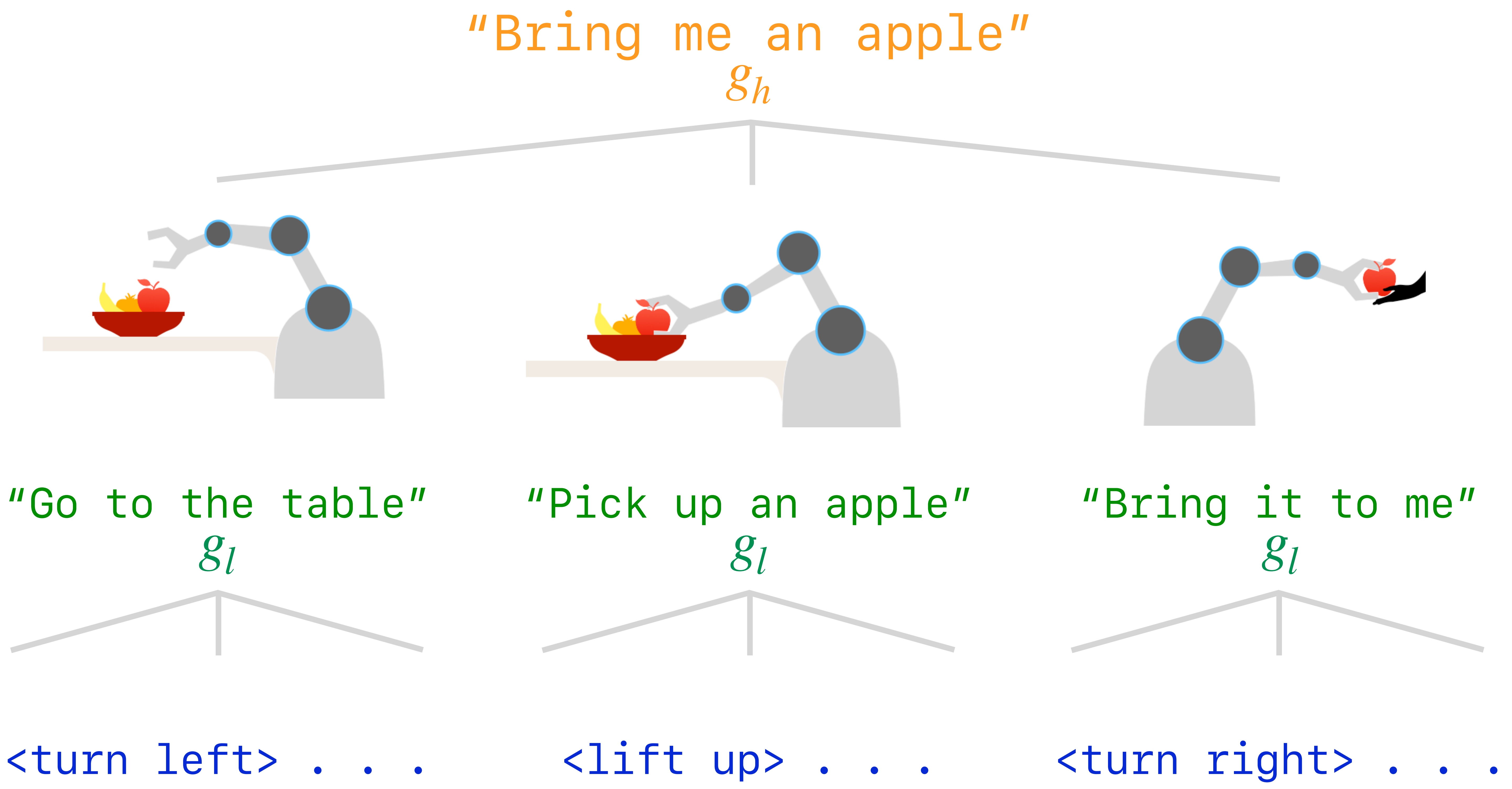
<turn right> . . .



“Bring me an apple”

“Go to the table” “Pick up an apple” “Bring it to me”

<turn left> . . . <lift up> . . . <turn right> . . .



“Bring me an apple”

g_h

“Go to the table”

g_l

“Pick up an apple”

g_l

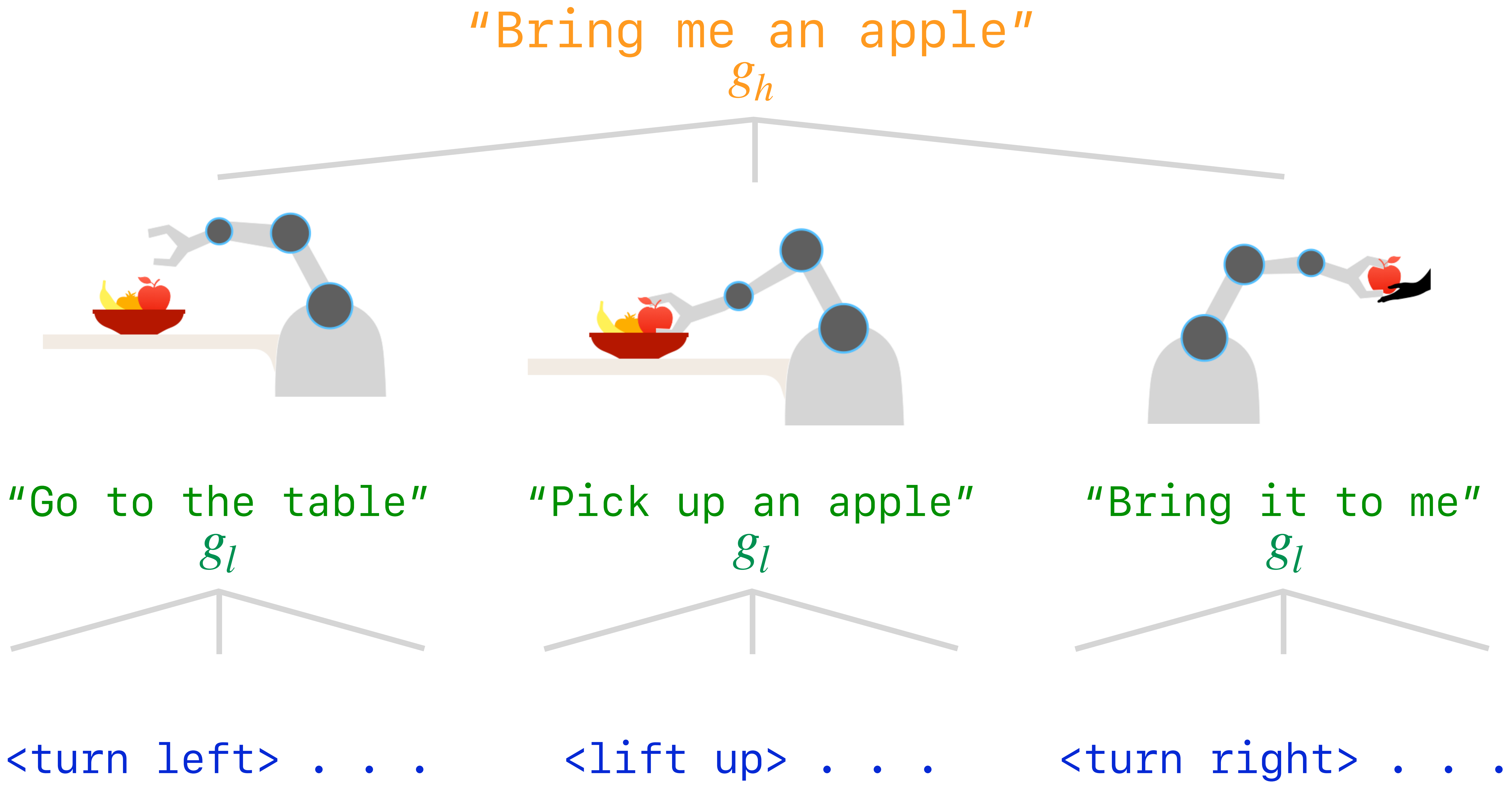
“Bring it to me”

g_l

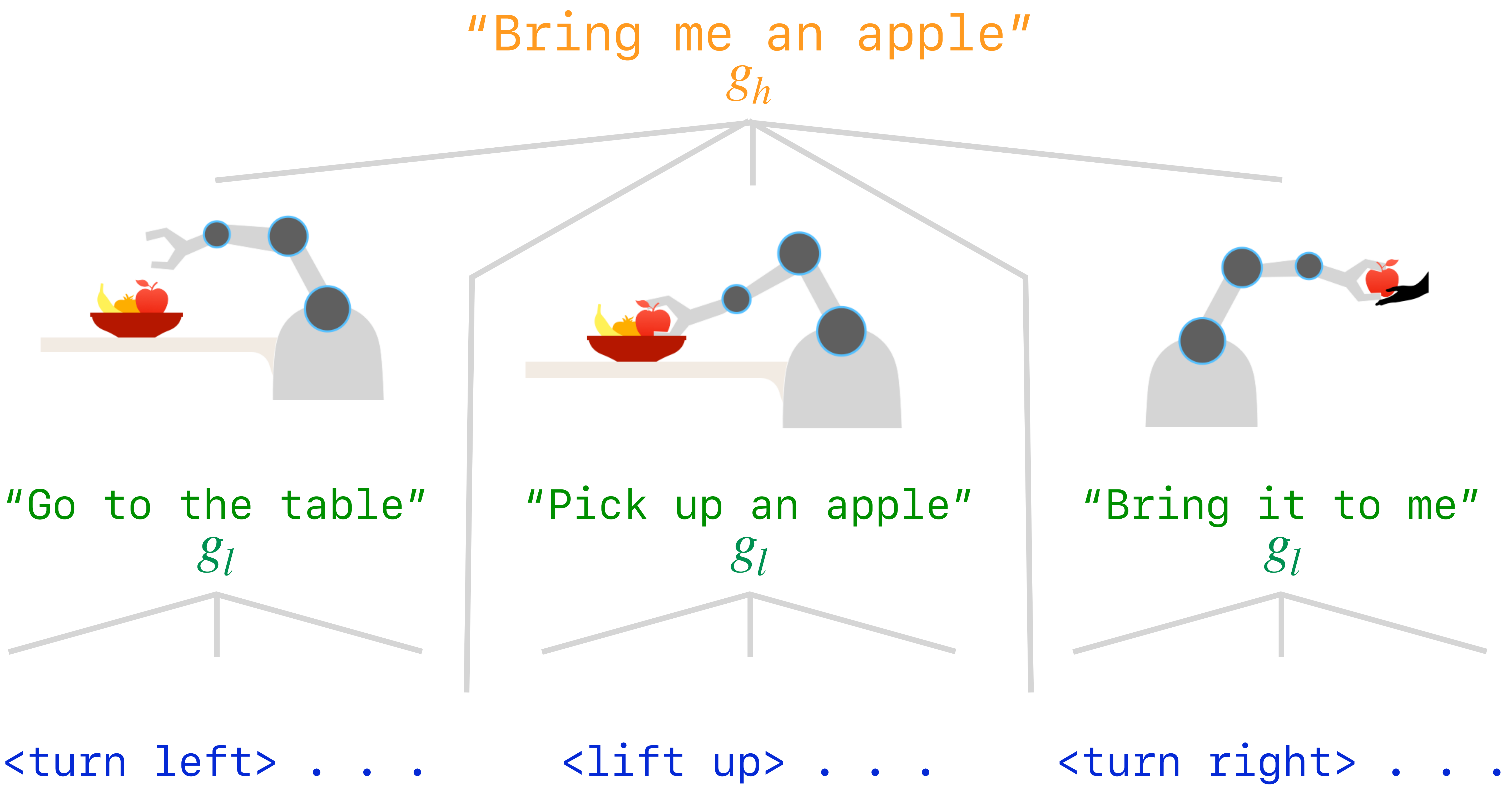
<turn left> . . .

<lift up> . . .

<turn right> . . .



- **Strict factorization**
  - Jiang et al. 2019
  - Das et al. 2018
  - **+ Explicit decomposition**
    - Andreas et al. 2017
    - Waytowich et al. 2019



“Bring me an apple”
g_h

“Go to the table”
g_l

“Pick up an apple”
g_l

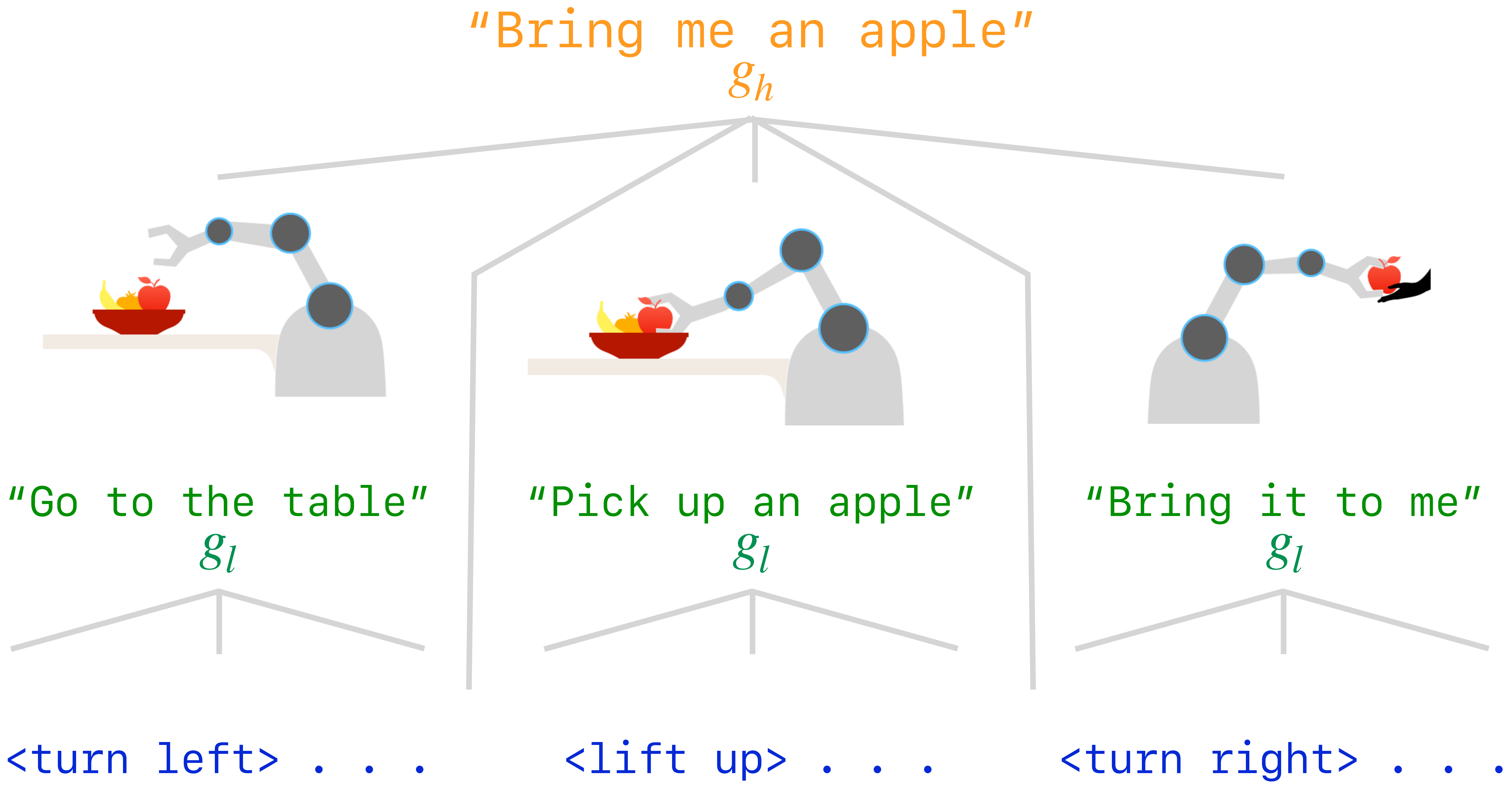
“Bring it to me”
g_l

`<turn left> . . .`

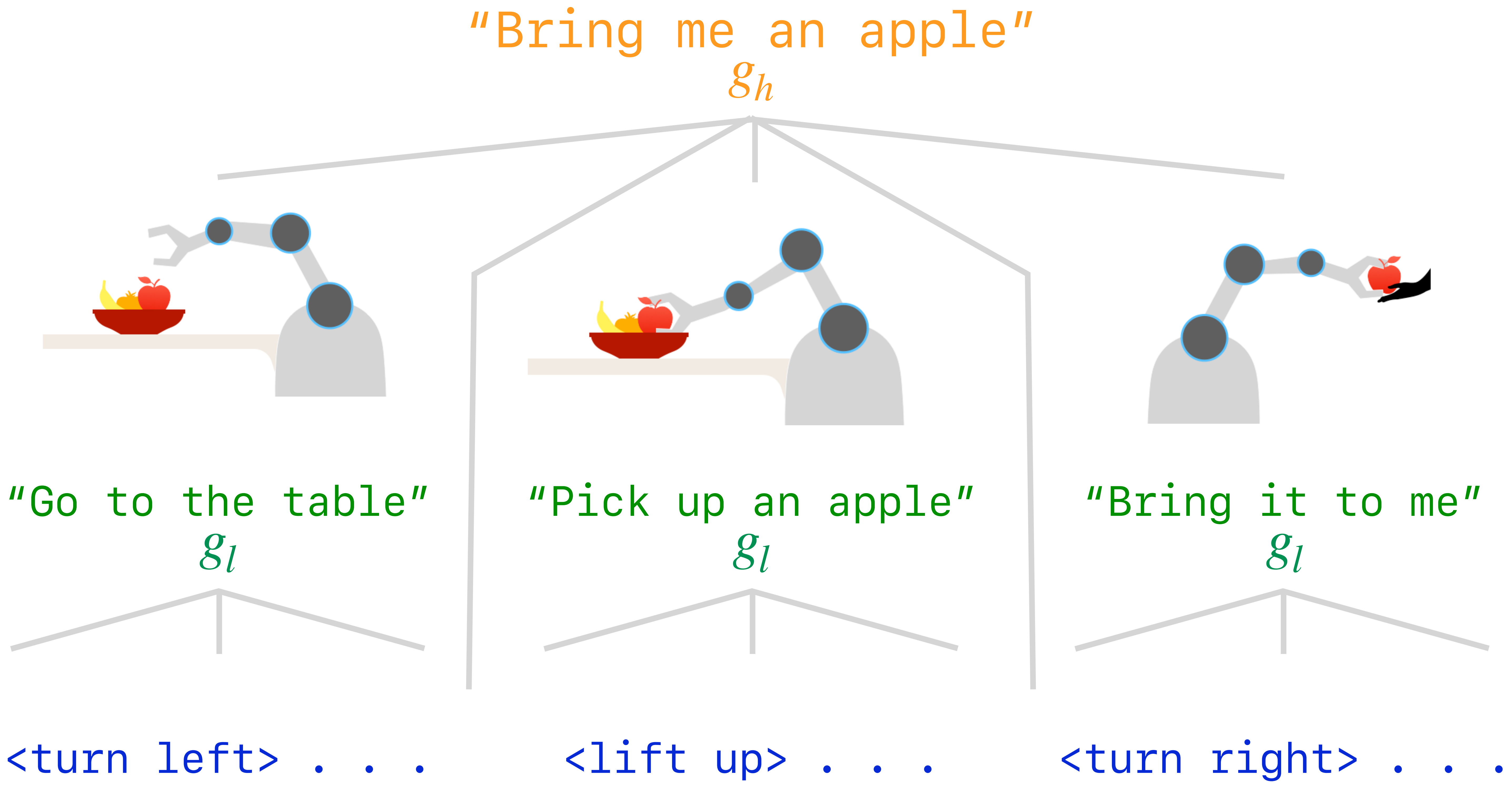
`<lift up> . . .`

`<turn right> . . .`

- **Strict factorization**
  - Jiang et al. 2019
  - Das et al. 2018
  - **+ Explicit decomposition**
    - Andreas et al. 2017
    - Waytowich et al. 2019

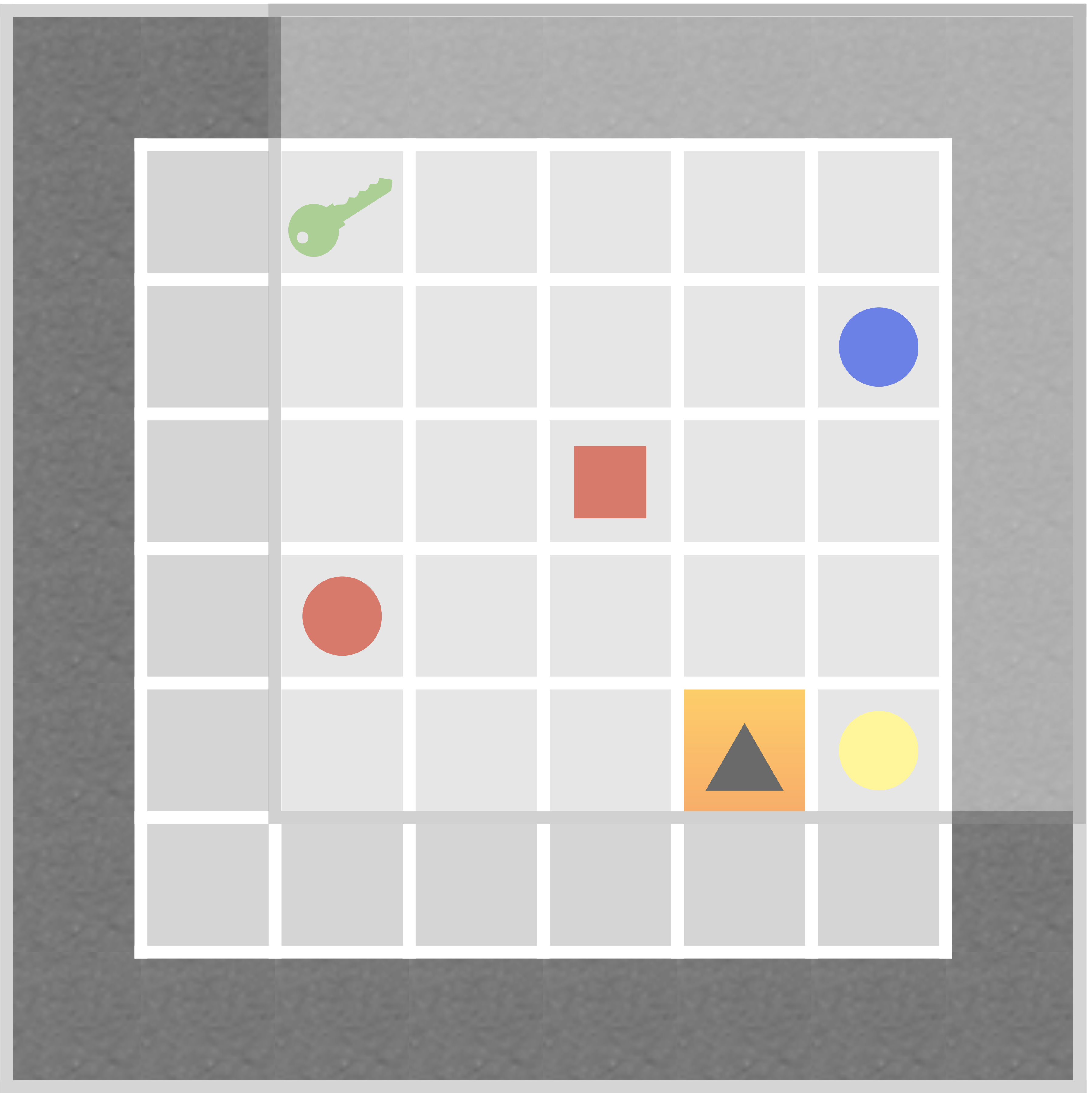


- **Strict factorization**
  - Jiang et al. 2019
  - Das et al. 2018
  - **+ Explicit decomposition**
    - Andreas et al. 2017
    - Waytowich et al. 2019
- **Non-abstractive**
  - Goyal et al. 2019



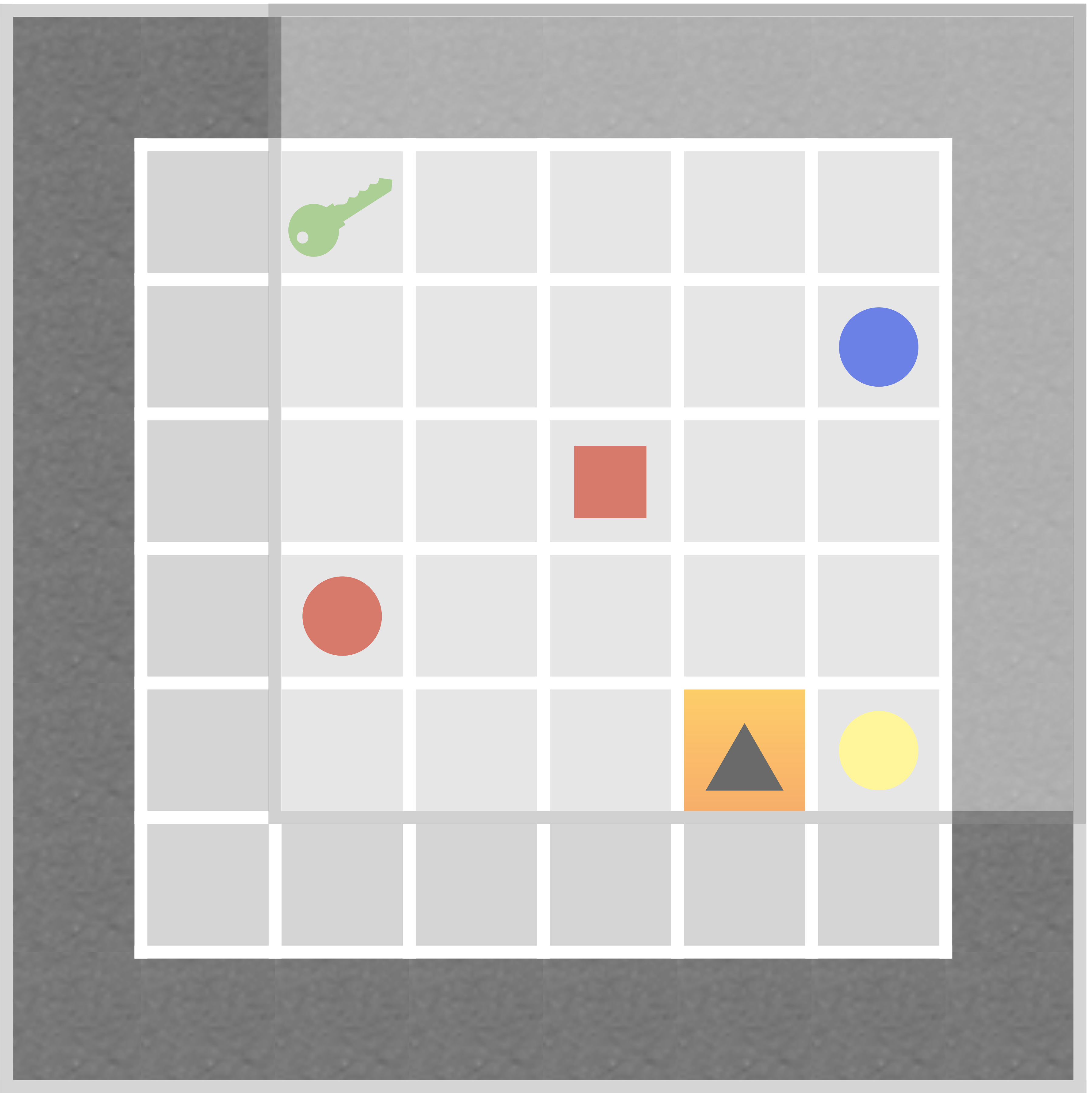
- **Strict factorization**
  - Jiang et al. 2019
  - Das et al. 2018
  - **+ Explicit decomposition**
    - Andreas et al. 2017
    - Waytowich et al. 2019
- **Non-abstractive**
  - Goyal et al. 2019
- **Intrinsic motivation & curiosity**
  - Burda et al. 2019

Formulation



$$(S, A, T, R, G, G_\ell, \gamma)$$

Formulation



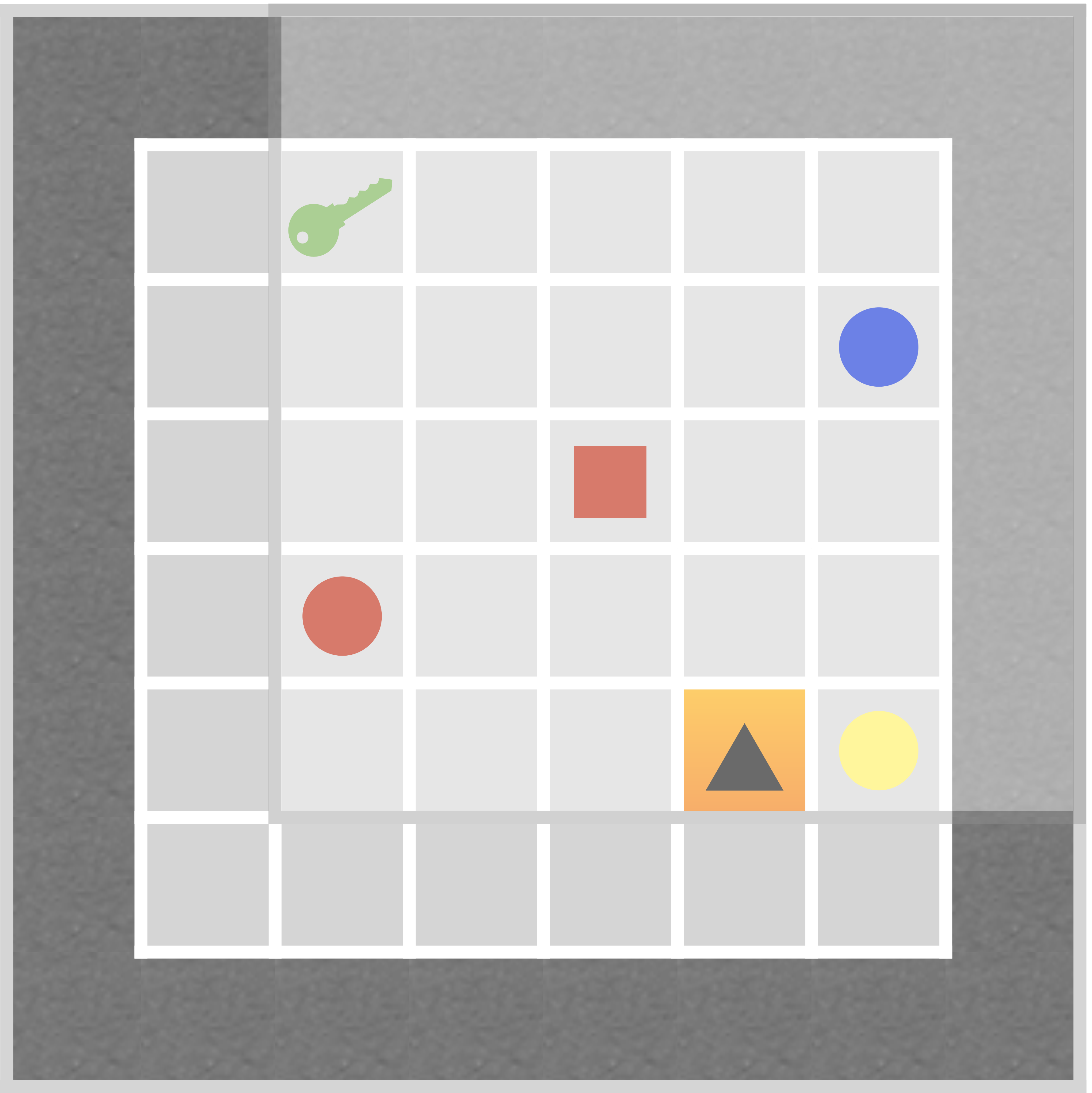
$$(S, A, T, R, G, G_\ell, \gamma)$$

A: Primitive Actions

`<turn left>`
`<turn right>`
`<move forward>`

`<pick up>`
`<put down>`
`<toggle>`
`<done>`

Formulation



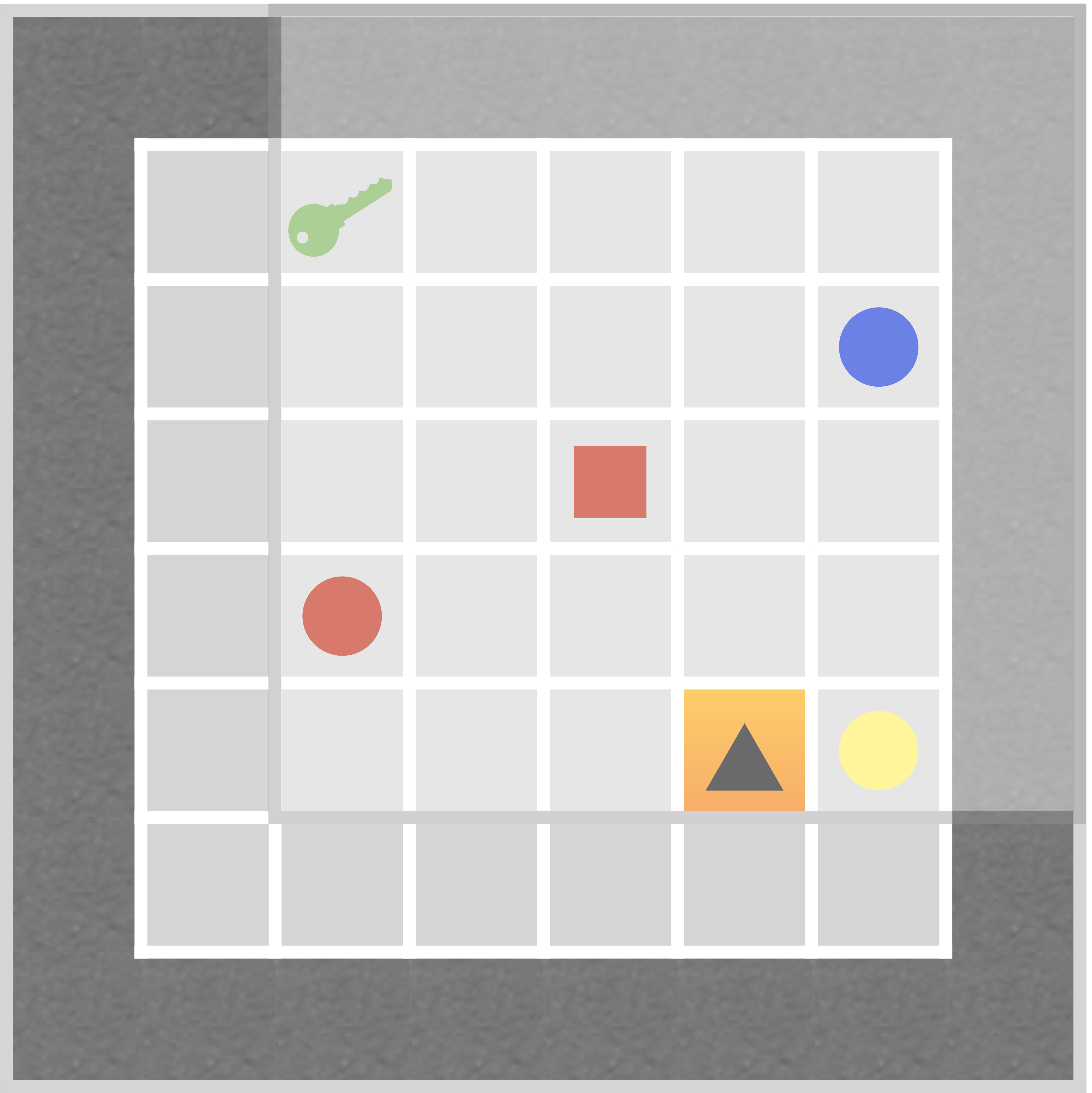
$$(S, A, T, R, G, G_\ell, \gamma)$$

A: Primitive Actions

<turn left>
<turn right>
<move forward>

<pick up>
<put down>
<toggle>
<done>

Formulation



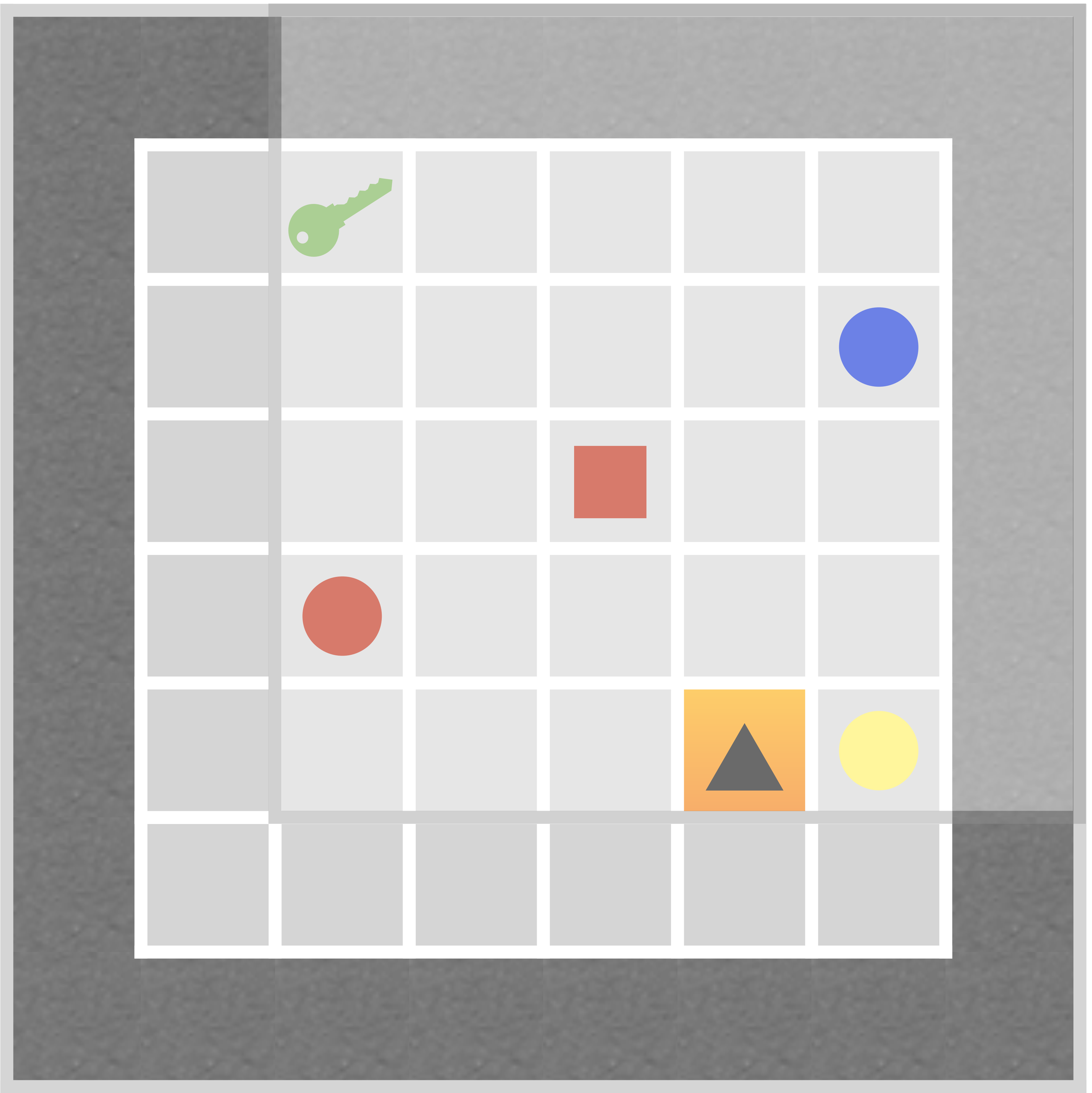
$$(S, A, T, R, G, G_\ell, \gamma)$$

A: Primitive Actions

```
<turn left>
<turn right>
<move forward>
```

```
<pick up>
<put down>
<toggle>
<done>
```

Formulation



$$(S, A, T, R, G, G_\ell, \gamma)$$

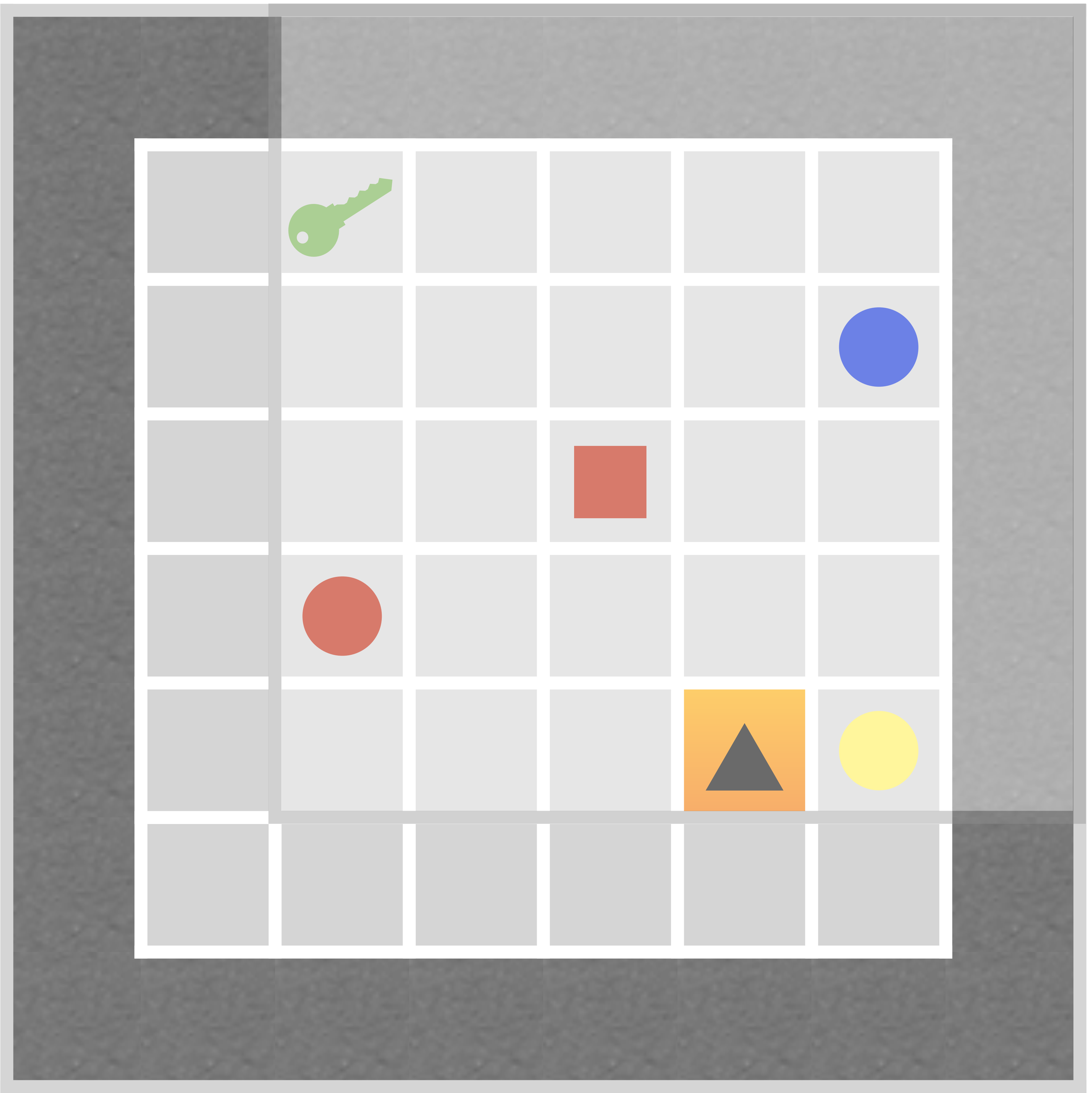
G: High-Level Instructions

```
“go to the red ball and
then to the blue ball”

“put the red ball next to
the blue ball”
```

A: Primitive Actions

```
<turn left>
<turn right>
<move forward>

<pick up>
<put down>
<toggle>
<done>
```

Formulation



$$(S, A, T, R, G, G_\ell, \gamma)$$

G: High-Level Instructions

“go to the red ball and then to the blue ball”

“put the red ball next to the blue ball”

G_ℓ: Low-Level Instructions

“go to the red ball”

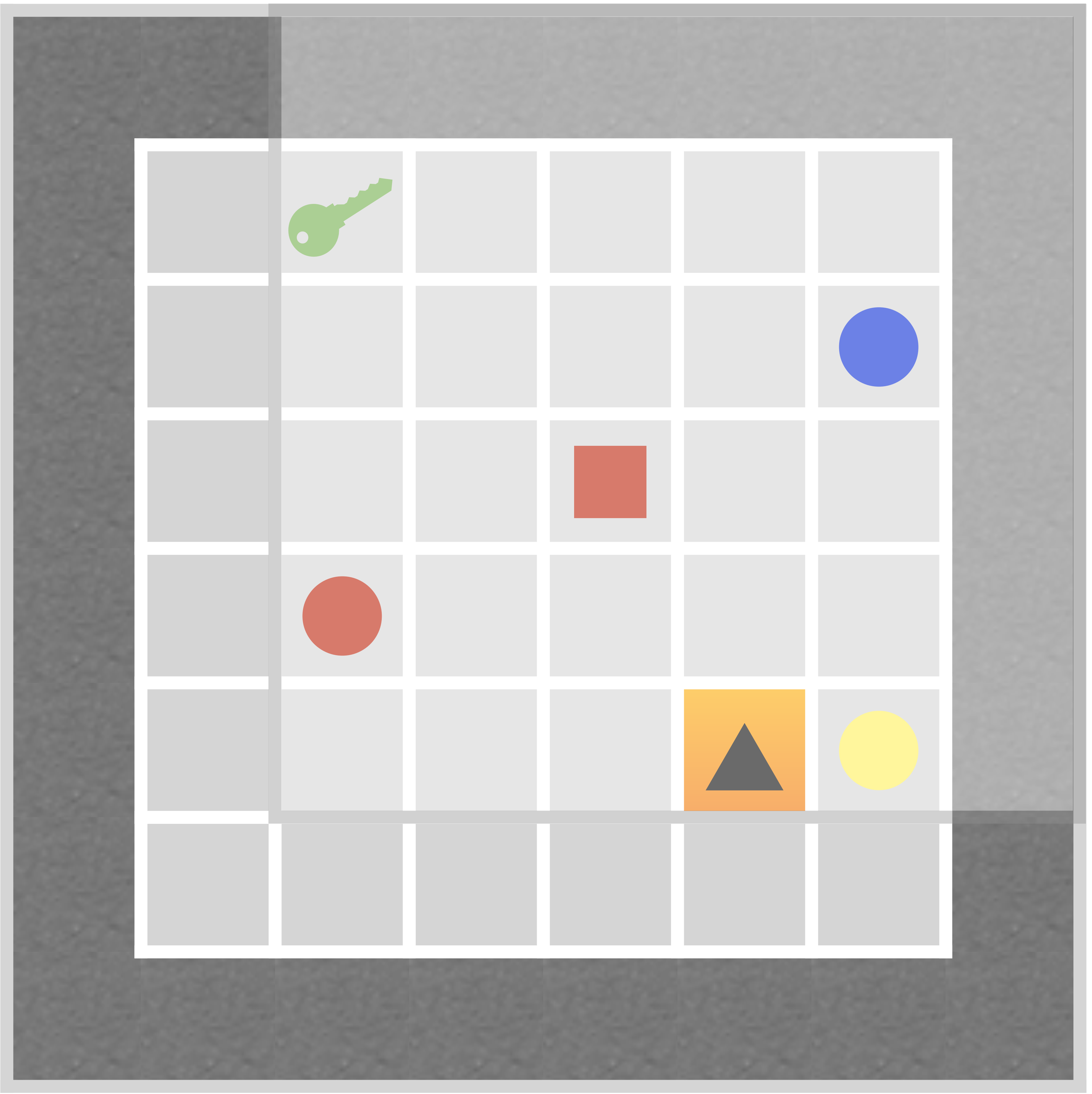
“go to the blue ball”

A: Primitive Actions

<turn left>
<turn right>
<move forward>

<pick up>
<put down>
<toggle>
<done>

Formulation



$$(S, A, T, R, G, G_\ell, \gamma)$$

G: High-Level Instructions

"go to the red ball and then to the blue ball"

"put the red ball next to the blue ball"

G_ℓ: Low-Level Instructions

"go to the red ball"

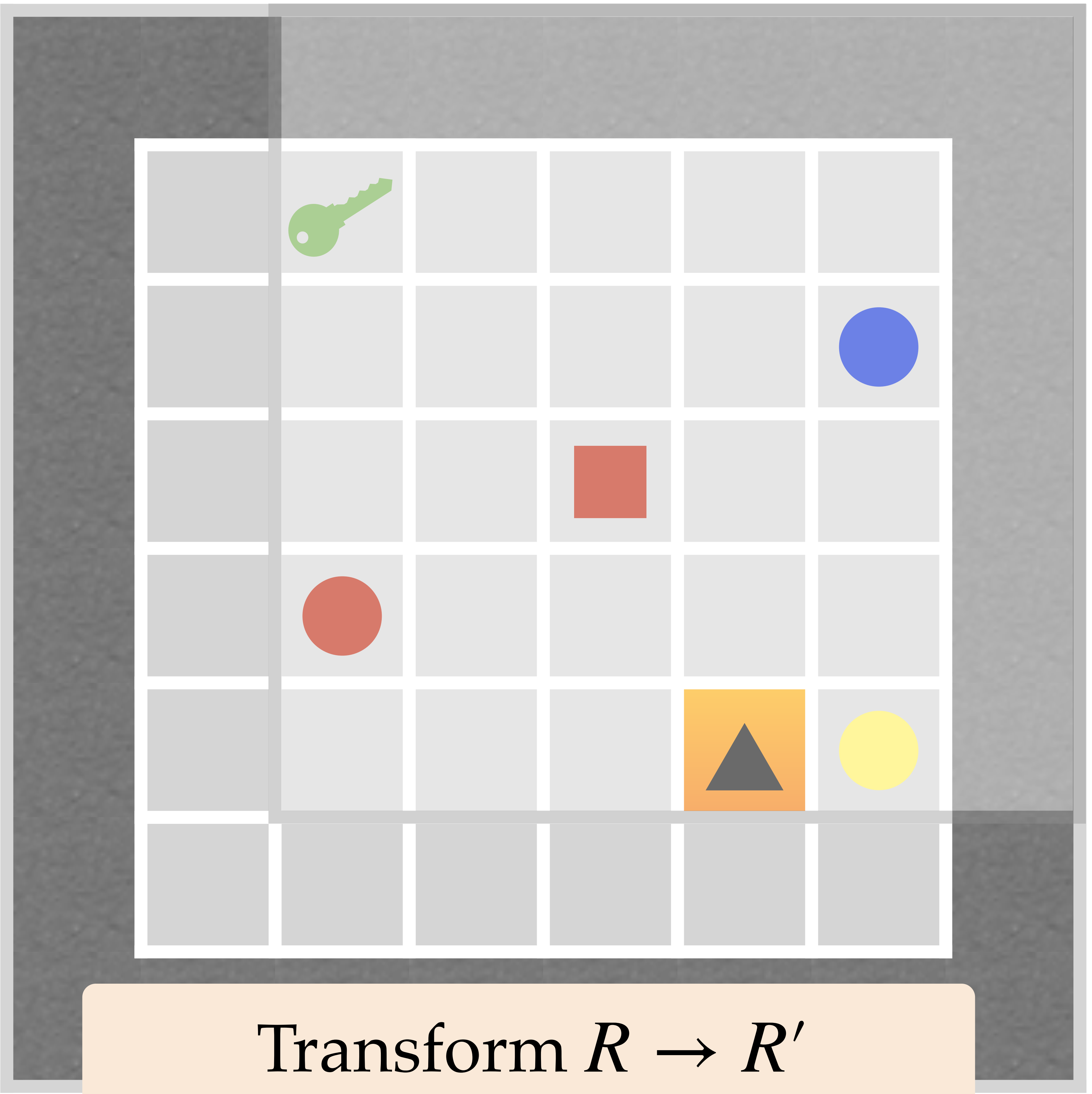
"go to the blue ball"

A: Primitive Actions

<turn left>
<turn right>
<move forward>

<pick up>
<put down>
<toggle>
<done>

Formulation



Transform $R \rightarrow R'$

$$(S, A, T, R, G, G_\ell, \gamma)$$

G: High-Level Instructions

“go to the red ball and then to the blue ball”

“put the red ball next to the blue ball”

G_ℓ: Low-Level Instructions

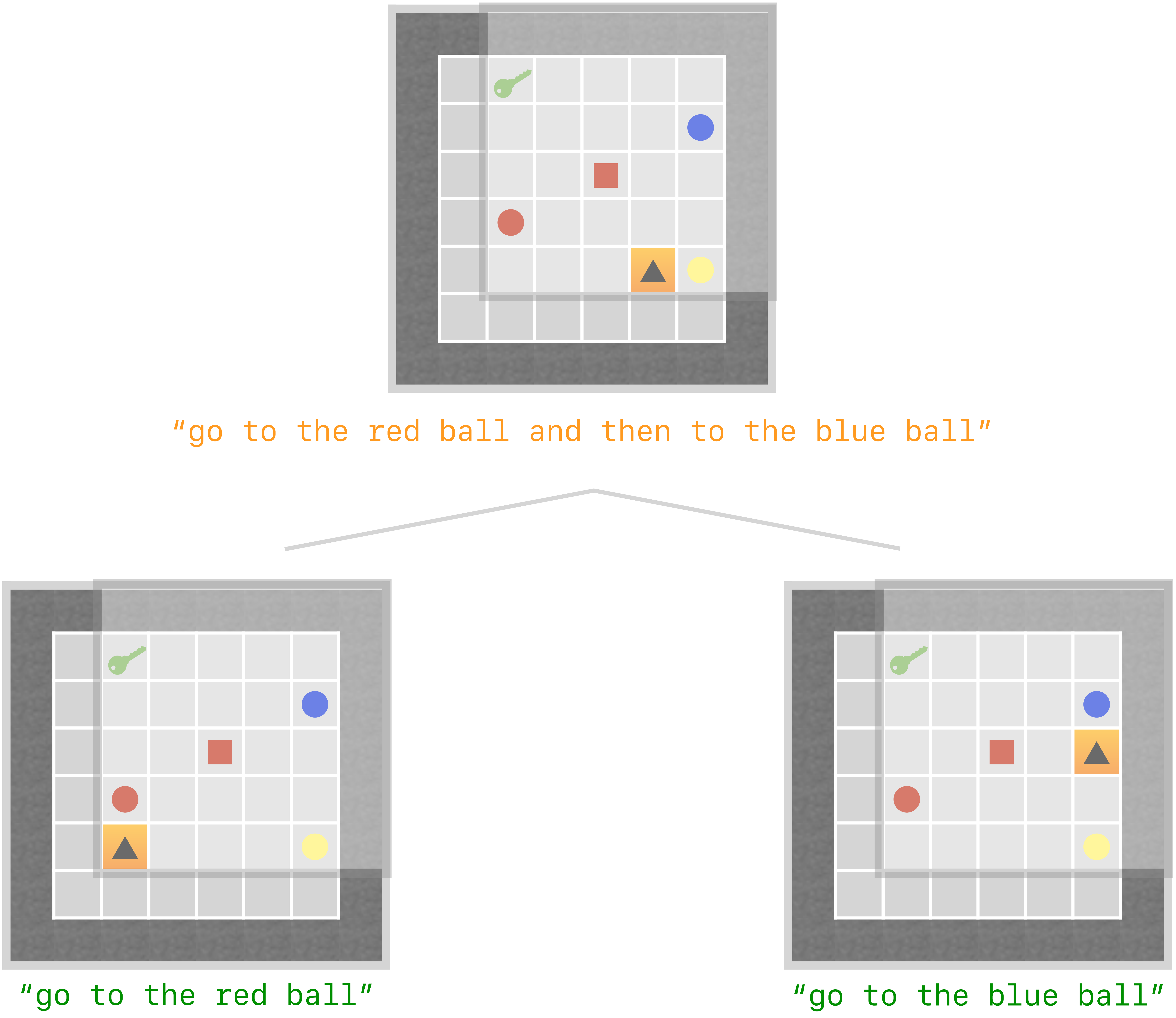
“go to the red ball”

“go to the blue ball”

A: Primitive Actions

<turn left>
<turn right>
<move forward>

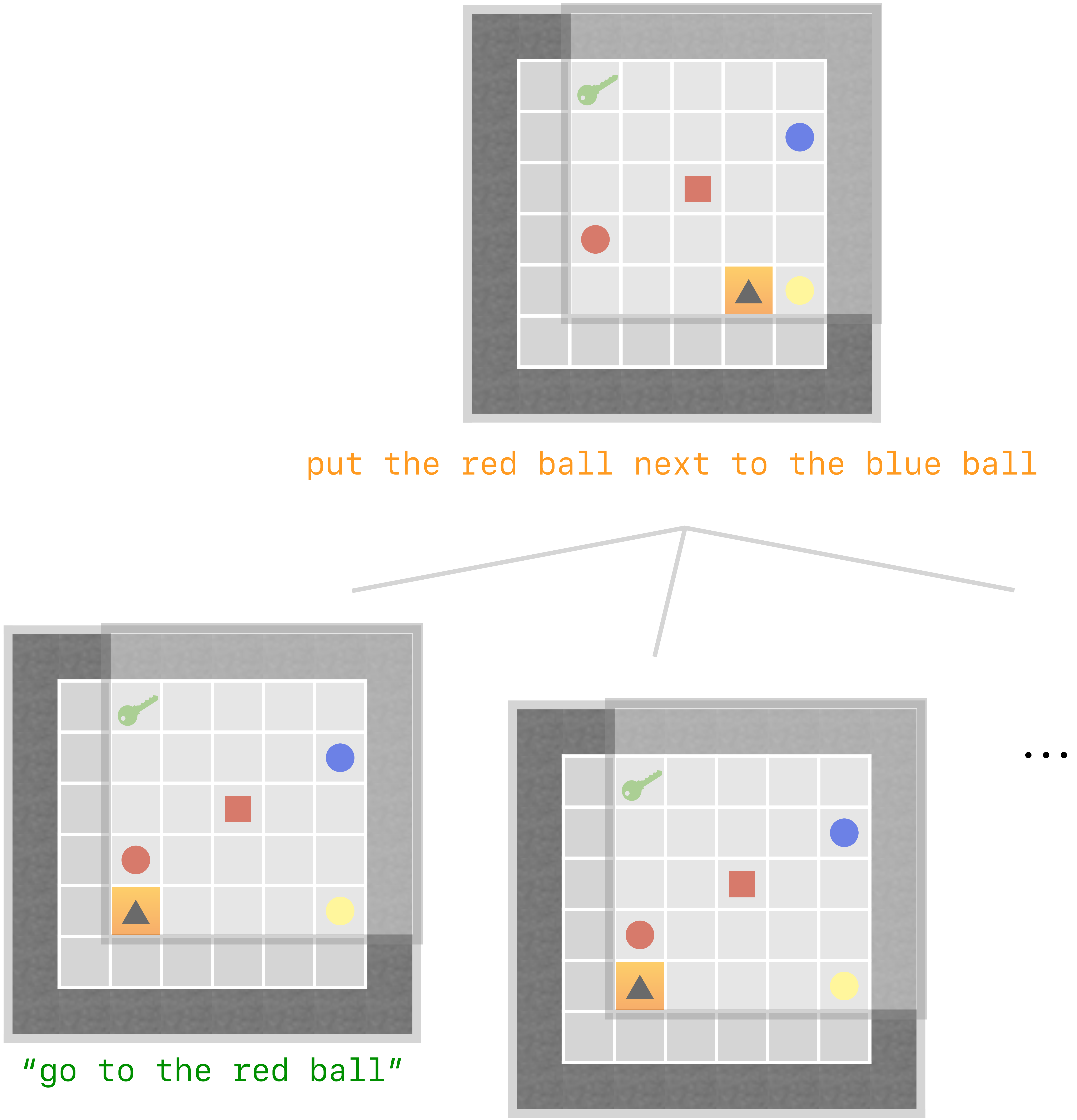
<pick up>
<put down>
<toggle>
<done>



“go to the red ball and then to the blue ball”

“go to the red ball”

“go to the blue ball”



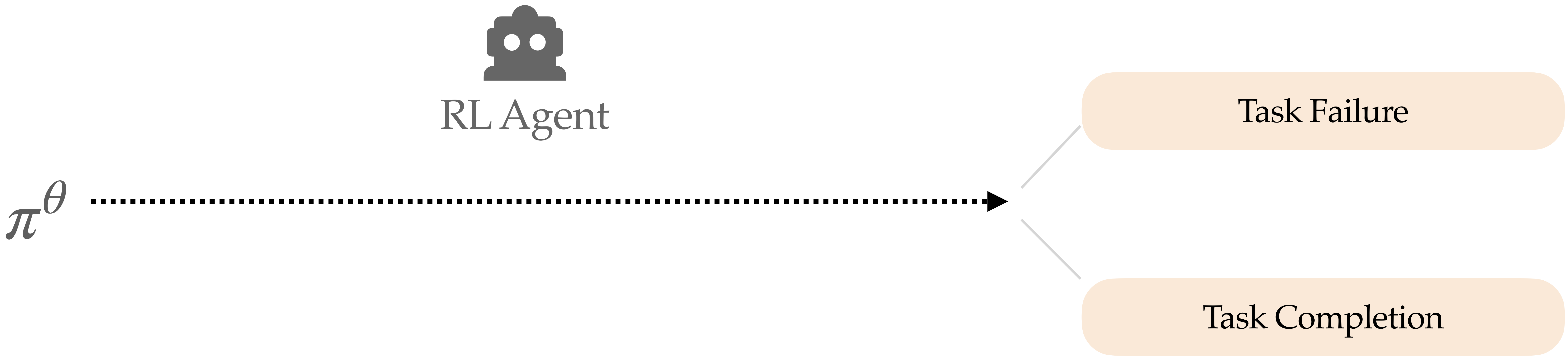
put the red ball next to the blue ball

"go to the red ball"

<pick up>

...

ELLA



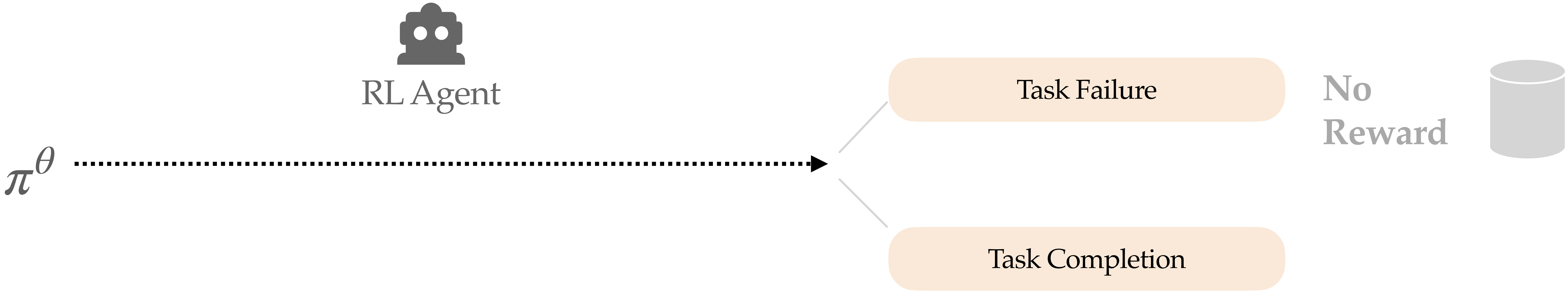
RL Agent

π^{θ}

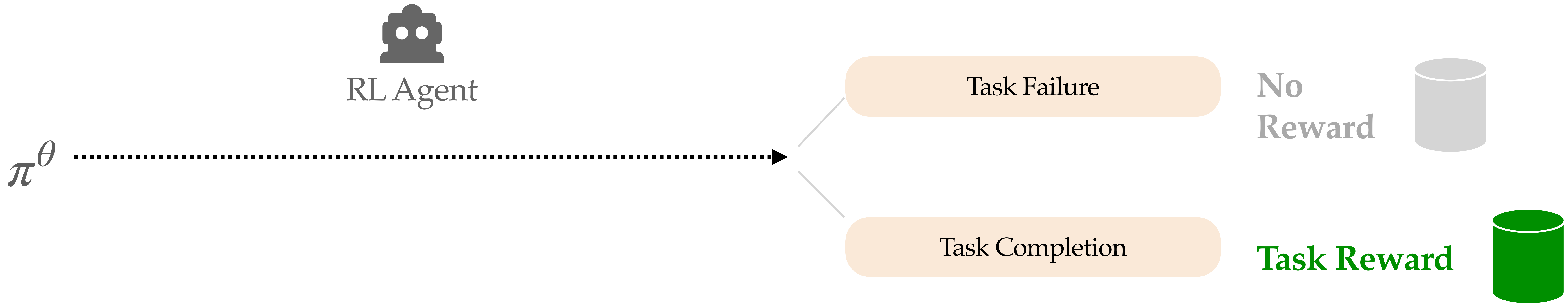
Task Failure

Task Completion

ELLA



ELLA



RL Agent

π^{θ}

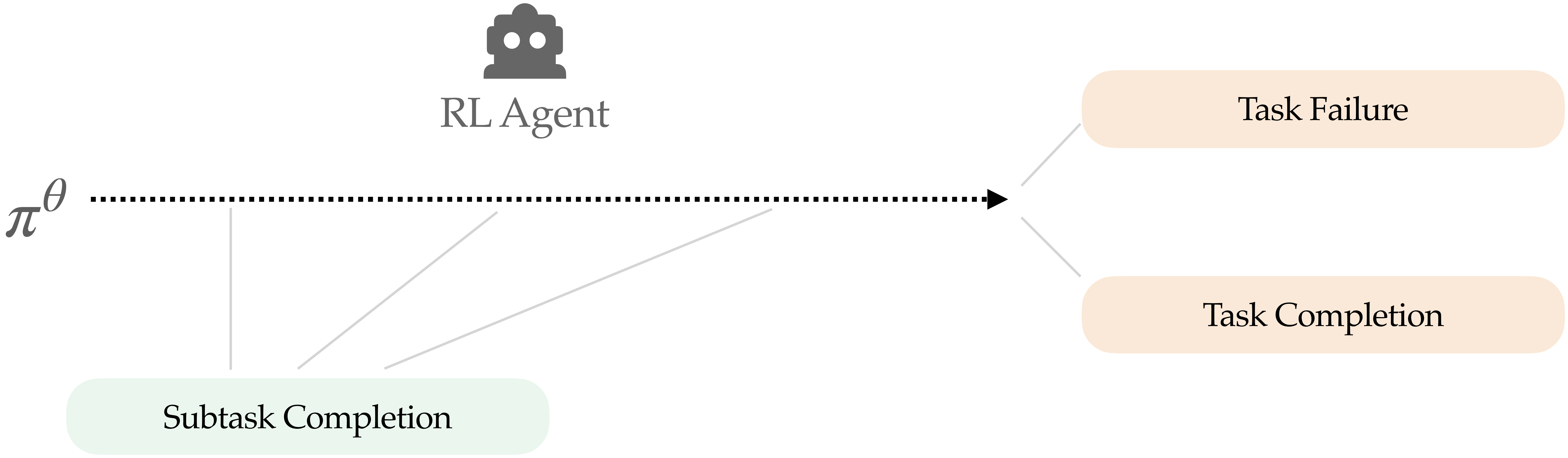
Task Failure

No Reward

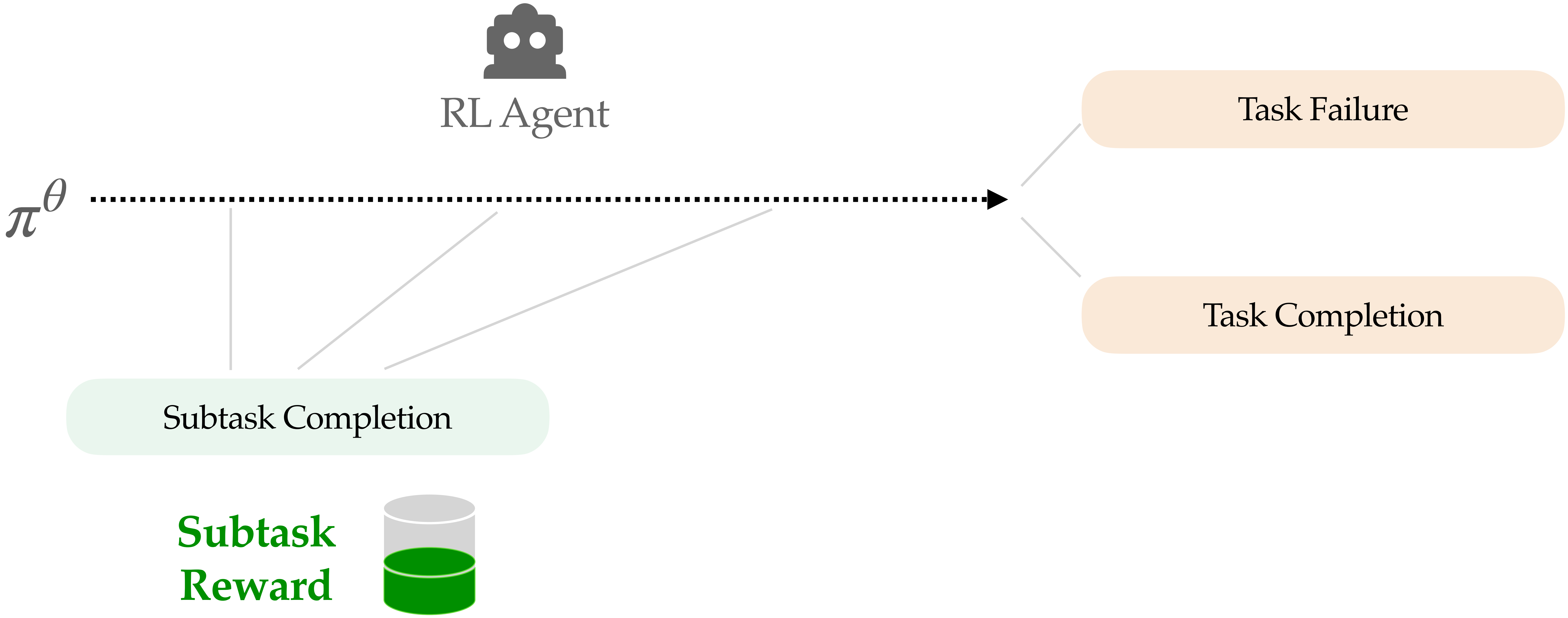
Task Completion

Task Reward

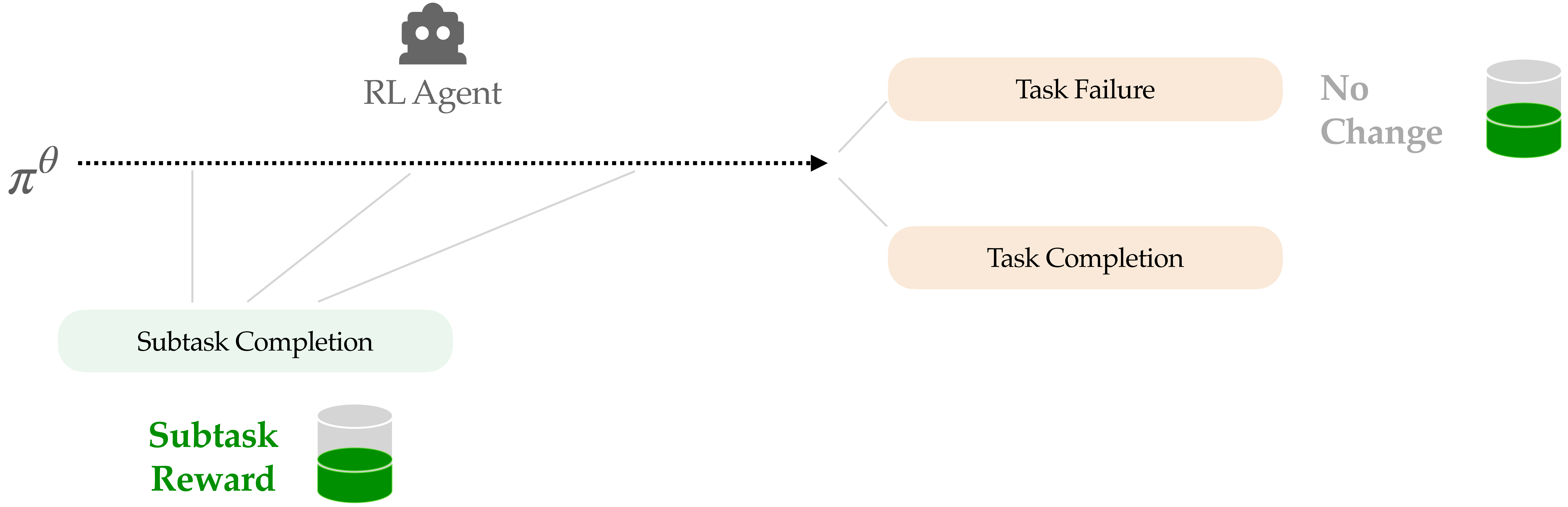
ELLA


RL Agent
π^{θ}
Task Failure
Task Completion
Subtask Completion

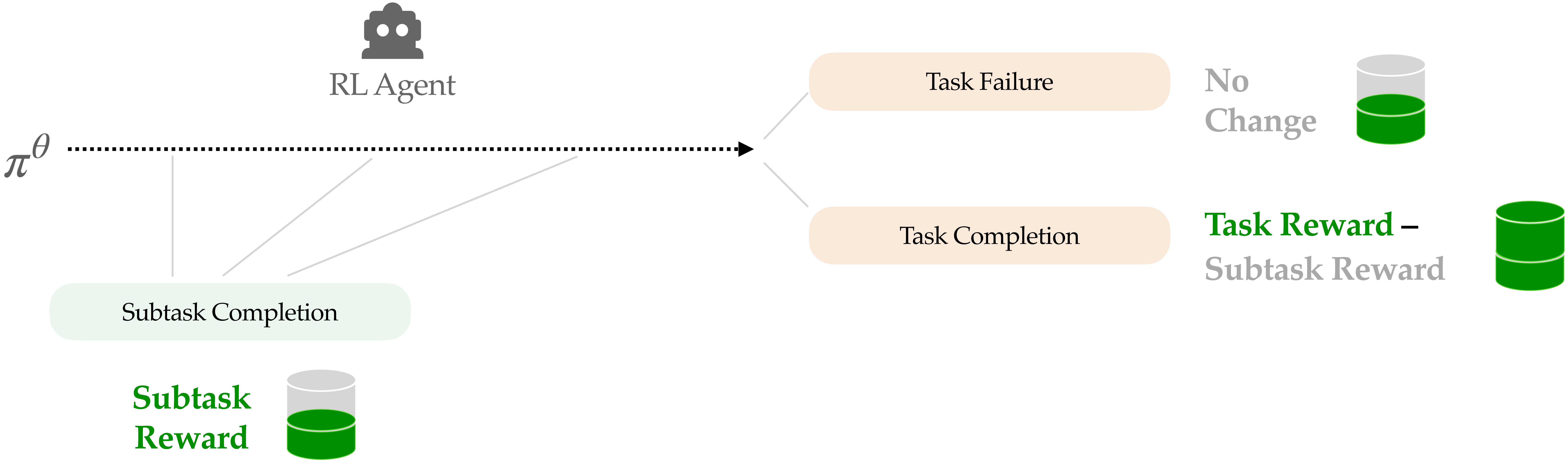
ELLA


RL Agent
π^θ
Task Failure
Task Completion
Subtask Completion
Subtask
Reward

ELLA


RL Agent
π^θ
Task Failure
No Change
Task Completion
Subtask Completion
Subtask Reward

ELLA


RL Agent

π^θ

Task Failure

No Change

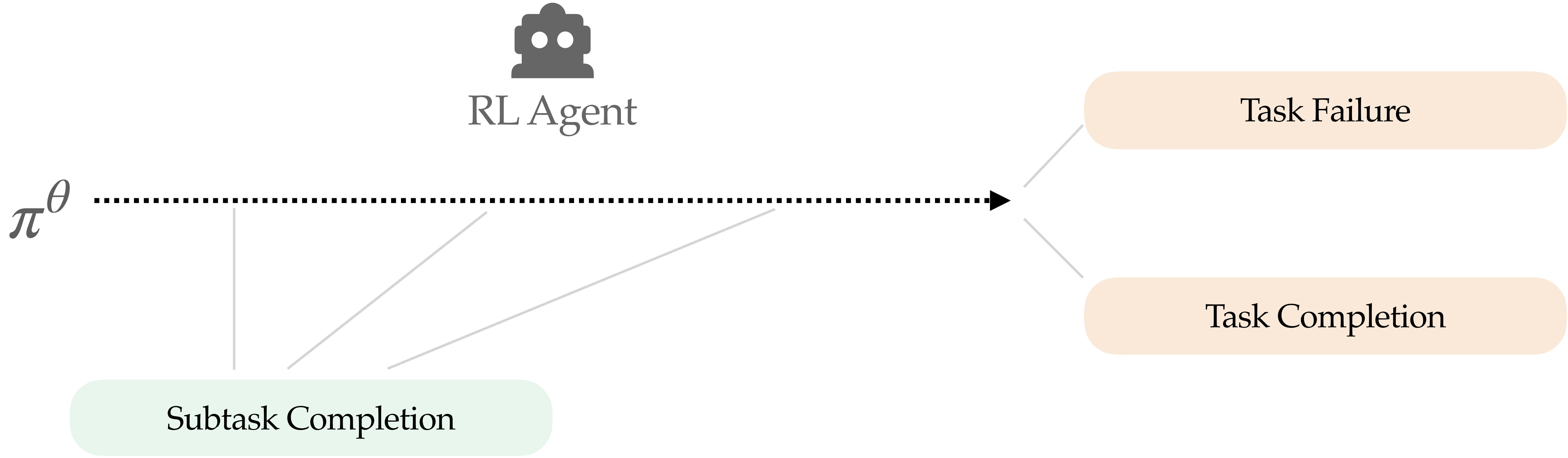
Task Completion

Task Reward – **Subtask Reward**

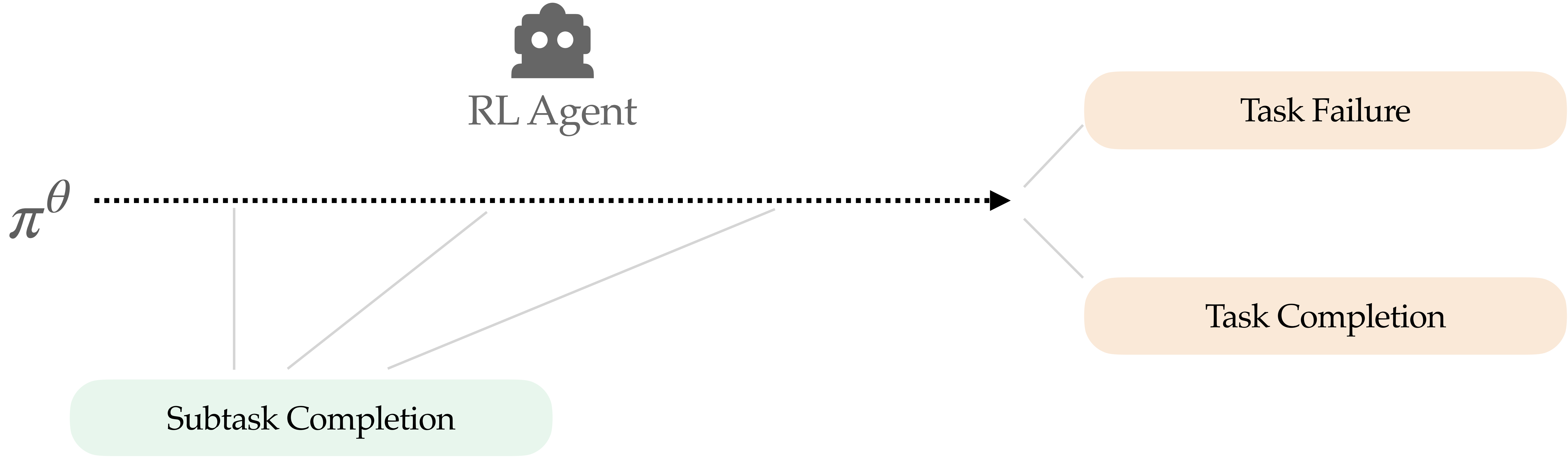
Subtask Completion

Subtask Reward

ELLA

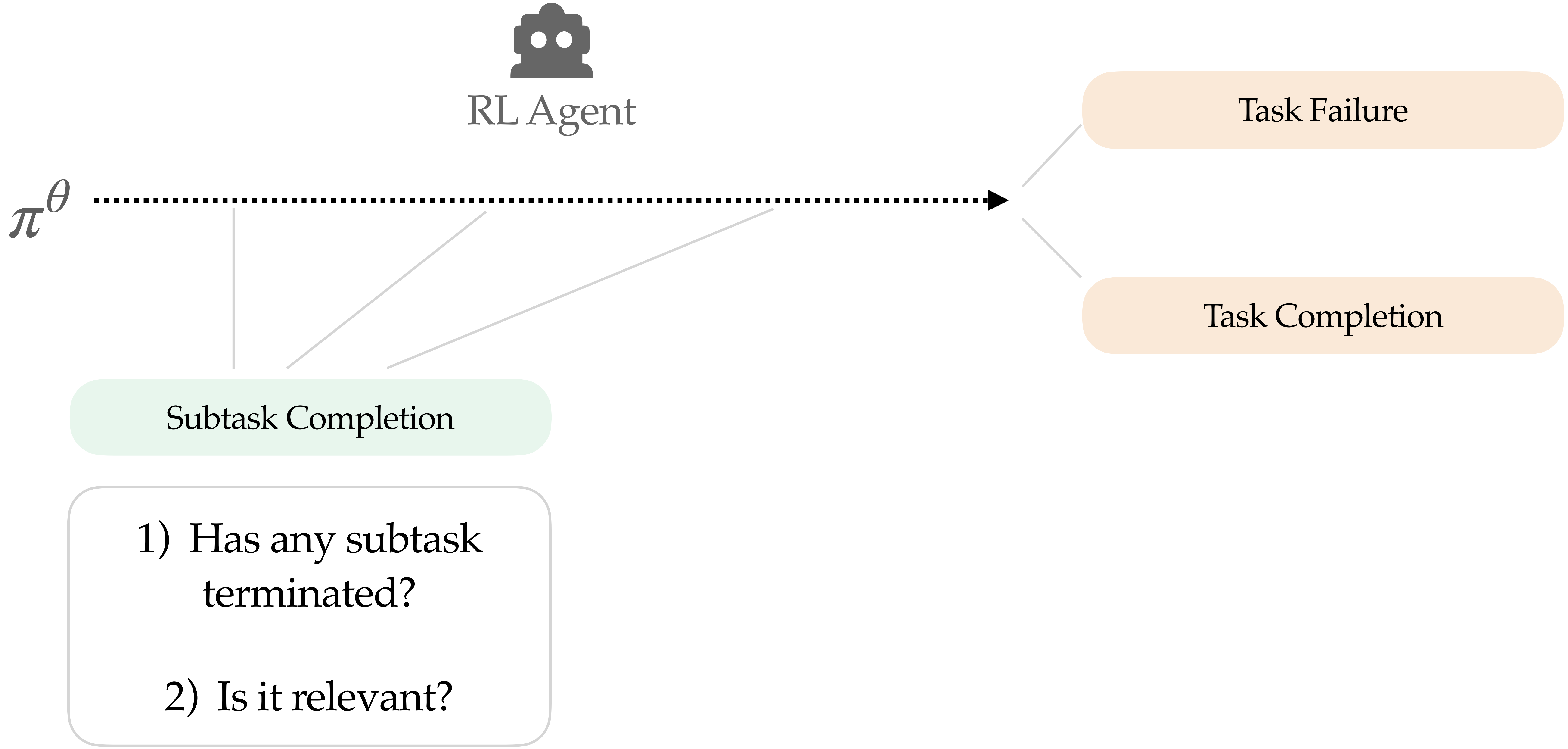

RL Agent
π^θ
Task Failure
Task Completion
Subtask Completion

ELLA



1) Has any subtask terminated?

ELLA



RL Agent

π^θ

Task Failure

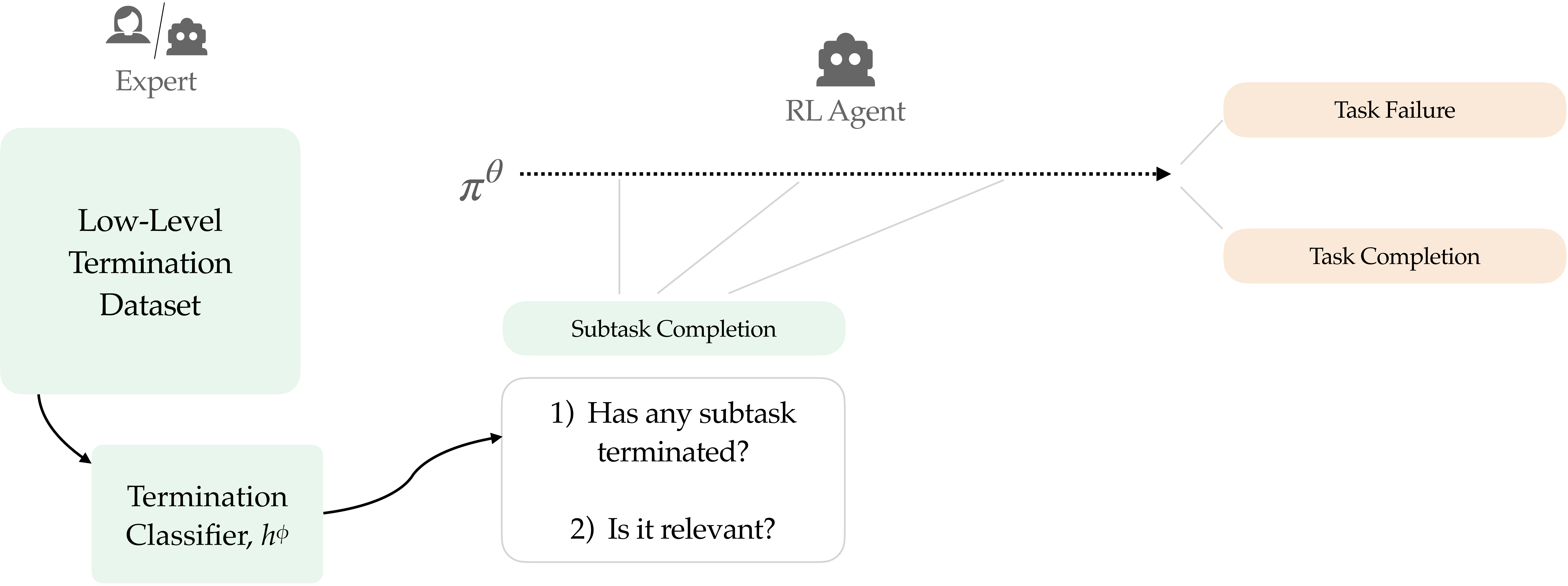
Task Completion

Subtask Completion

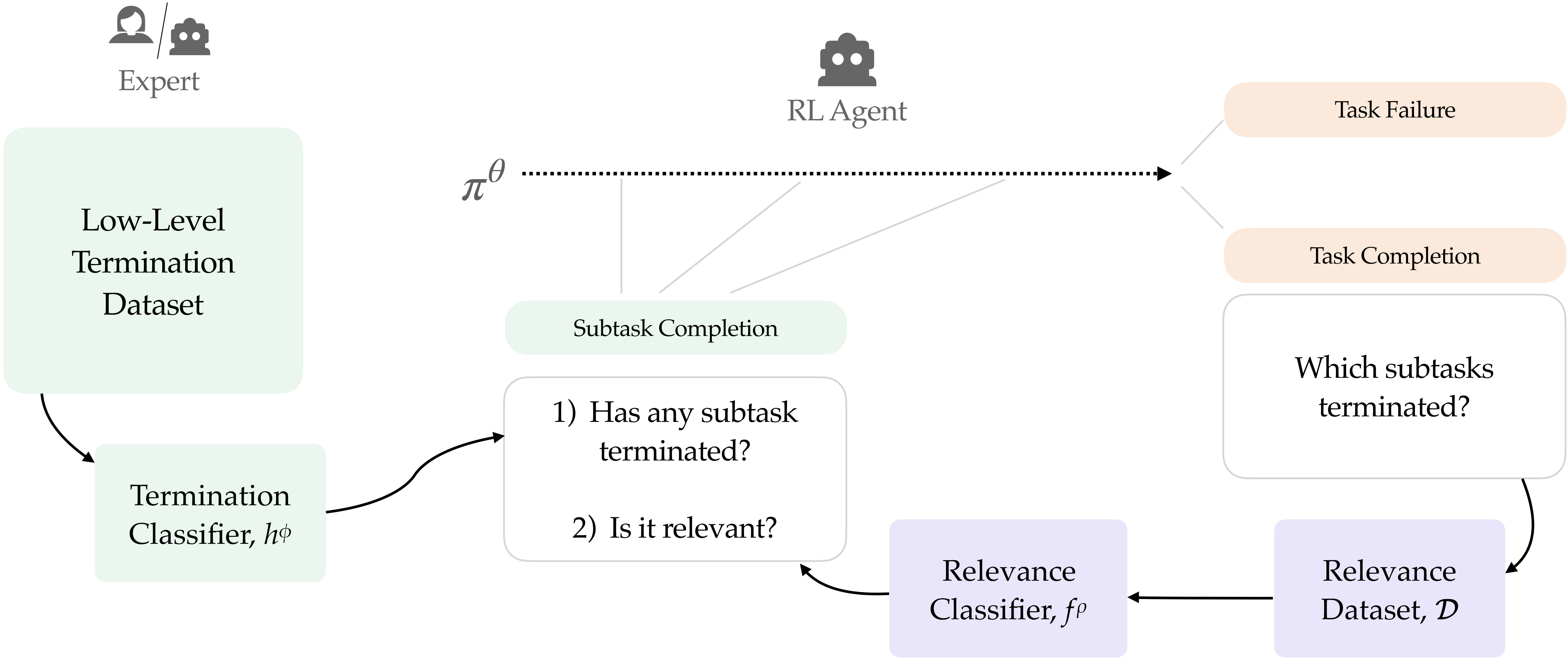
1) Has any subtask terminated?

2) Is it relevant?

ELLA


Expert
Low-Level Termination Dataset
Termination Classifier, h^ϕ
RL Agent
π^θ
Task Failure
Task Completion
Subtask Completion
1) Has any subtask terminated?
2) Is it relevant?

ELLA



Expert

Low-Level Termination Dataset

Termination Classifier, h^{ϕ}

RL Agent

π^{θ}

Task Failure

Task Completion

Subtask Completion

1) Has any subtask terminated?

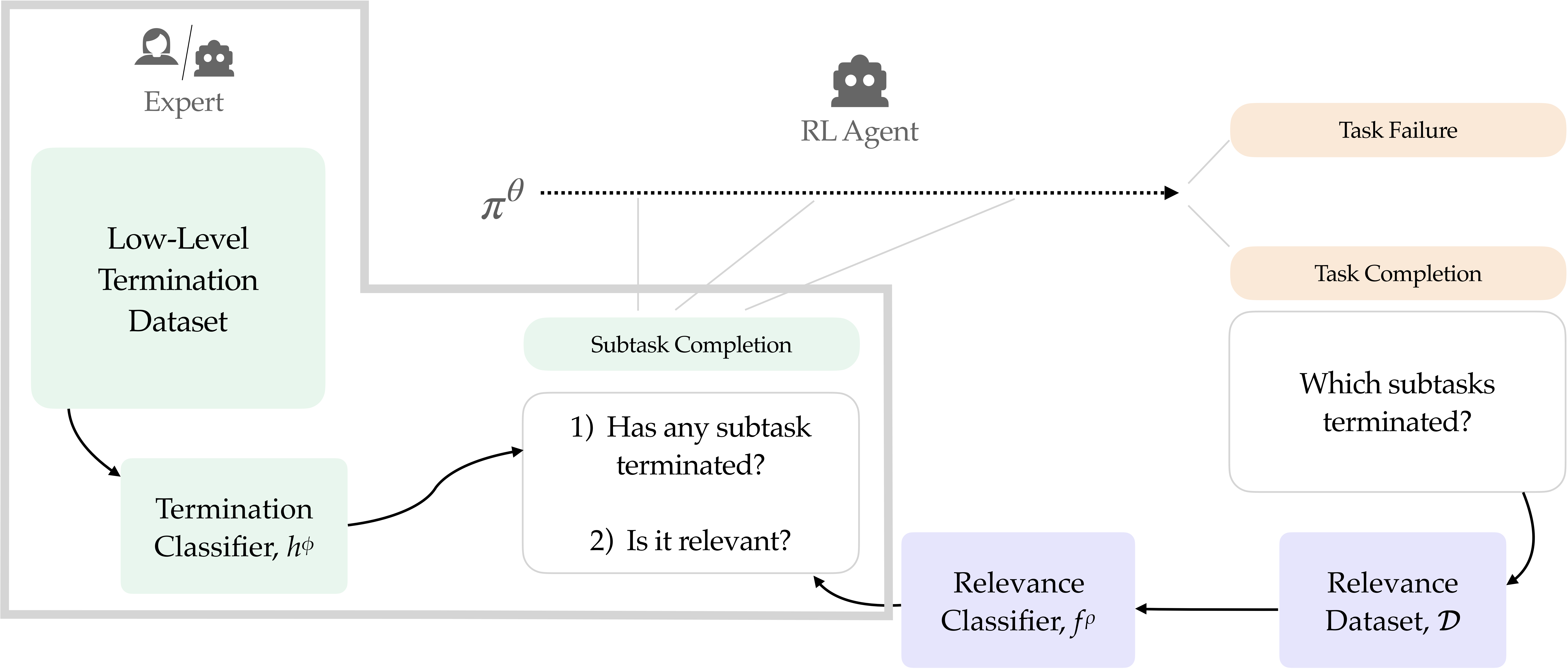
2) Is it relevant?

Which subtasks terminated?

Relevance Classifier, f^{ρ}

Relevance Dataset, $\mathcal{D}$

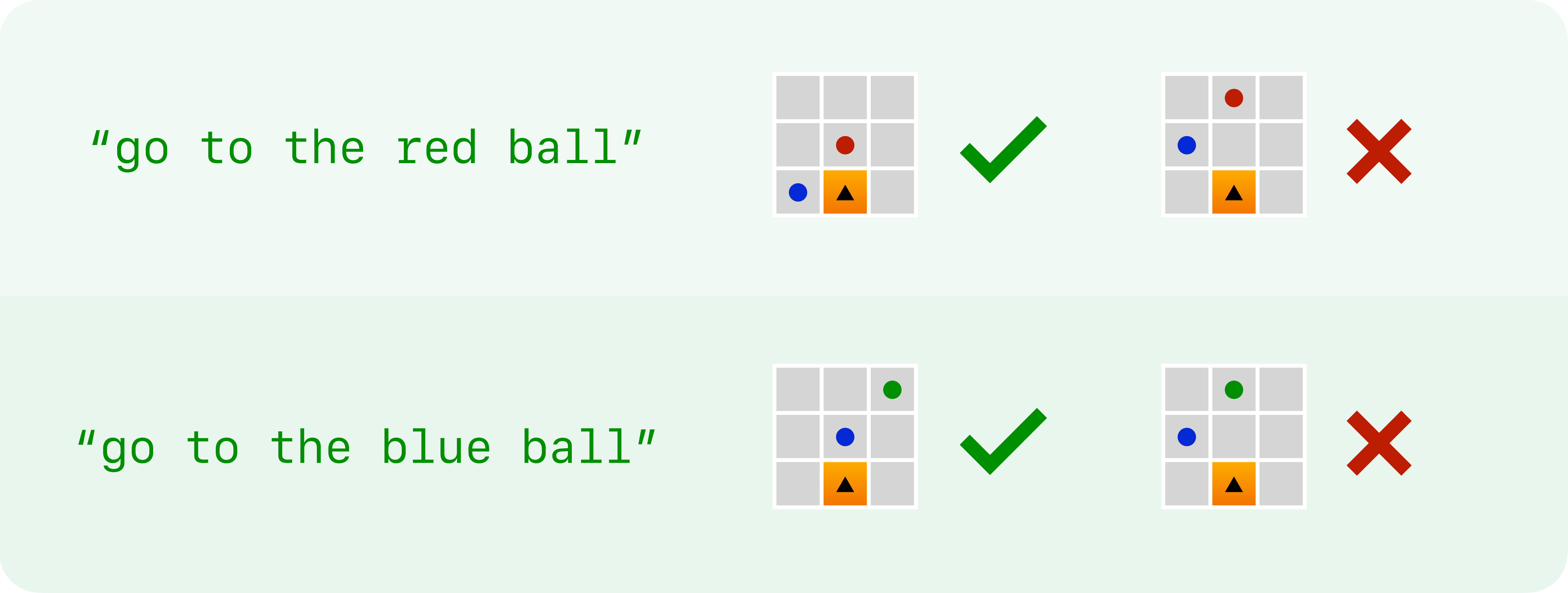
ELLA



Expert
RL Agent
π^{θ}
Task Failure
Task Completion
Low-Level Termination Dataset
Subtask Completion
1) Has any subtask terminated?
2) Is it relevant?
Which subtasks terminated?
Termination Classifier, h^{ϕ}
Relevance Classifier, f^{ρ}
Relevance Dataset, $\mathcal{D}$

Low-Level Termination Classifier

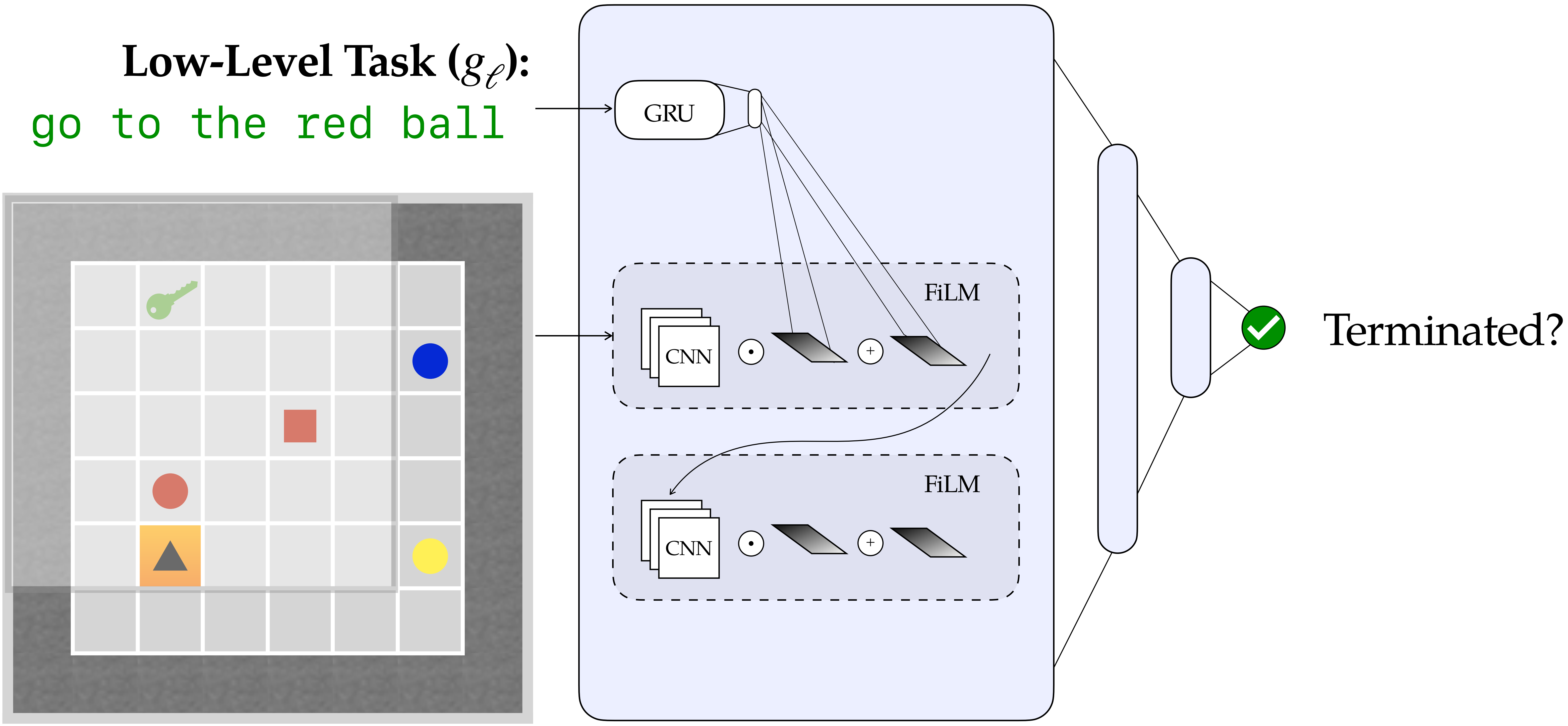
- $h^{\phi} : S \times G_{\ell} \rightarrow \{0,1\}$



“go to the red ball”

“go to the blue ball”

Low-Level Termination Classifier



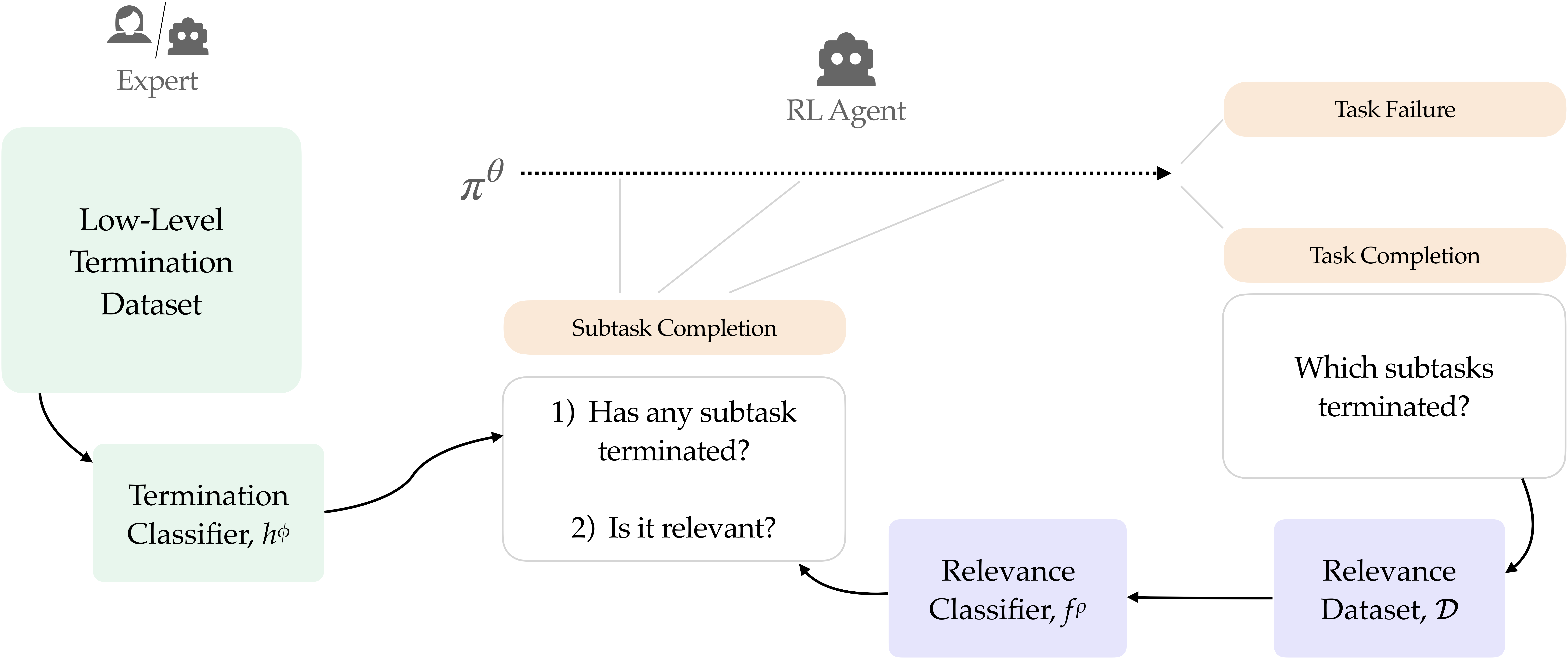
Low-Level Task (g_ℓ):

go to the red ball

Terminated?

Multimodal
State & Goal Encoder

Approach



Expert

Low-Level Termination Dataset

Termination Classifier, h^ϕ

RL Agent

π^θ

Subtask Completion

1) Has any subtask terminated?

2) Is it relevant?

Task Failure

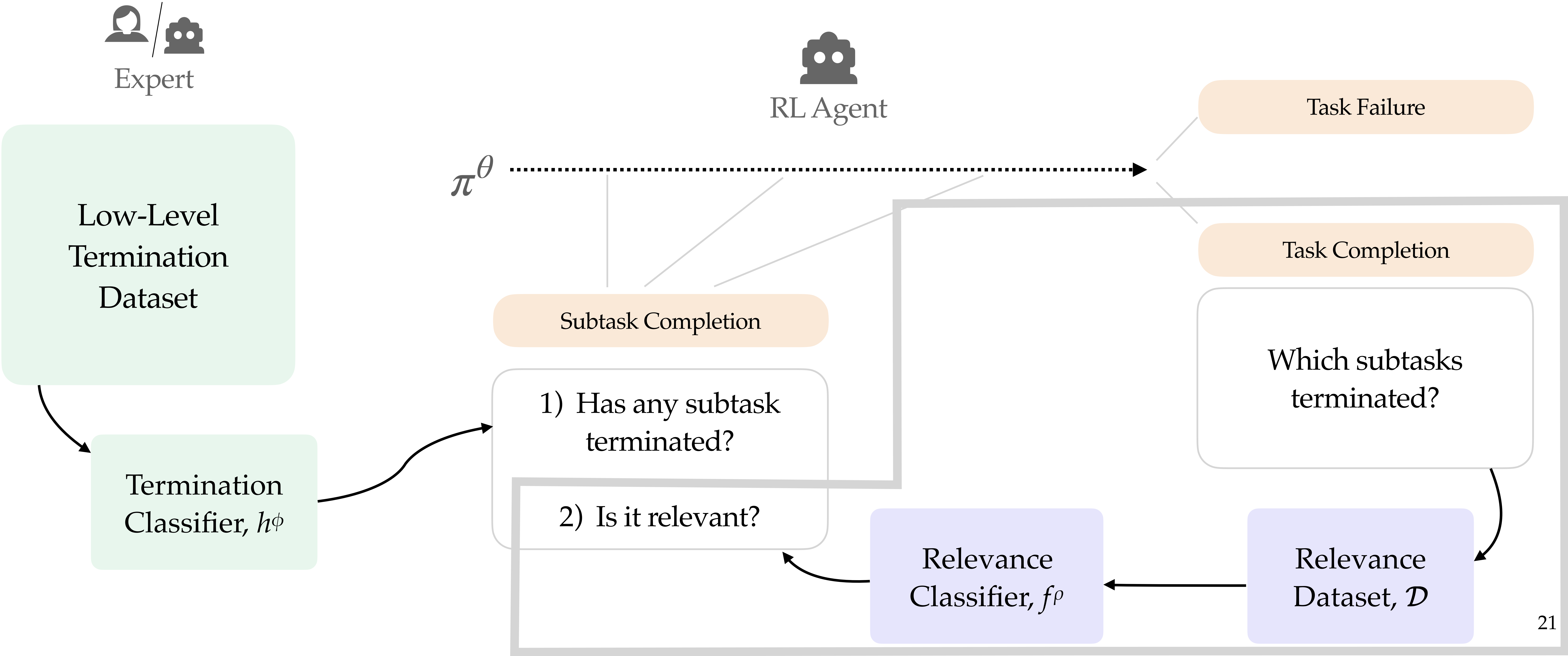
Task Completion

Which subtasks terminated?

Relevance Dataset, $\mathcal{D}$

Relevance Classifier, f^ρ

Approach


Expert
RL Agent
π^θ
Task Failure
Task Completion
Low-Level Termination Dataset
Subtask Completion
1) Has any subtask terminated?
2) Is it relevant?
Which subtasks terminated?
Termination Classifier, h^ϕ
Relevance Classifier, f^ρ
Relevance Dataset, $\mathcal{D}$

Relevance Classifier

Relevance Classifier

- Mapping $G \to \mathscr{P}(G_\ell)$ is initially unknown

Relevance Classifier

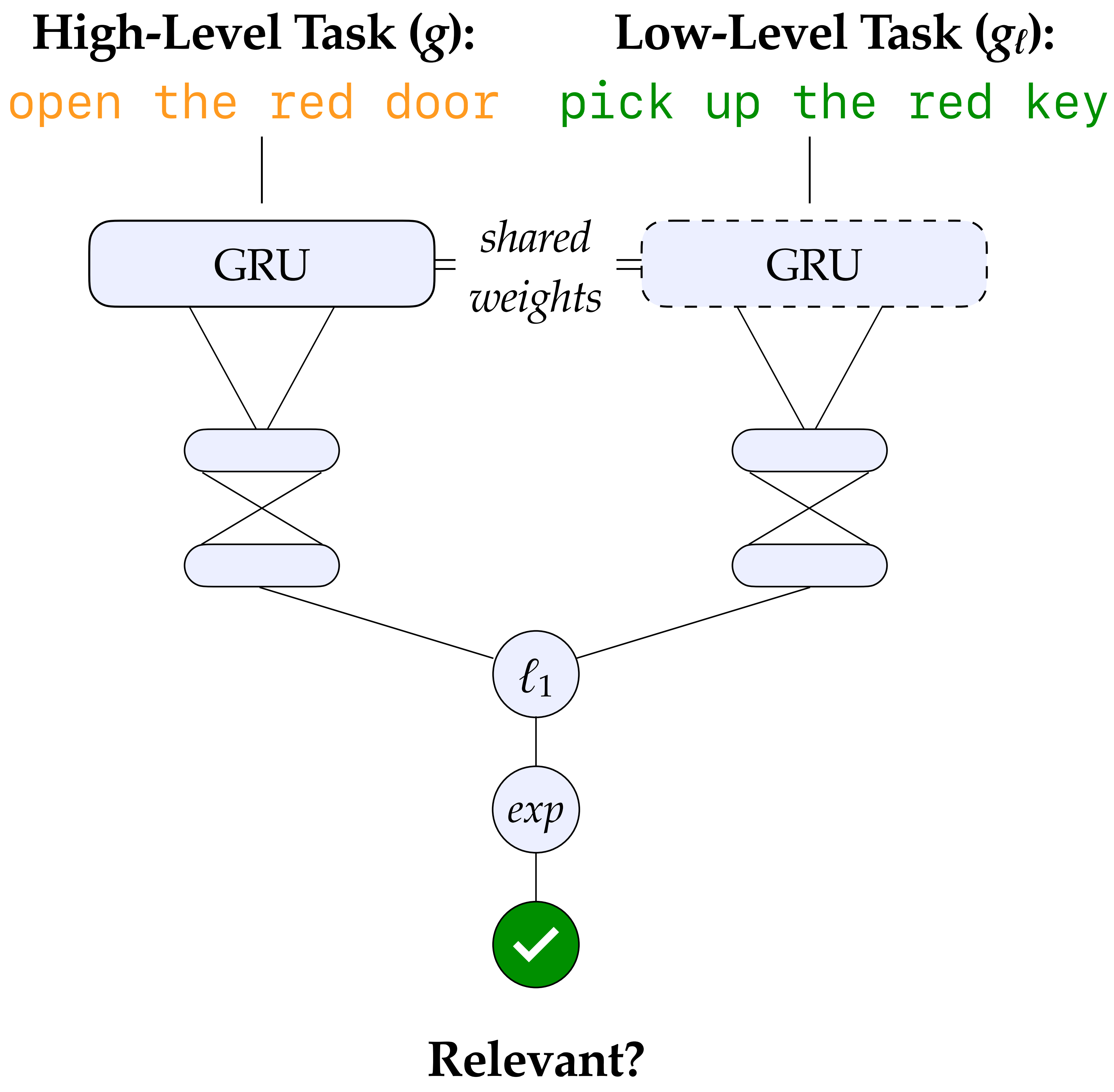
- Mapping $G \to \mathscr{P}(G_\ell)$ is initially unknown
  - Example: `open the red door` → { `pick up the red key`, `go to the red door` }

Relevance Classifier

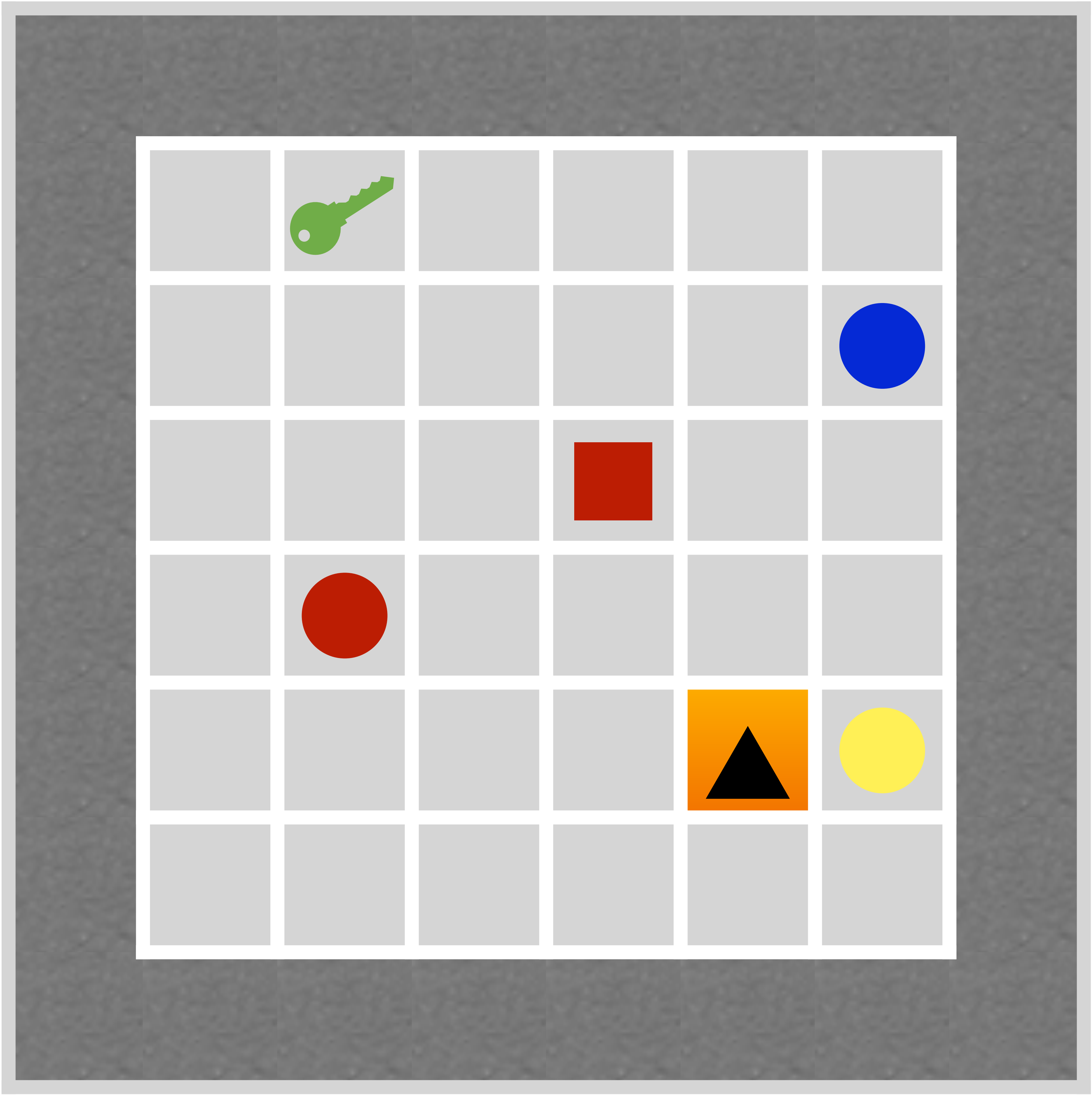
- Mapping $G \to \mathscr{P}(G_\ell)$ is initially unknown
    - Example: `open the red door` → { `pick up the red key`, `go to the red door` }
- $f^\rho : G \times G_\ell \to \{0,1\}$

Relevance Classifier

- Mapping $G \rightarrow \mathscr{P}(G_\ell)$ is initially unknown
  - Example: `open the red door` → { `pick up the red key`, `go to the red door` }
- $f^p : G \times G_\ell \rightarrow \{0,1\}$

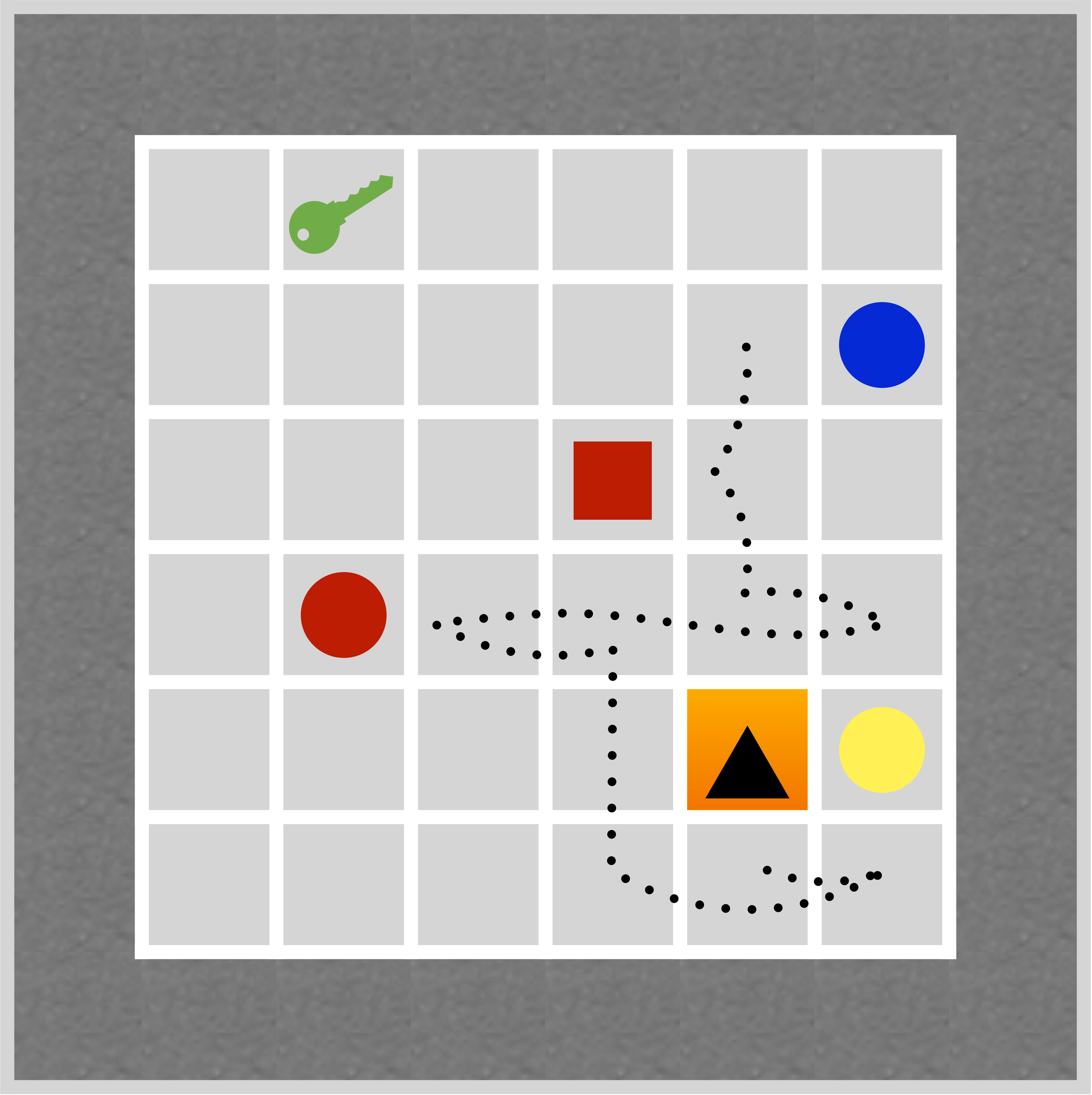


Collecting Relevance Data Online



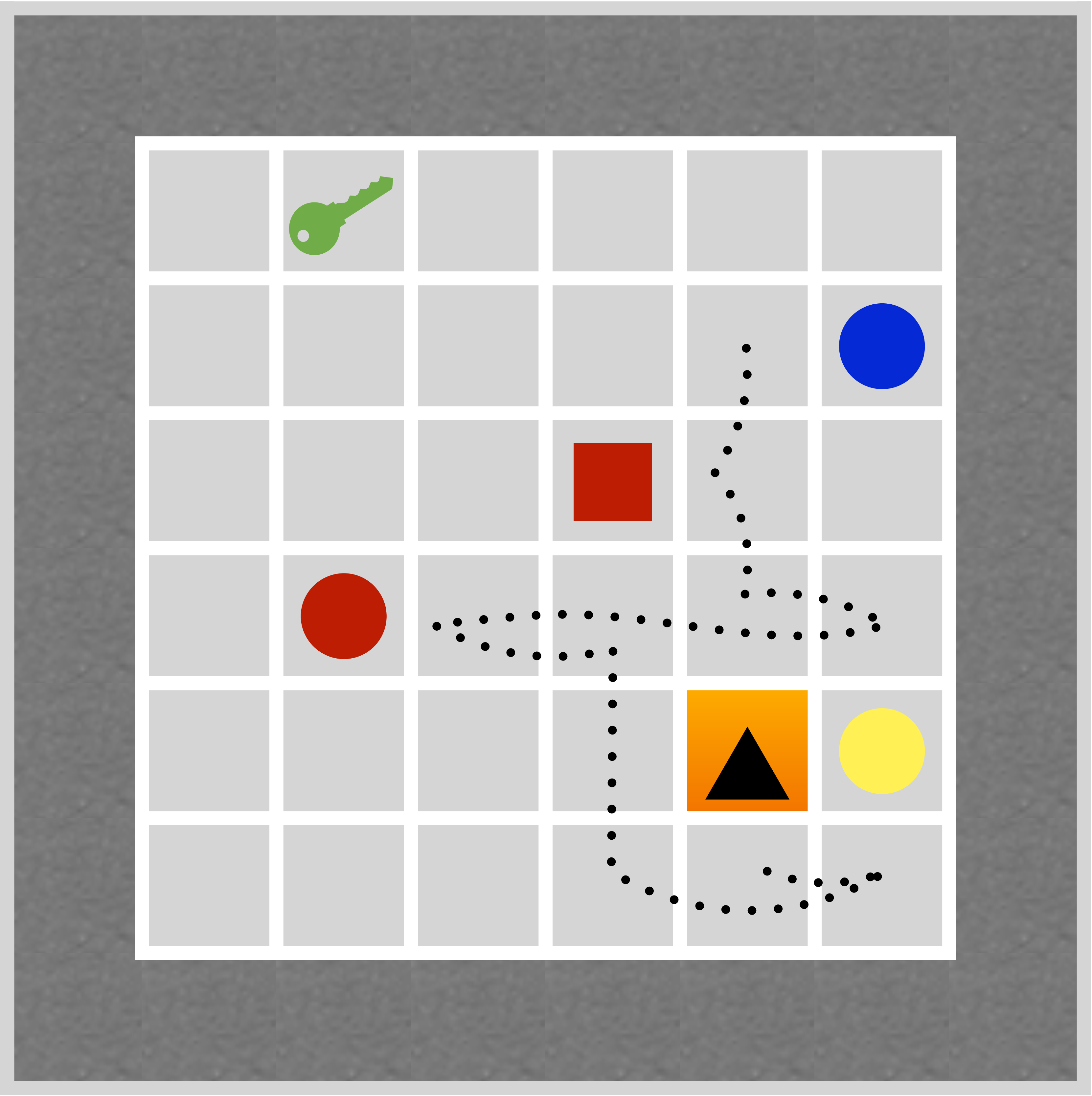
put the red ball next
to the blue ball

Collecting Relevance Data Online



put the red ball next to the blue ball

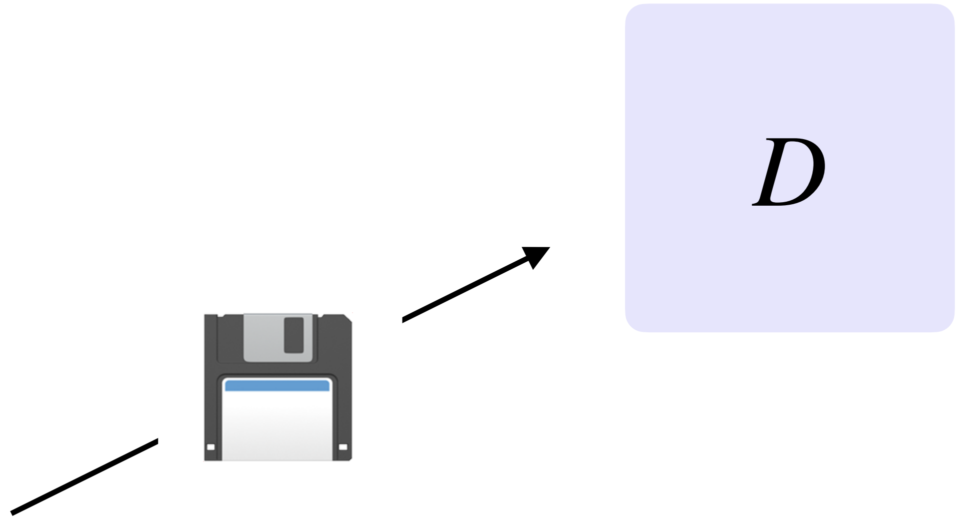
Collecting Relevance Data Online

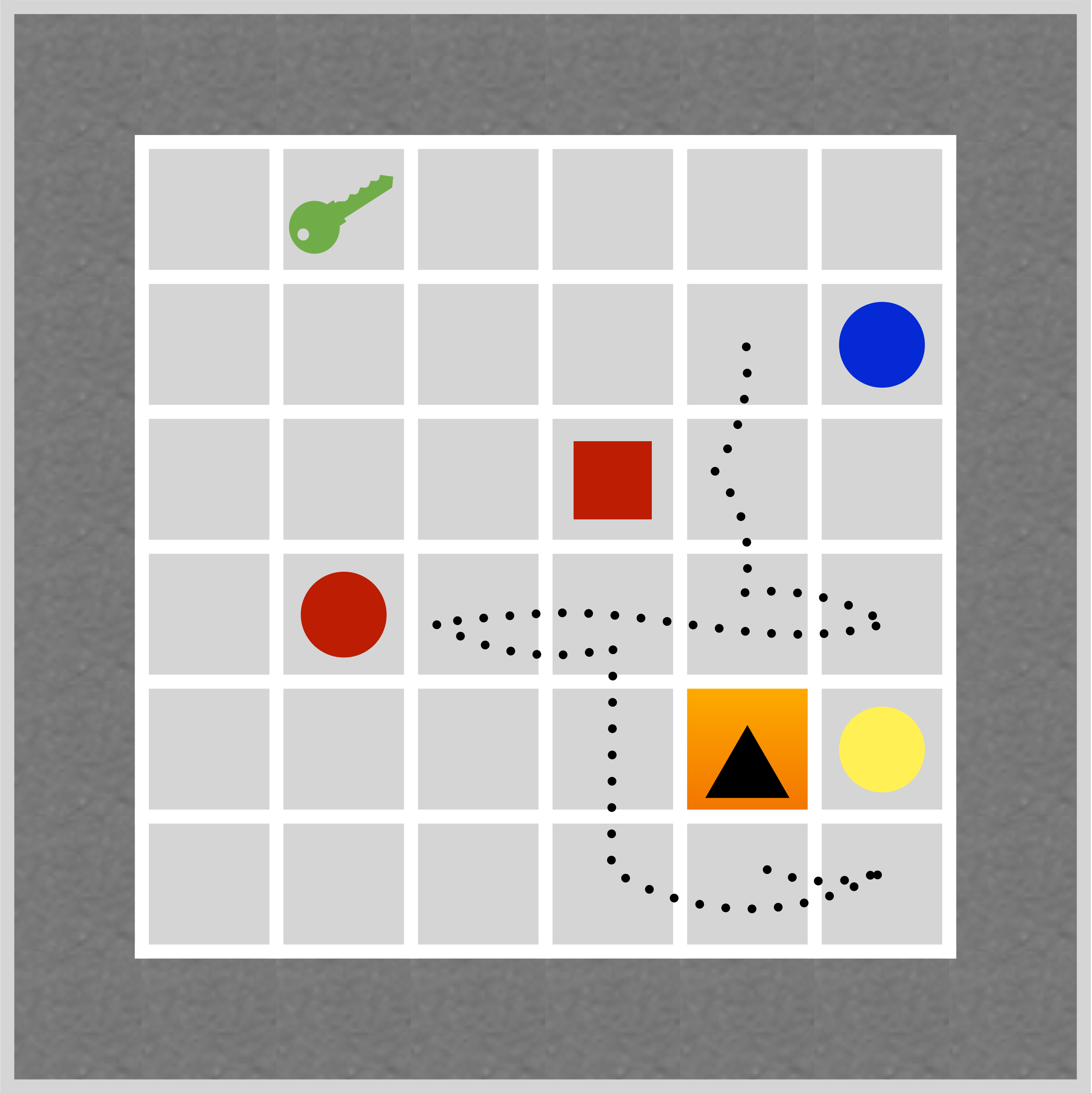


put the red ball next to the blue ball

- go to the yellow ball
- go to the red ball
- go to the red square
- go to the blue ball

Collecting Relevance Data Online



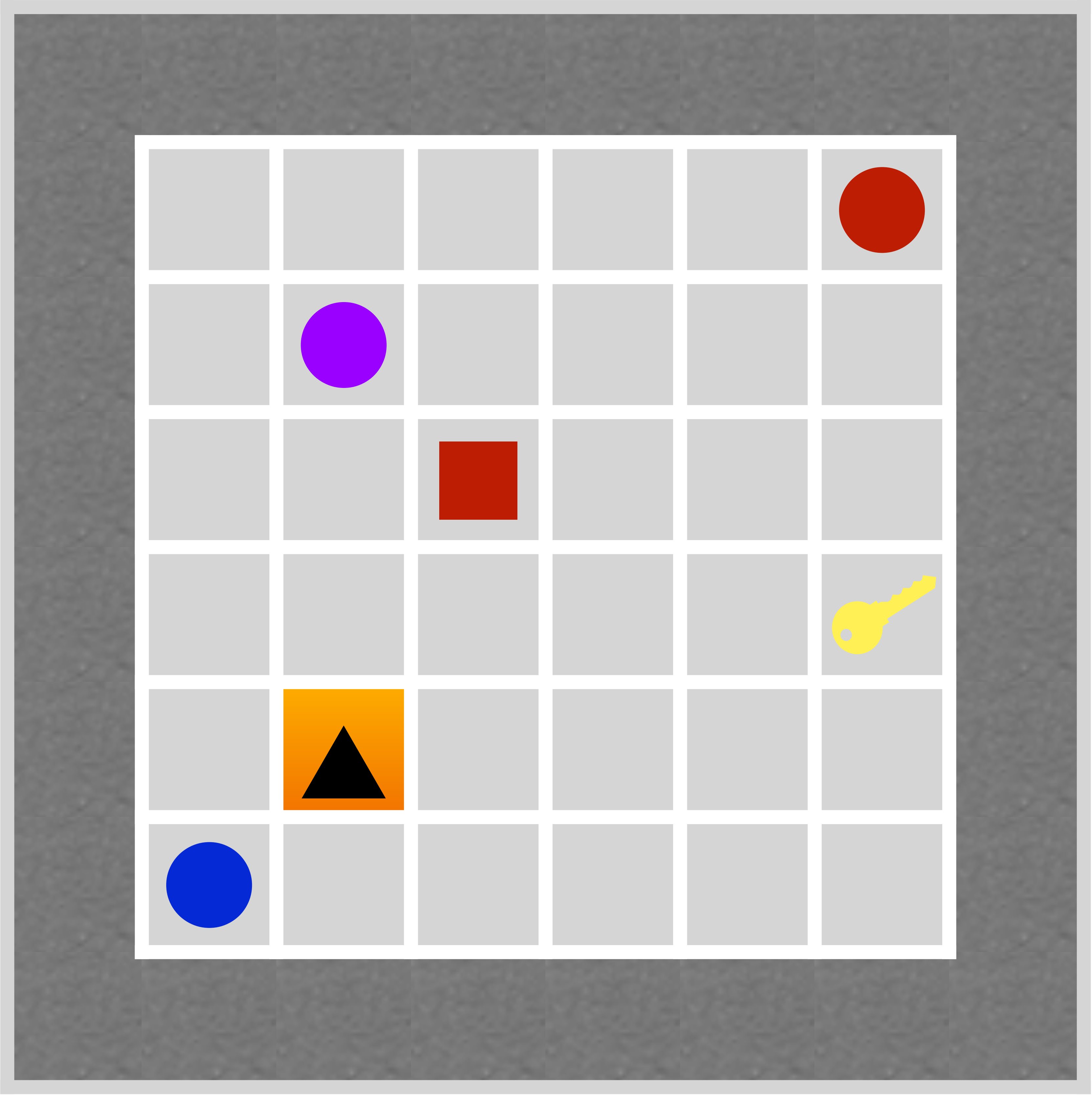


put the red ball next to the blue ball

- go to the yellow ball
- go to the red ball
- go to the red square
- go to the blue ball

Deduplication

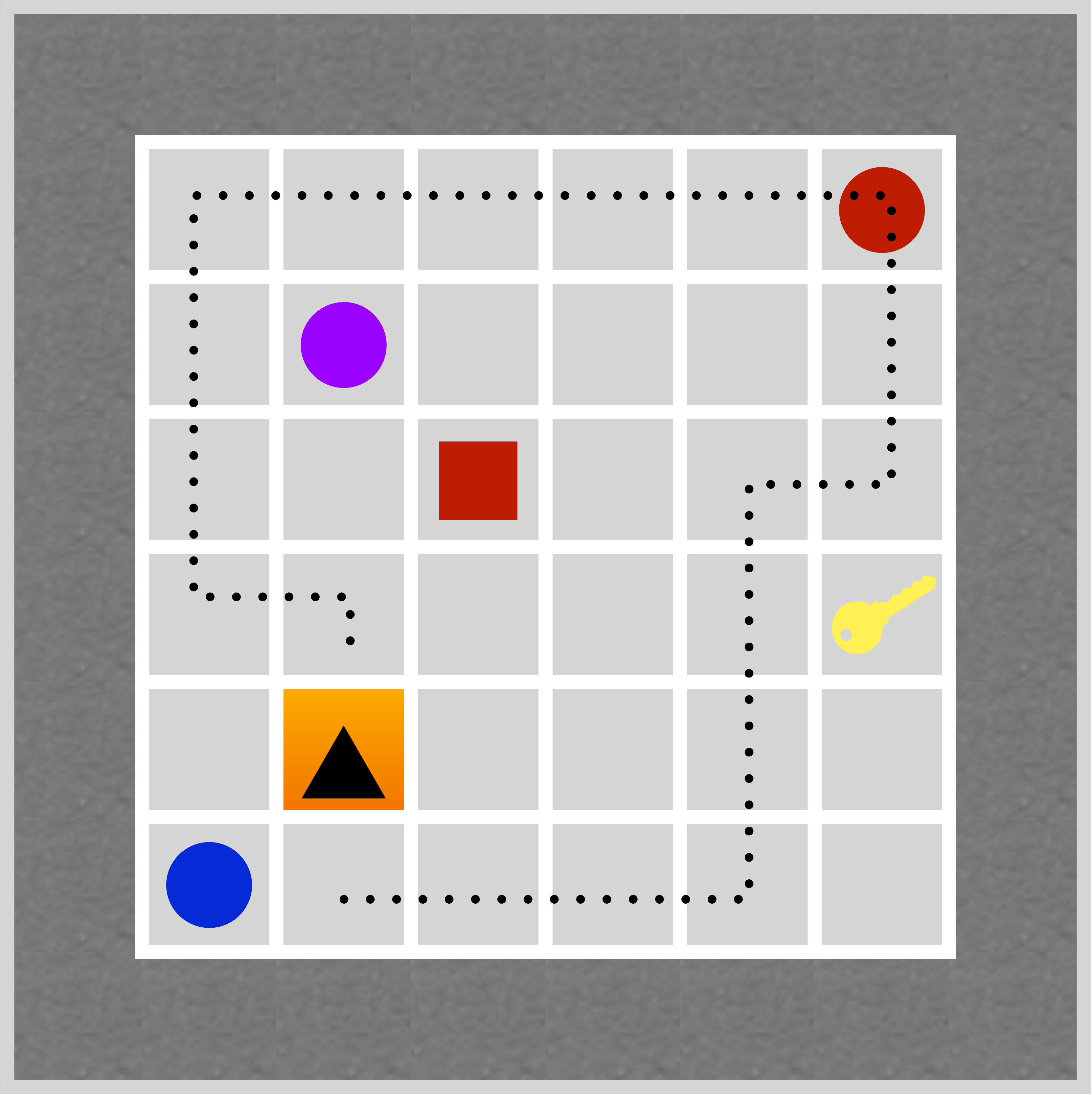
D



put the red ball next to the blue ball

Deduplication

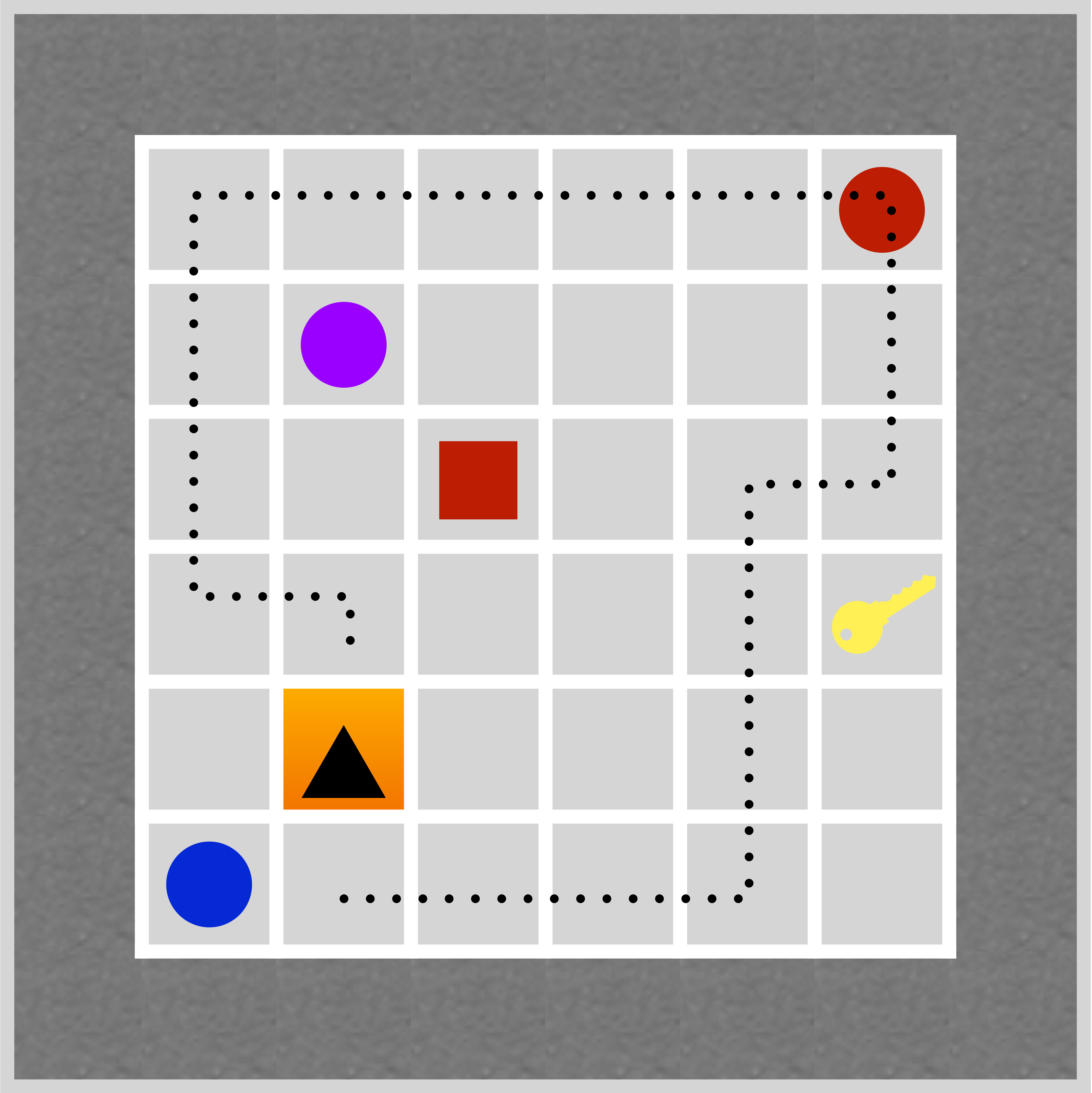
D



put the red ball next to the blue ball

Deduplication

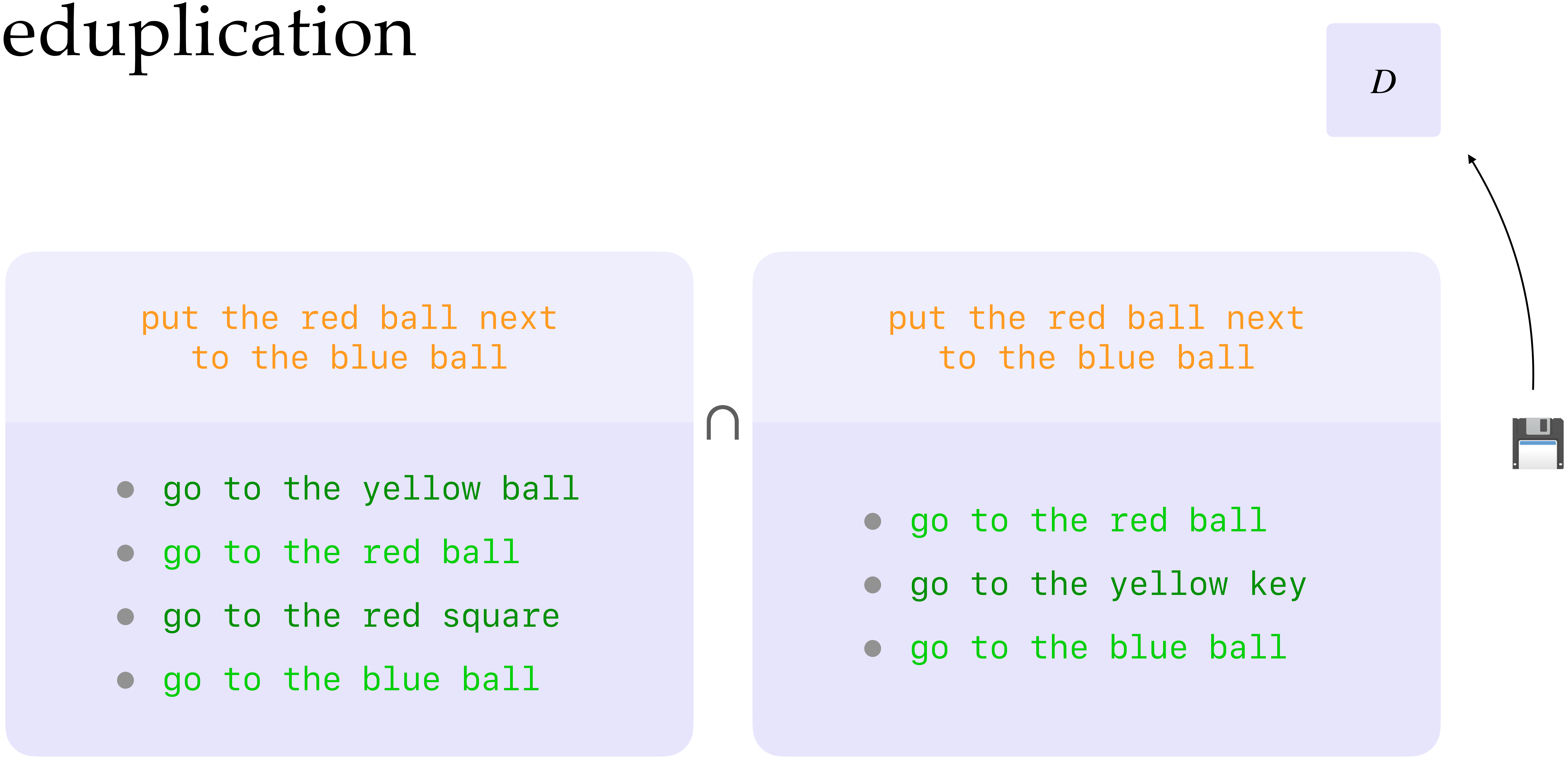
D



put the red ball next to the blue ball

- go to the red ball
- go to the yellow key
- go to the blue ball

Deduplication



put the red ball next to the blue ball

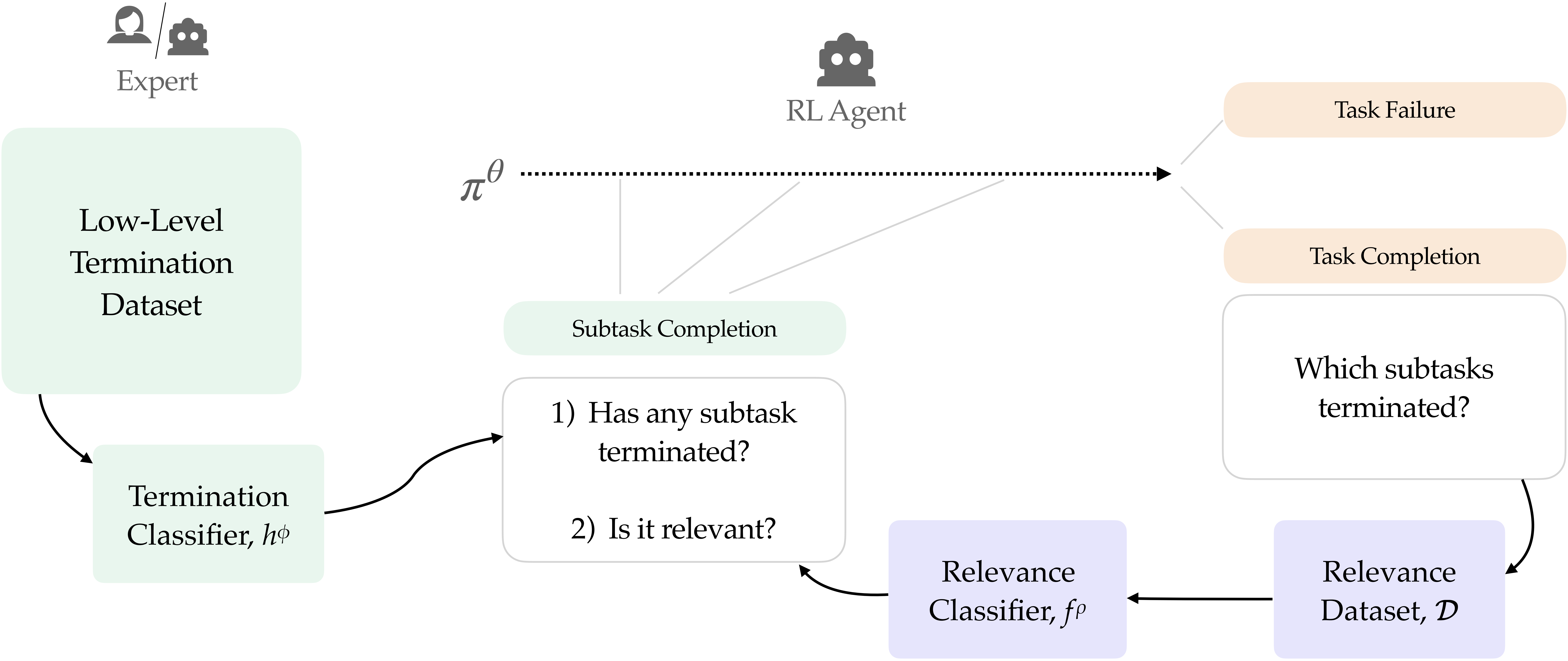
- go to the yellow ball
- go to the red ball
- go to the red square
- go to the blue ball

∩

put the red ball next to the blue ball

- go to the red ball
- go to the yellow key
- go to the blue ball

ELLA



Expert

RL Agent

π^{θ}

Low-Level Termination Dataset

Termination Classifier, h^{ϕ}

Subtask Completion

1) Has any subtask terminated?

2) Is it relevant?

Task Failure

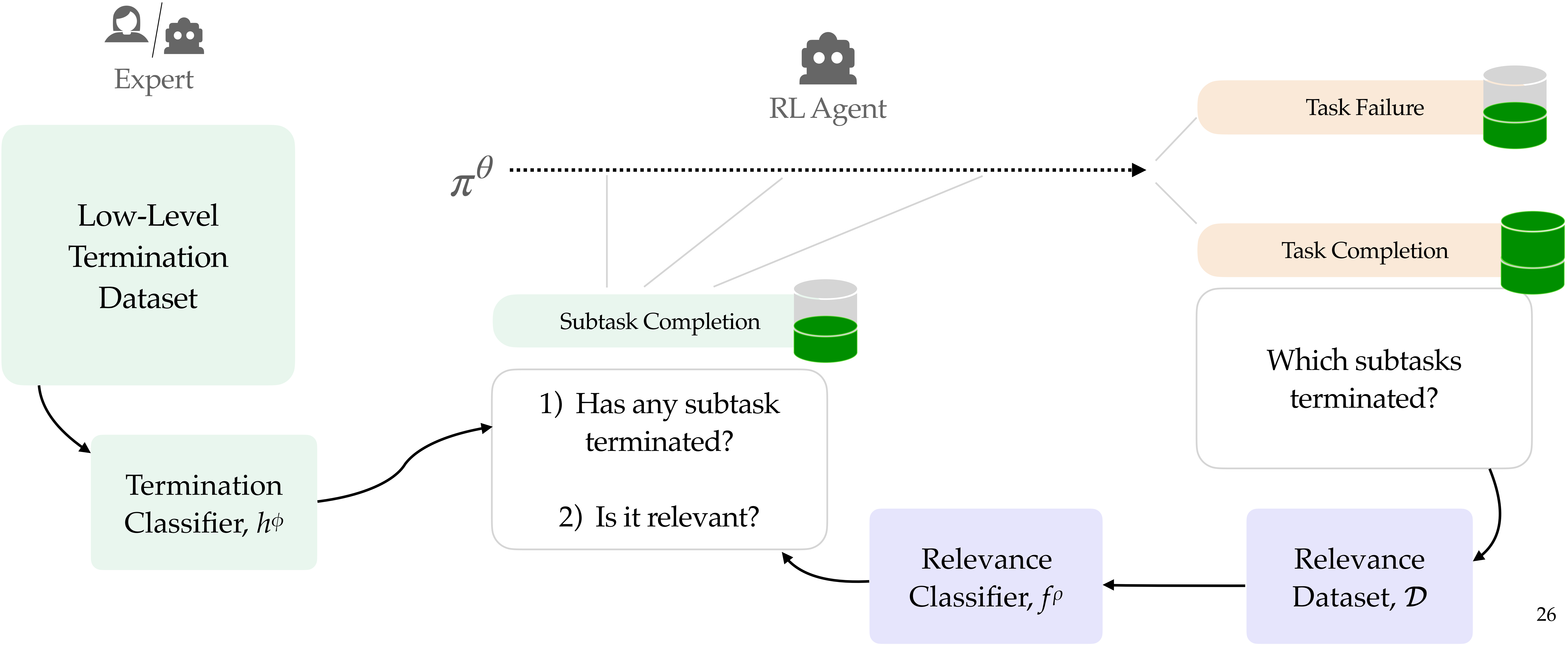
Task Completion

Which subtasks terminated?

Relevance Dataset, $\mathcal{D}$

Relevance Classifier, f^{ρ}

ELLA



Expert

Low-Level Termination Dataset

Termination Classifier, h^{ϕ}

RL Agent

π^{θ}

Subtask Completion

1) Has any subtask terminated?

2) Is it relevant?

Task Failure

Task Completion

Which subtasks terminated?

Relevance Classifier, f^{ρ}

Relevance Dataset, $\mathcal{D}$

Reward Shaping

Reward Shaping

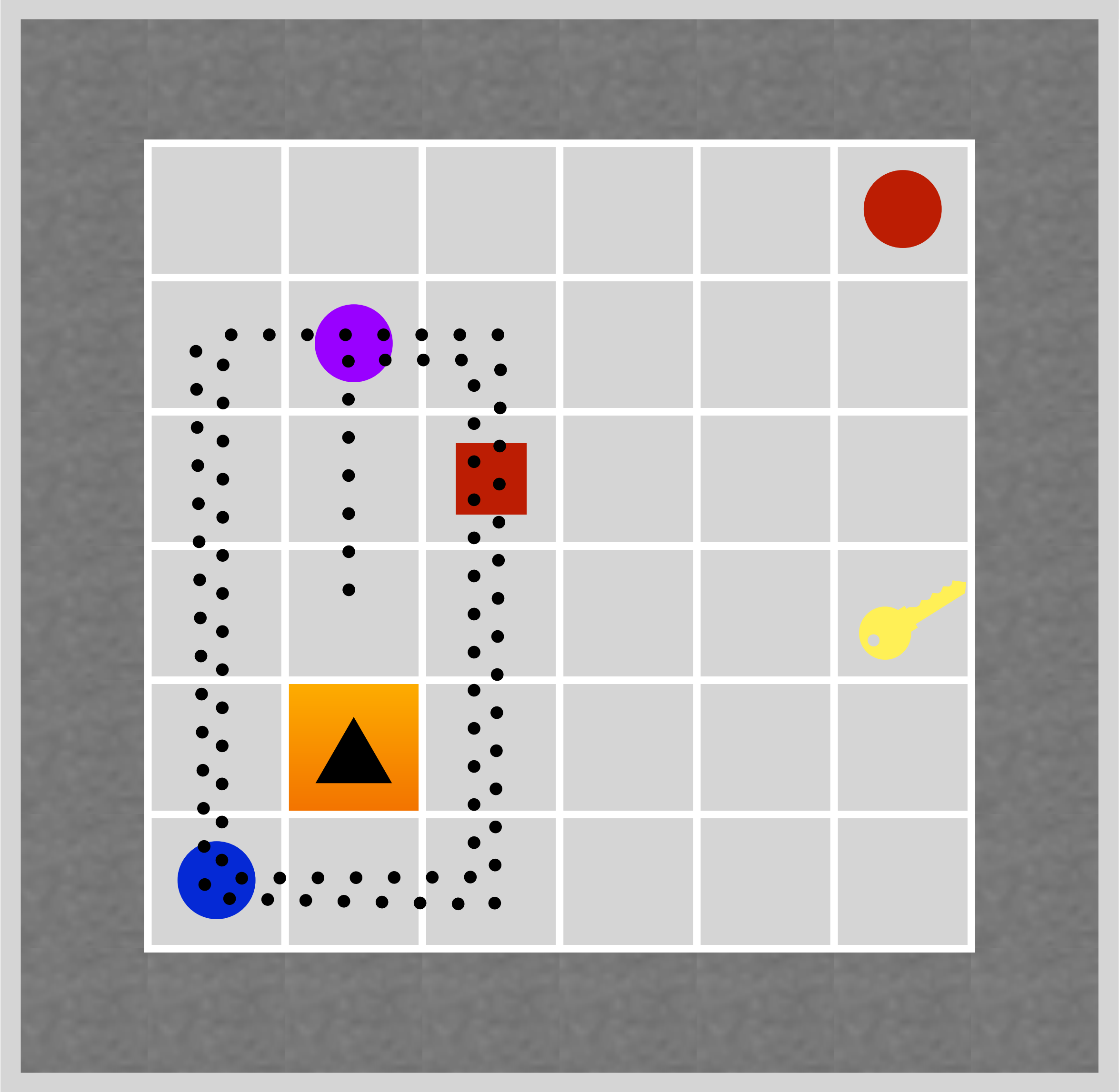
- Bonus λ for relevant low-level tasks

Reward Shaping

- Bonus λ for relevant low-level tasks
- However, we do not want “distraction” by subtasks

Reward Shaping

- Bonus λ for relevant low-level tasks
- However, we do not want "distraction" by subtasks



Reward Shaping

Reward Shaping

- Neutralize intermediate rewards in successful trajectories

Reward Shaping

- Neutralize intermediate rewards in successful trajectories
  - Set $r'_N = r_N - \sum_{t \in T_{\mathbb{S}}} \gamma^{t-N} \lambda$ where $T_{\mathbb{S}}$ is the set of time steps when bonuses were applied

Reward Shaping

- Neutralize intermediate rewards in successful trajectories
  - Set $r'_N = r_N - \sum_{t \in T_{\mathbb{S}}} \gamma^{t-N} \lambda$ where $T_{\mathbb{S}}$ is the set of time steps when bonuses were applied
- Limit return in unsuccessful trajectories by tuning λ

Reward Shaping

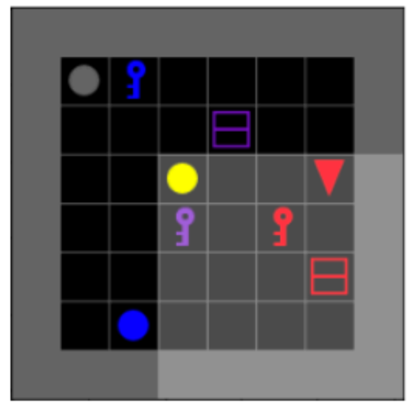
- Neutralize intermediate rewards in successful trajectories
  - Set $r'_N = r_N - \sum_{t \in T_{\mathbb{S}}} \gamma^{t-N} \lambda$ where $T_{\mathbb{S}}$ is the set of time steps when bonuses were applied
- Limit return in unsuccessful trajectories by tuning λ
  - $\lambda < \frac{\gamma^H r_H}{|G_\ell|}$

Experiments

Experiments

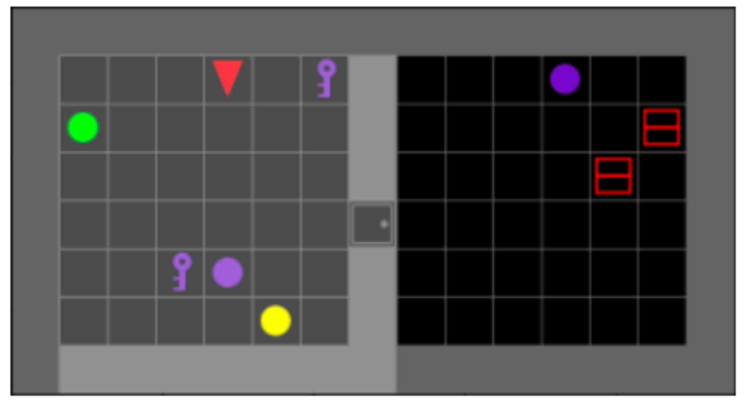
High-Level Tasks

PutNext-Room



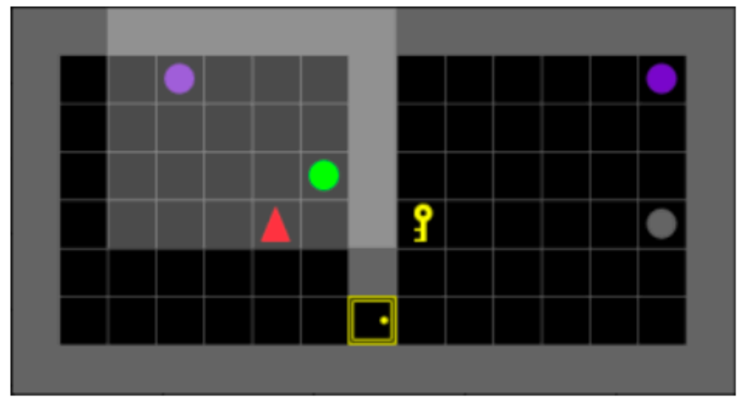
put the yellow ball next to a purple key

PutNext-Maze



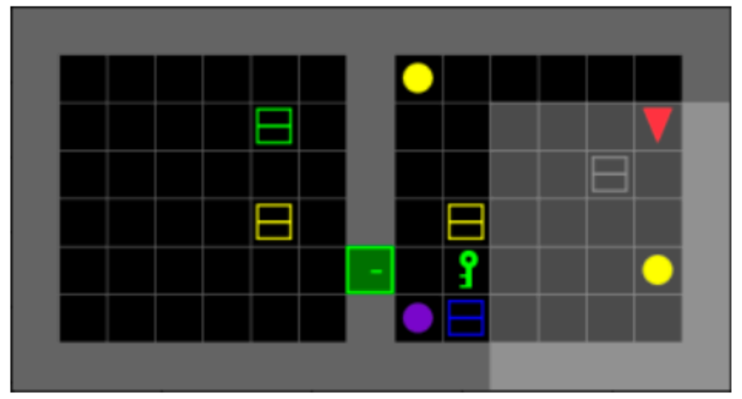
put the blue key next to the yellow ball

Open&Pick-Maze



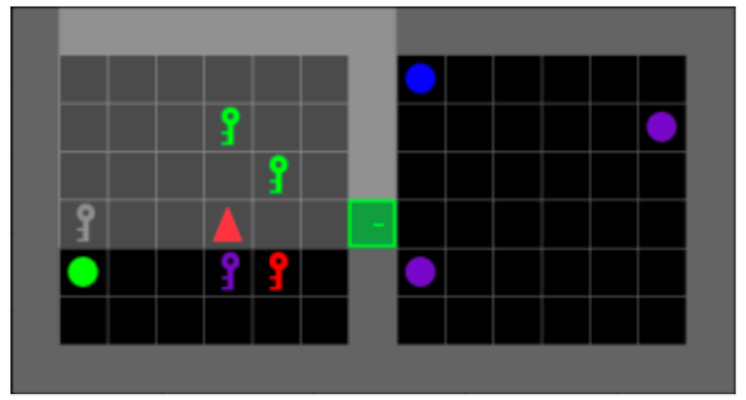
open the yellow door and pick up the grey ball

Unlock-Maze



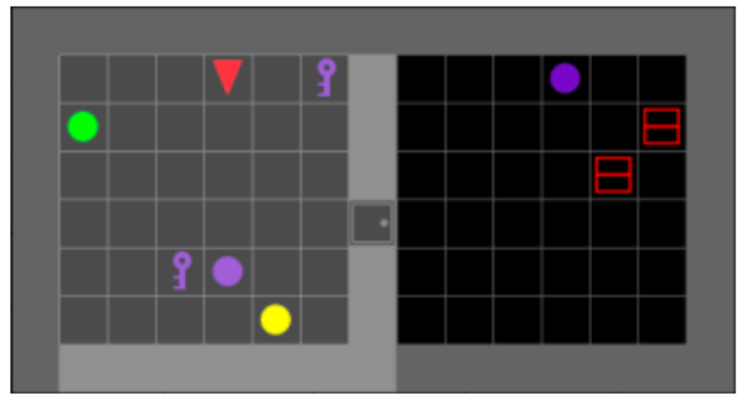
open the green door

Combo-Maze



pick up the green ball

Sequence-Maze

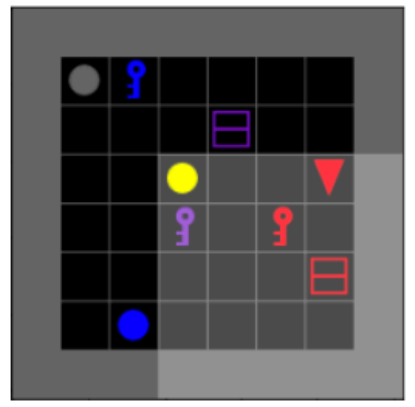


open the grey door after you put the yellow ball next to a purple key

Experiments

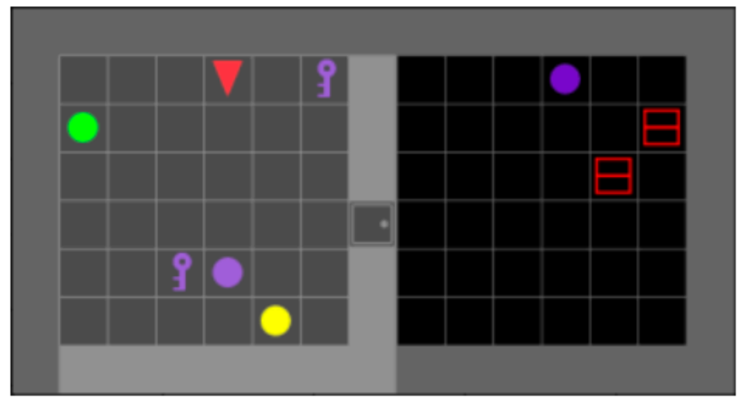
High-Level Tasks

PutNext-Room



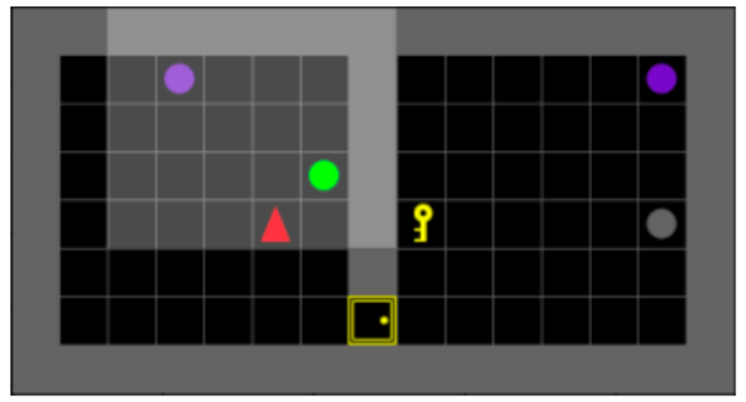
put the yellow ball next to a purple key

PutNext-Maze



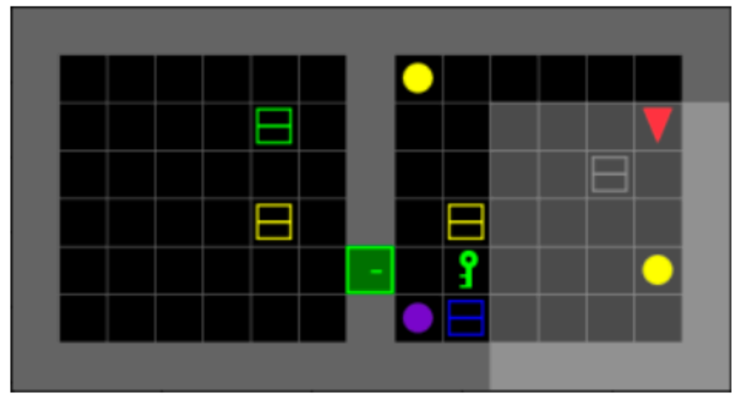
put the blue key next to the yellow ball

Open&Pick-Maze



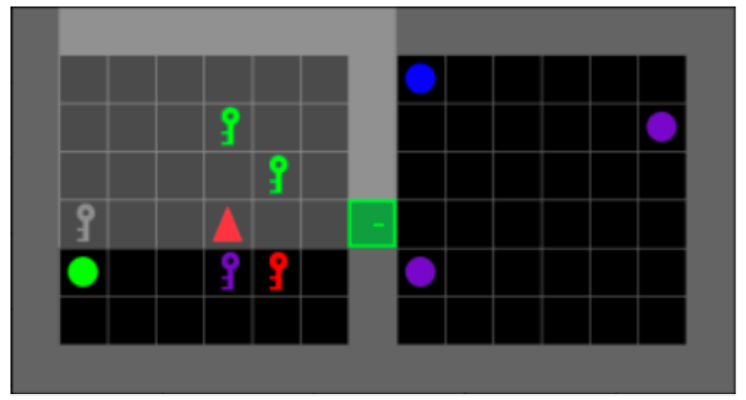
open the yellow door and pick up the grey ball

Unlock-Maze



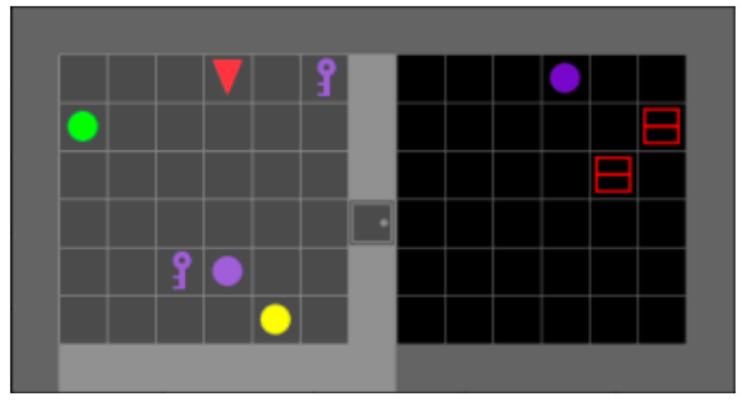
open the green door

Combo-Maze



pick up the green ball

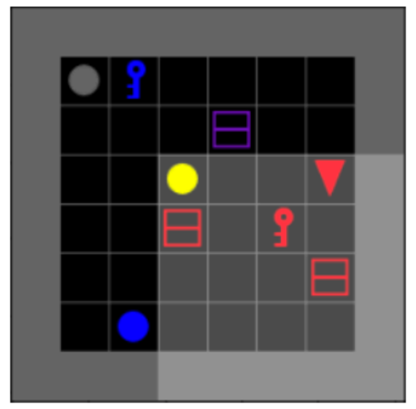
Sequence-Maze



open the grey door after you put the yellow ball next to a purple key

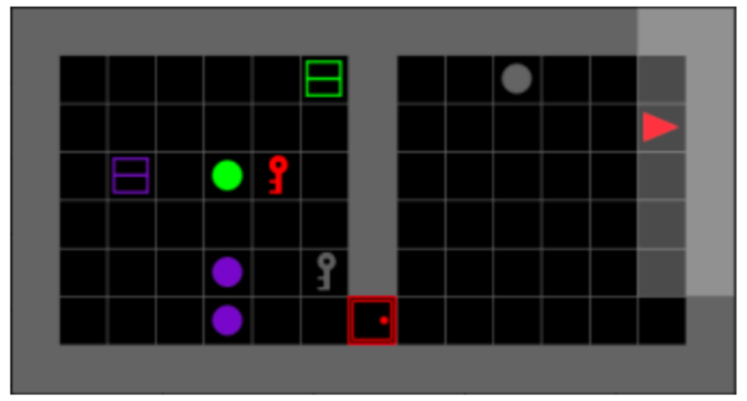
Low-Level Tasks

GoTo-Room



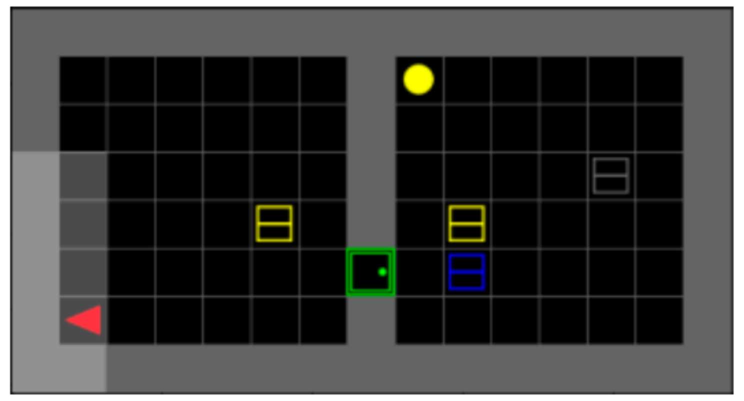
go to a yellow ball

GoTo-Maze



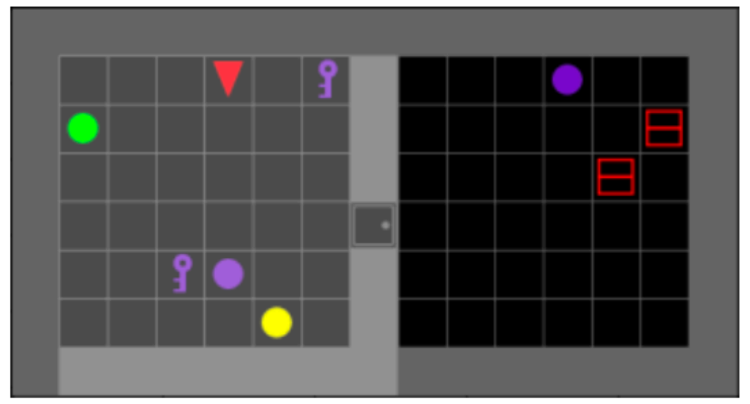
go to a red key

Open-Maze



open the green door

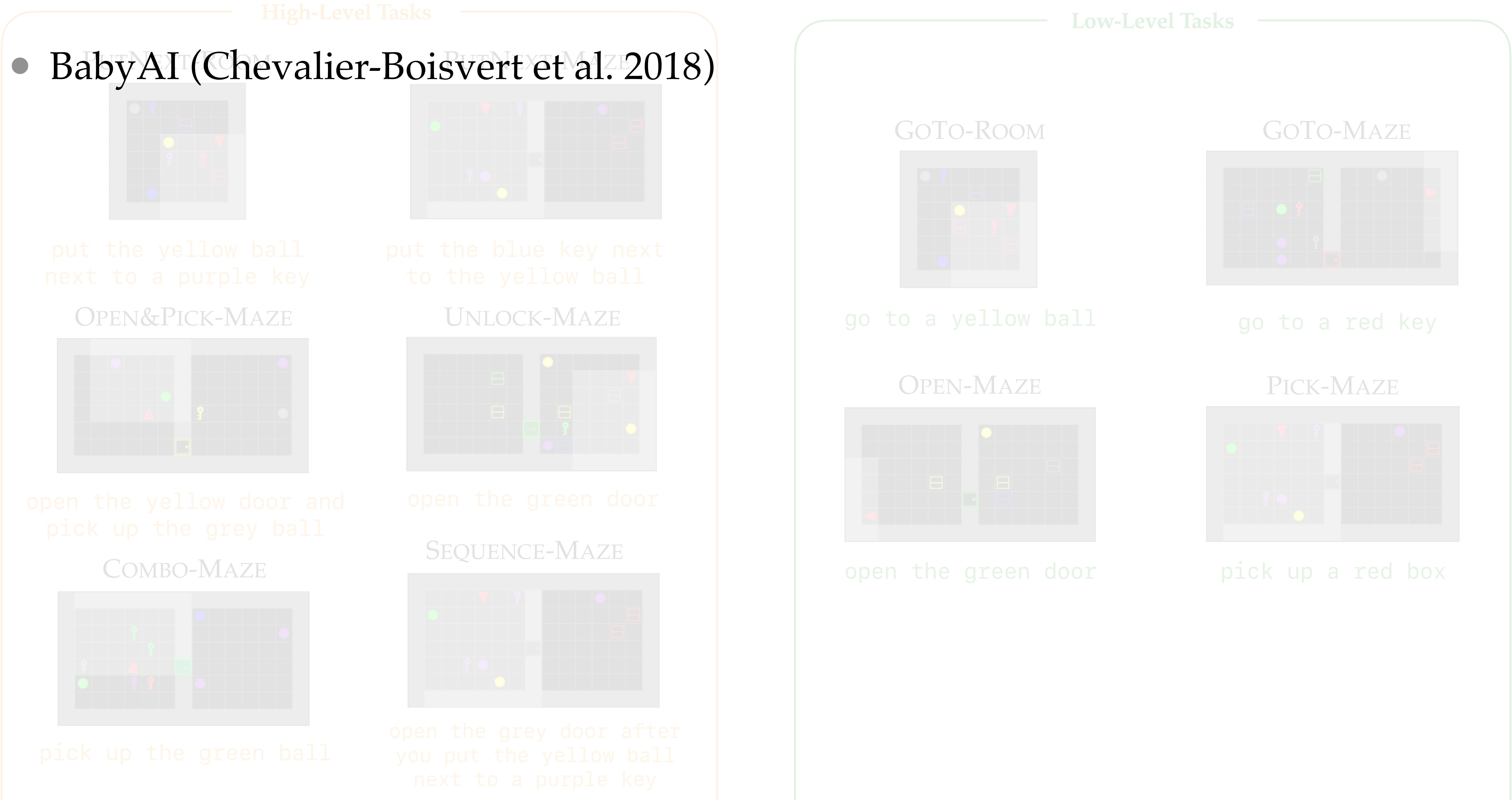
Pick-Maze



pick up a red box

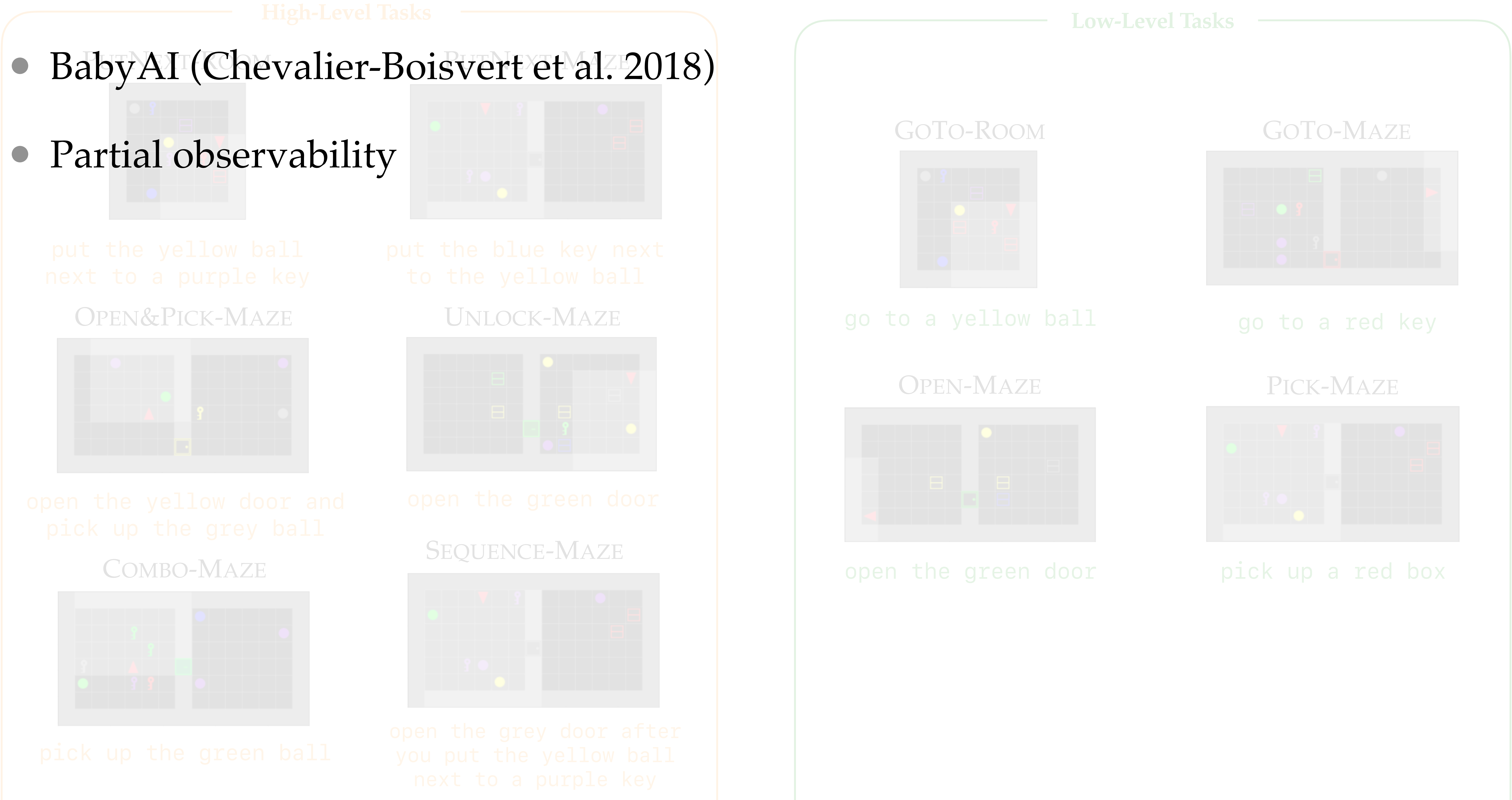
Experiments

- BabyAI (Chevalier-Boisvert et al. 2018)



Experiments

- BabyAI (Chevalier-Boisvert et al. 2018)
- Partial observability



Experiments

- BabyAI (Chevalier-Boisvert et al. 2018)
- Partial observability
- Distractor objects

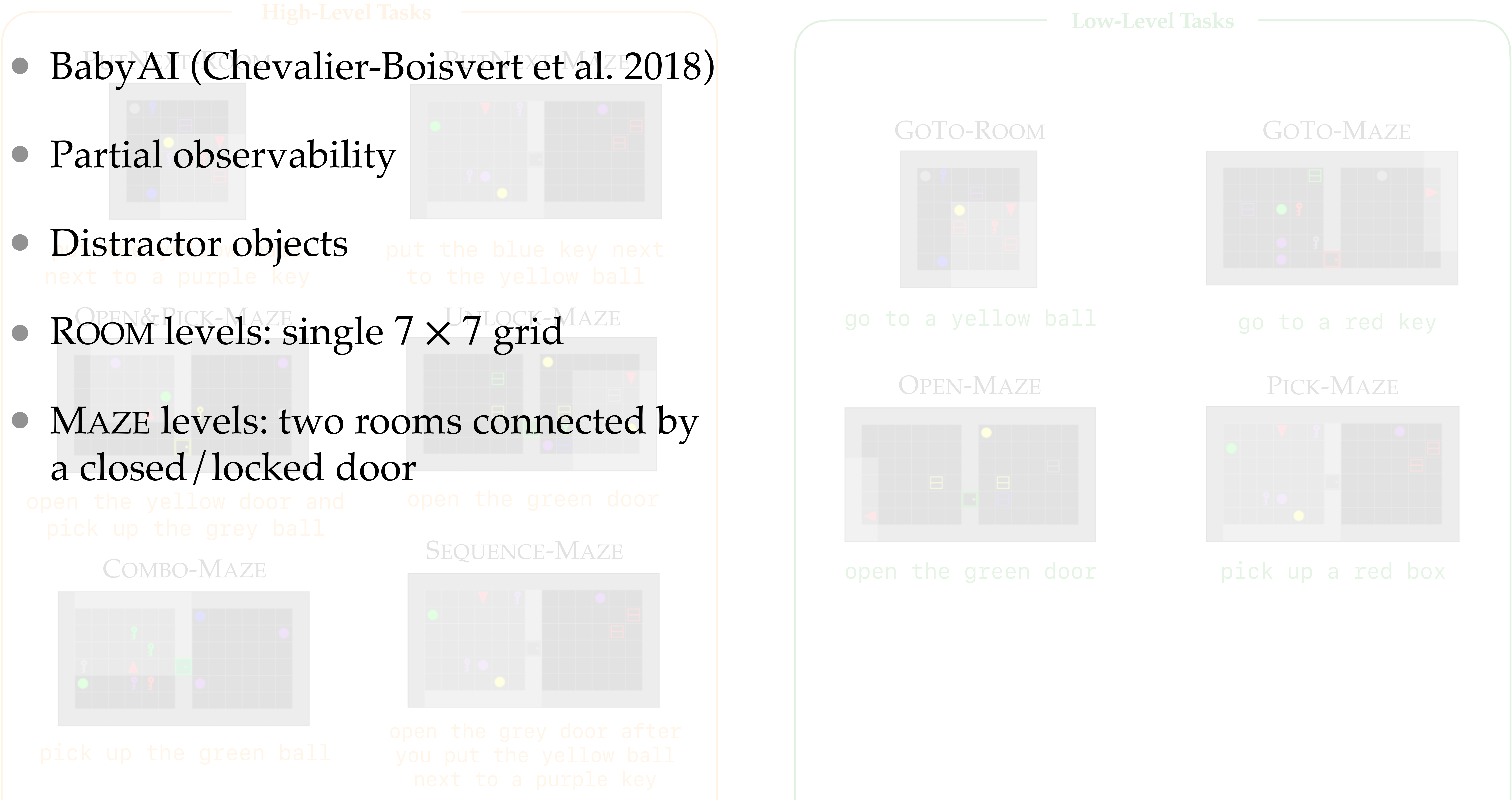
Experiments

- BabyAI (Chevalier-Boisvert et al. 2018)
- Partial observability
- Distractor objects
- ROOM levels: single 7 × 7 grid

High-Level Tasks
PUTNEXT-ROOM
PUTNEXT-MAZE
put the blue key next to the yellow ball
next to a purple key
OPEN&PICK-MAZE
UNLOCK-MAZE
open the yellow door and pick up the grey ball
open the green door
COMBO-MAZE
SEQUENCE-MAZE
pick up the green ball
open the grey door after you put the yellow ball next to a purple key
Low-Level Tasks
GOTO-ROOM
GOTO-MAZE
go to a yellow ball
go to a red key
OPEN-MAZE
PICK-MAZE
open the green door
pick up a red box

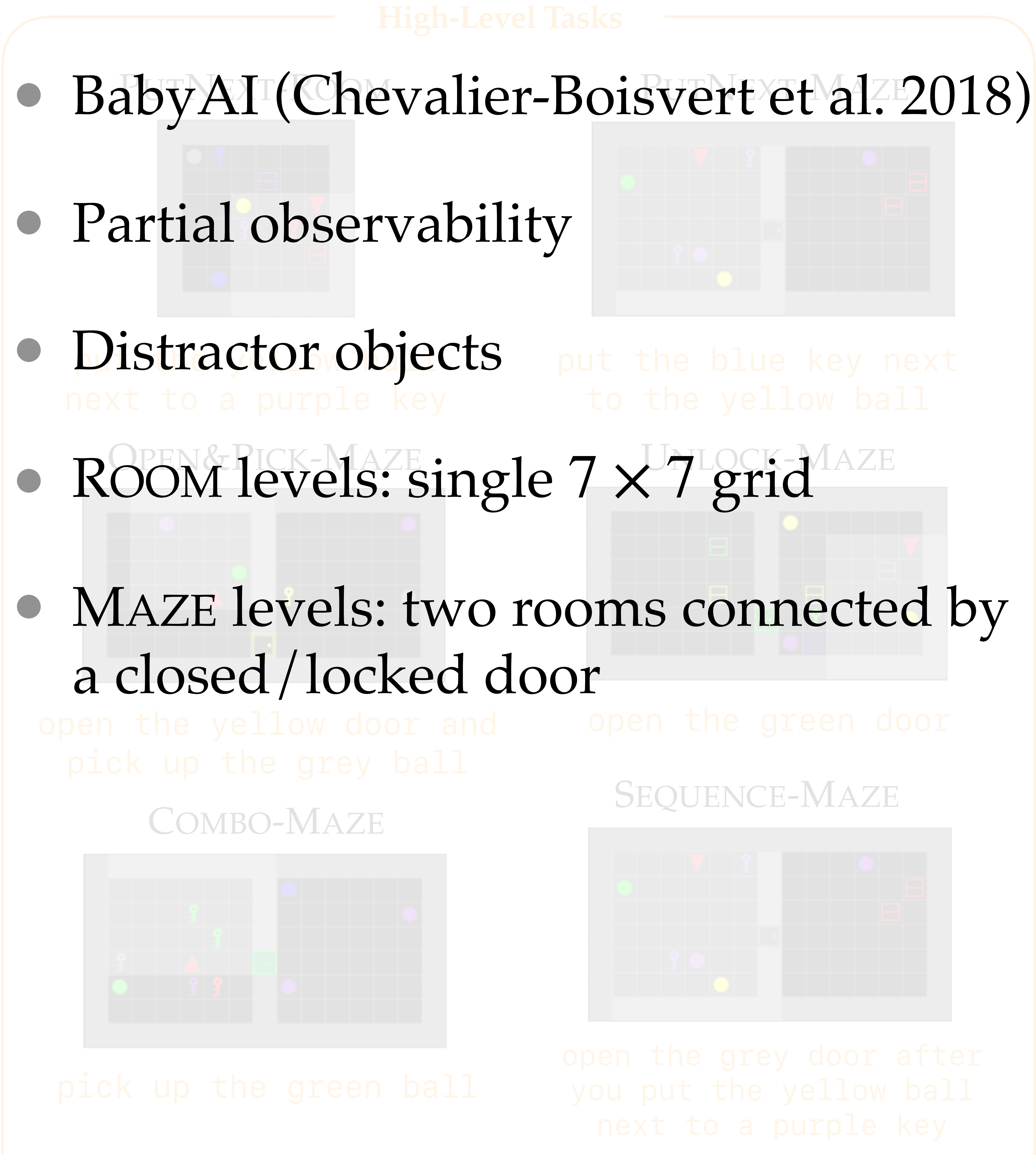
Experiments

- BabyAI (Chevalier-Boisvert et al. 2018)
- Partial observability
- Distractor objects
- ROOM levels: single 7×7 grid
- MAZE levels: two rooms connected by a closed / locked door

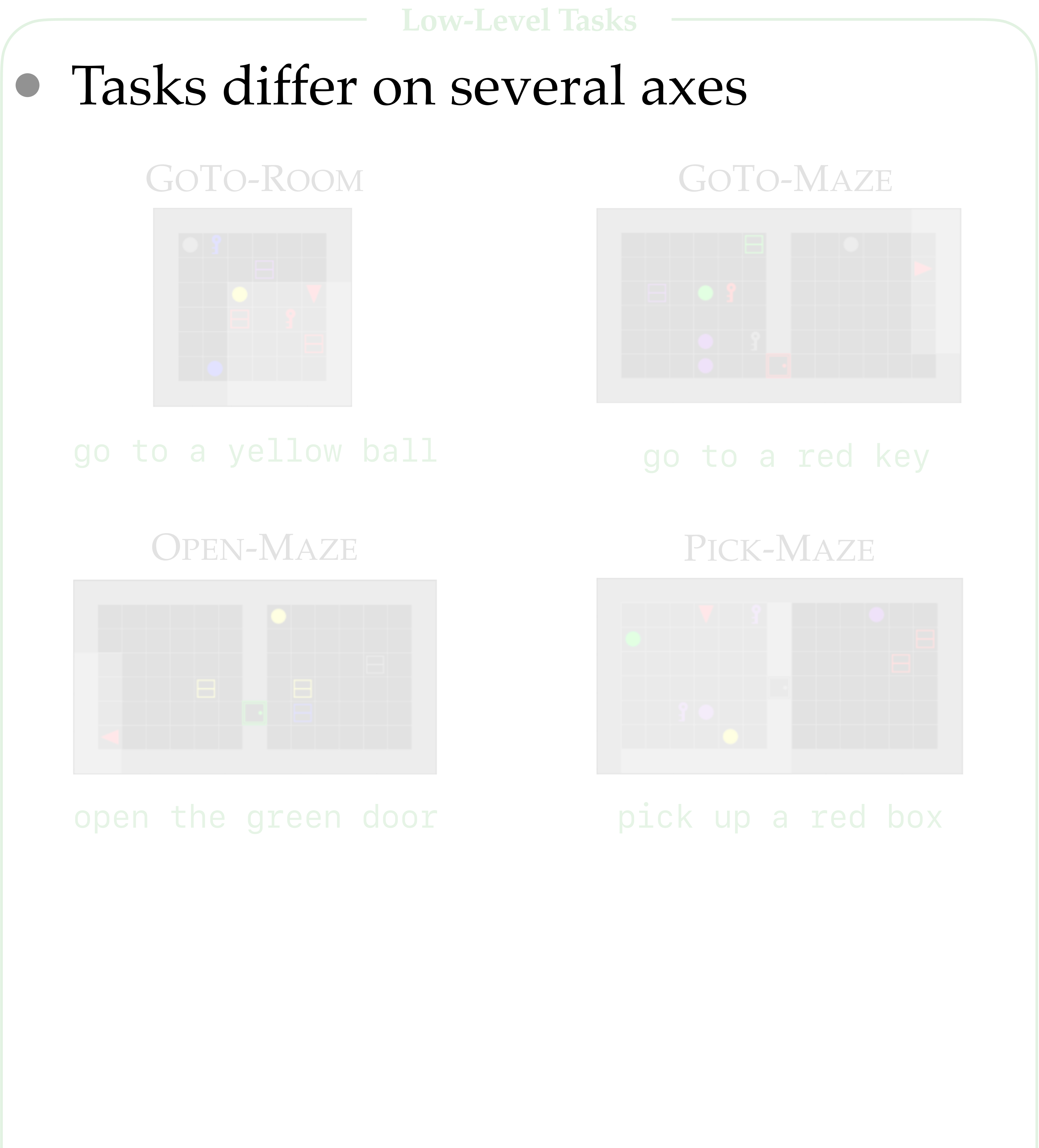


Experiments

- BabyAI (Chevalier-Boisvert et al. 2018)
- Partial observability
- Distractor objects
- ROOM levels: single 7 × 7 grid
- MAZE levels: two rooms connected by a closed / locked door


High-Level Tasks
PutNext-Room
PutNext-Maze
next to a purple key
put the blue key next to the yellow ball
Open&Pick-Maze
Unlock-Maze
open the yellow door and pick up the grey ball
open the green door
Combo-Maze
Sequence-Maze
pick up the green ball
open the grey door after you put the yellow ball next to a purple key

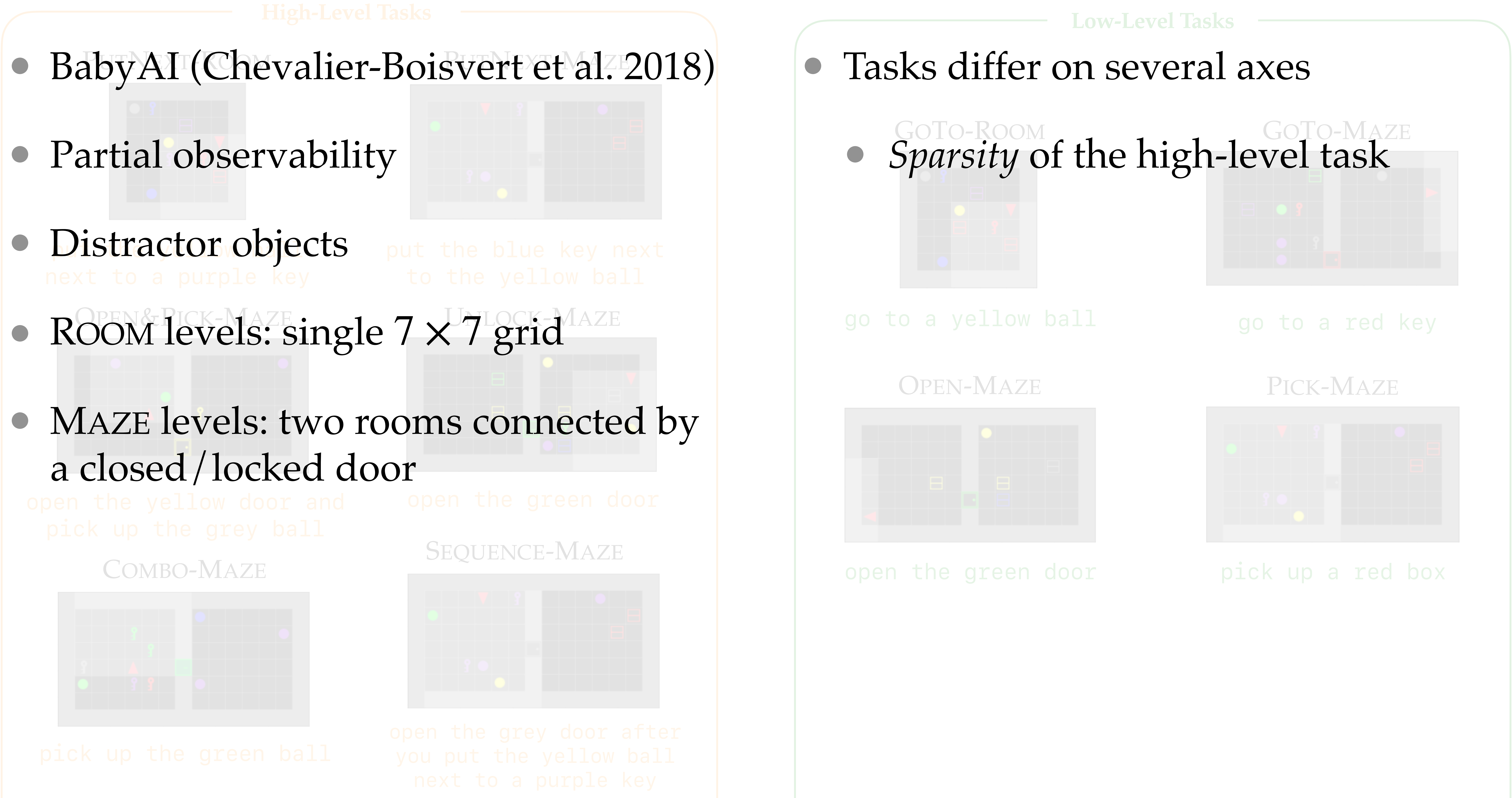
- Tasks differ on several axes



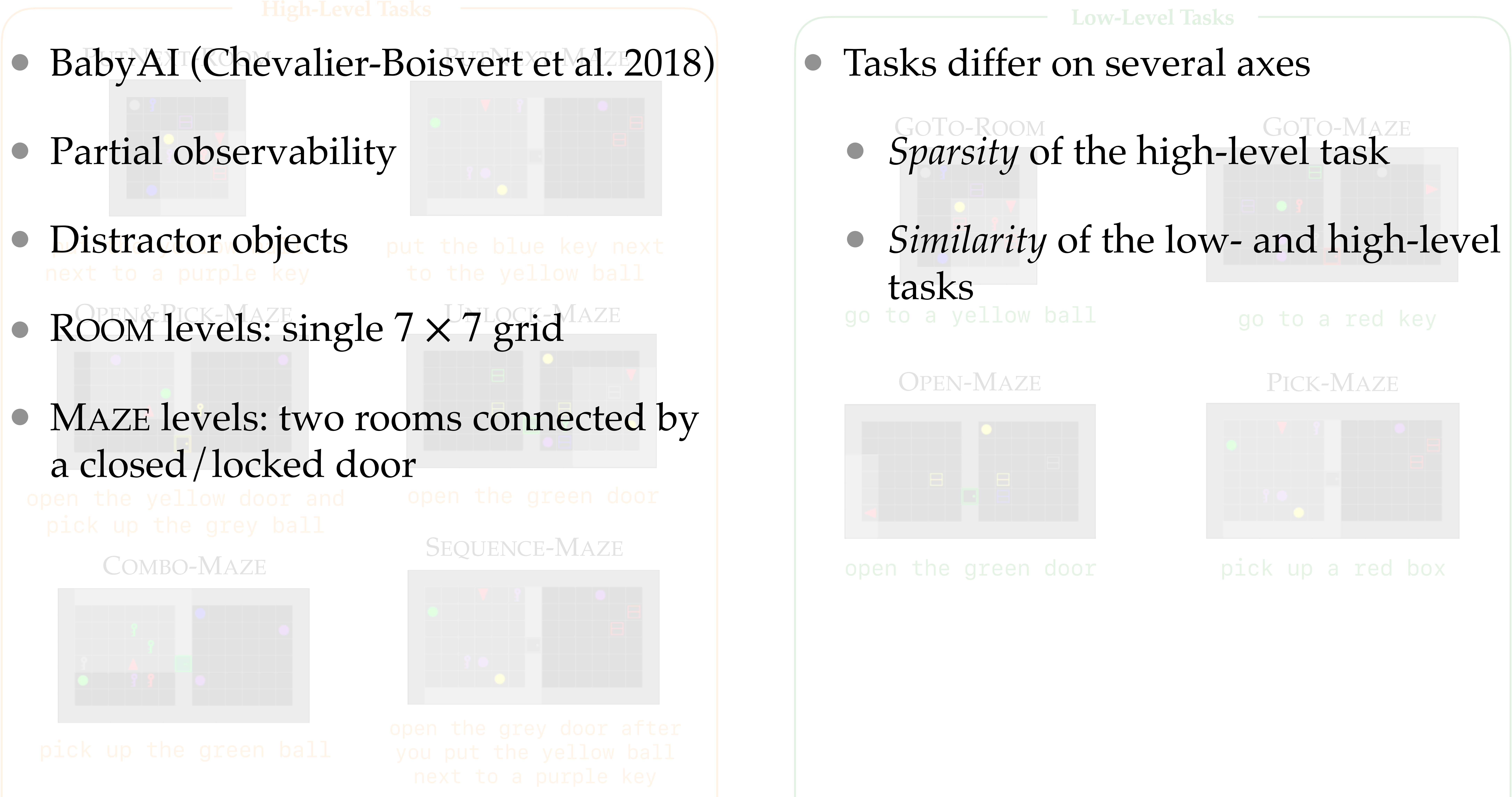
Experiments

- BabyAI (Chevalier-Boisvert et al. 2018)
- Partial observability
- Distractor objects
- ROOM levels: single 7×7 grid
- MAZE levels: two rooms connected by a closed / locked door

- Tasks differ on several axes
    - *Sparsity* of the high-level task



Experiments



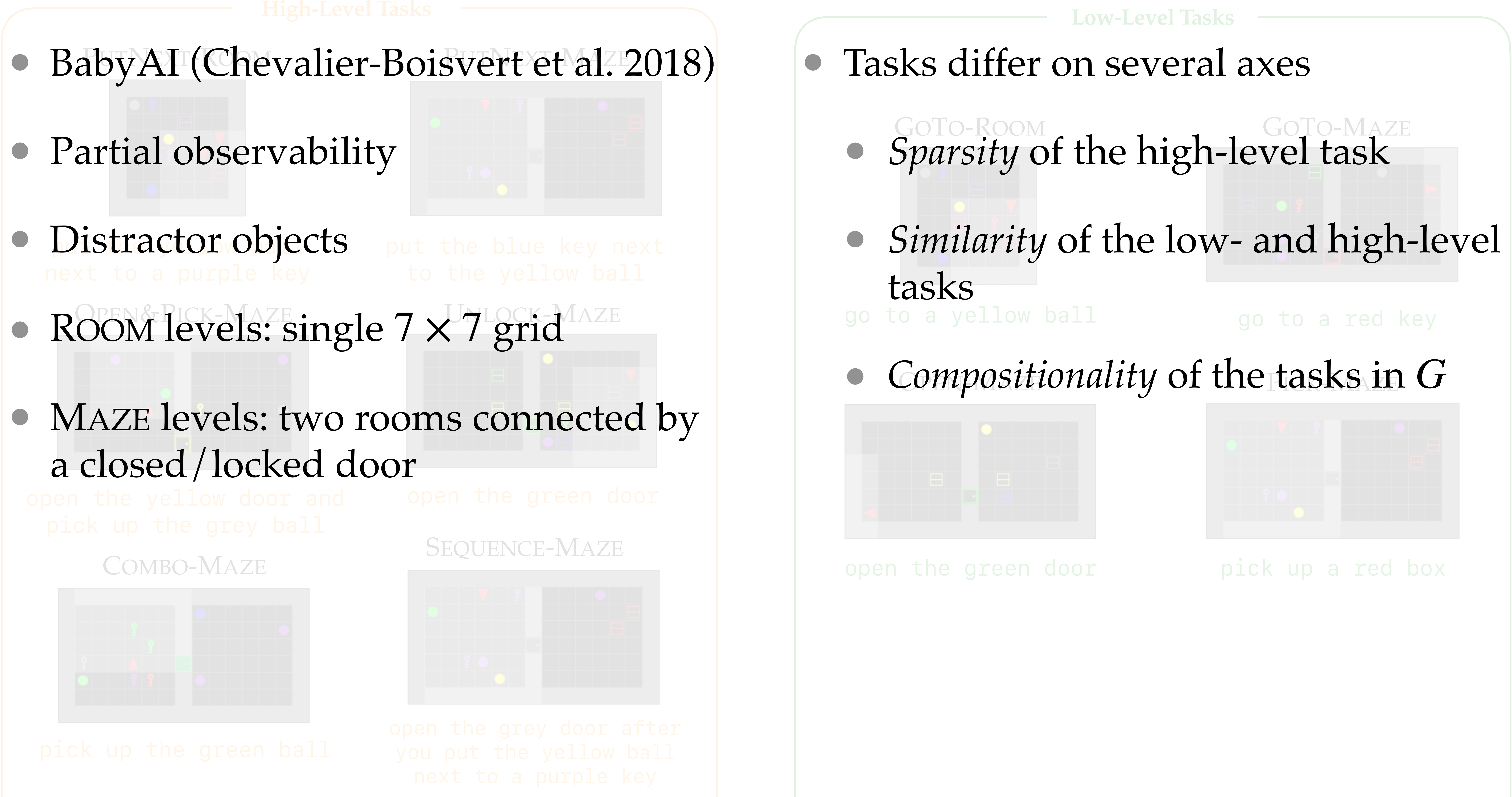
- BabyAI (Chevalier-Boisvert et al. 2018)
- Partial observability
- Distractor objects
- Room levels: single 7 × 7 grid
- Maze levels: two rooms connected by a closed / locked door

- Tasks differ on several axes
  - *Sparsity* of the high-level task
  - *Similarity* of the low- and high-level tasks

Experiments

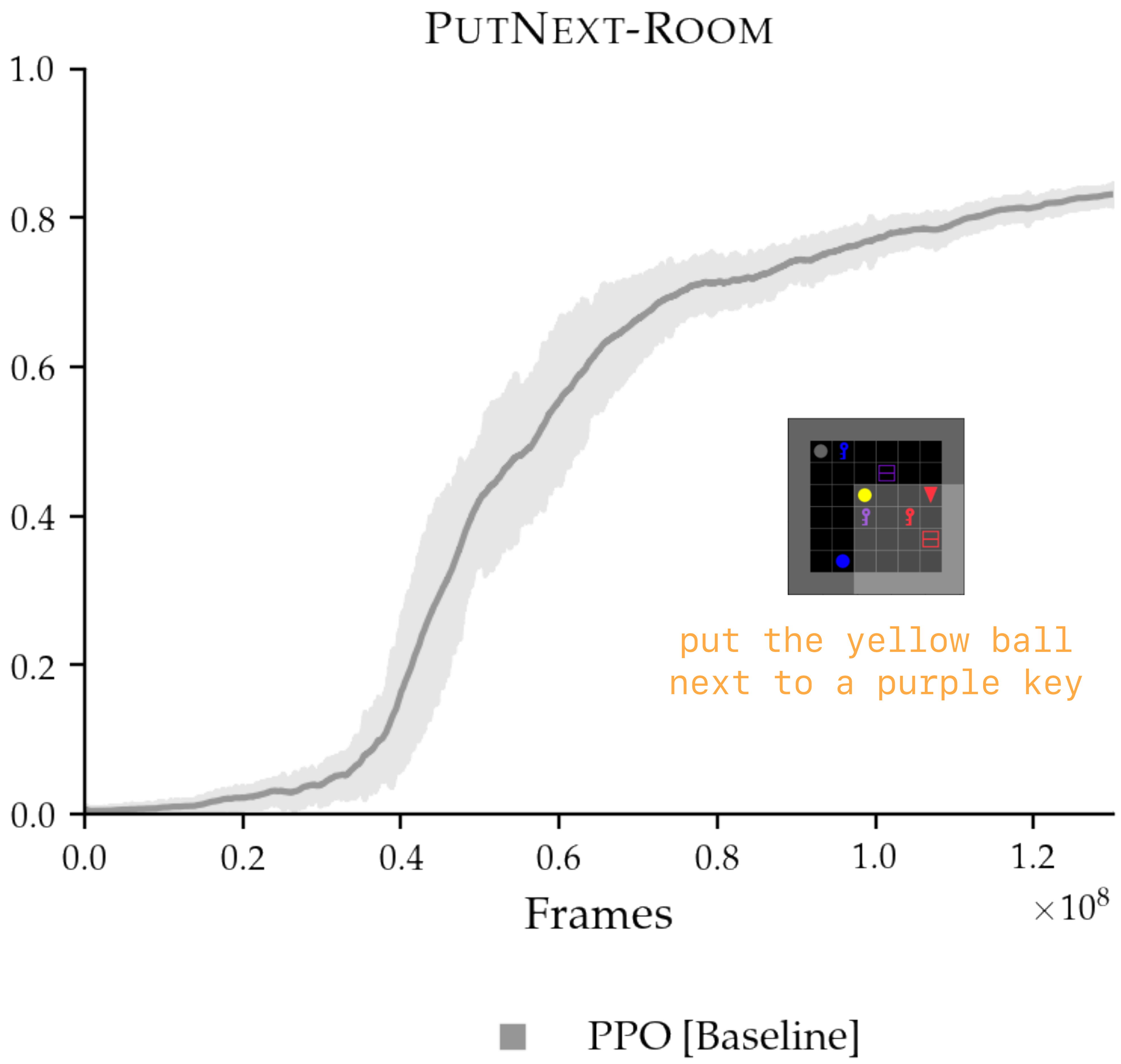
- BabyAI (Chevalier-Boisvert et al. 2018)
- Partial observability
- Distractor objects
- Room levels: single 7 × 7 grid
- Maze levels: two rooms connected by a closed / locked door

- Tasks differ on several axes
  - *Sparsity* of the high-level task
  - *Similarity* of the low- and high-level tasks
  - *Compositionality* of the tasks in G

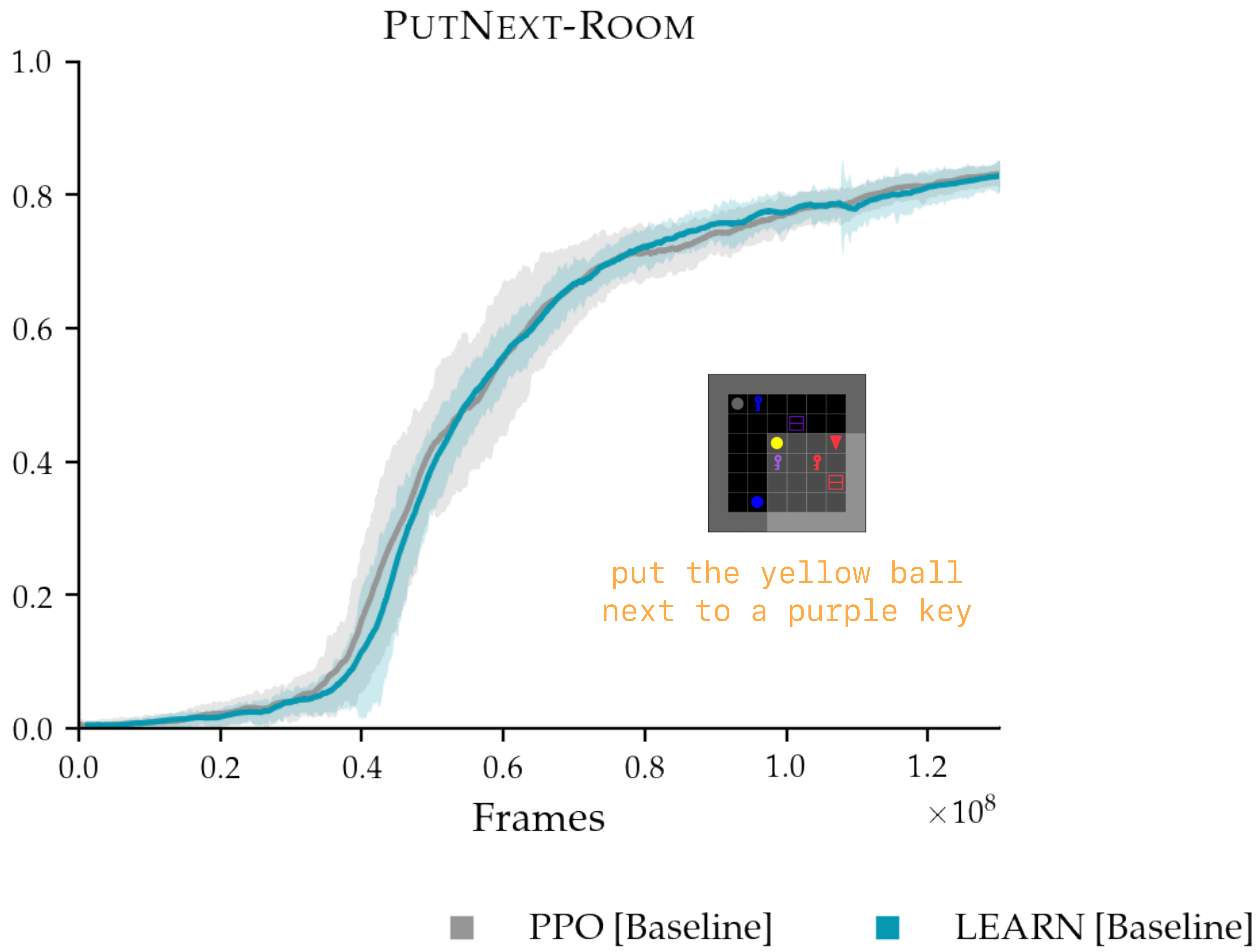


Results: Sparsity

Results: Sparsity

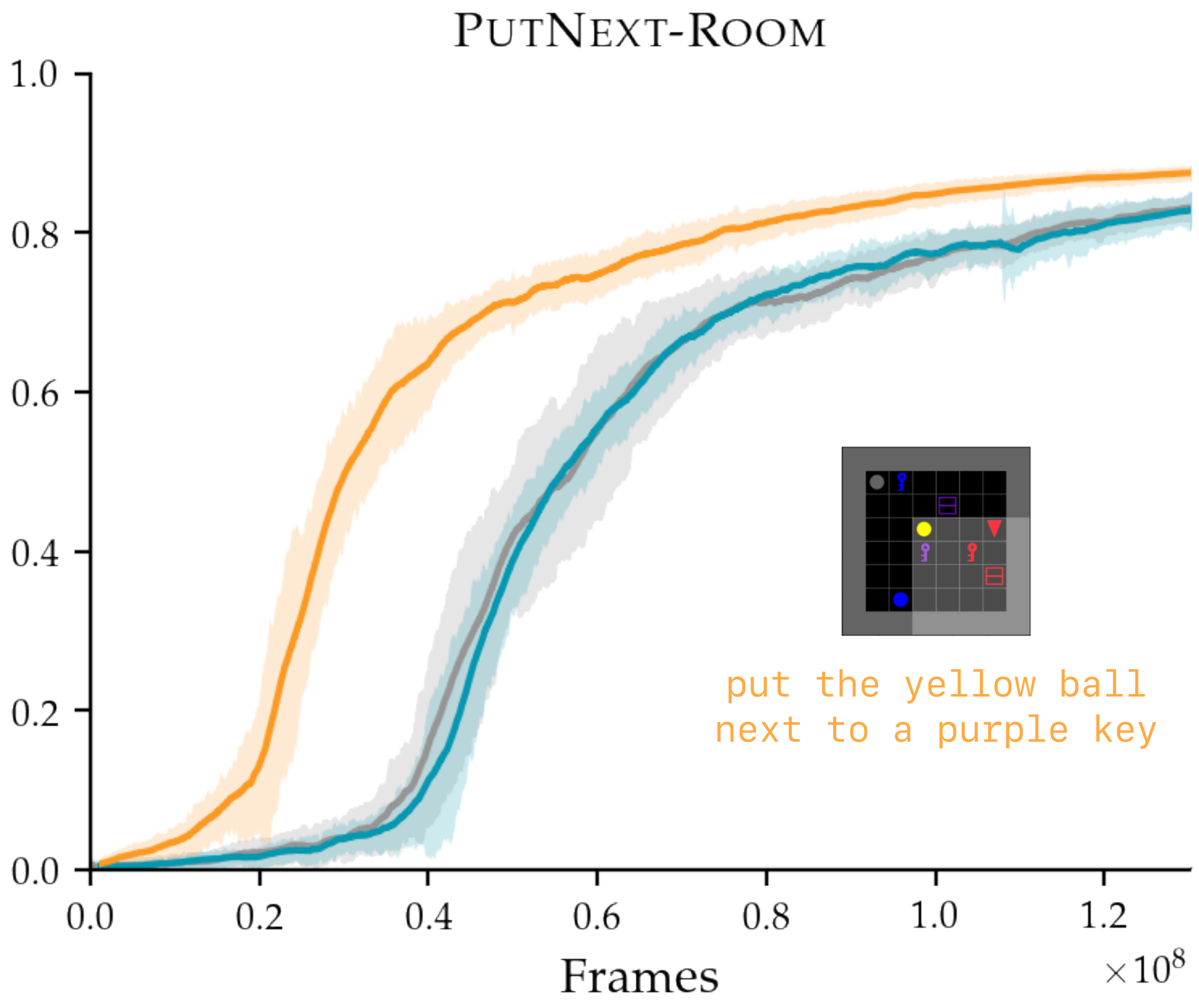


Results: Sparsity



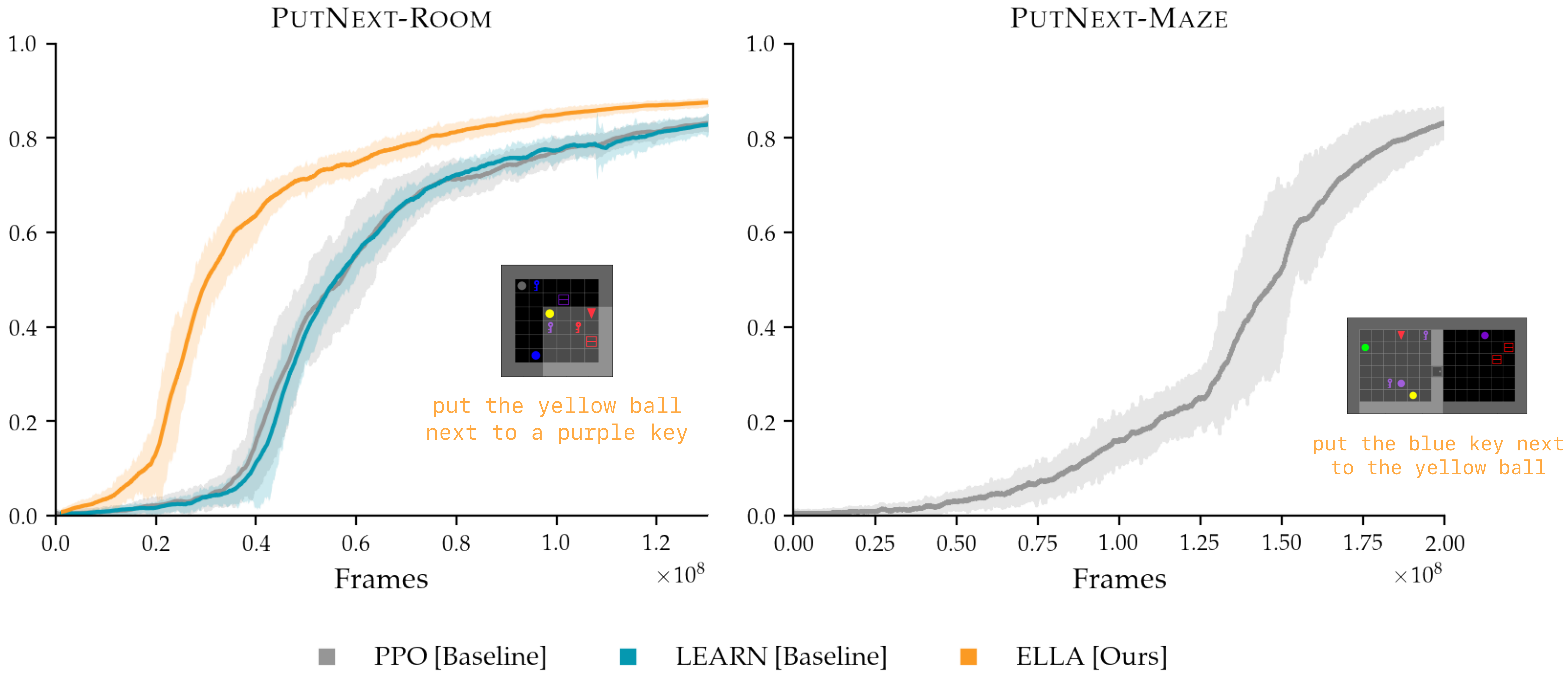
(Goyal et al. 2019)

Results: Sparsity



■ PPO [Baseline] ■ LEARN [Baseline] (Goyal et al. 2019) ■ ELLA [Ours]

Results: Sparsity

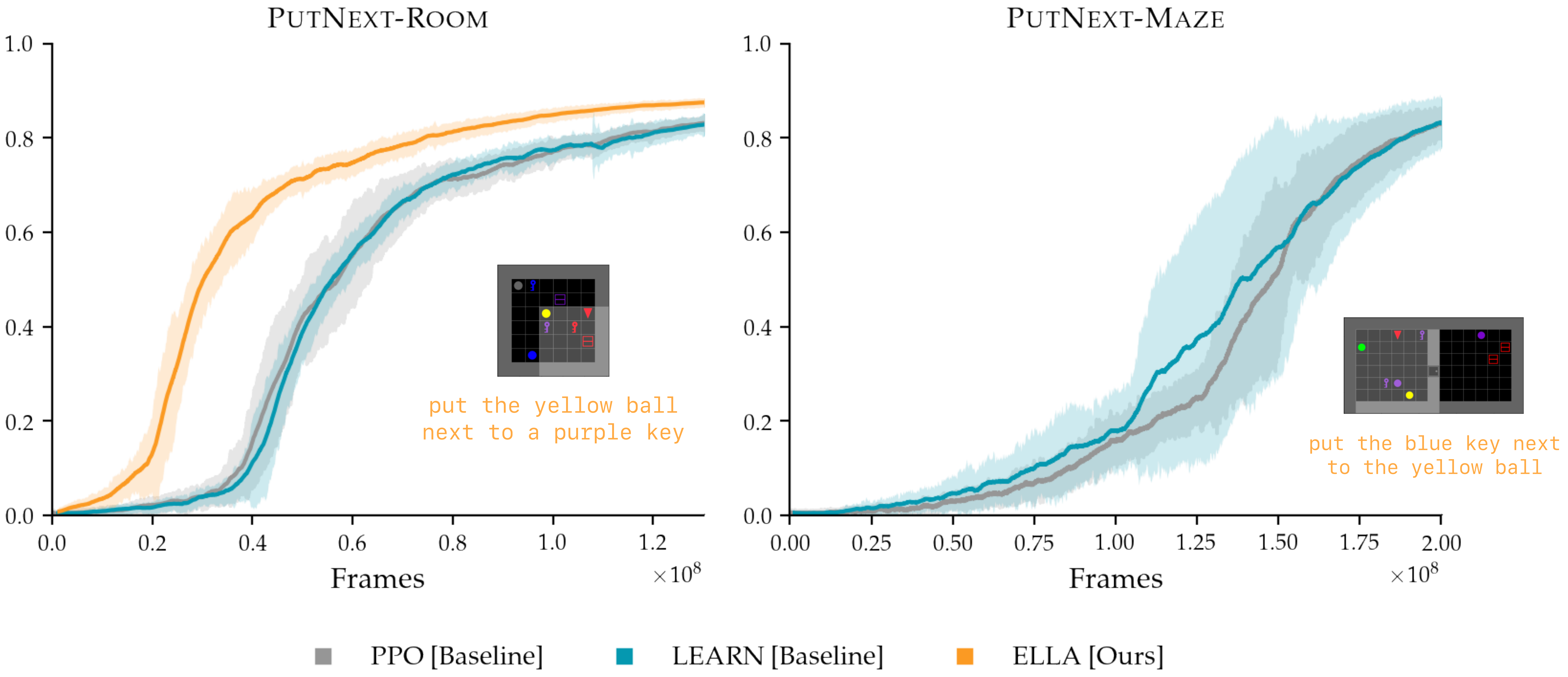


put the yellow ball next to a purple key

put the blue key next to the yellow ball

(Goyal et al. 2019)

Results: Sparsity

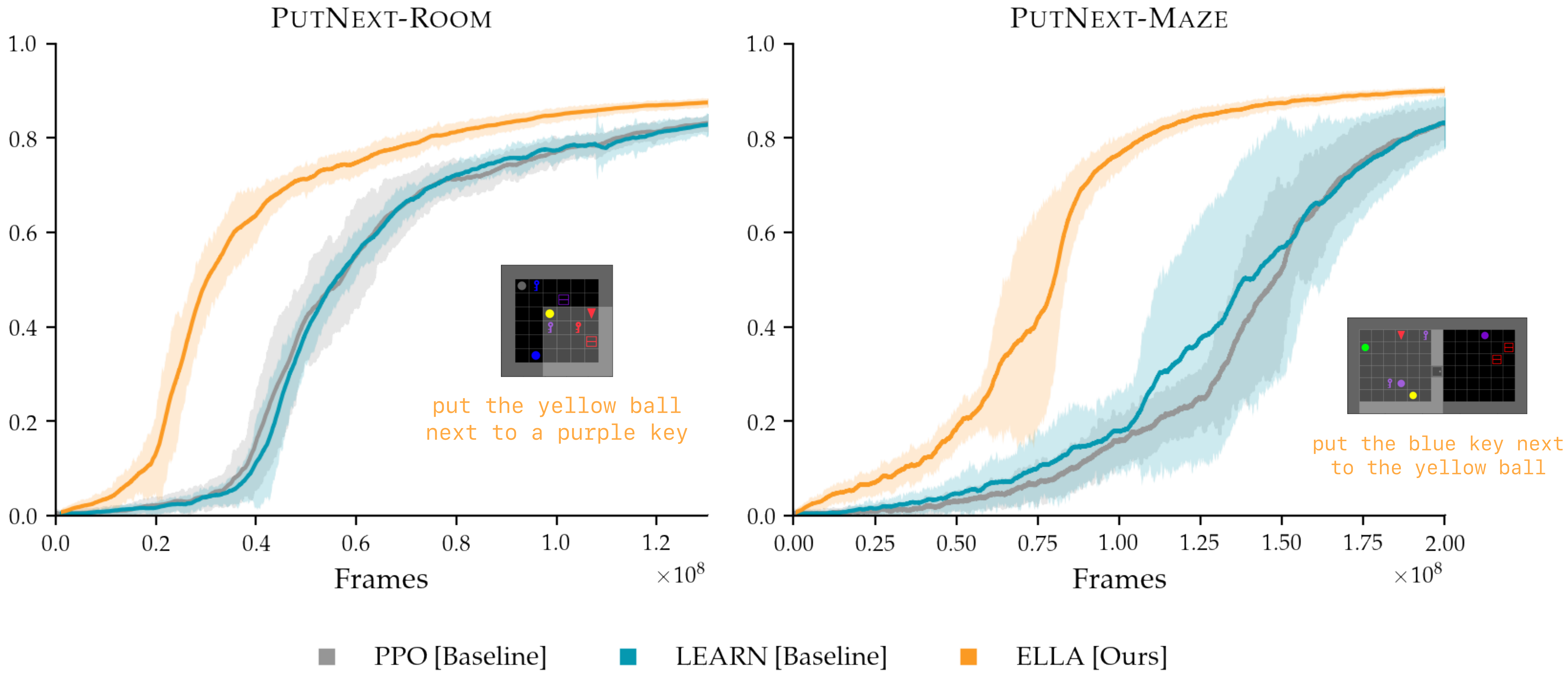


put the yellow ball next to a purple key

put the blue key next to the yellow ball

(Goyal et al. 2019)

Results: Sparsity

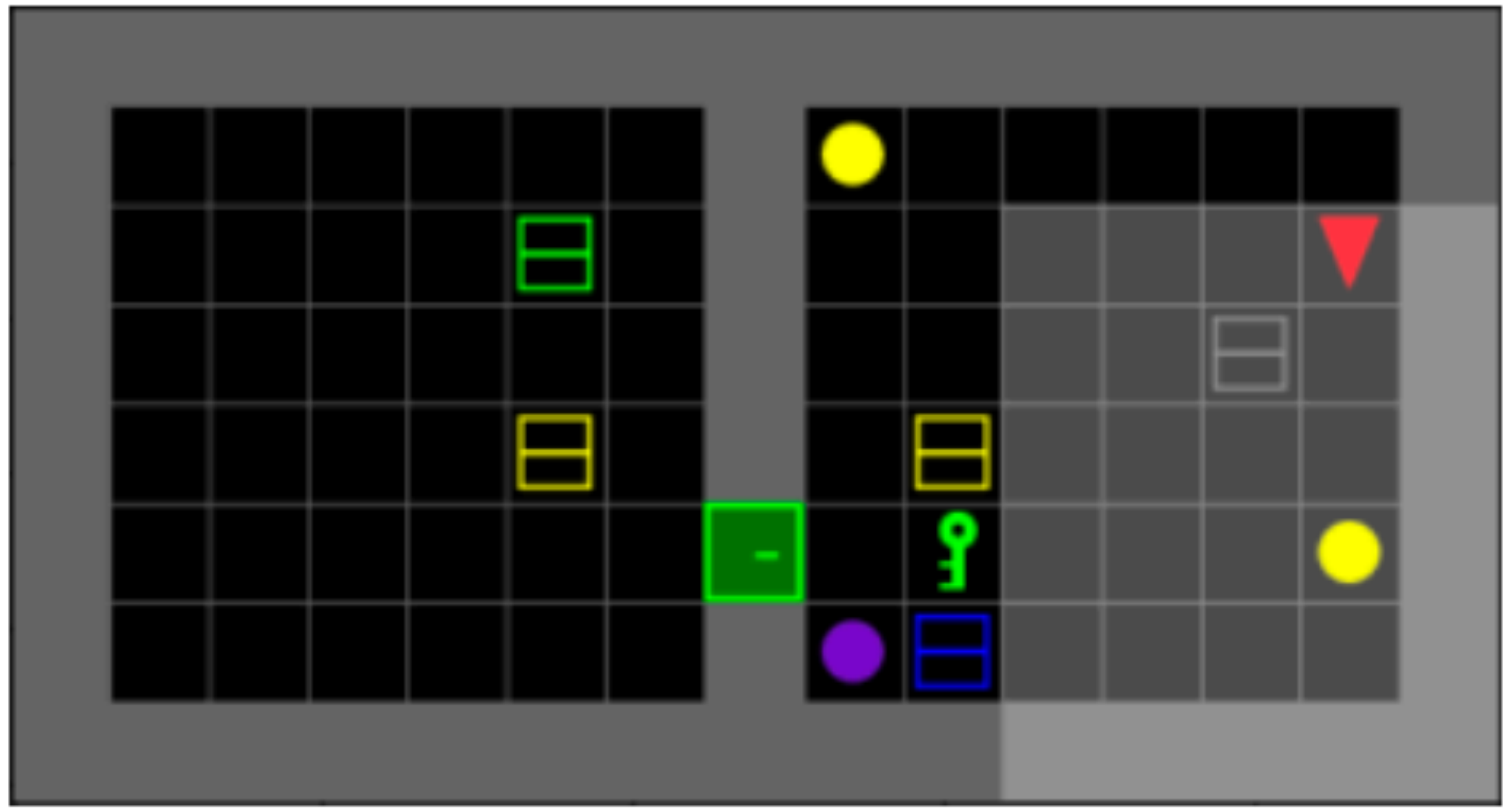


put the yellow ball next to a purple key

put the blue key next to the yellow ball

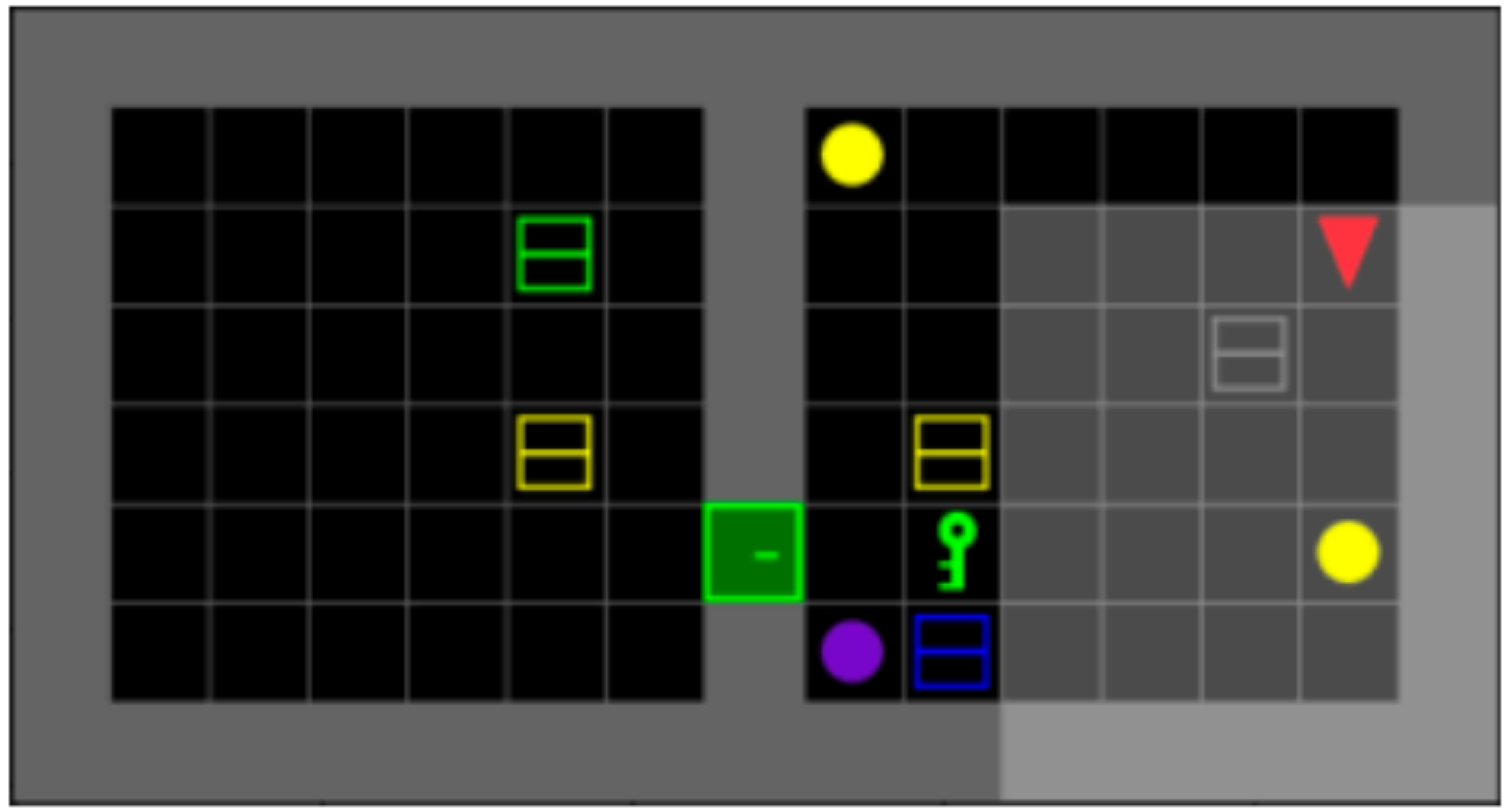
(Goyal et al. 2019)

Results: Sparsity

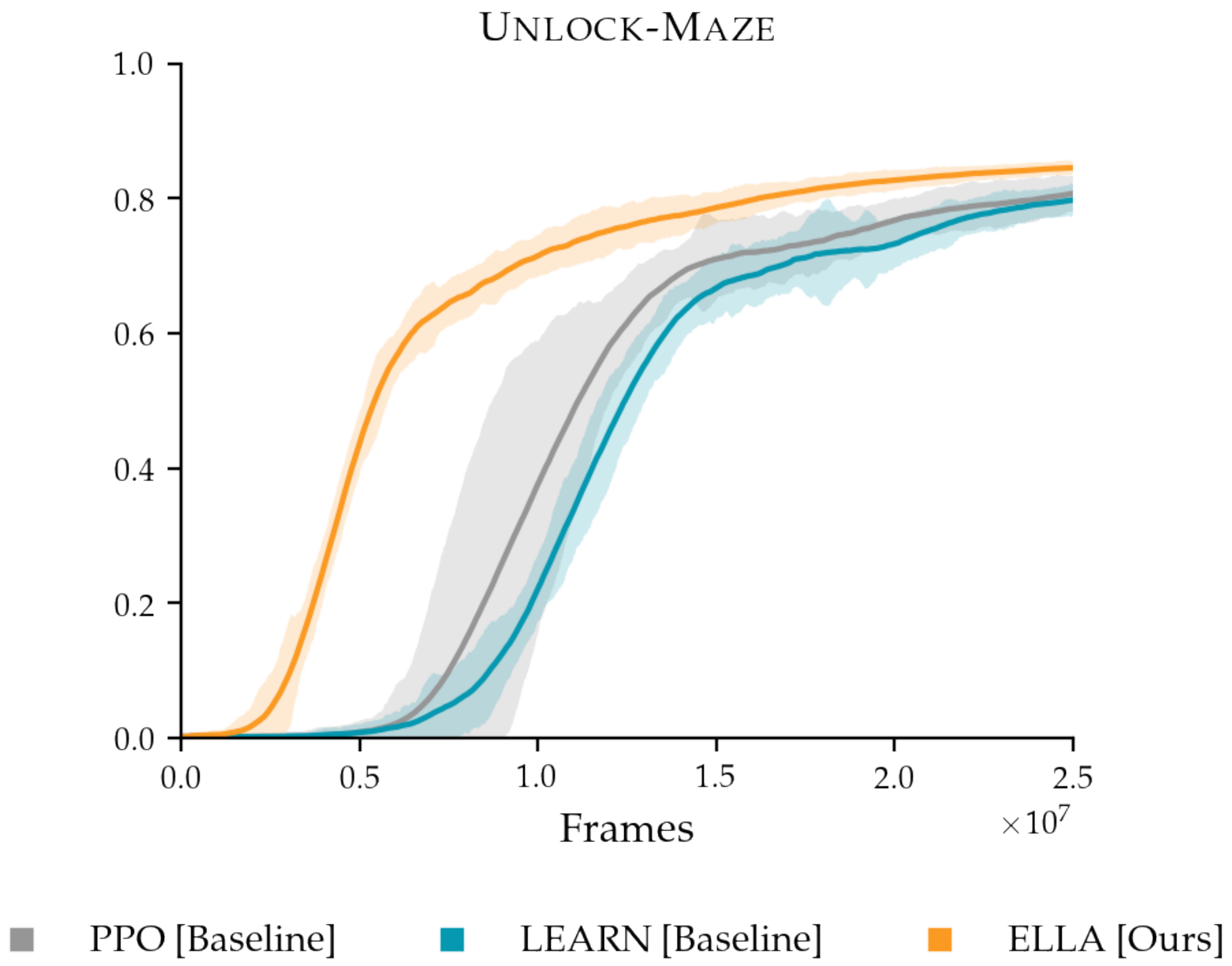


open the green door

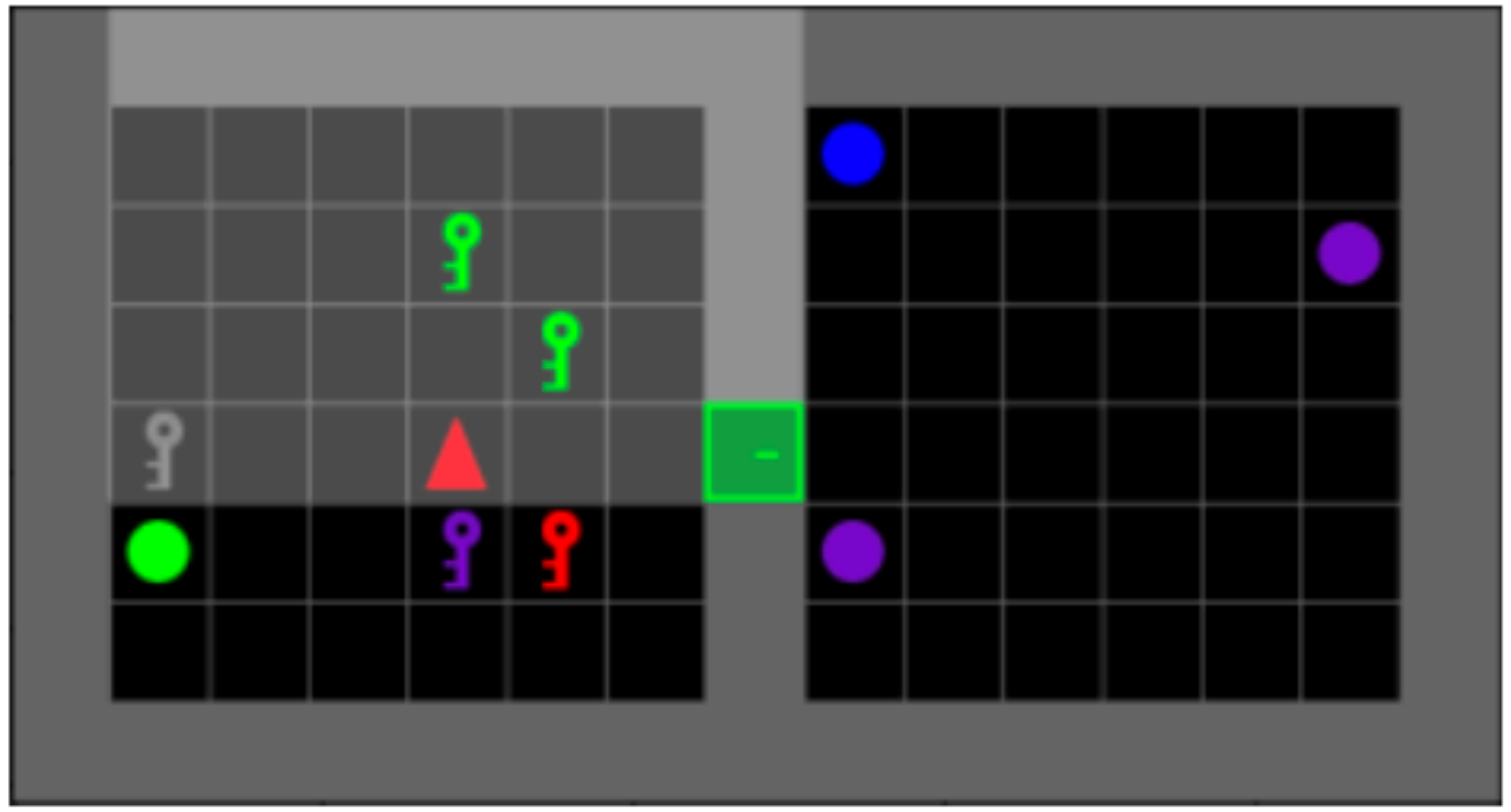
Results: Sparsity



open the green door



Results: Similarity

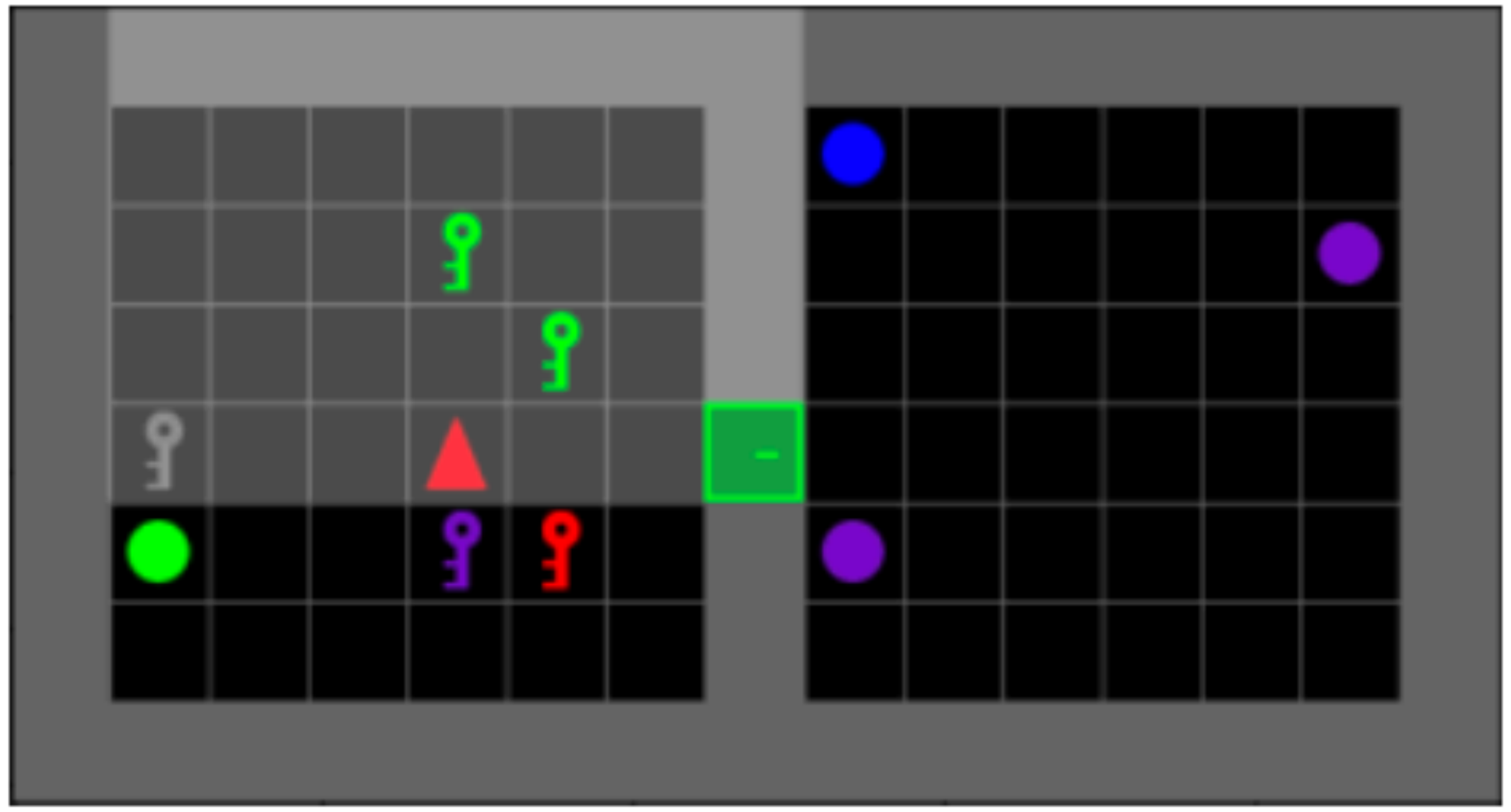


open the green door

pickup the red key

put the green ball next
to the purple ball

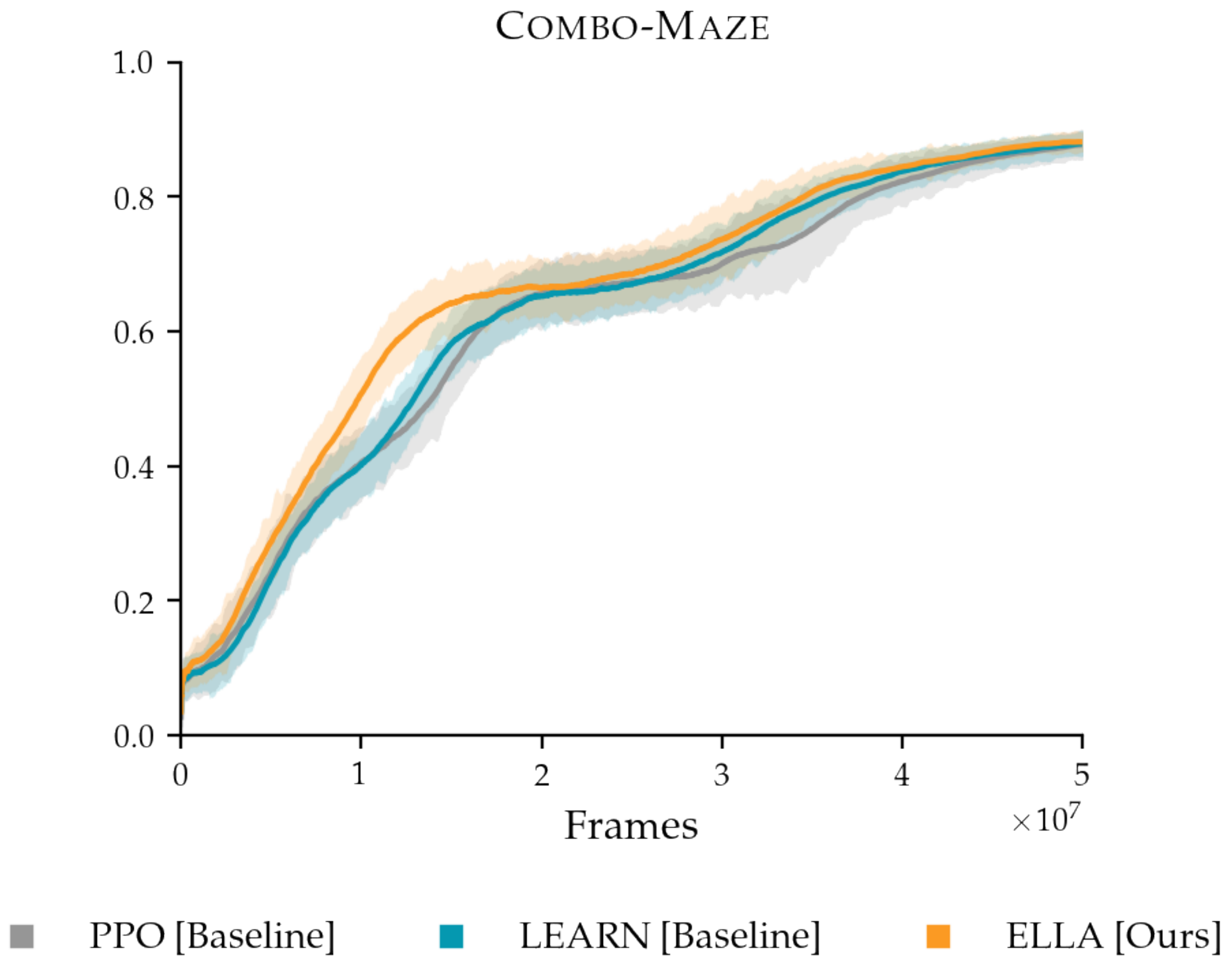
Results: Similarity



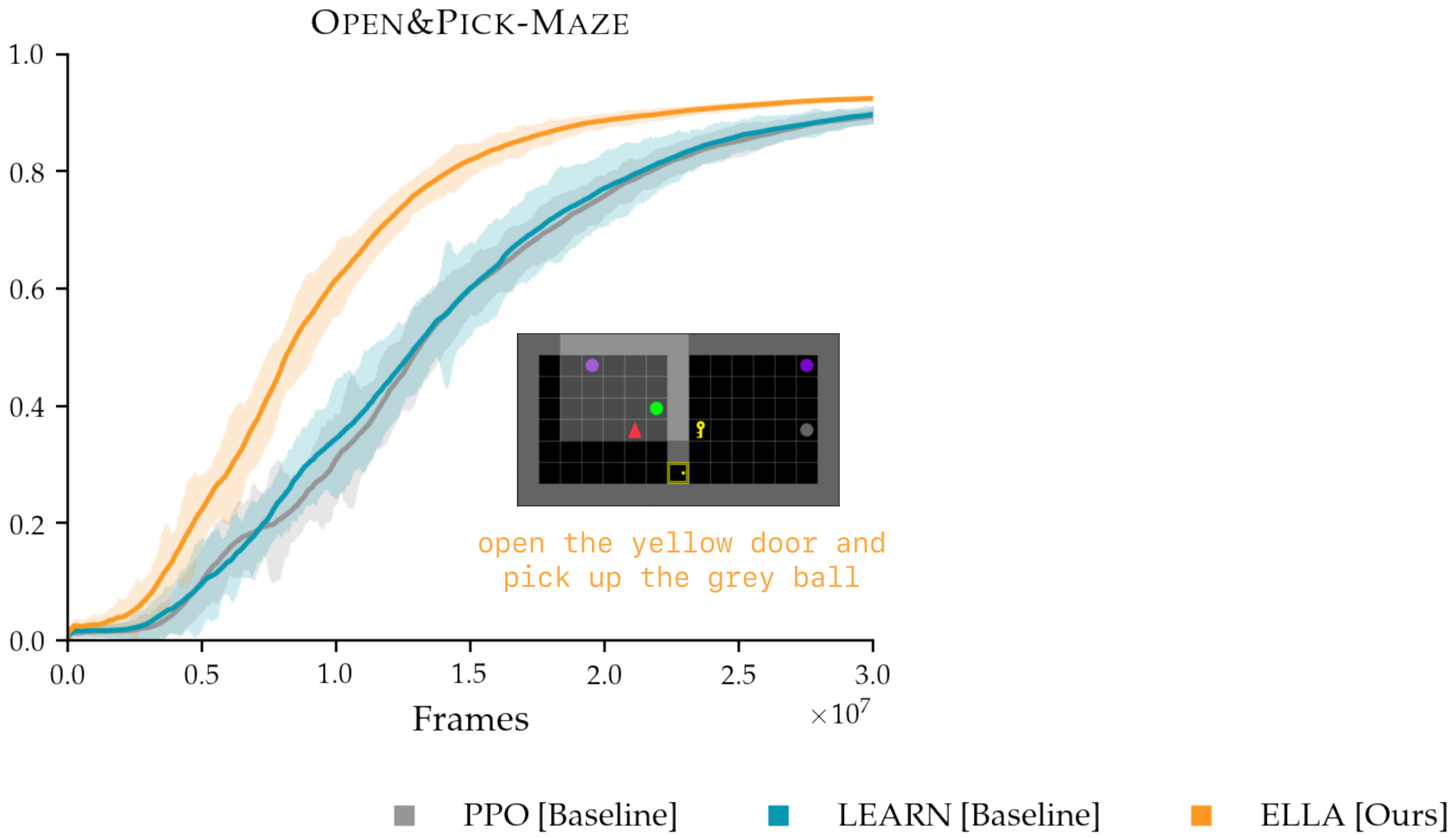
open the green door

pickup the red key

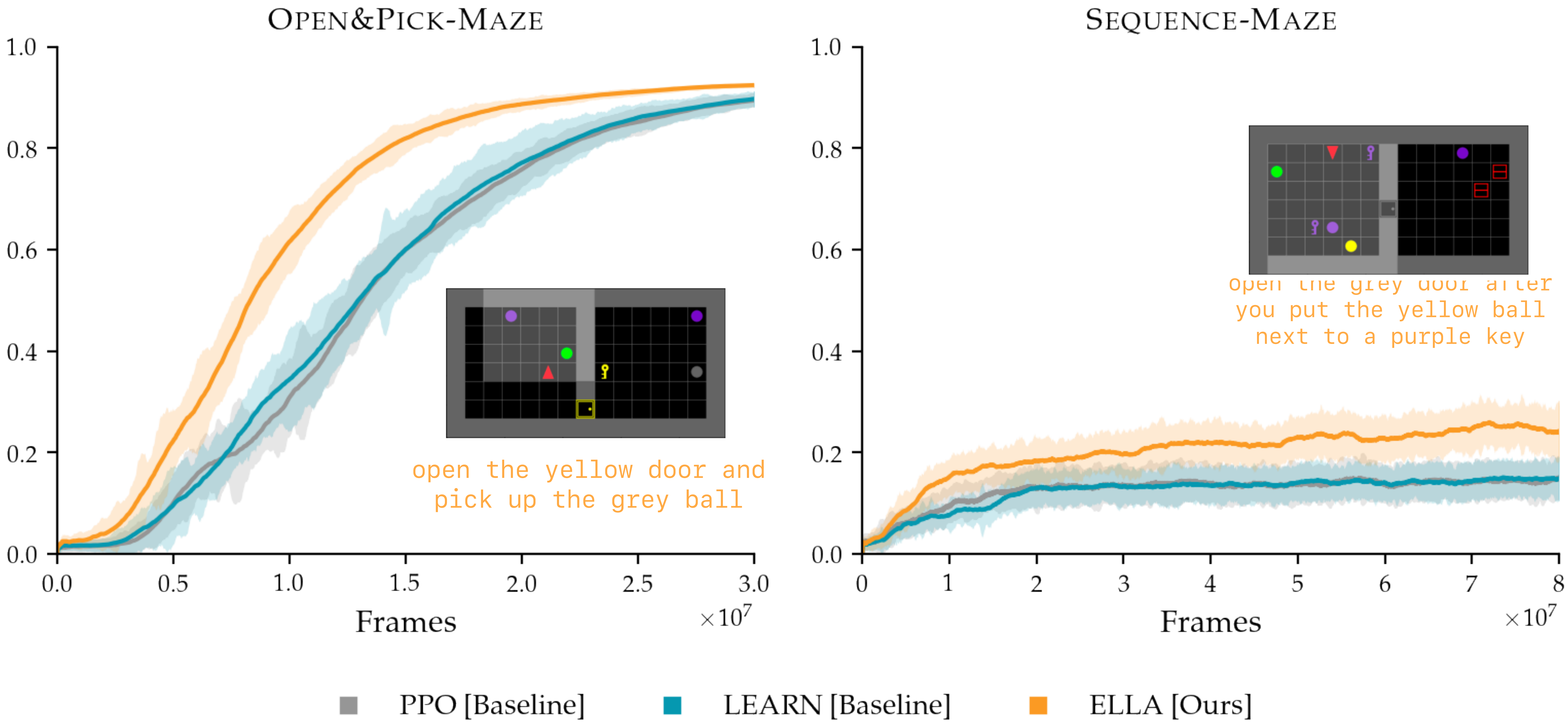
put the green ball next to the purple ball



Results: Compositionality



Results: Compositionality



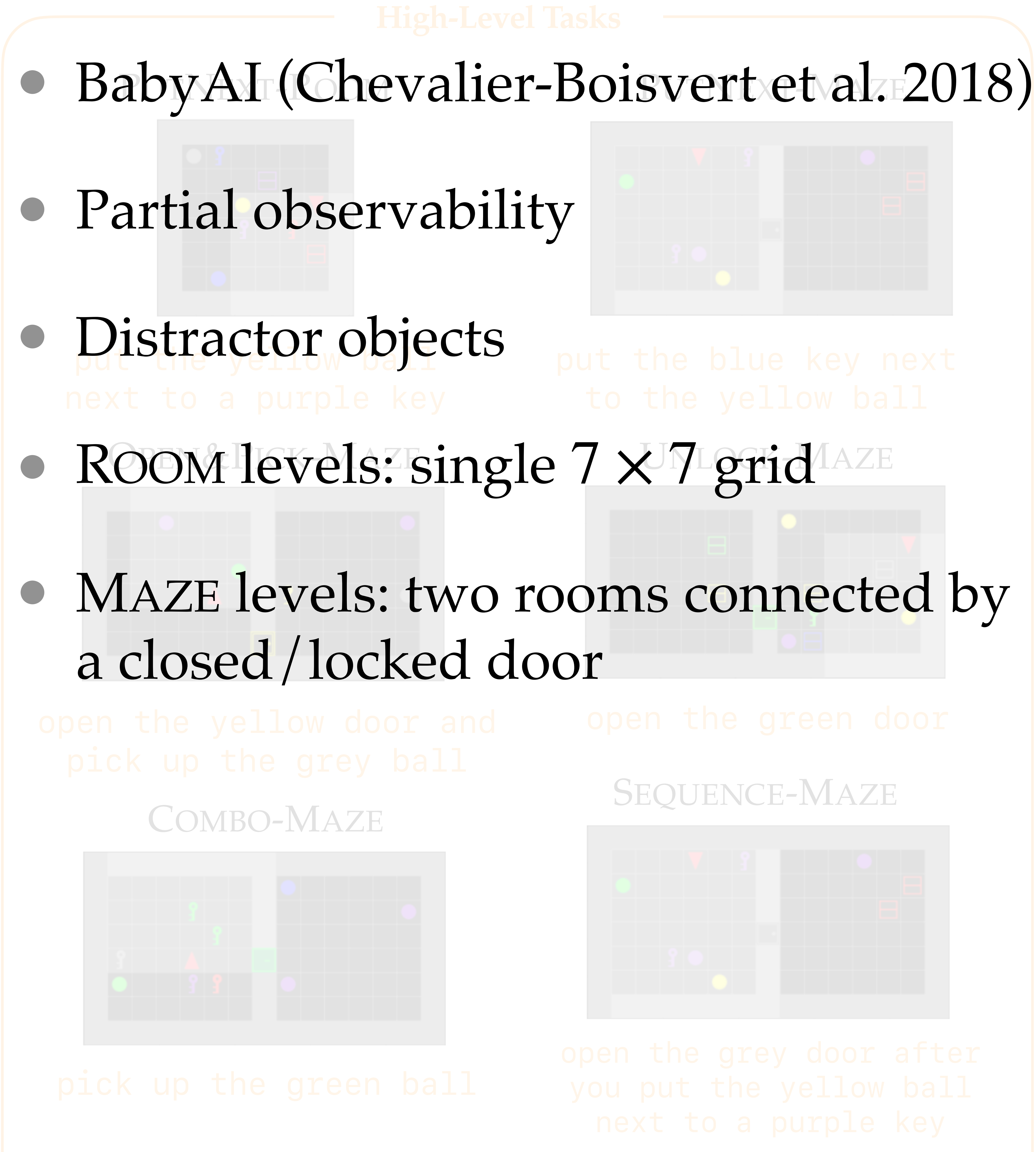
open the yellow door and pick up the grey ball

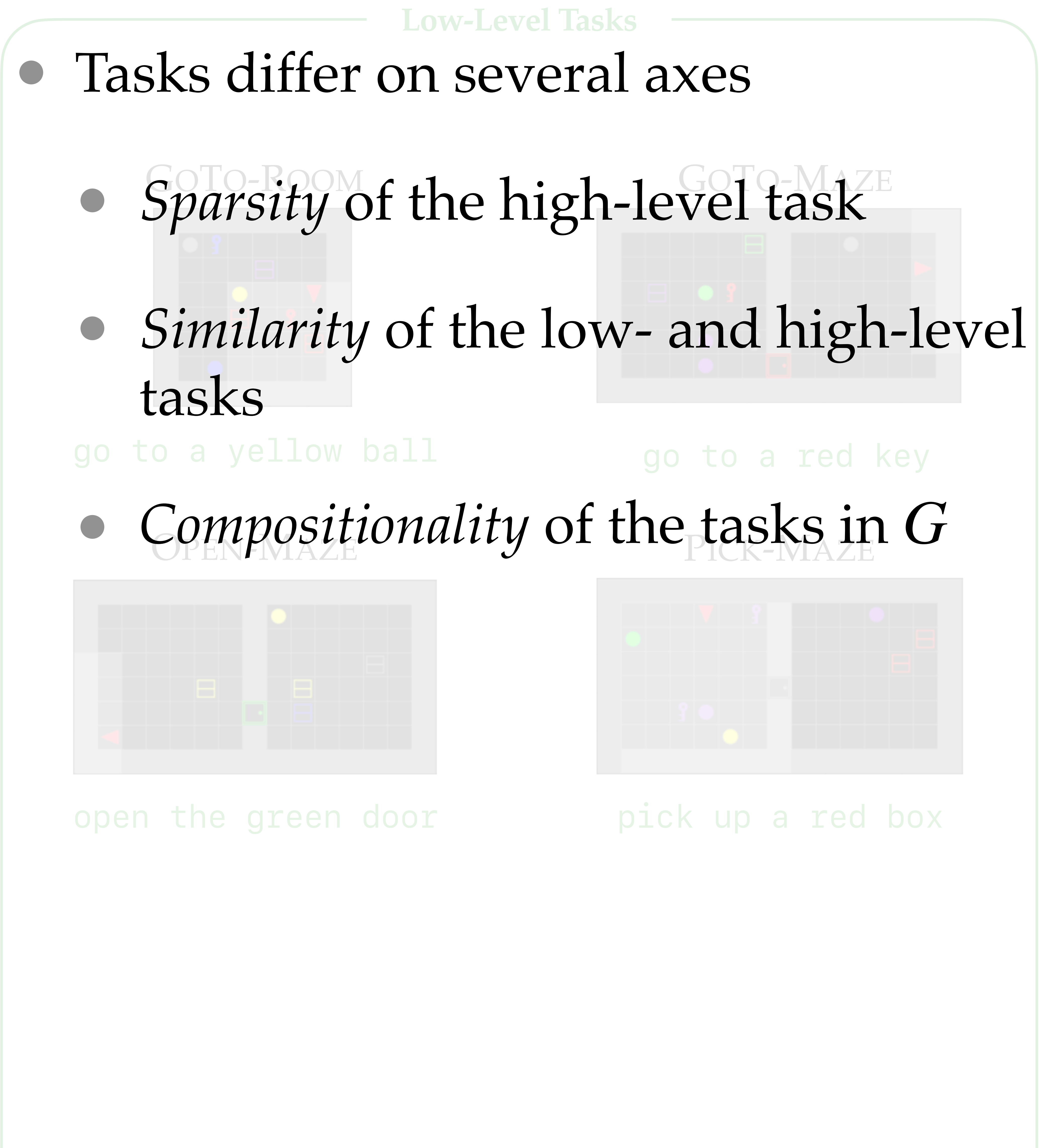
open the grey door after you put the yellow ball next to a purple key

Experiments

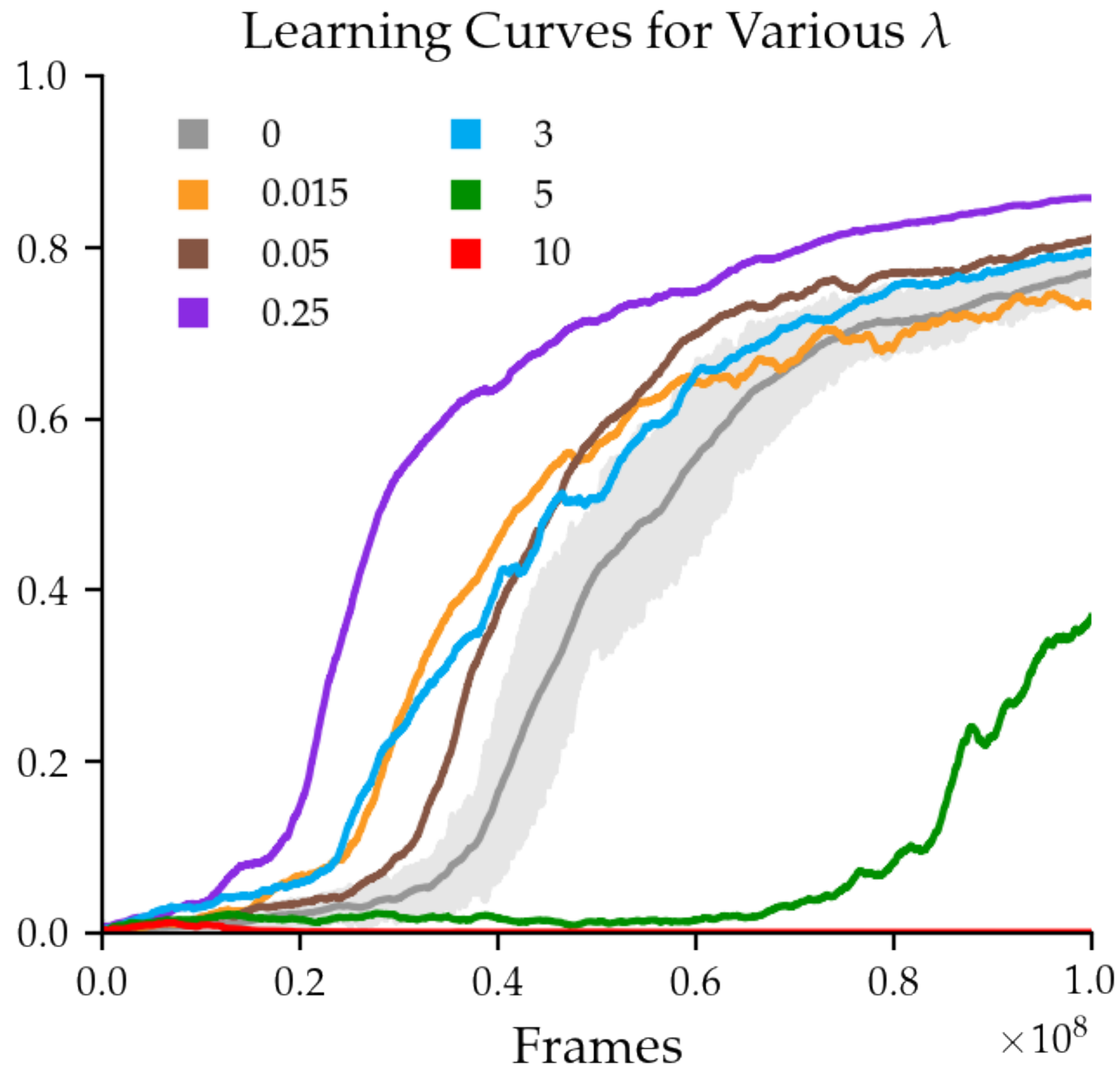
- BabyAI (Chevalier-Boisvert et al. 2018)
- Partial observability
- Distractor objects
- ROOM levels: single 7 × 7 grid
- MAZE levels: two rooms connected by a closed / locked door

- Tasks differ on several axes
  - *Sparsity* of the high-level task
  - *Similarity* of the low- and high-level tasks
  - *Compositionality* of the tasks in G


High-Level Tasks
PutNext-Room
PutNext-Maze
put the yellow ball next to a purple key
put the blue key next to the yellow ball
Open&Pick-Maze
Unlock-Maze
open the yellow door and pick up the grey ball
open the green door
Combo-Maze
Sequence-Maze
pick up the green ball
open the grey door after you put the yellow ball next to a purple key

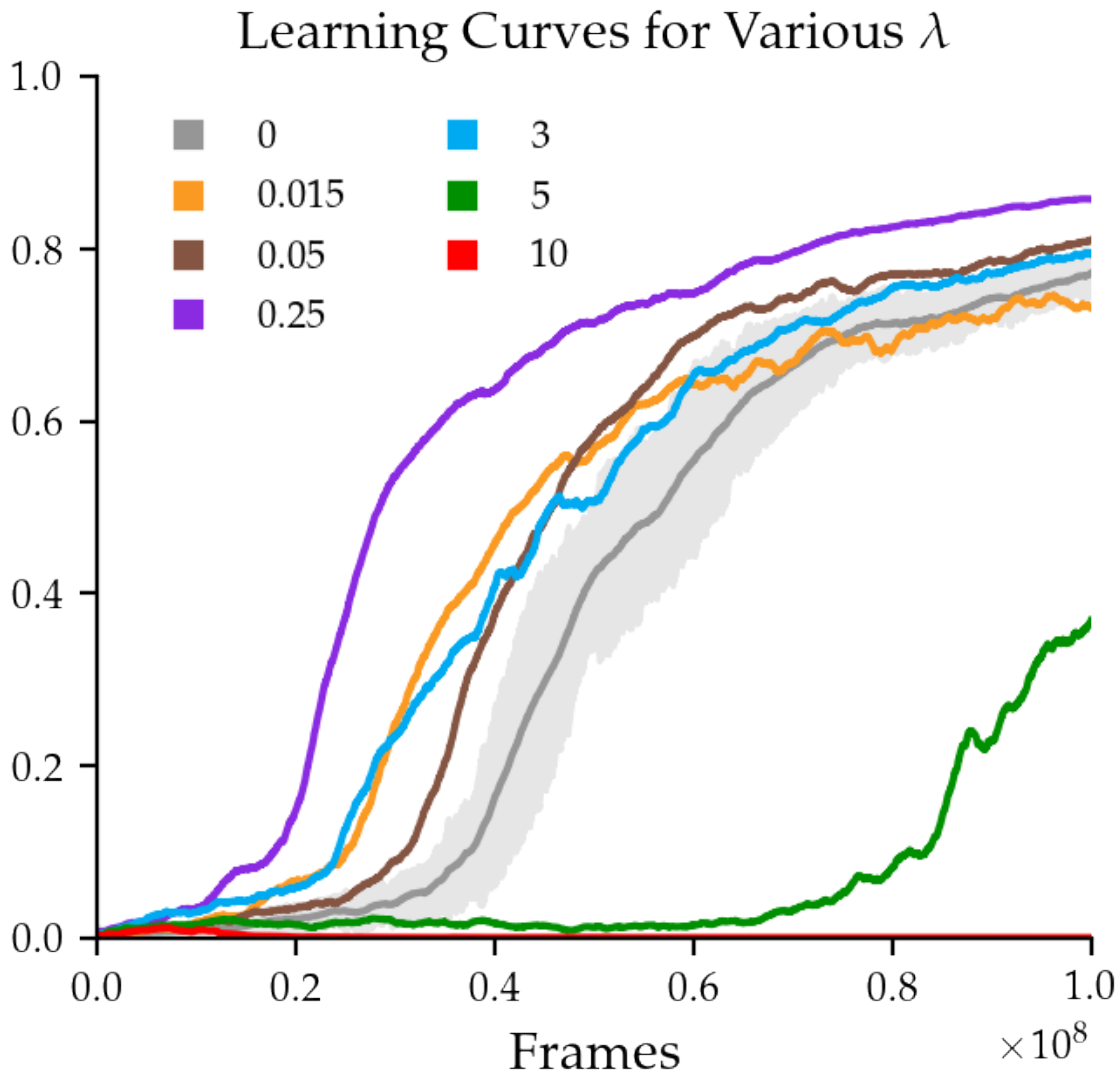


Effect of λ



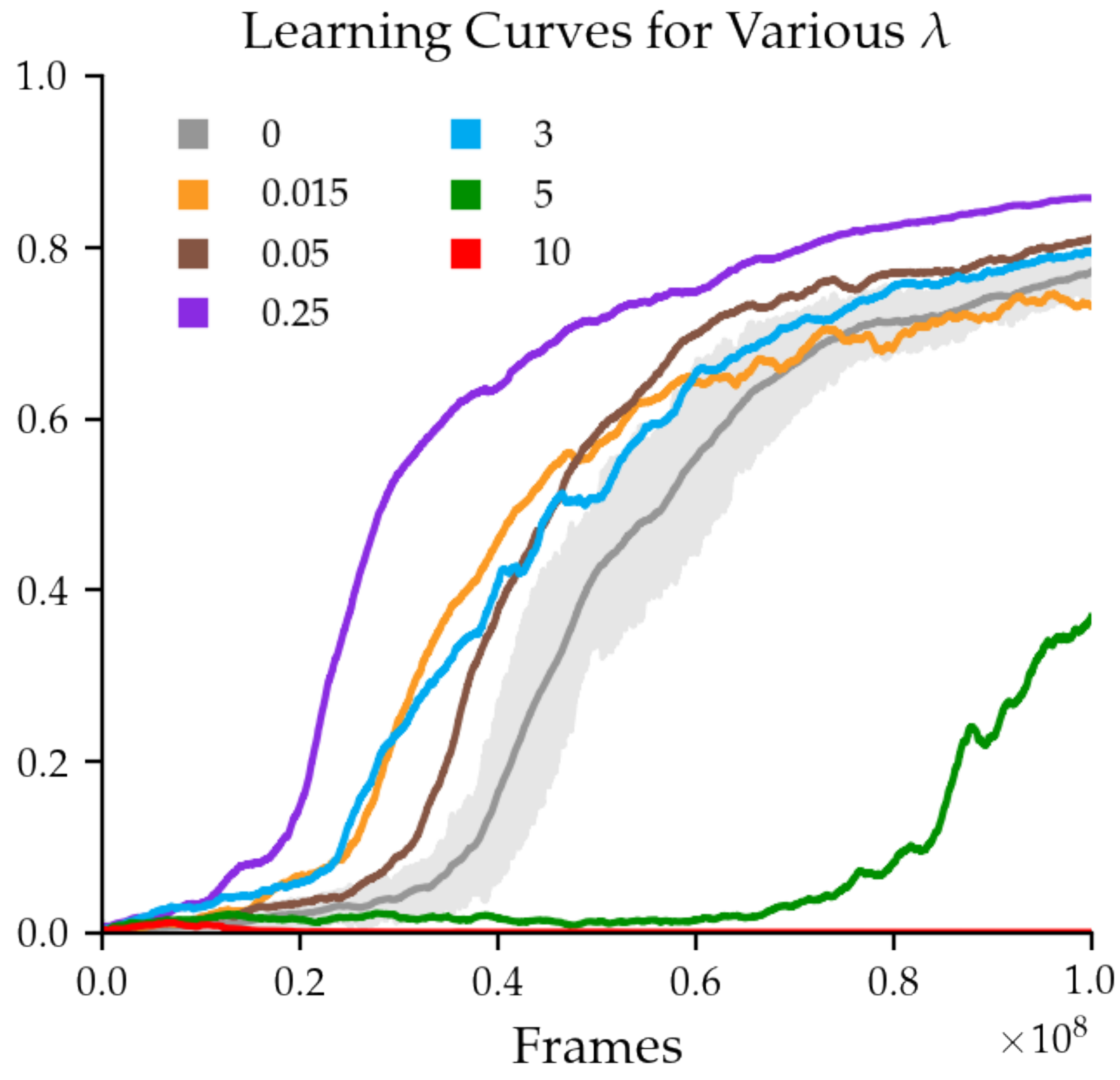
Effect of λ

- Set λ intelligently to increase sample efficiency

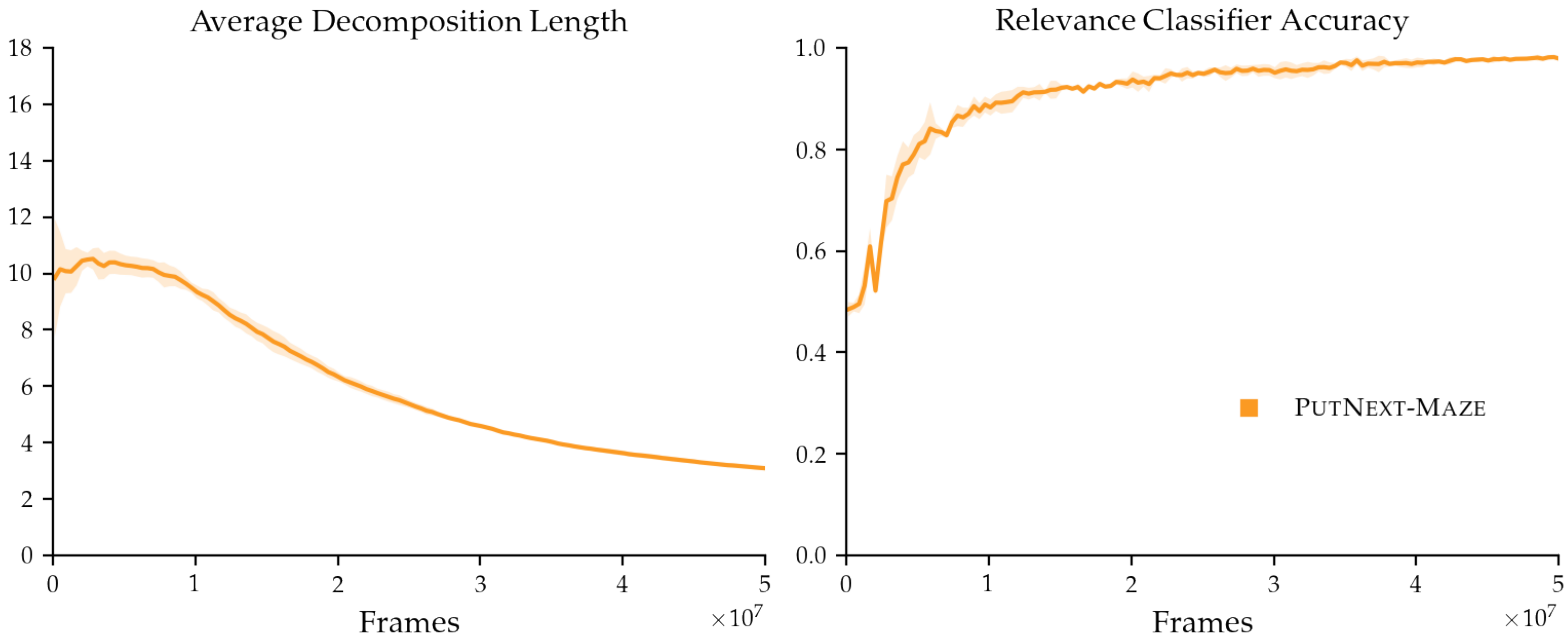


Effect of λ

- Set λ intelligently to increase sample efficiency
- Large values of λ lead to unstable learning



Relevance Classifier Performance



Relevance Classifier Performance

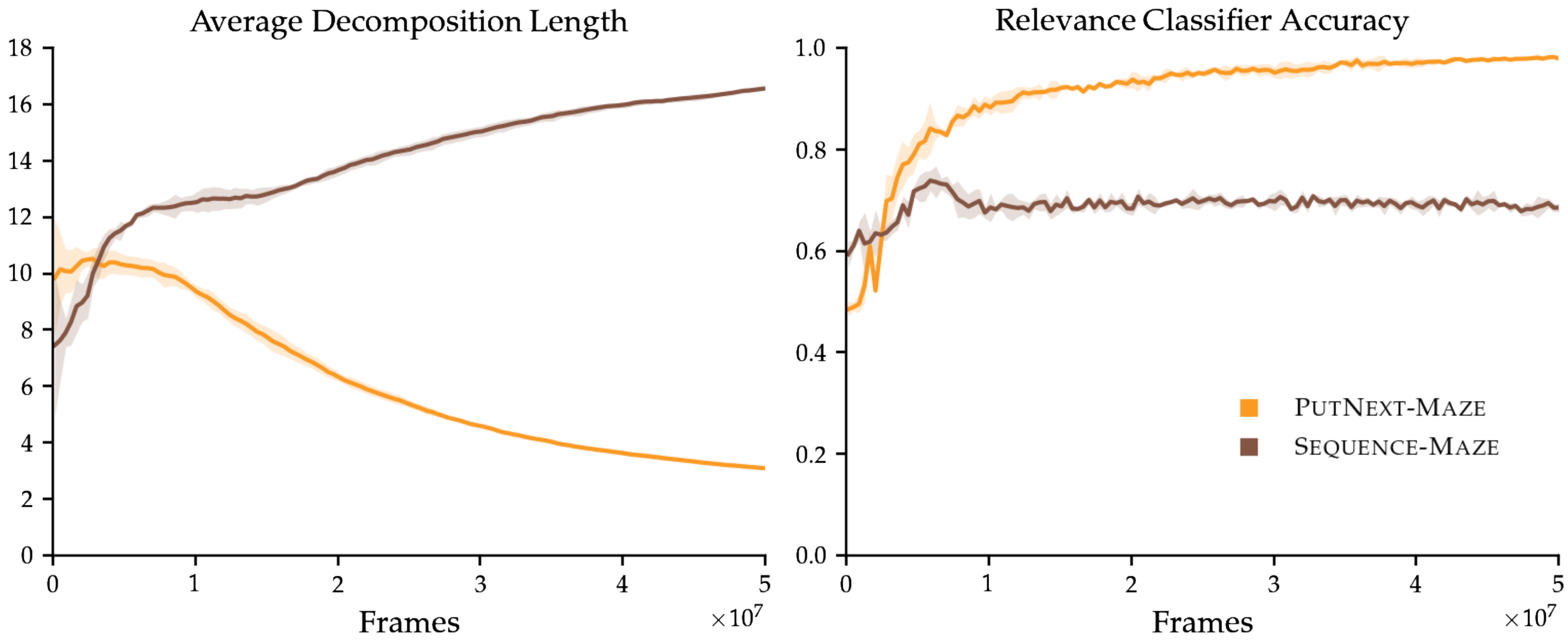
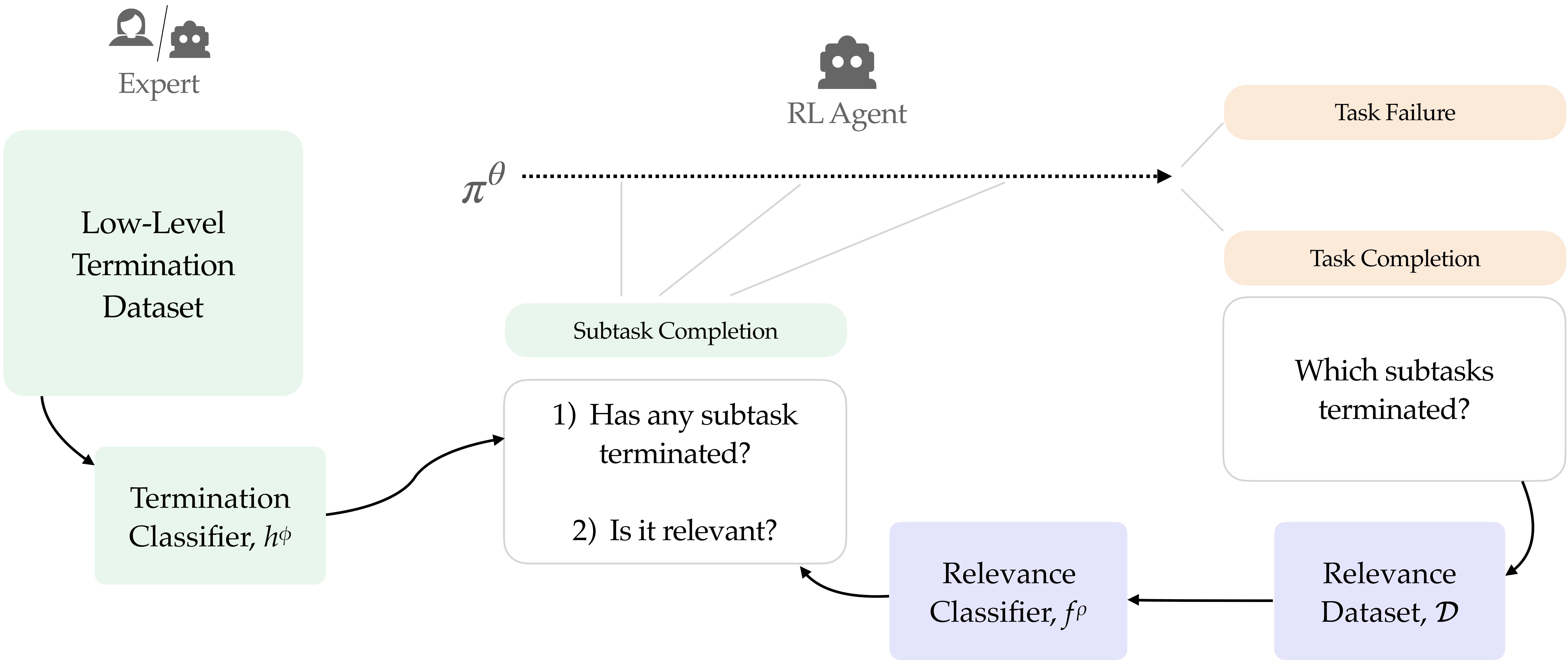


Image: Szot et al. 2021

ELLA



Expert

Low-Level Termination Dataset

Termination Classifier, h^{ϕ}

RL Agent

π^{θ}

Task Failure

Task Completion

Subtask Completion

1) Has any subtask terminated?

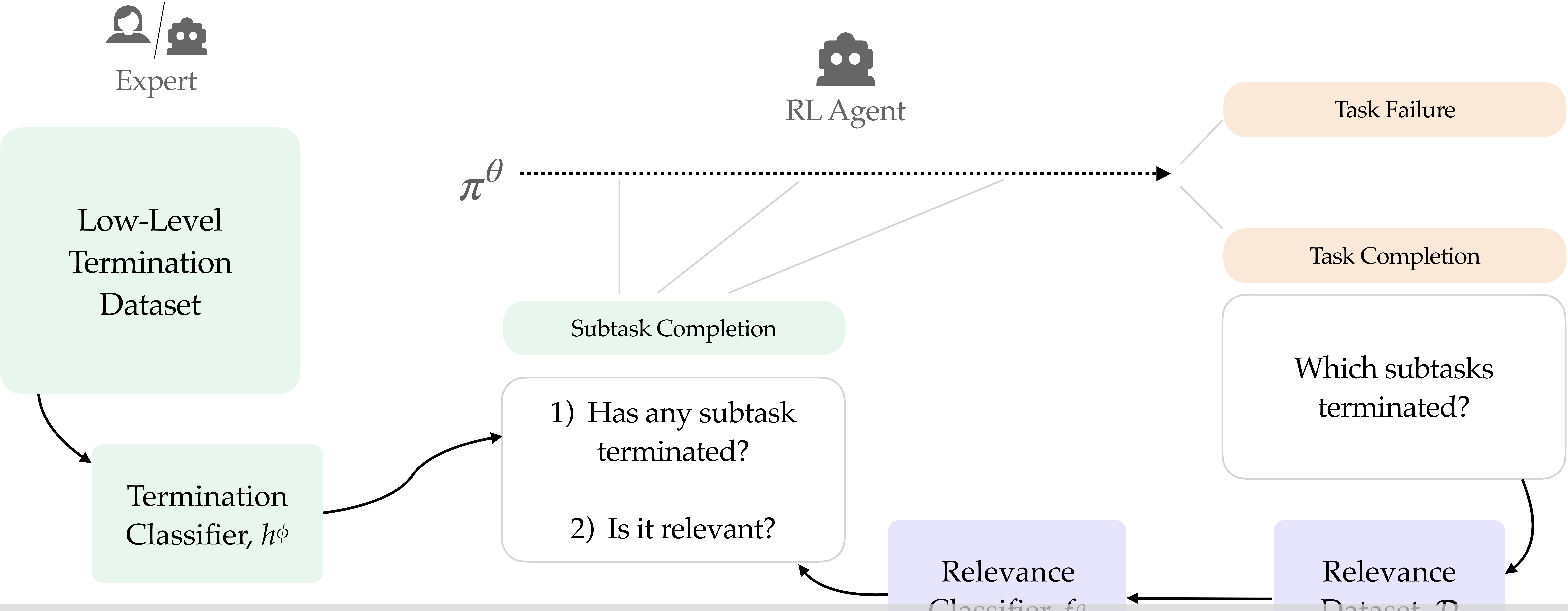
2) Is it relevant?

Which subtasks terminated?

Relevance Dataset, $\mathcal{D}$

Relevance Classifier, f^{ρ}

ELLA



Expert

Low-Level Termination Dataset

Termination Classifier, h^ϕ

RL Agent

π^θ

Task Failure

Task Completion

Subtask Completion

1) Has any subtask terminated?

2) Is it relevant?

Which subtasks terminated?

Relevance Classifier, f^ρ

Relevance Dataset, $\mathcal{D}$