

Language as a scaffold for RL

Jacob Andreas

MIT CSAIL

LINGO.csail.mit.edu

Language as a scaffold for RL

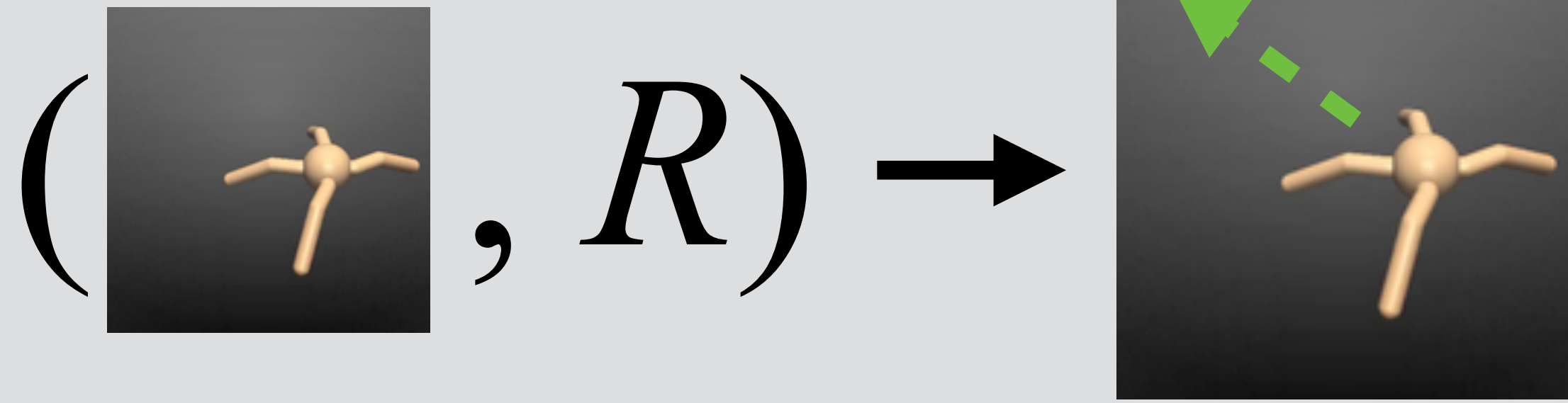
(what can language do for reinforcement learning?)

Jacob Andreas

MIT CSAIL

LINGO.csail.mit.edu

An NLP'er's view of RL

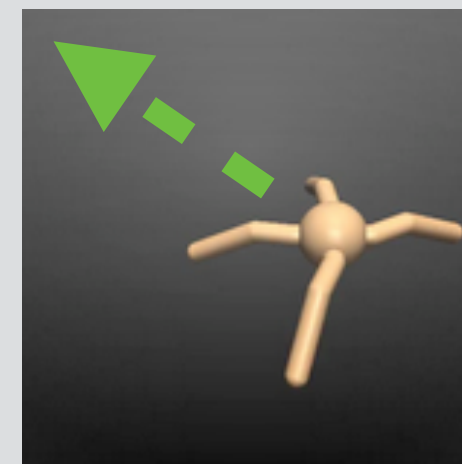


memorize 1 reward fn

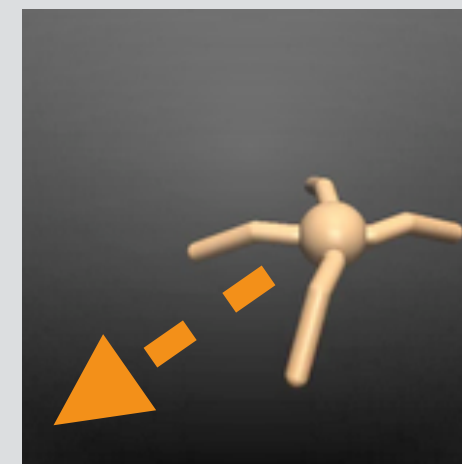
An NLP'er's view of RL

$(\text{Image}, R) \rightarrow$
Learn to accomplish
new goals

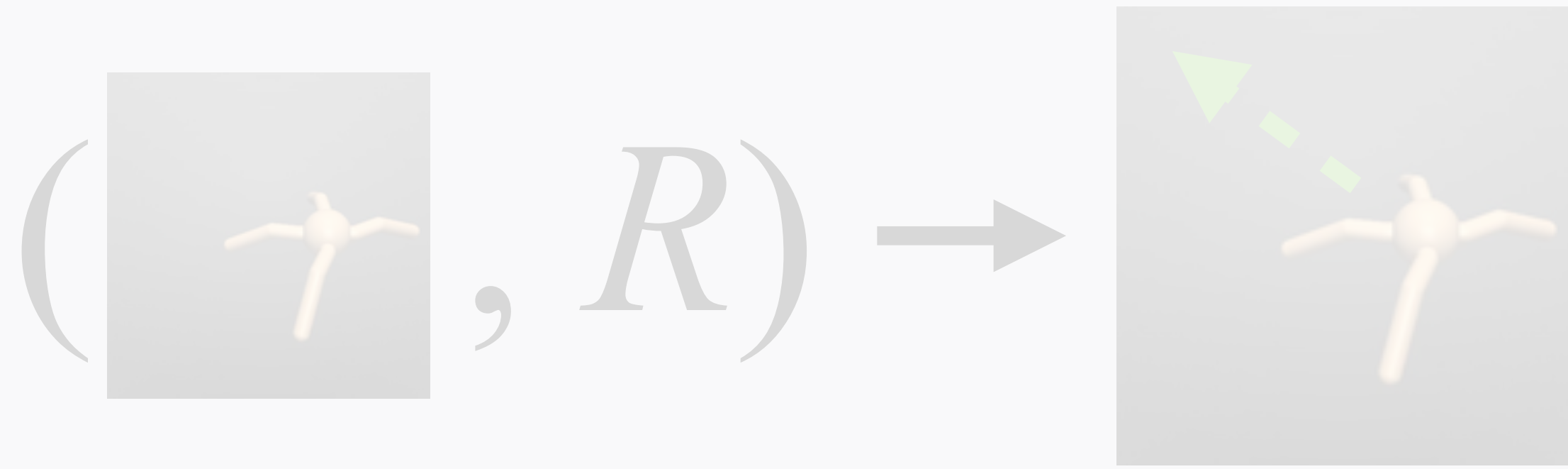
(Image, R_1)
 $(-2, 3)$



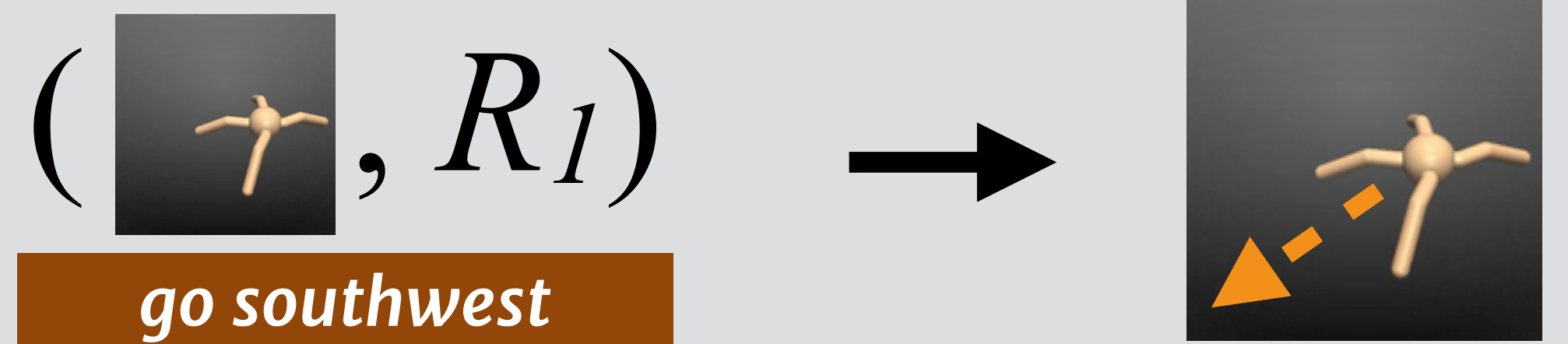
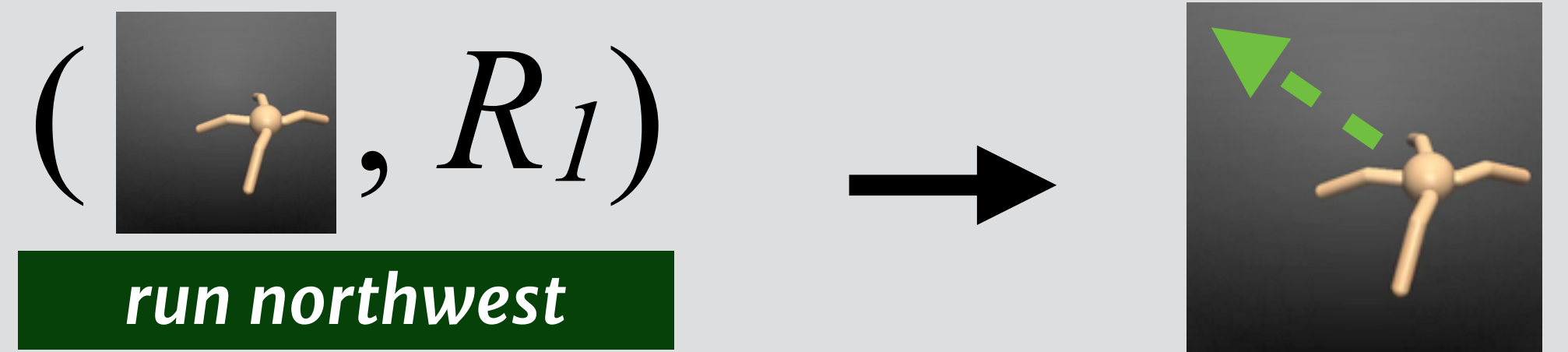
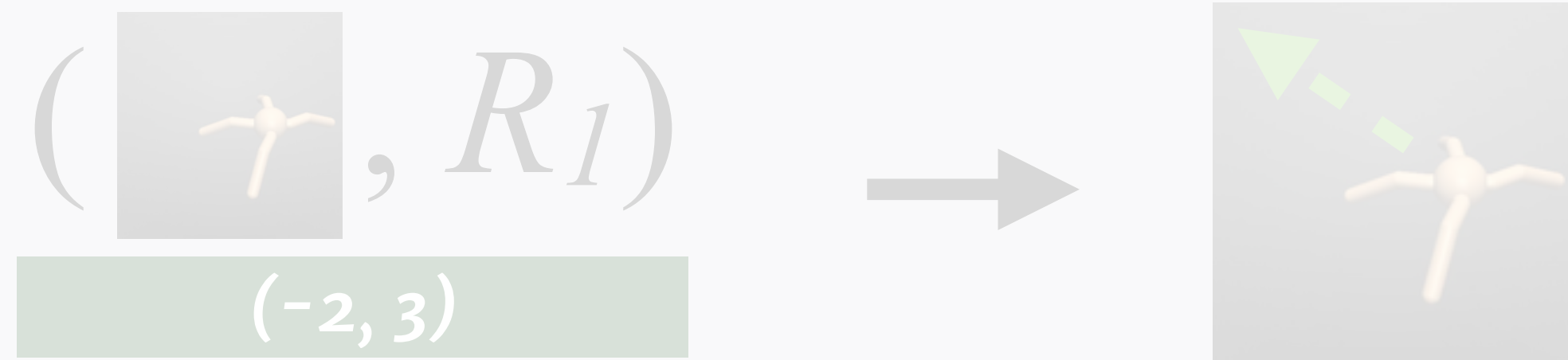
(Image, R_1)
 $(-2, -2)$



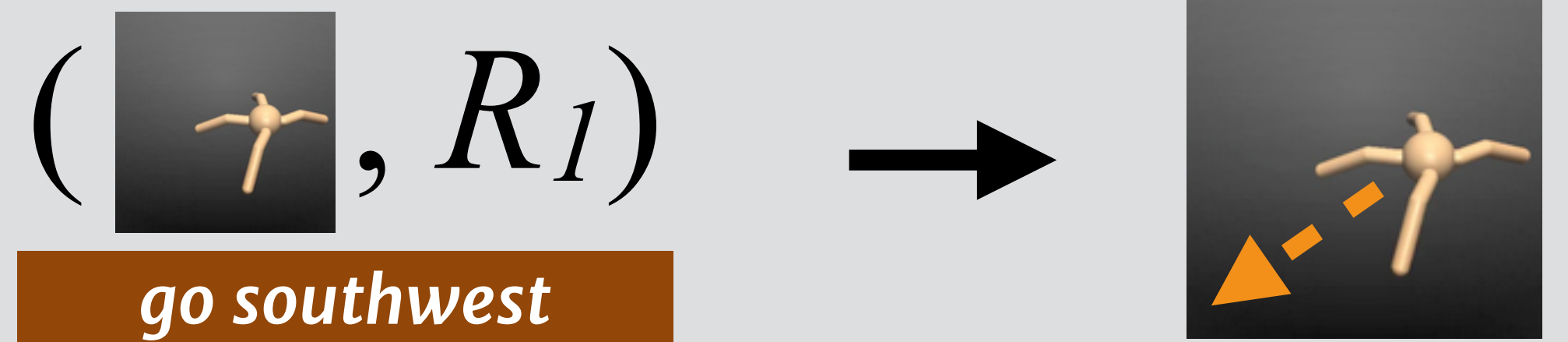
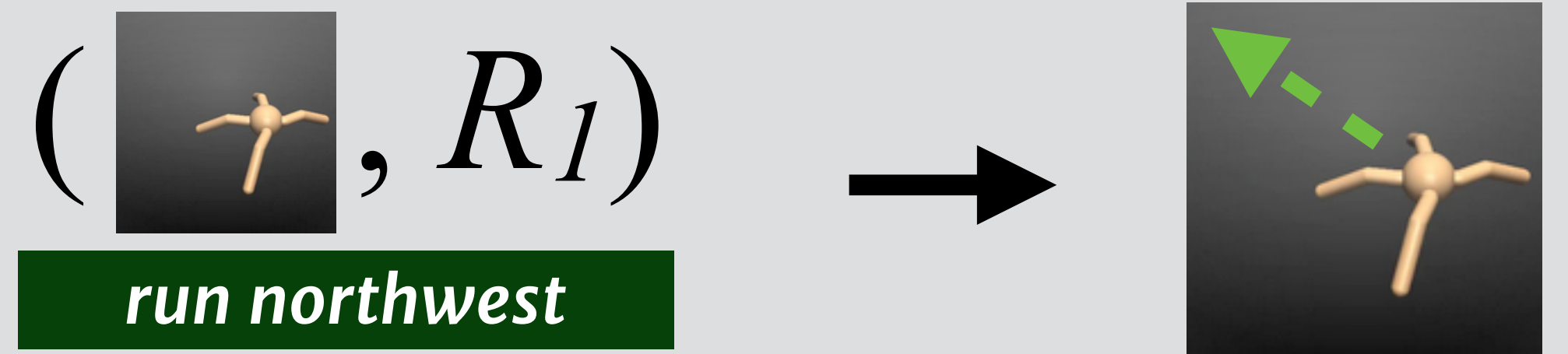
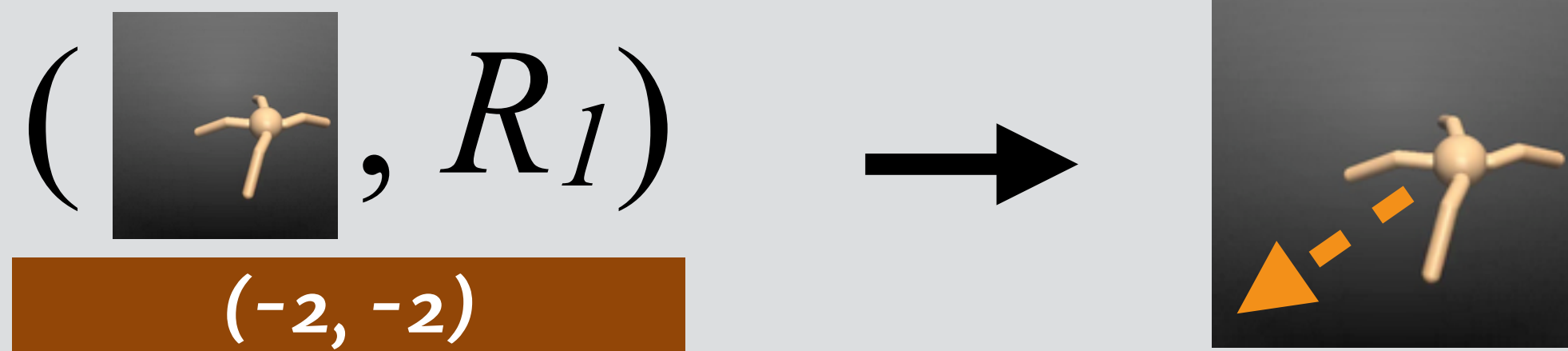
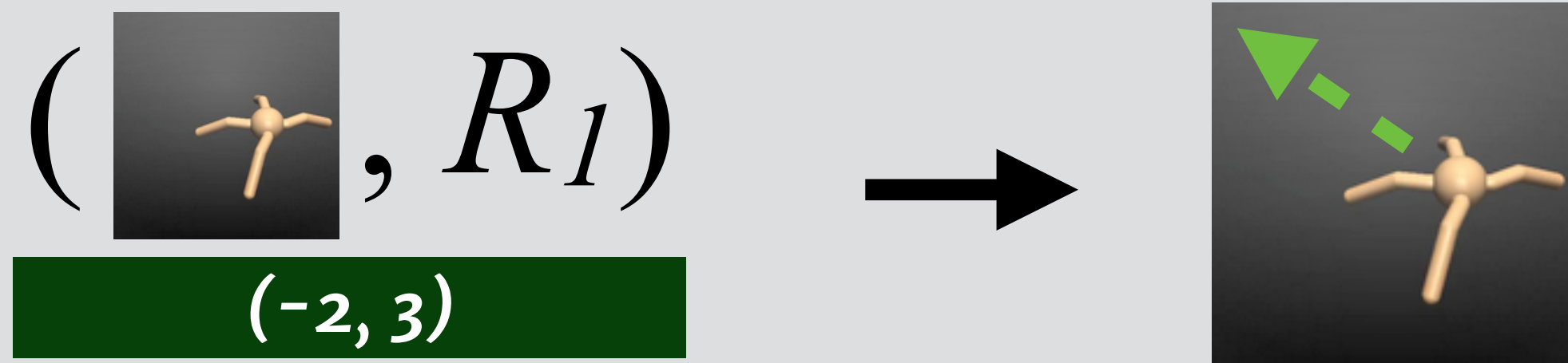
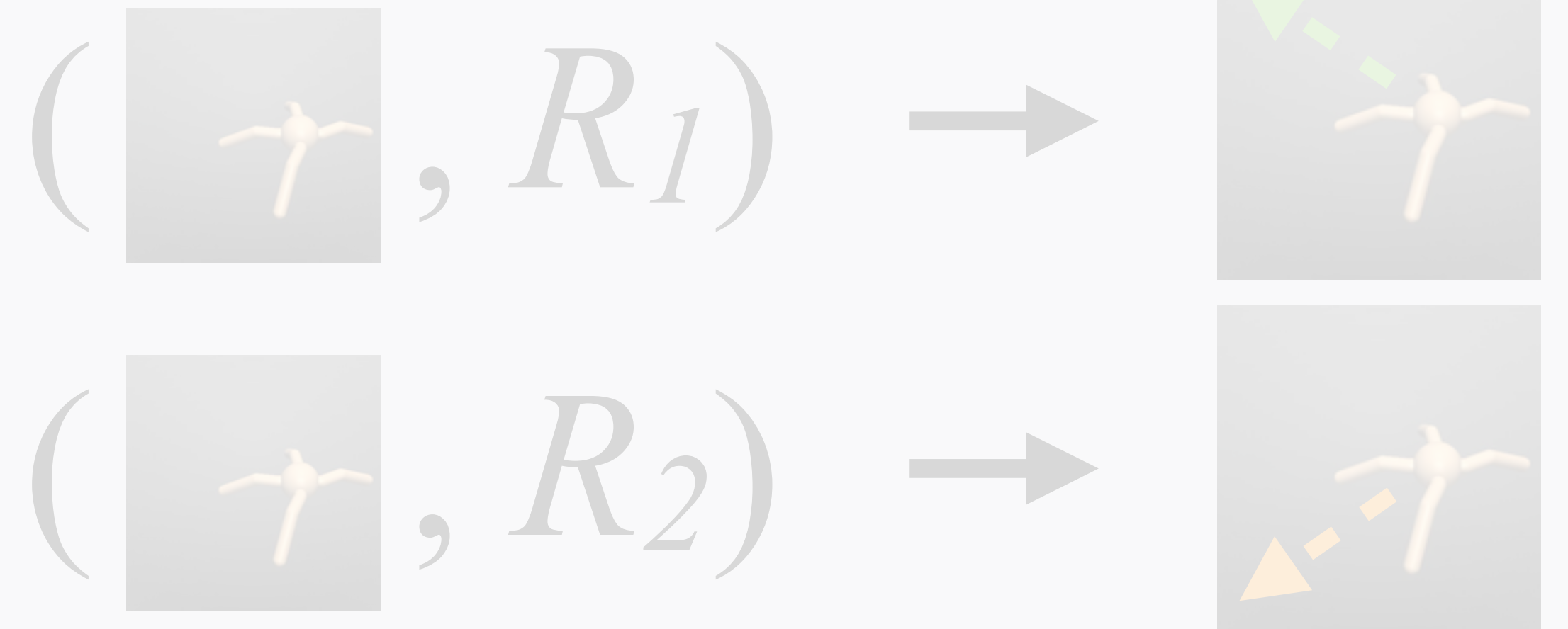
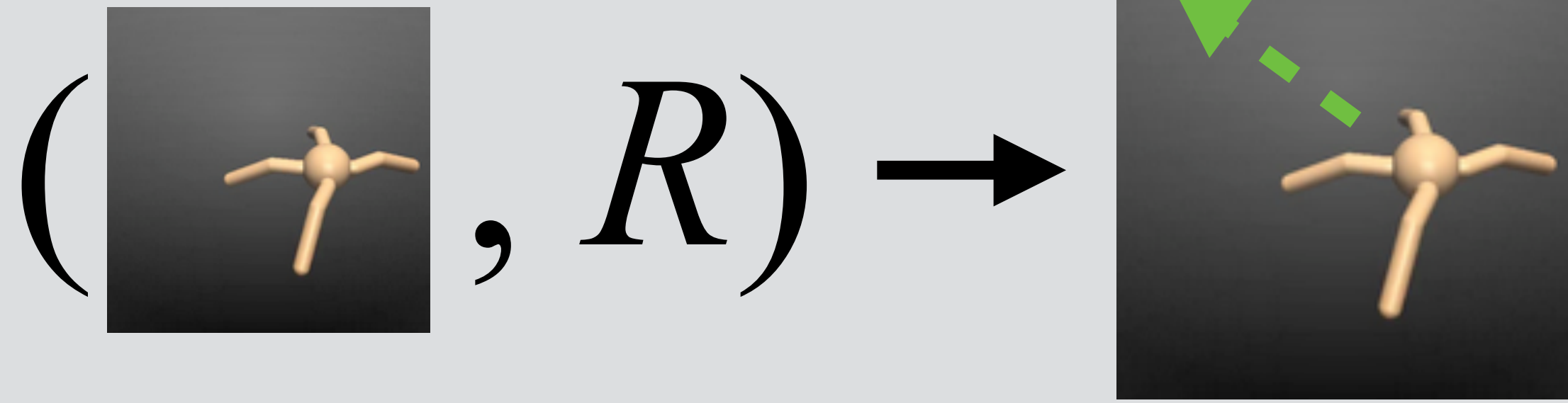
An NLPer's view of RL



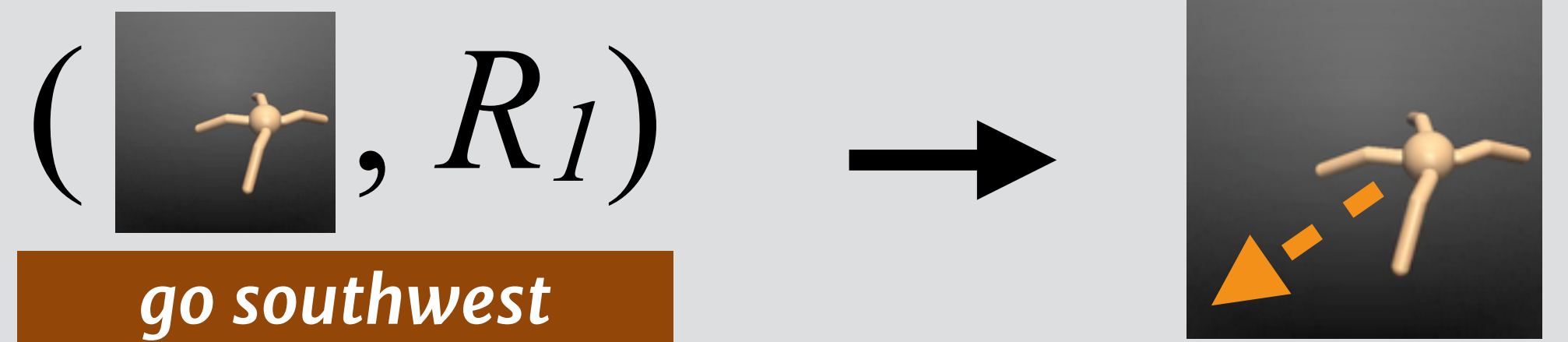
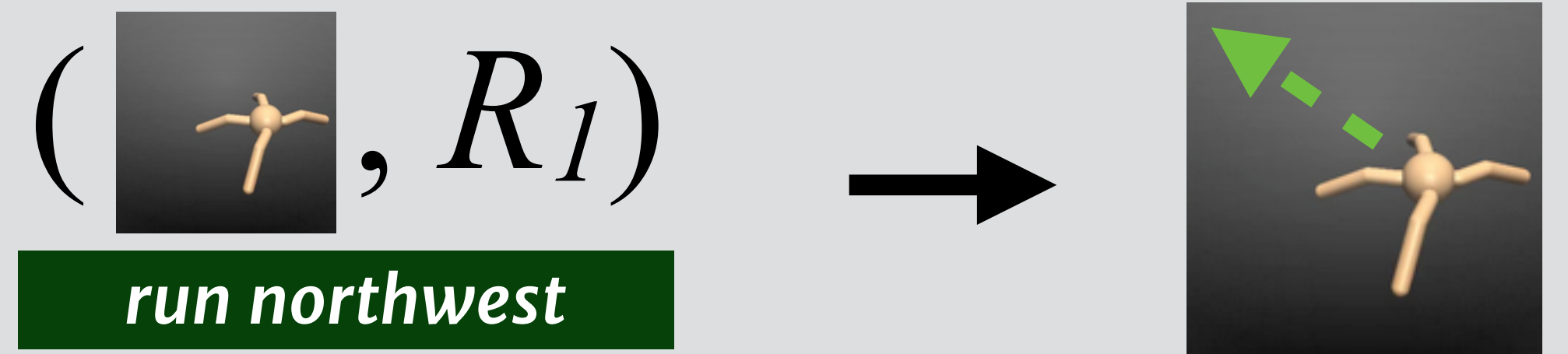
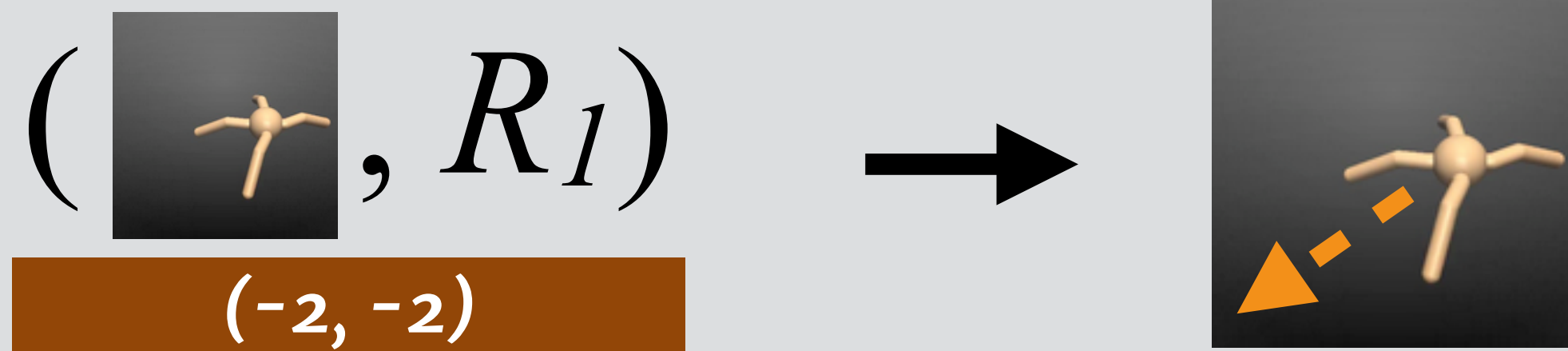
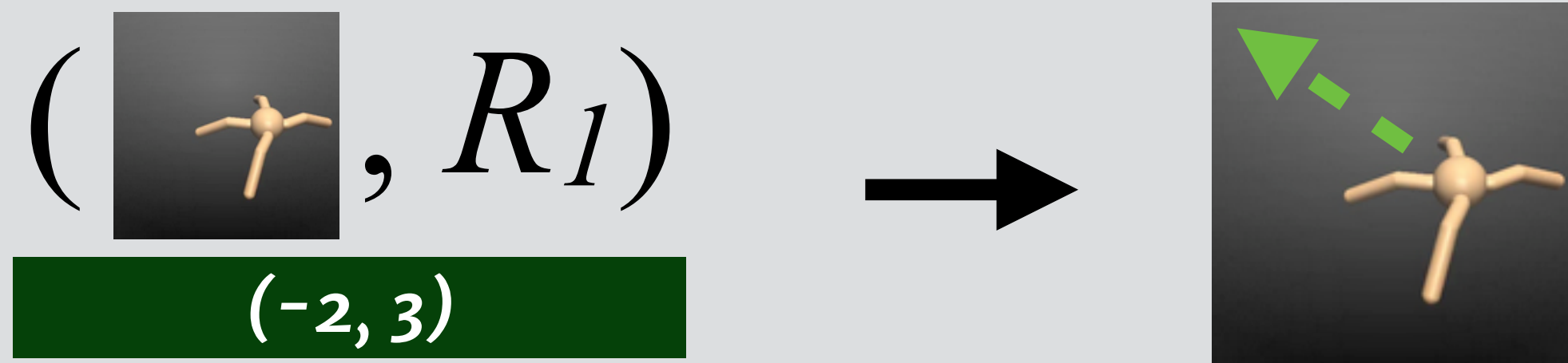
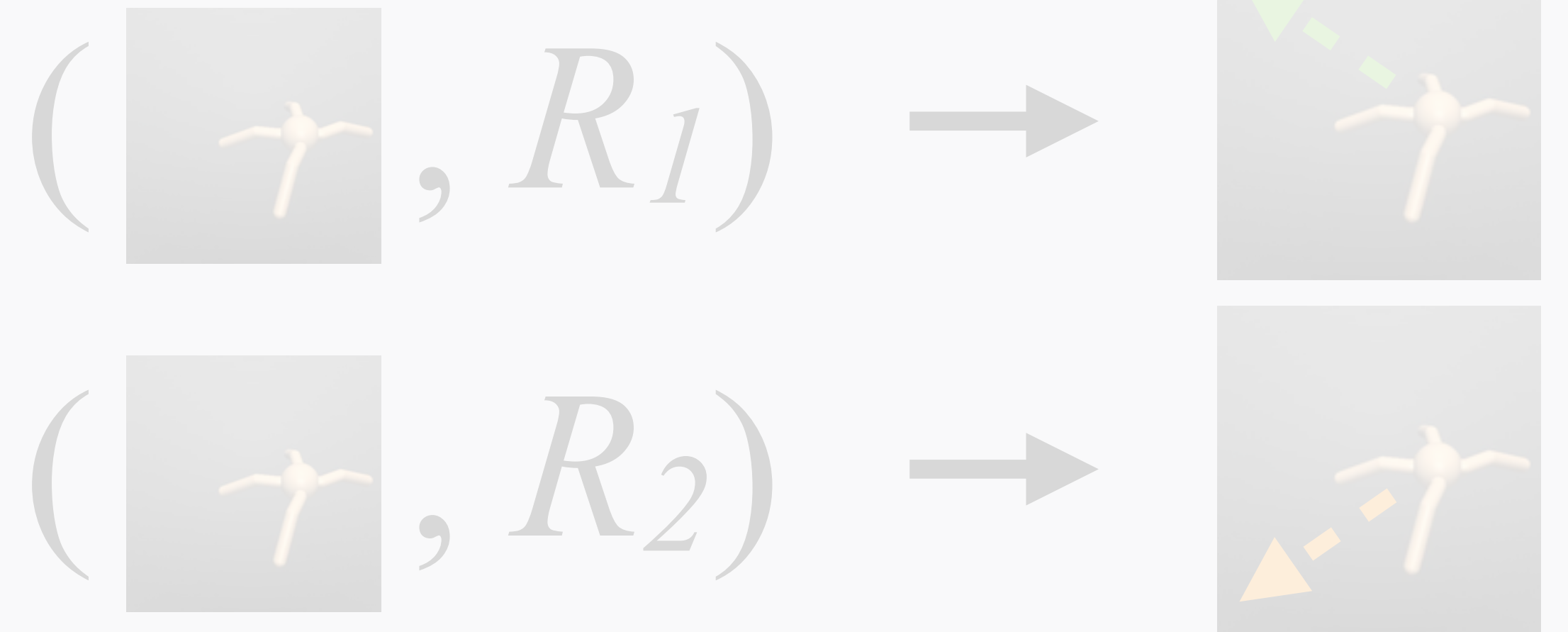
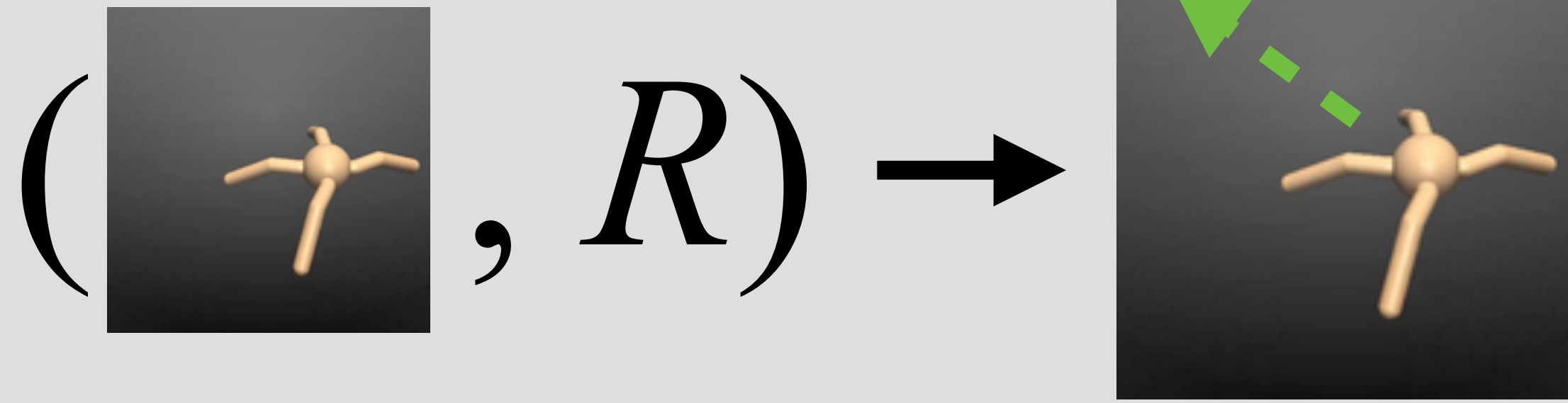
Learn to follow instructions



Instructions as observations



Instructions as observations

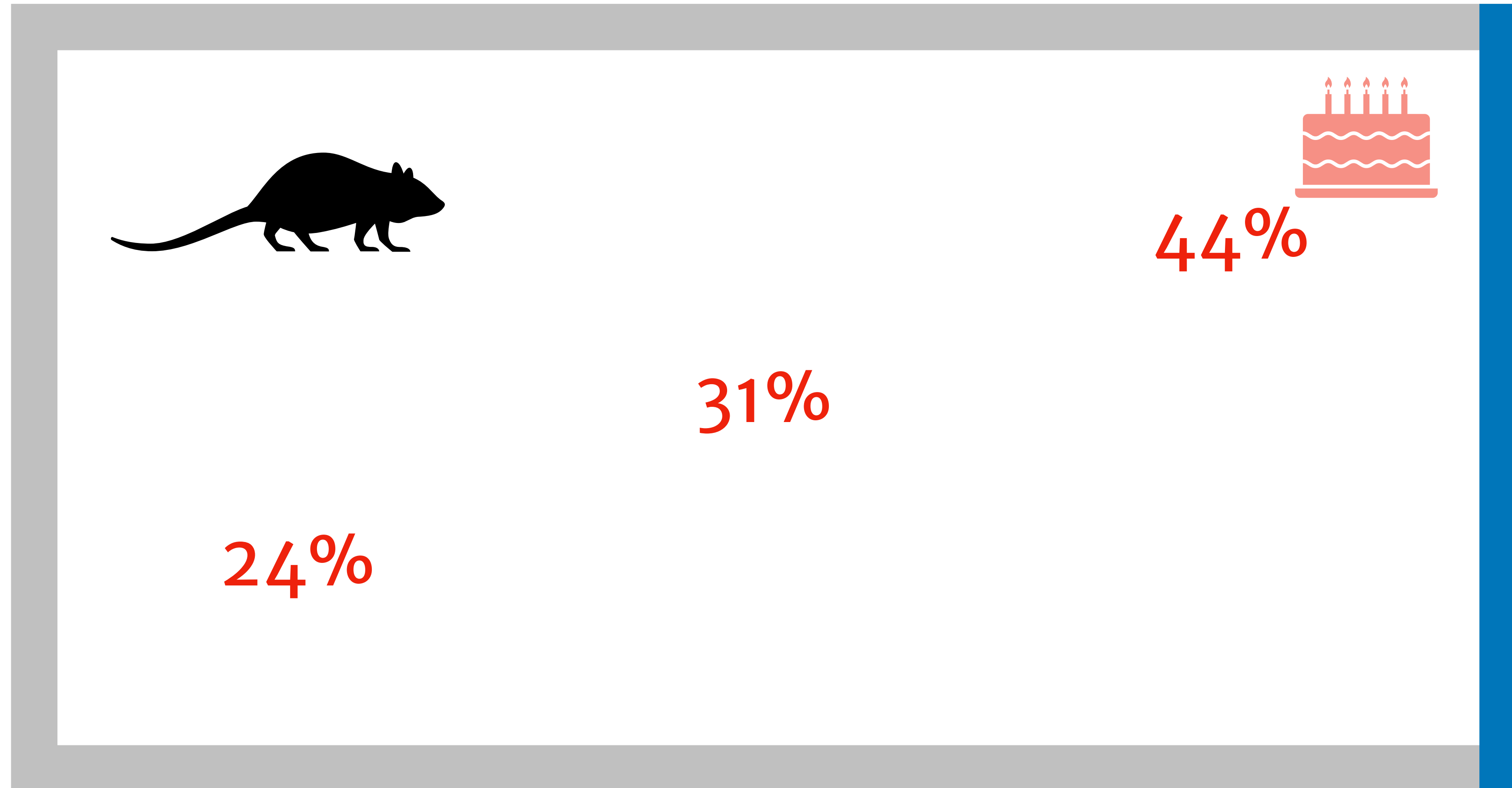


Language and goals in (human) cognition



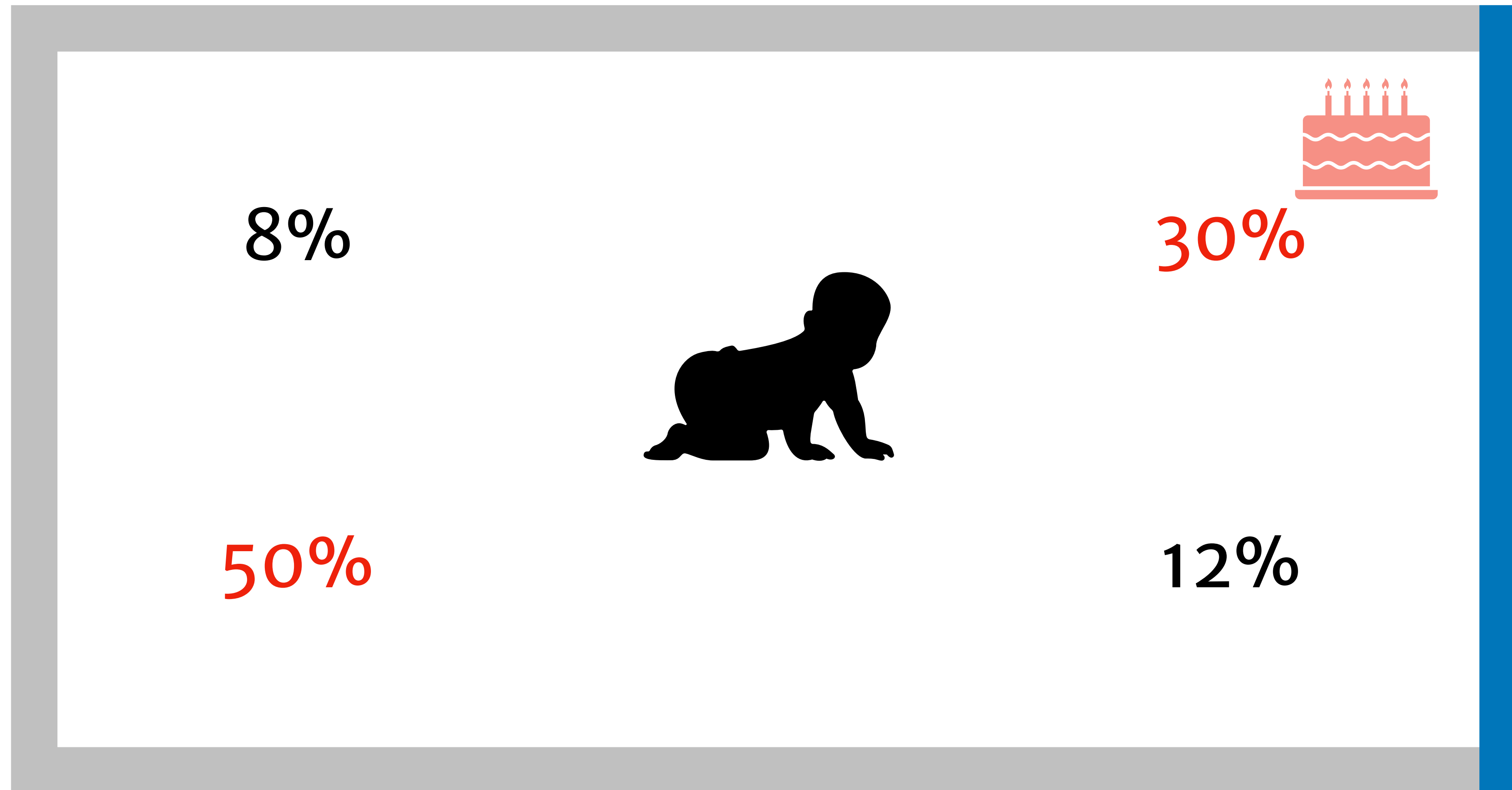
[Hermer-Vazquez, Spelke, Katznelson 1999]

Language and goals in (human) cognition



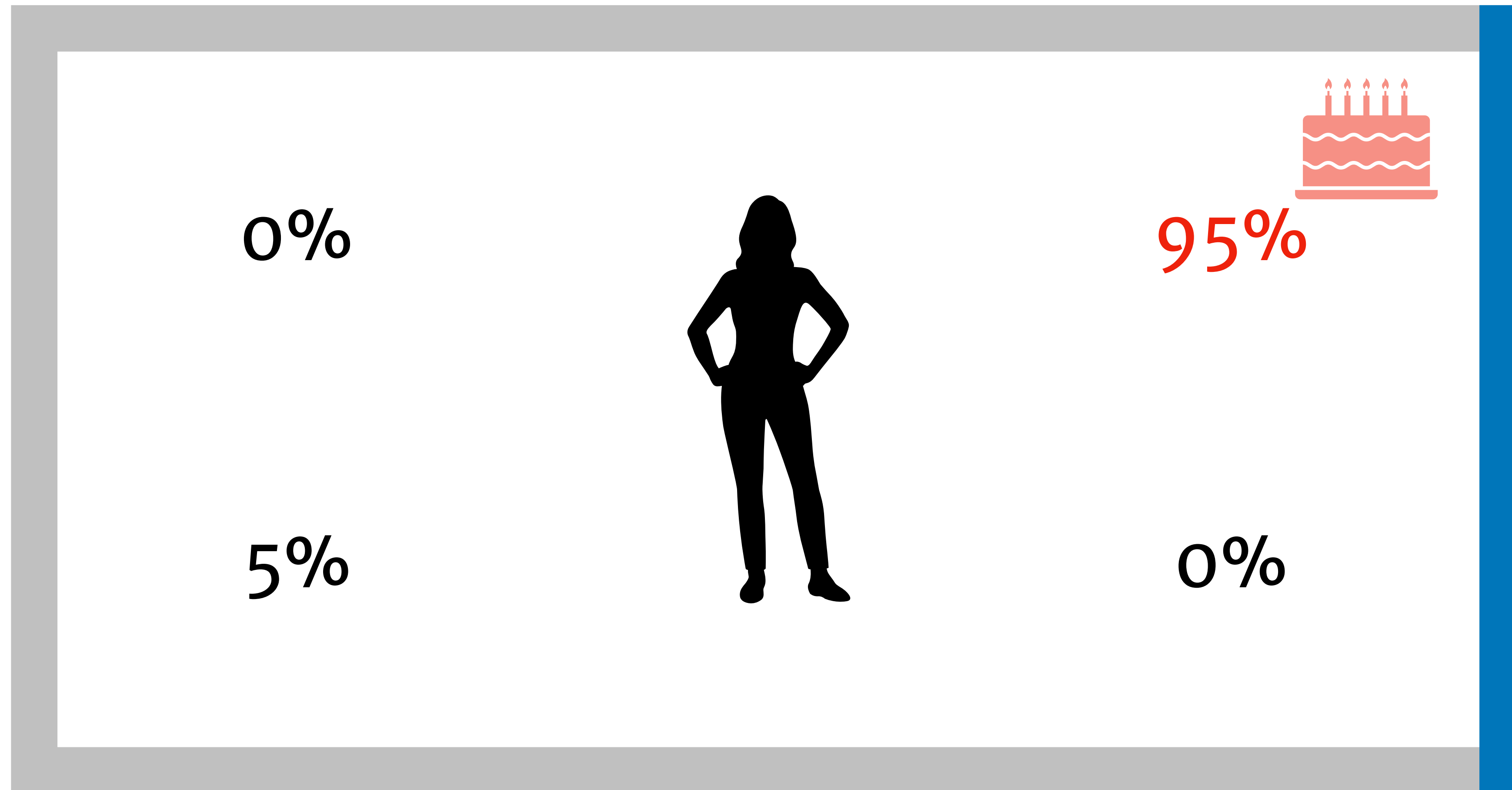
[Hermer-Vazquez, Spelke, Katznelson 1999]

Language and goals in (human) cognition



[Hermer-Vazquez, Spelke, Katznelson 1999]

Language and goals in (human) cognition



[Hermer-Vazquez, Spelke, Katznelson 1999]

Language and goals in (human) cognition



[Hermer-Vazquez, Spelke, Katznelson 1999]

Language and goals in (human) cognition

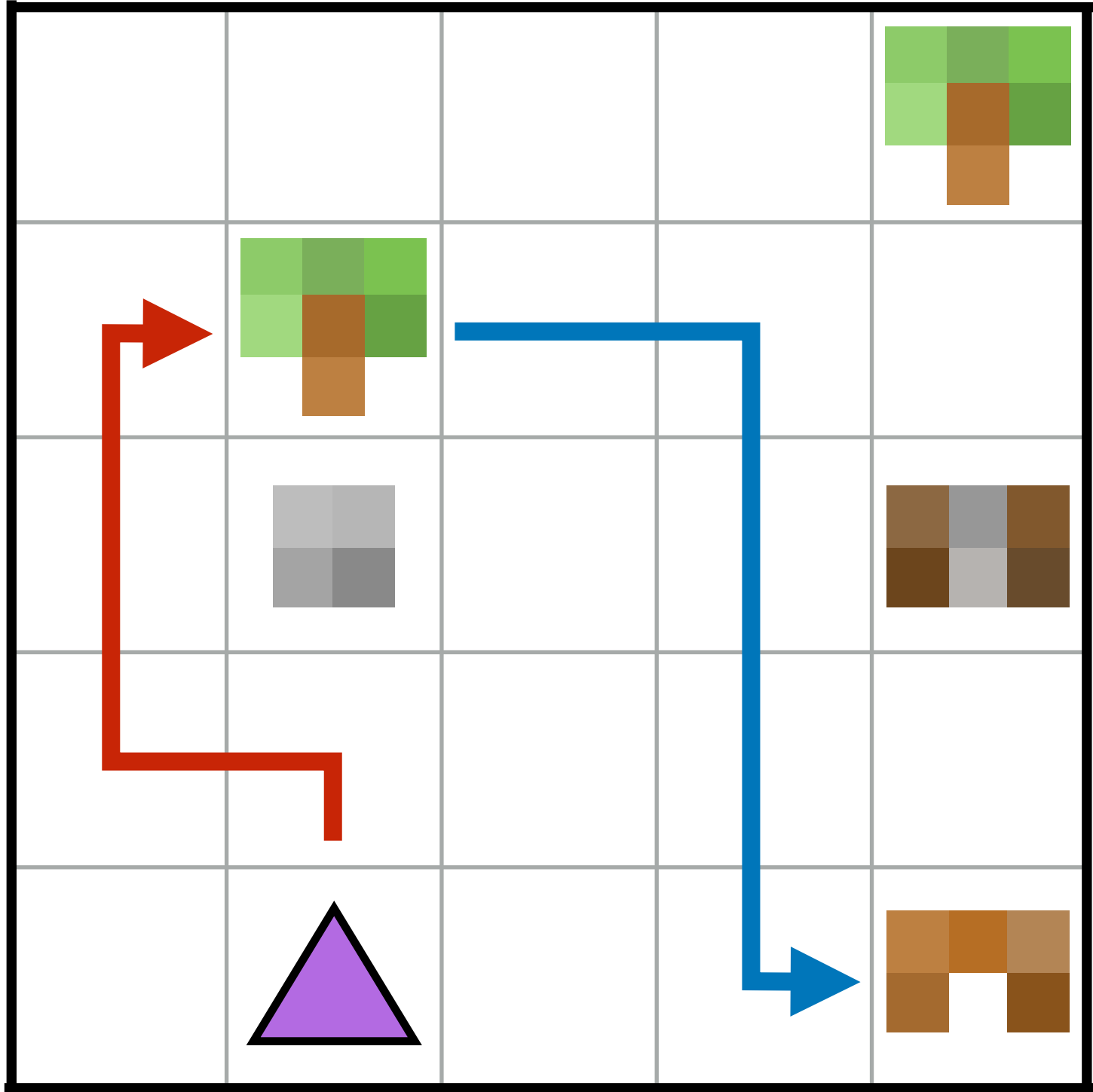
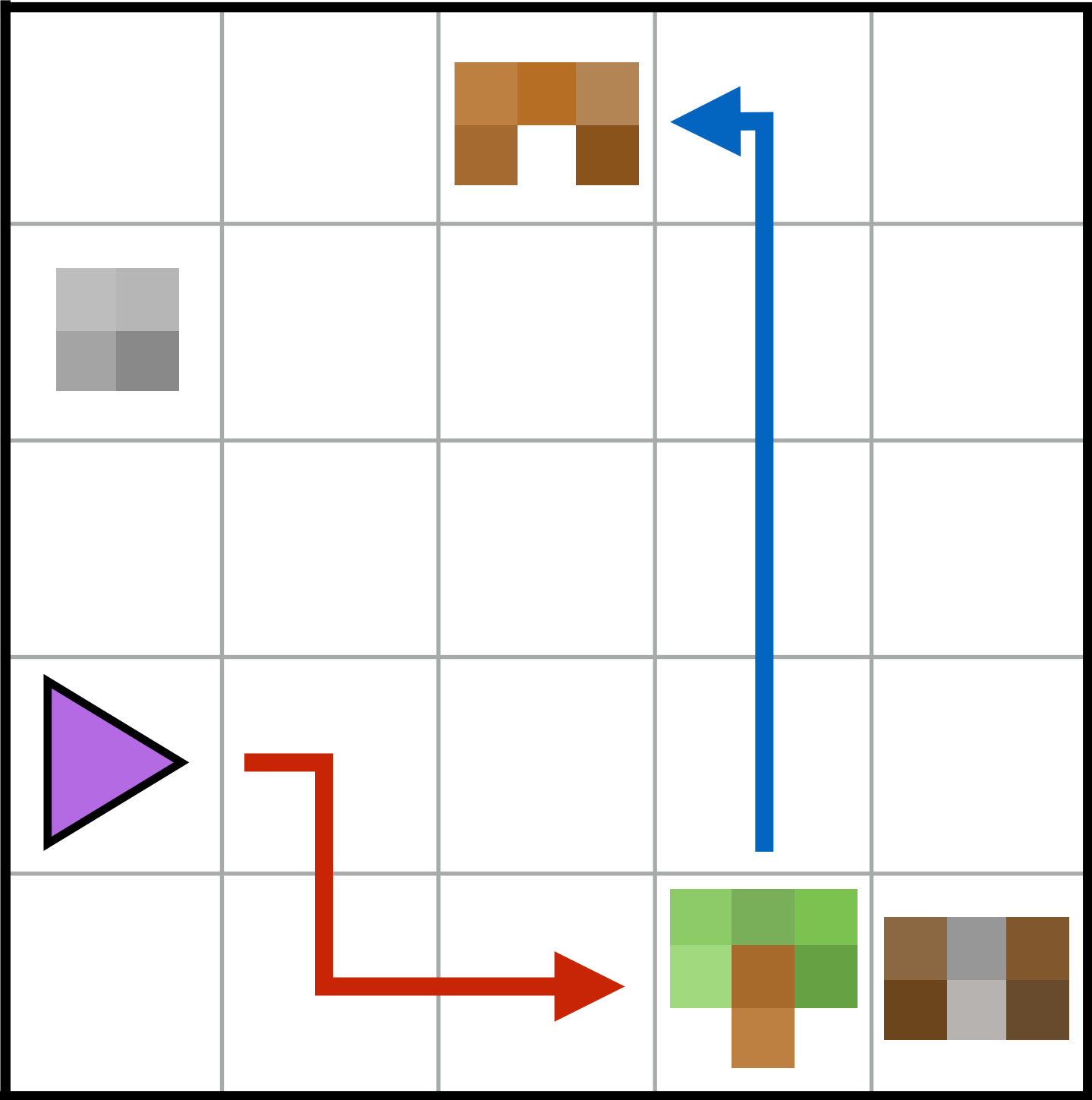


[Hermer-Vazquez, Spelke, Katznelson 1999]

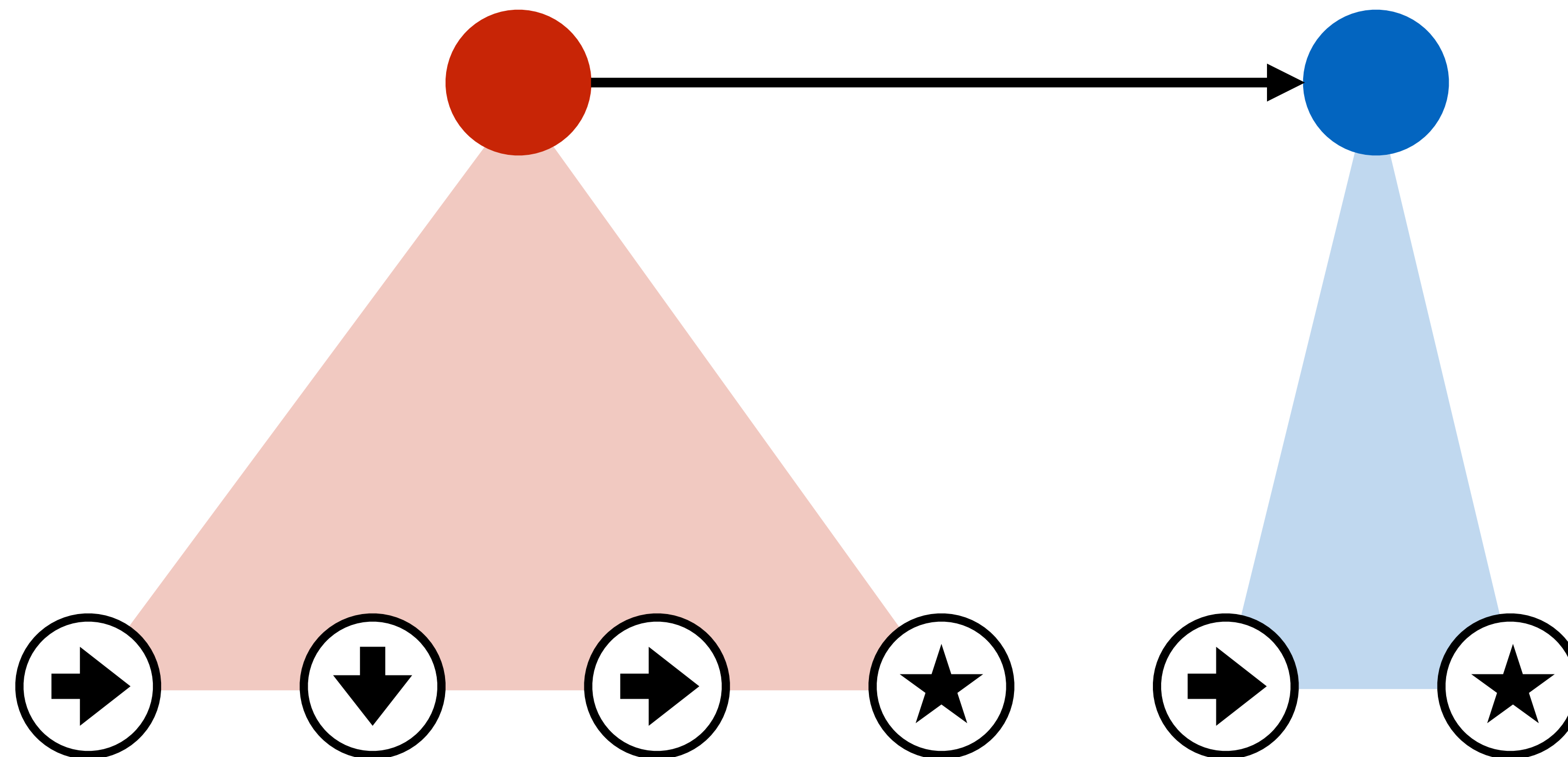
Language as a representation of options

Tasks & subtasks

make planks



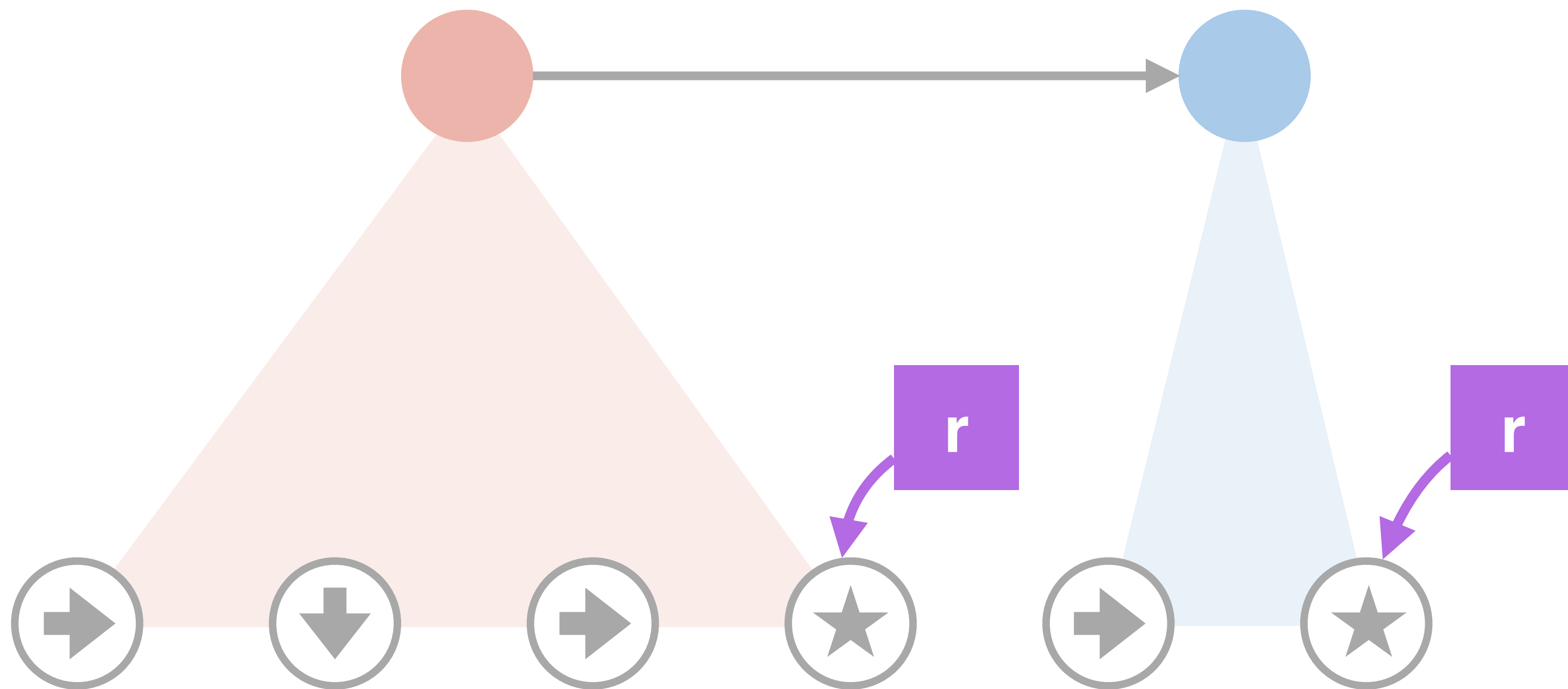
The options framework



[Sutton et al. 99, Bacon & Precup 16]

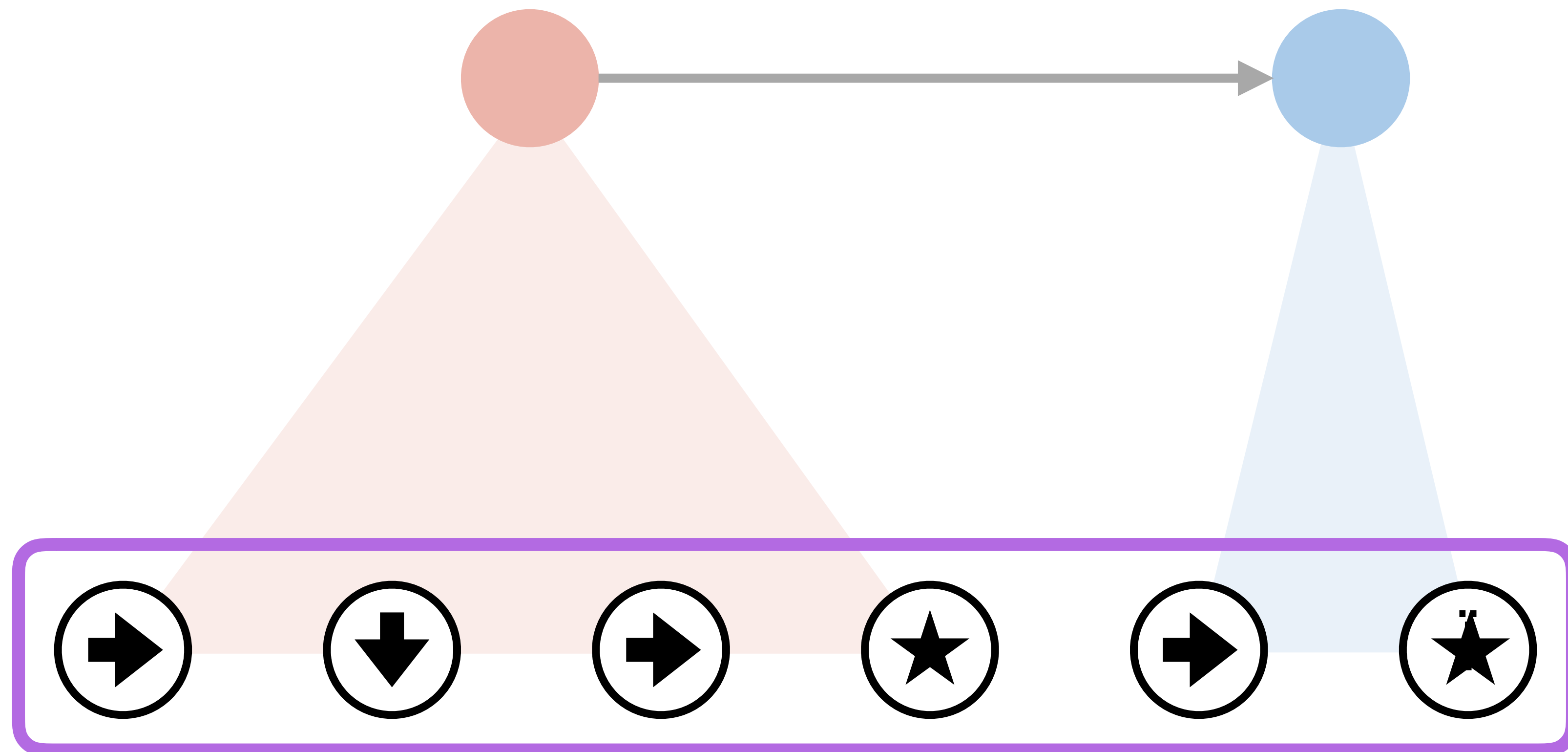


Learning from intermediate rewards



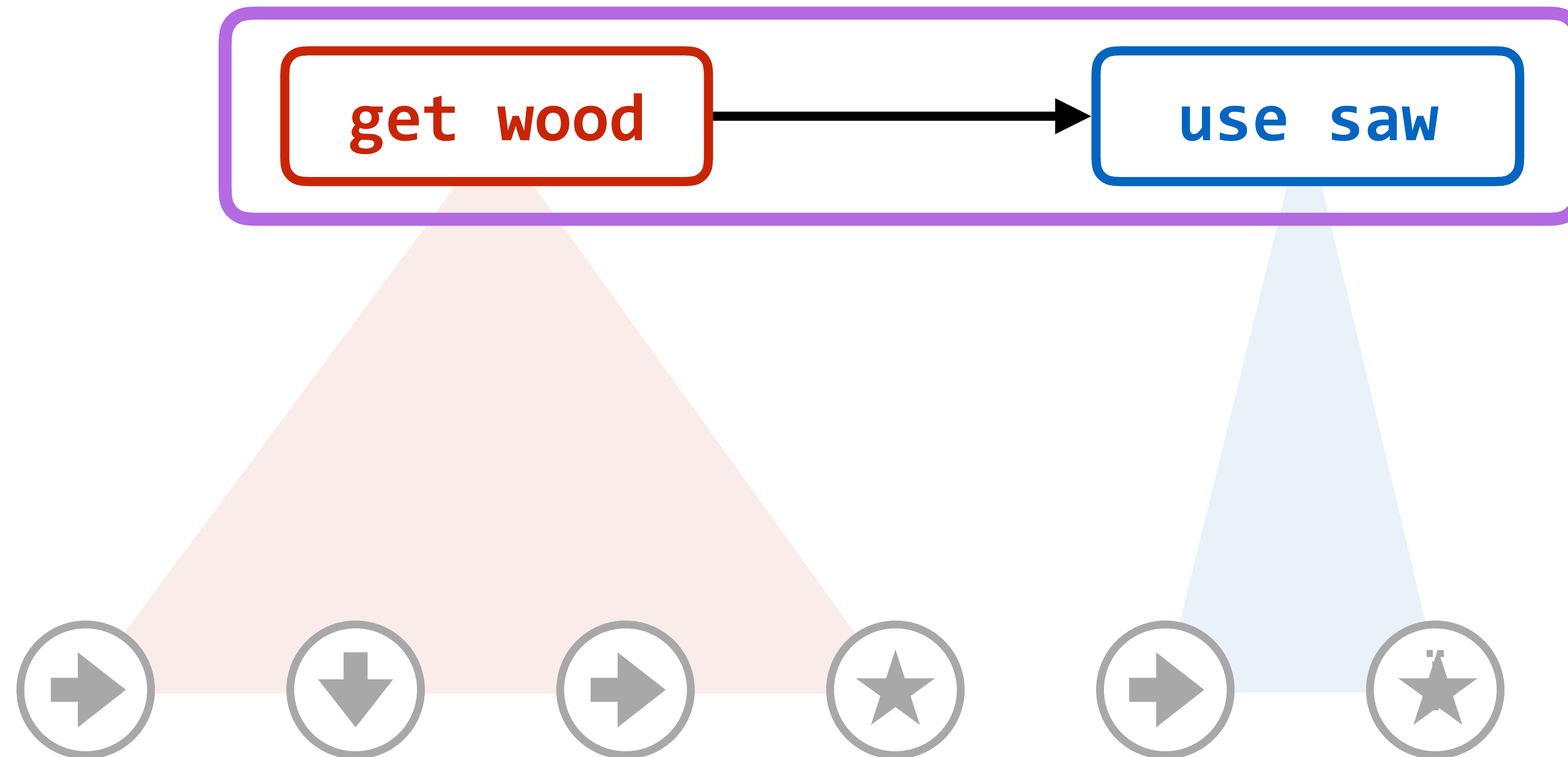
[Kearns & Singh 02, Kulkarni et al. 16]

Learning from demonstrations



[Stolle & Precup 02, Fox & Krishnan et al. 16]

Learning from policy sketches

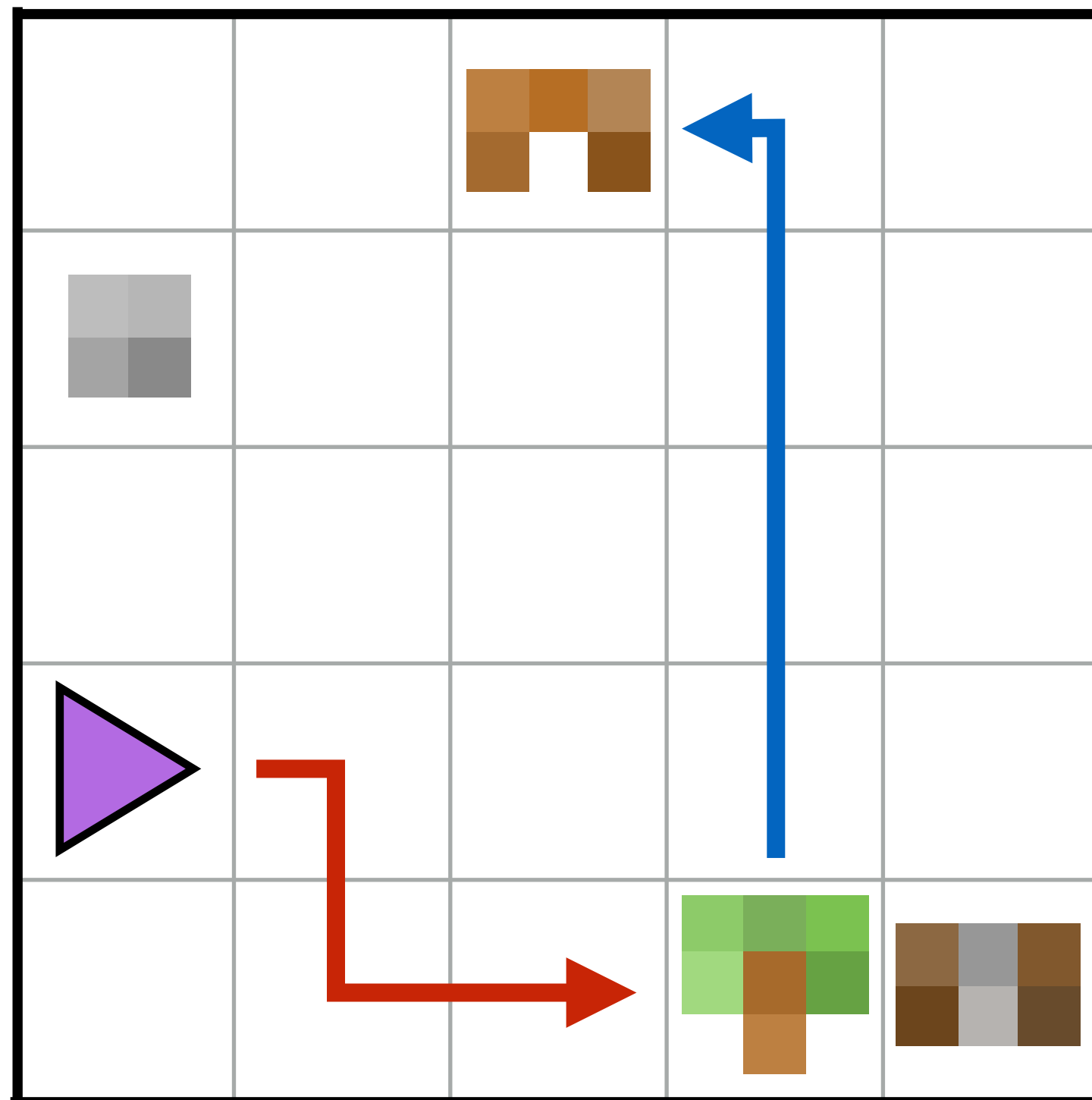


[A, Klein and Levine. "Modular Multitask Reinforcement Learning with Policy Sketches."]

Learning from policy sketches

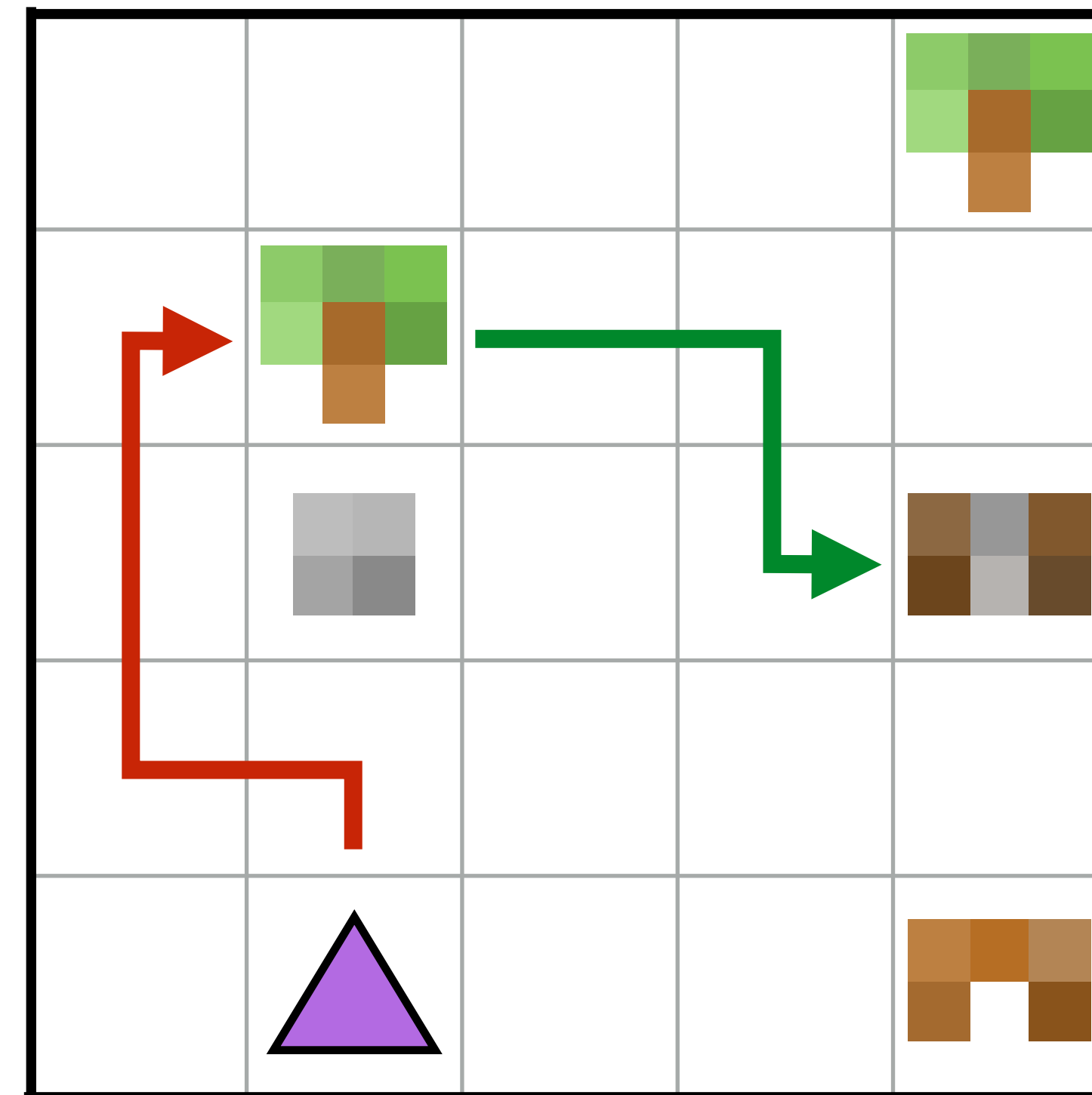
get wood

use saw

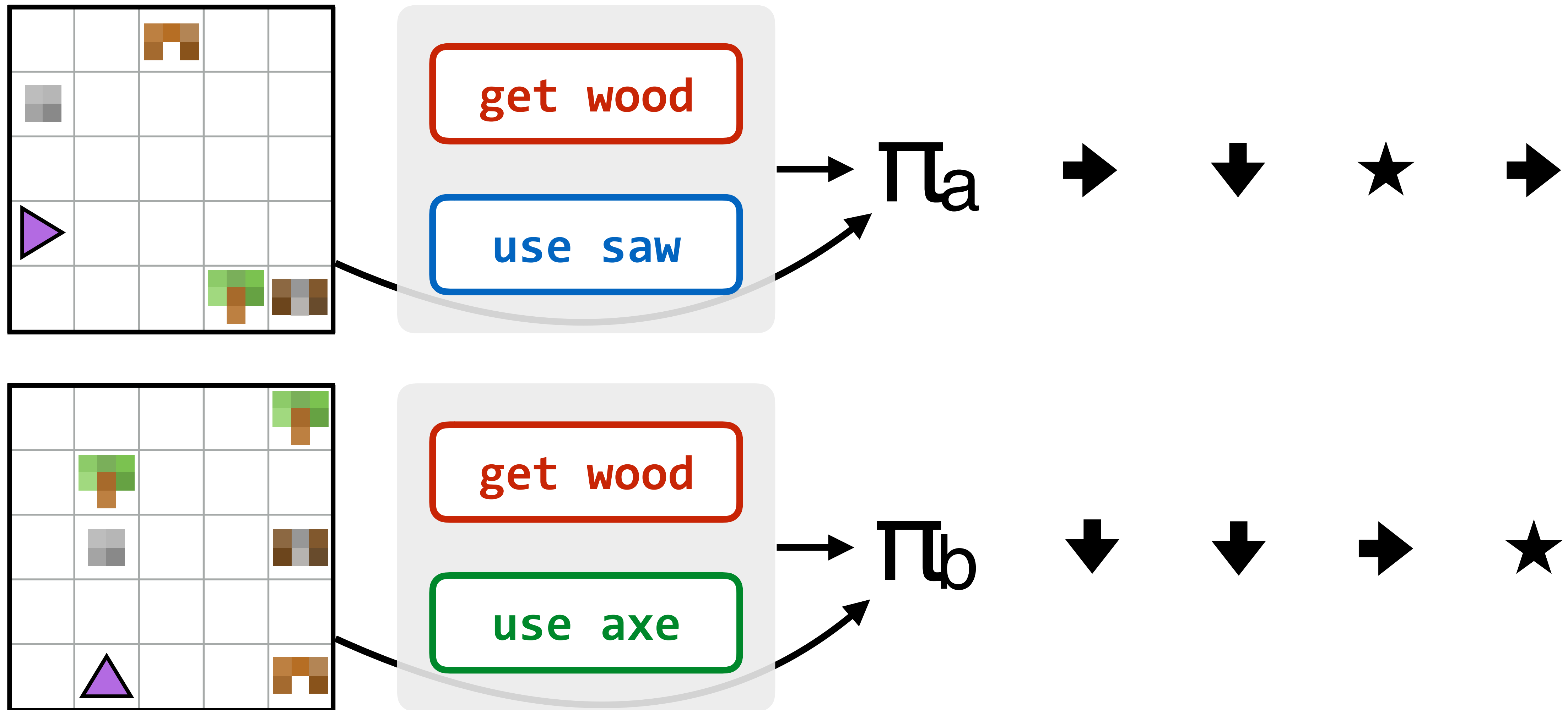


get wood

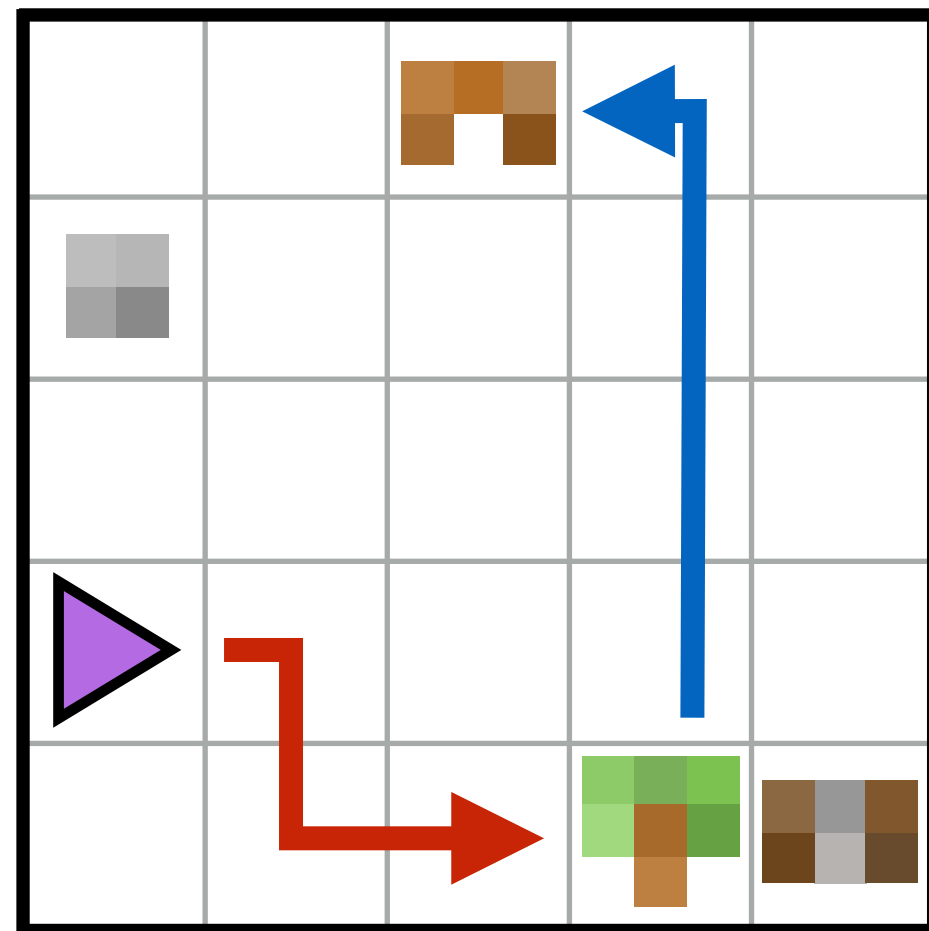
use axe



Learning from policy sketches



Learning from policy sketches

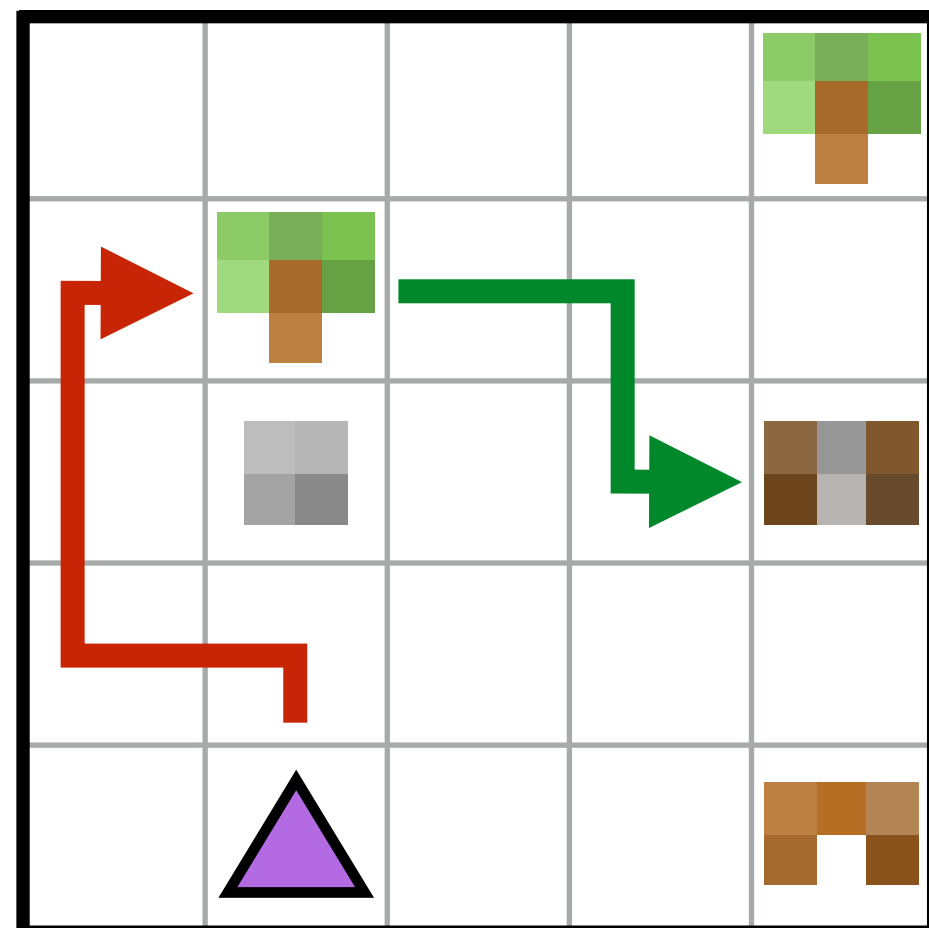
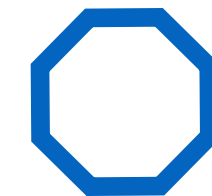
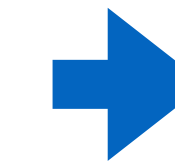
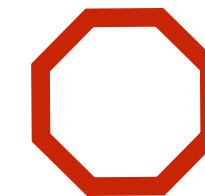
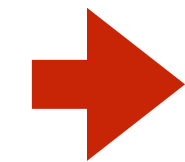


get wood

use saw

π_1

π_2

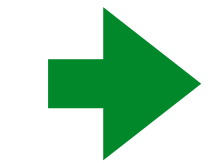
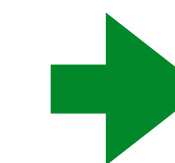
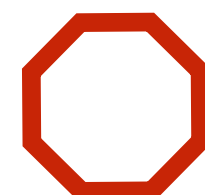
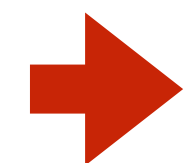


get wood

use axe

π_1

π_3



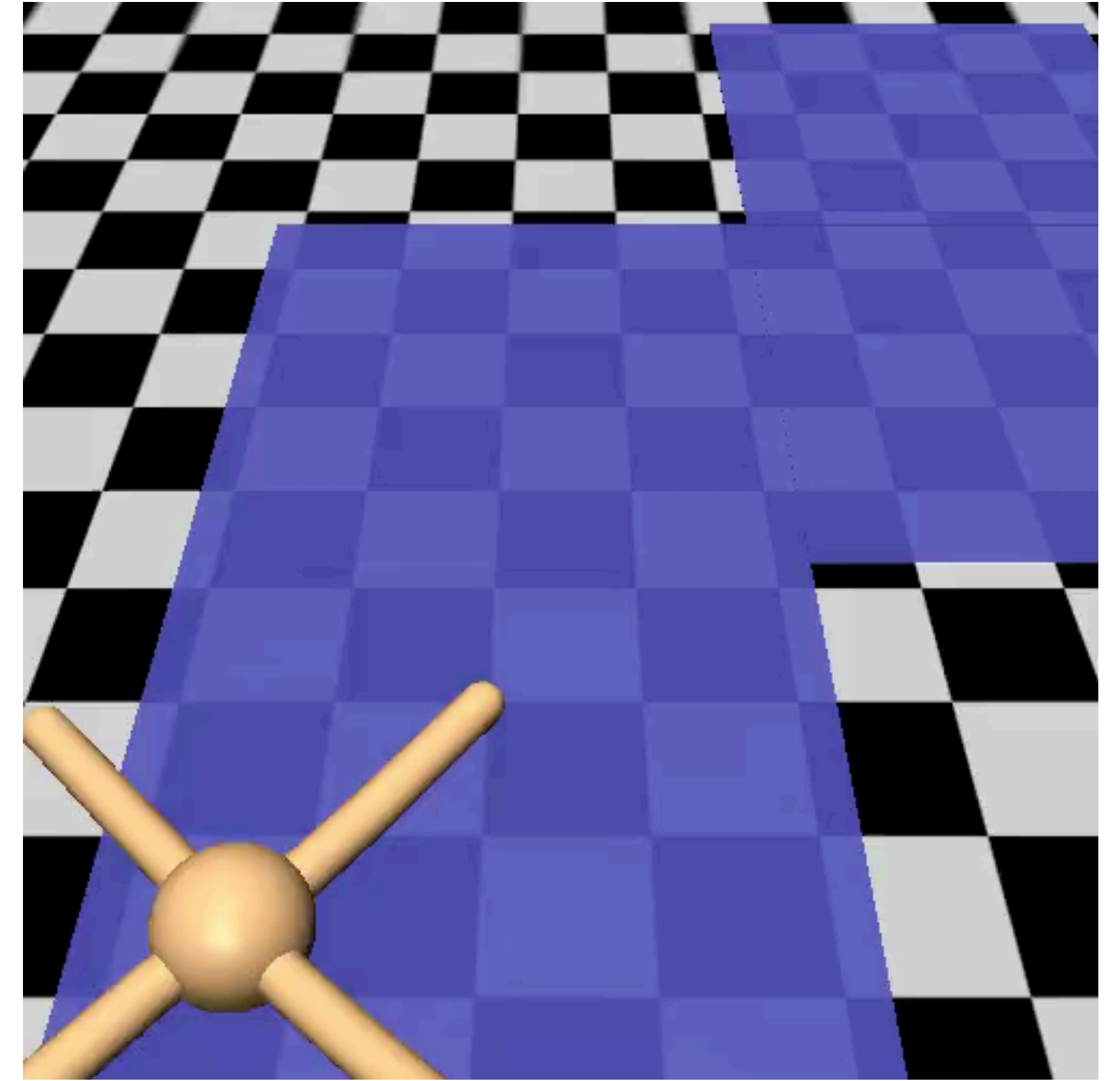
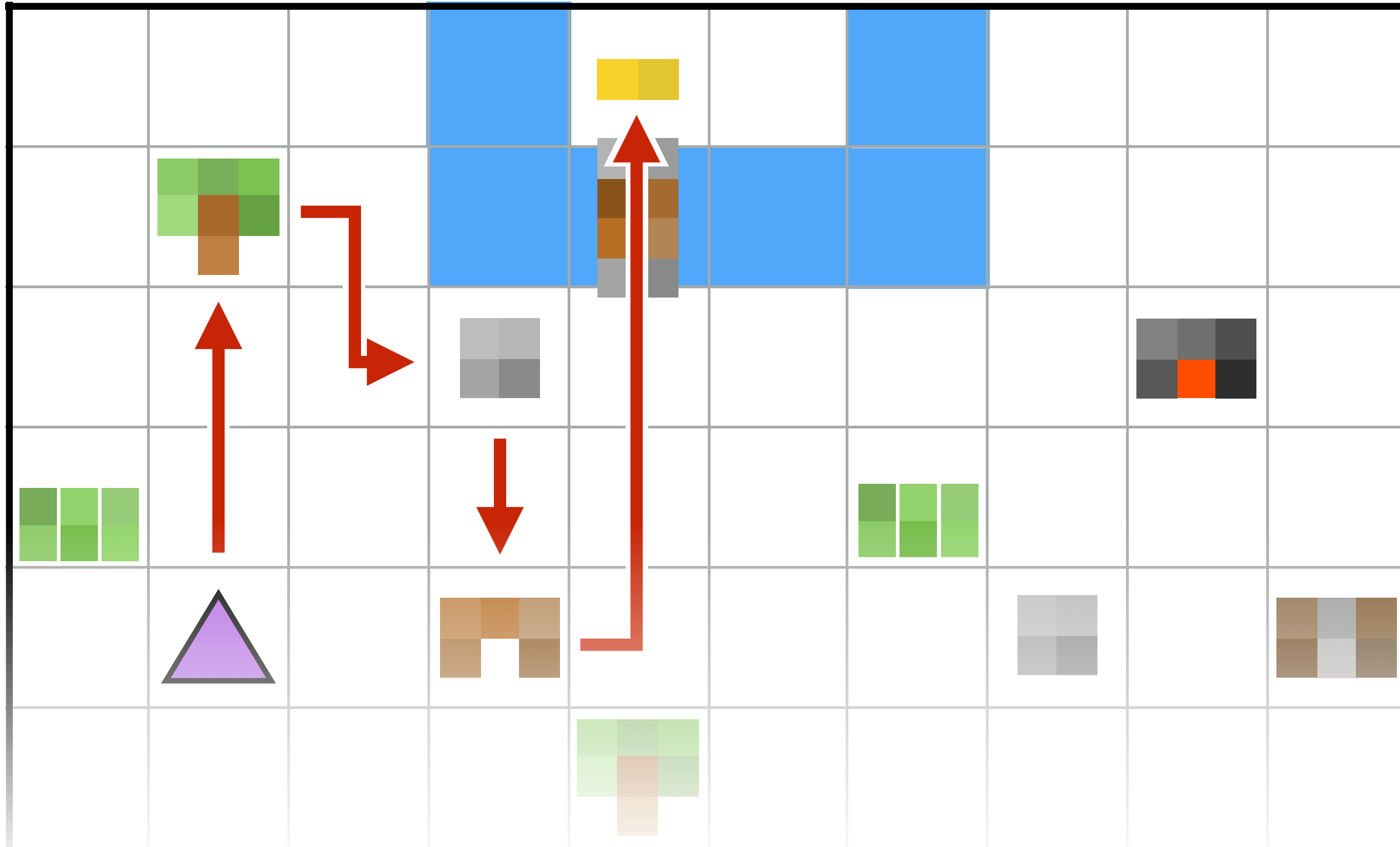


A tiny bit of data goes a long way

make plank	get wood	use toolshed			
make stick	get wood	use workbench			
make cloth	get grass	use factory			
make rope	get grass	use toolshed			
make bridge	get iron	get wood	use factory		
make bed*	get wood	use toolshed	get grass	use workbench	
make axe*	get wood	use workbench	get iron	use toolshed	
make shears	get wood	use workbench	get iron	use workbench	
get gold	get iron	get wood	use factory	use bridge	
get gem	get wood	use workbench	get iron	use toolshed	use axe



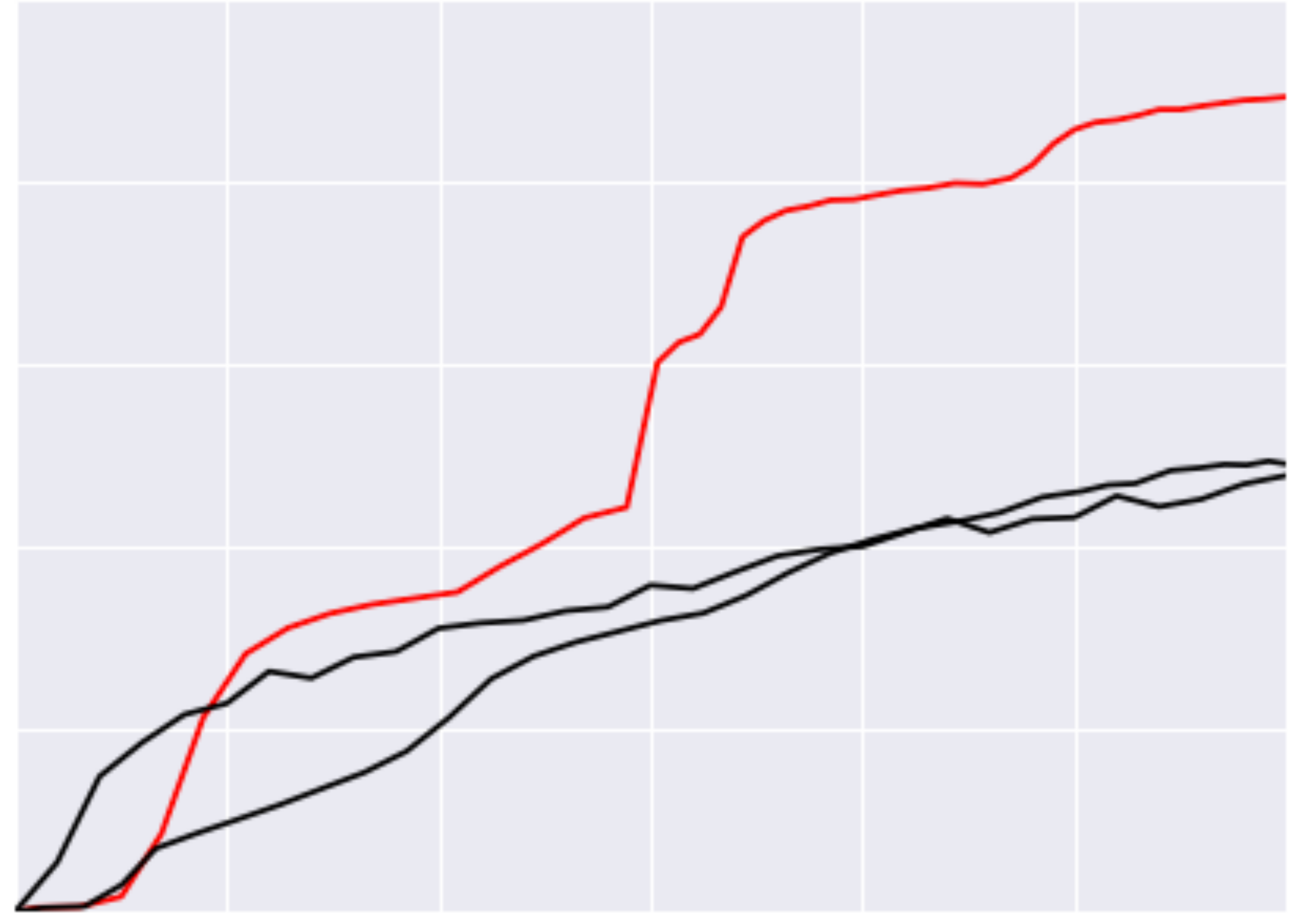
Tasks



The mini-craft task



Reward



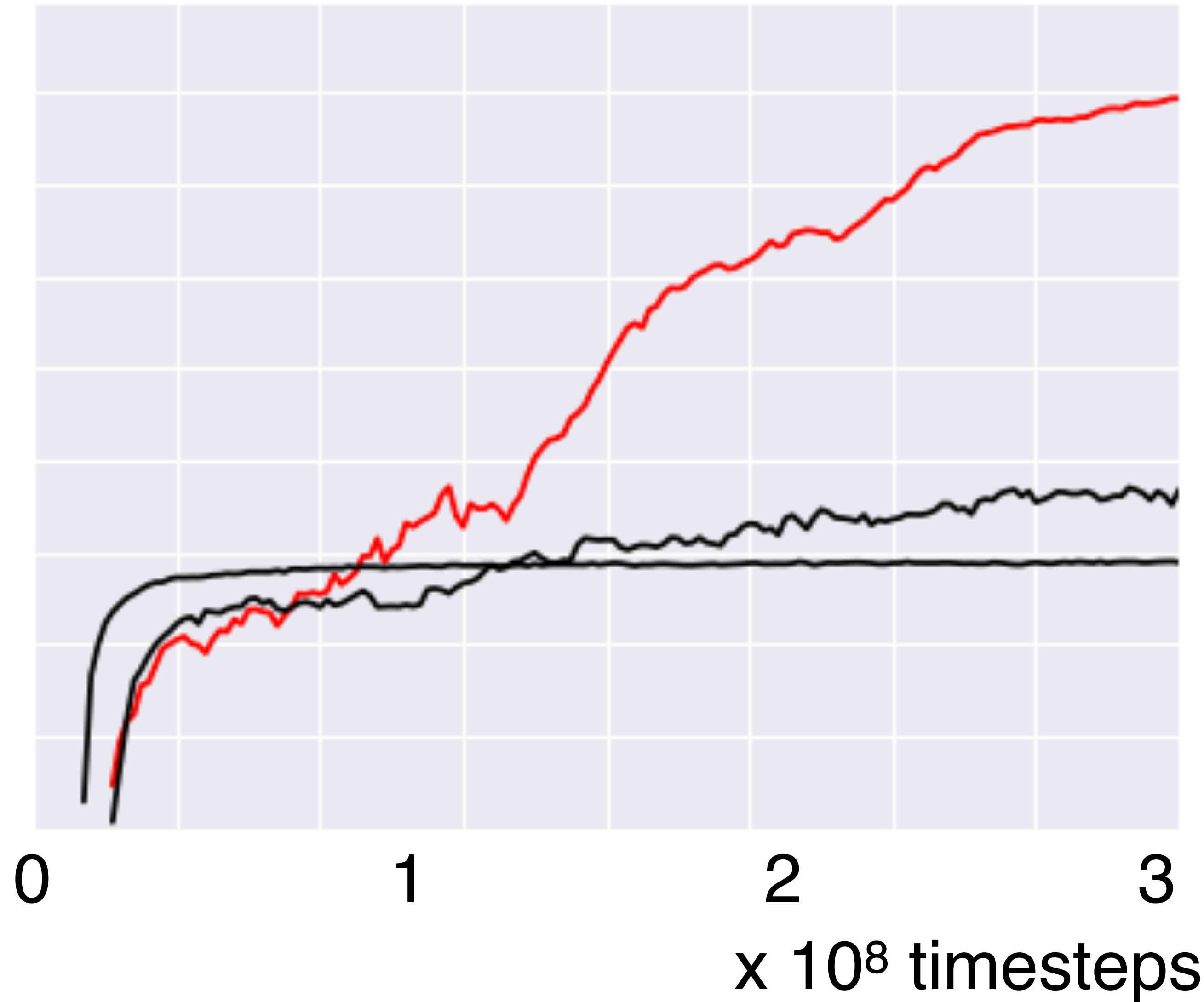
Sketches:
modular

Sketches: joint
Unsupervised

0 1 2 3
x 10⁶ episodes

The path-walking task

Reward

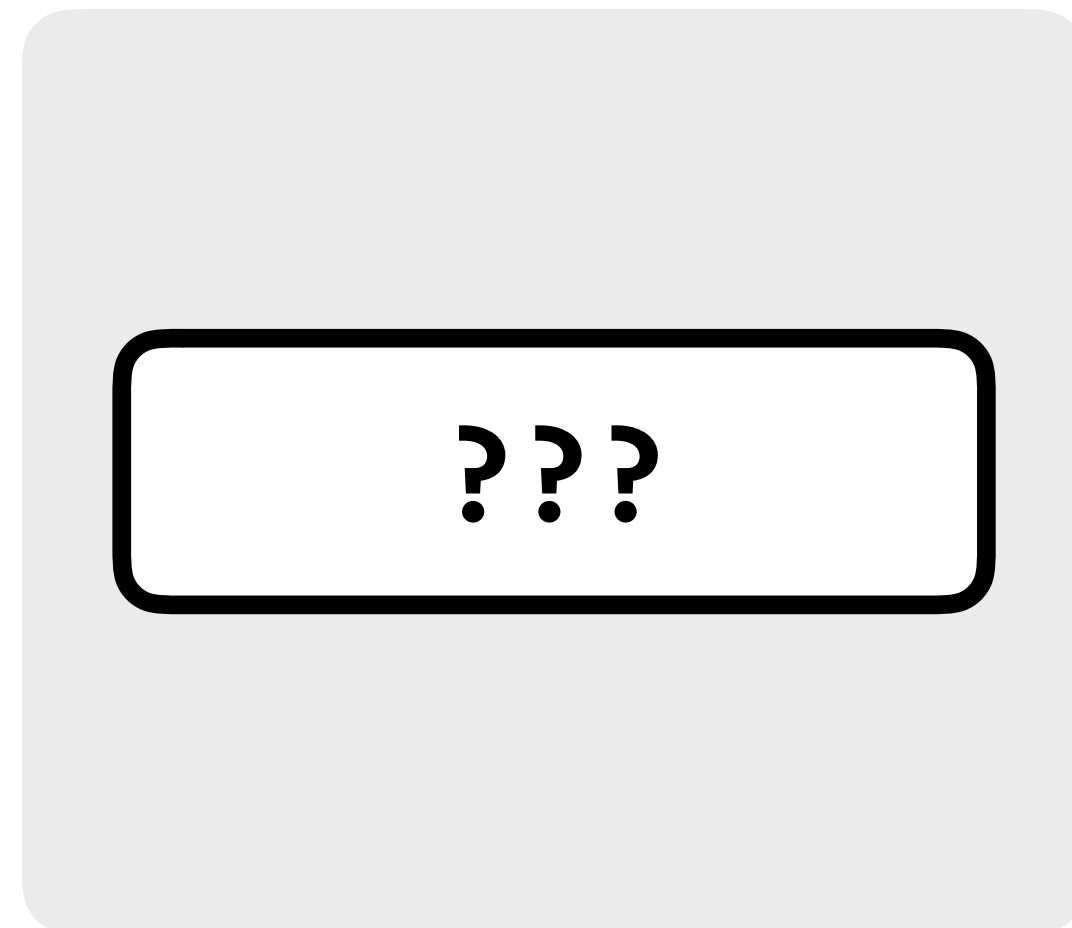


Sketches:
modular

Sketches: joint
Unsupervised

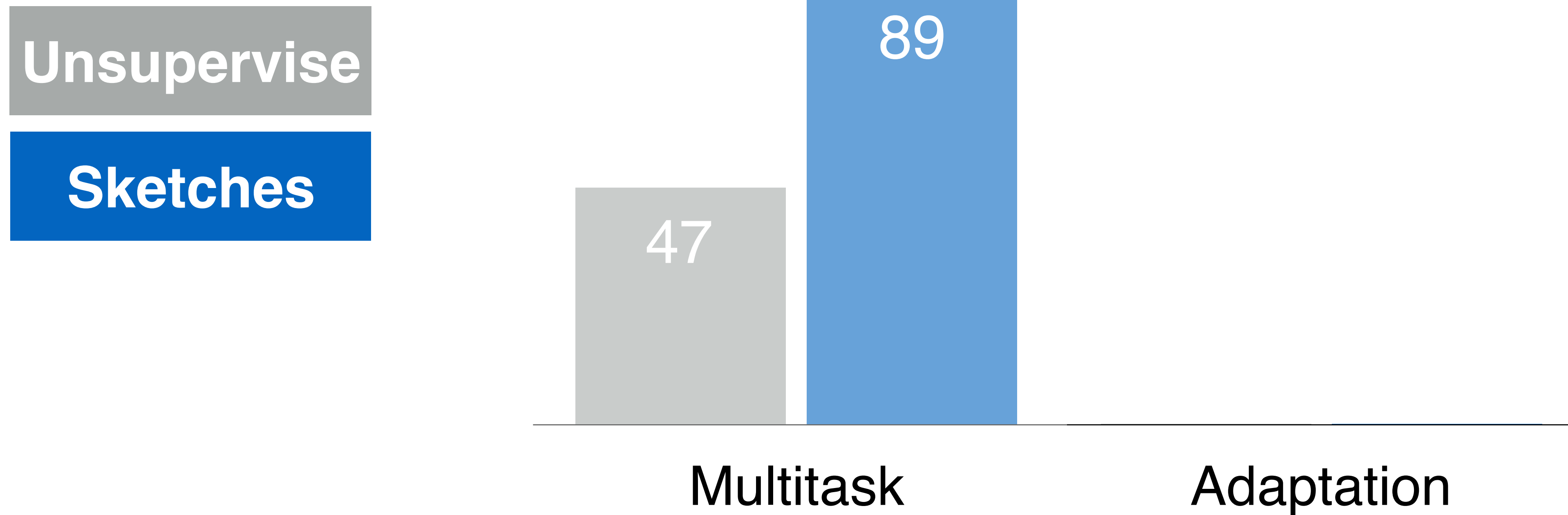
Fast adaptation

What if I don't get a sketch at test time?



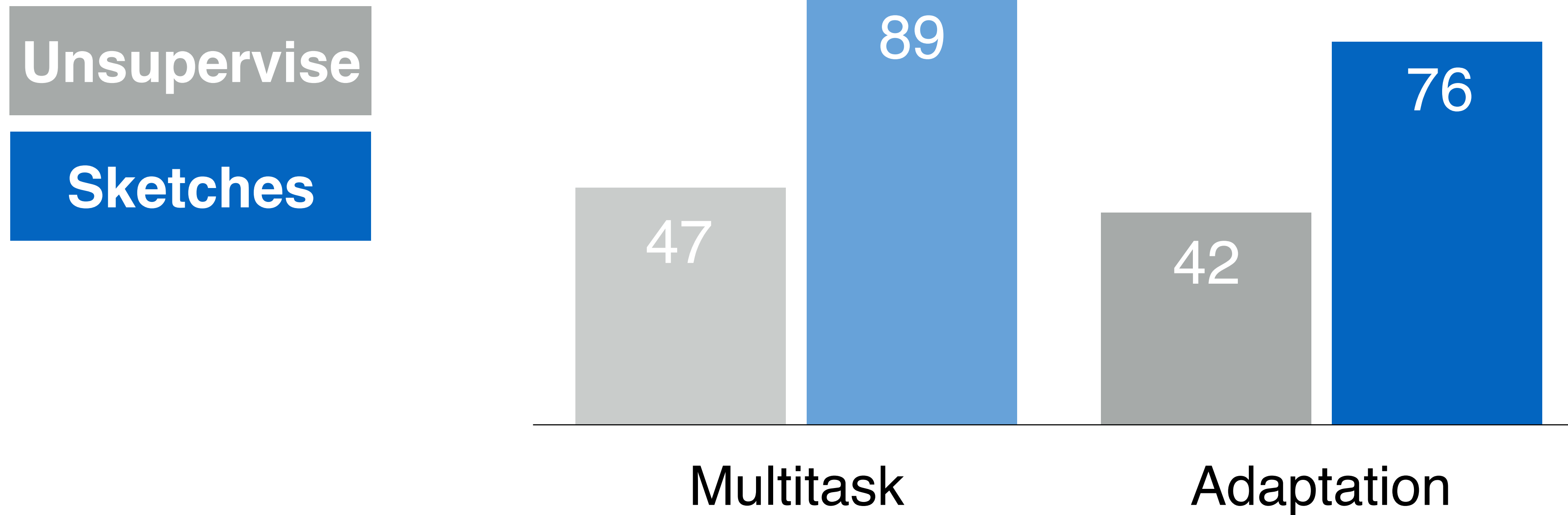
Fast adaptation

What if I don't get a sketch at test time?

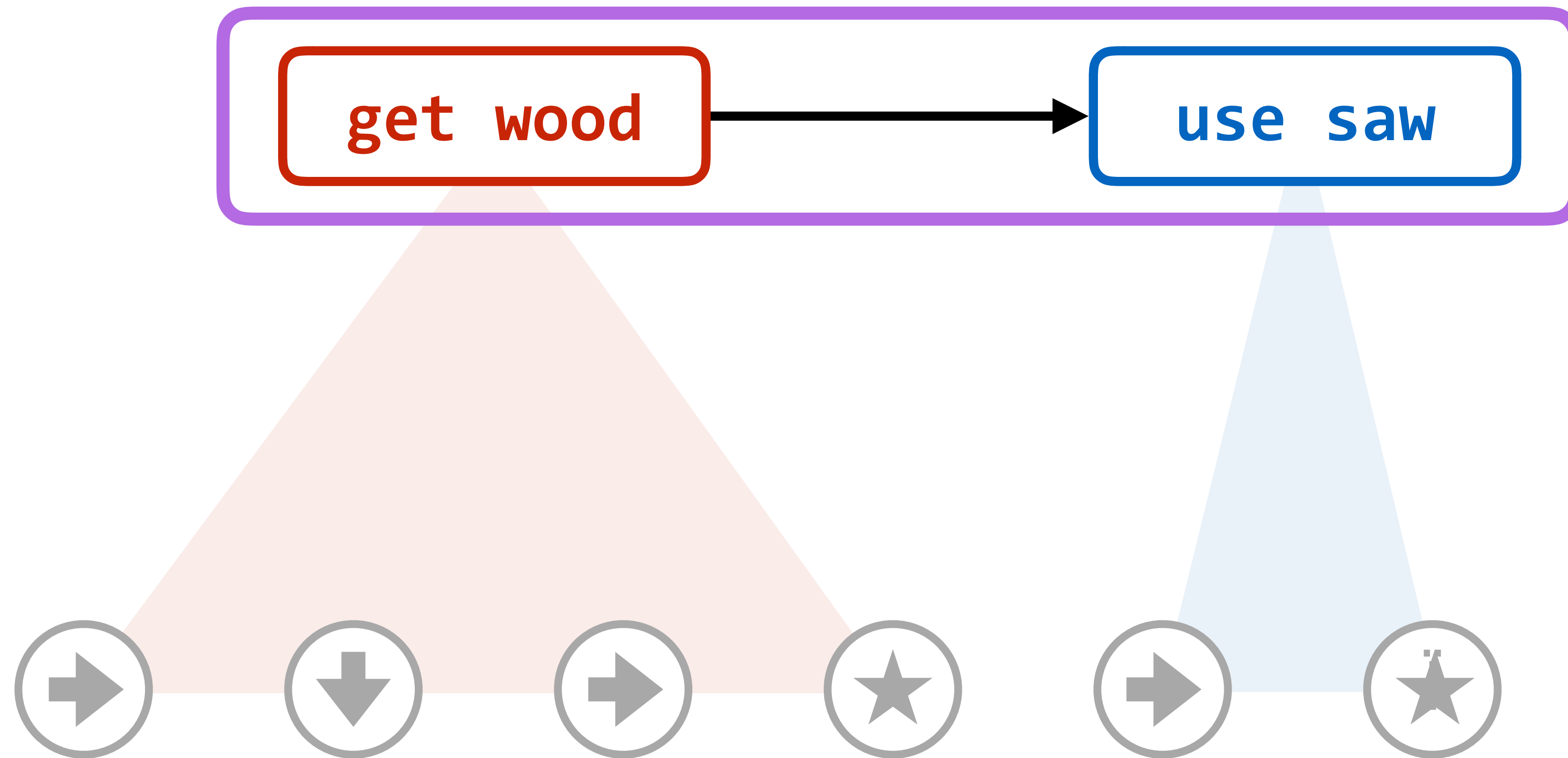


Fast adaptation

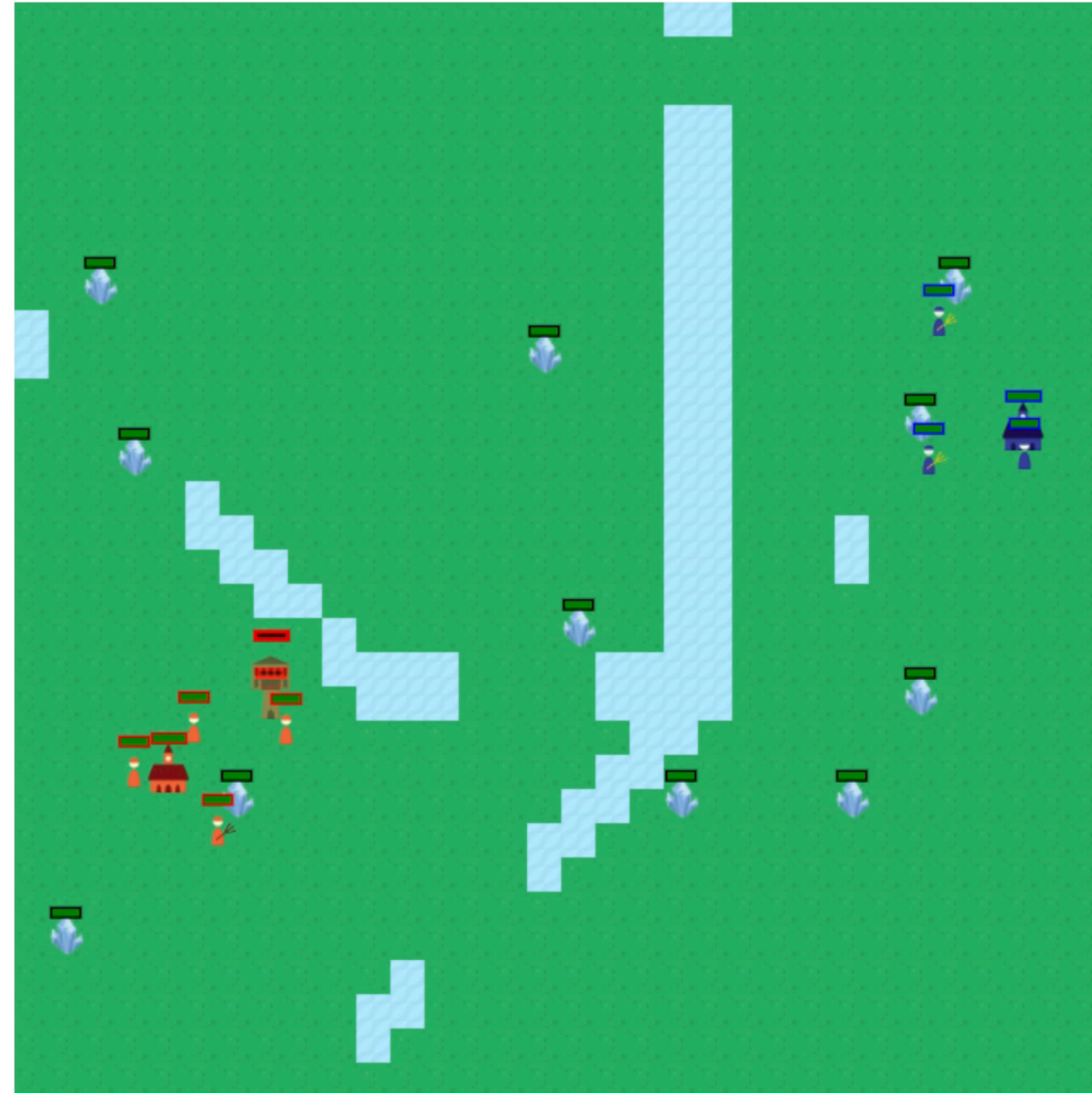
What if I don't get a sketch at test time?



Learning from policy sketches



Natural language options

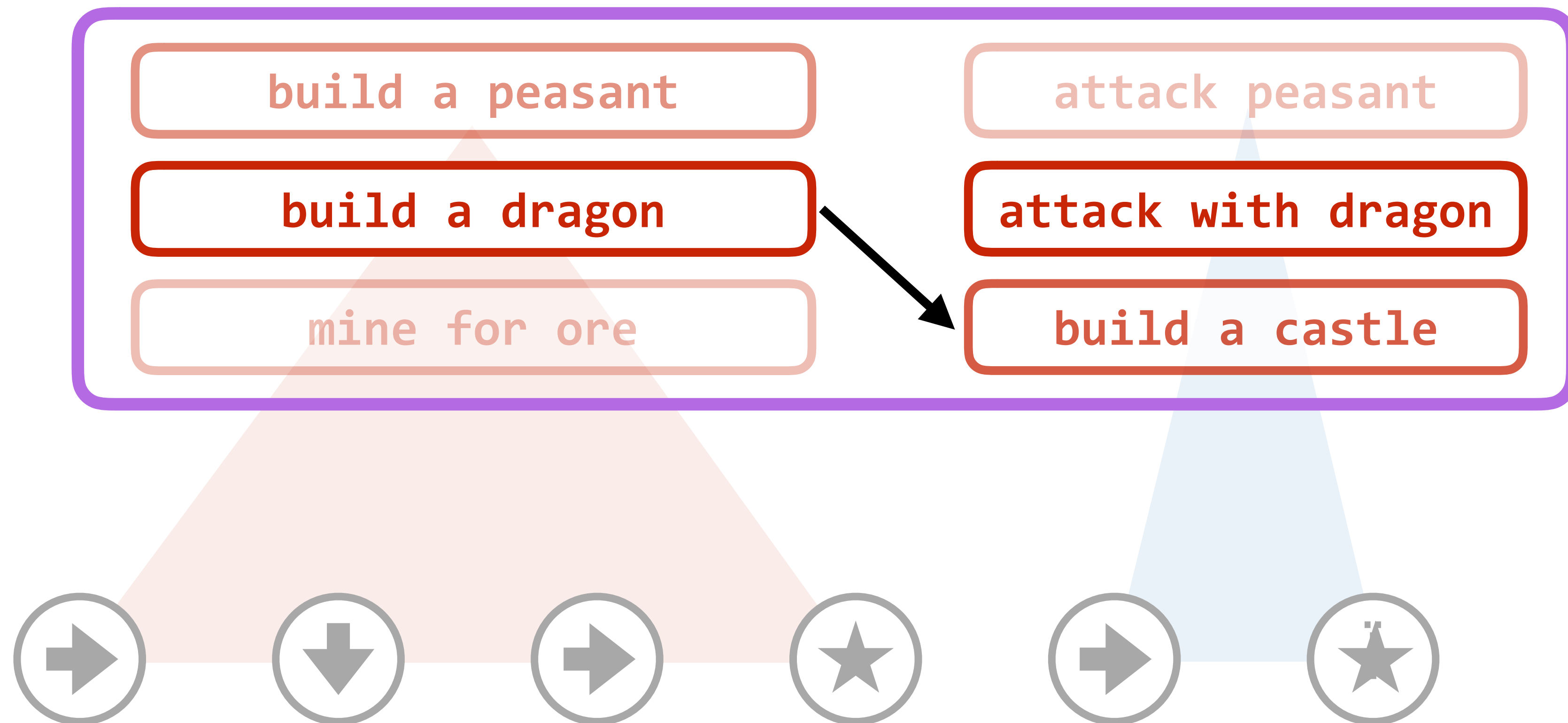


Current order to execute on:

build 6 peasant

[Hu et al. 2019, “Hierarchical Decision Making by Generating and Following Natural Language Instructions”]

Learning with natural language options

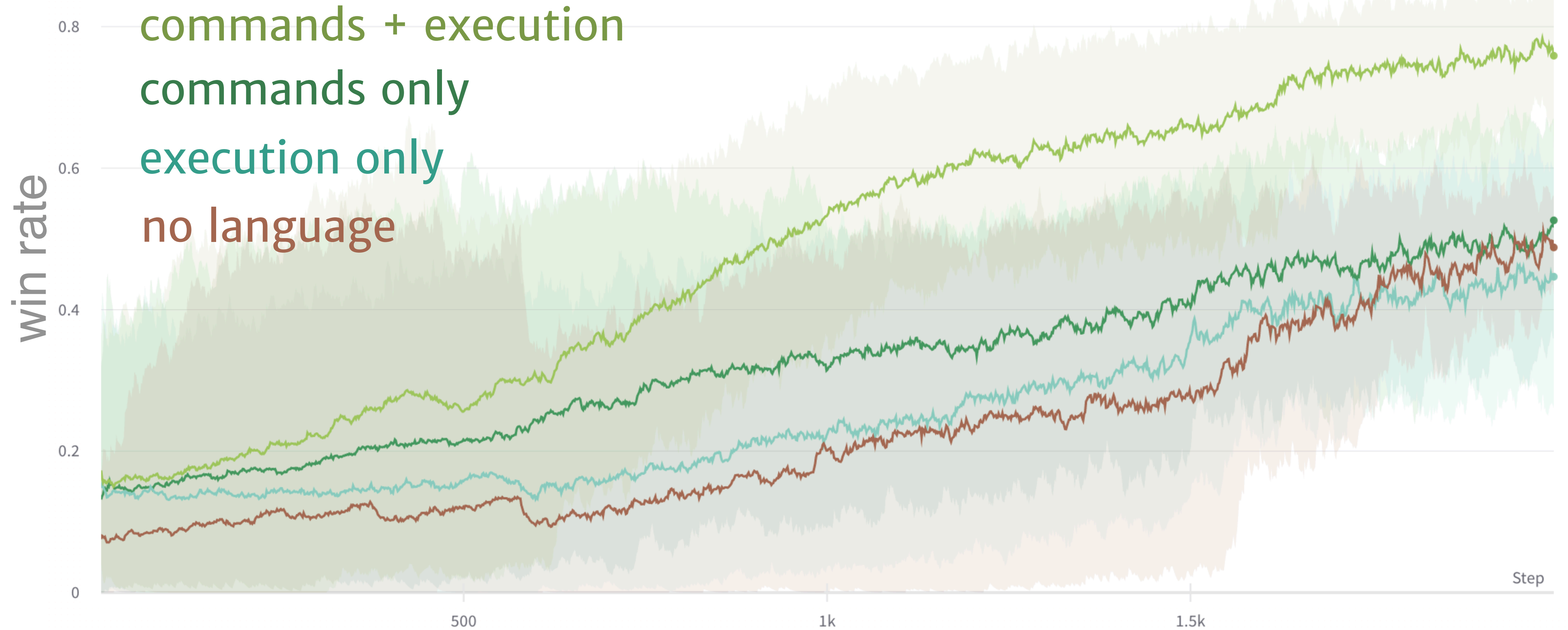


[Jacob and Andreas. "Adaptable RL with natural language hierarchies." In prep.]

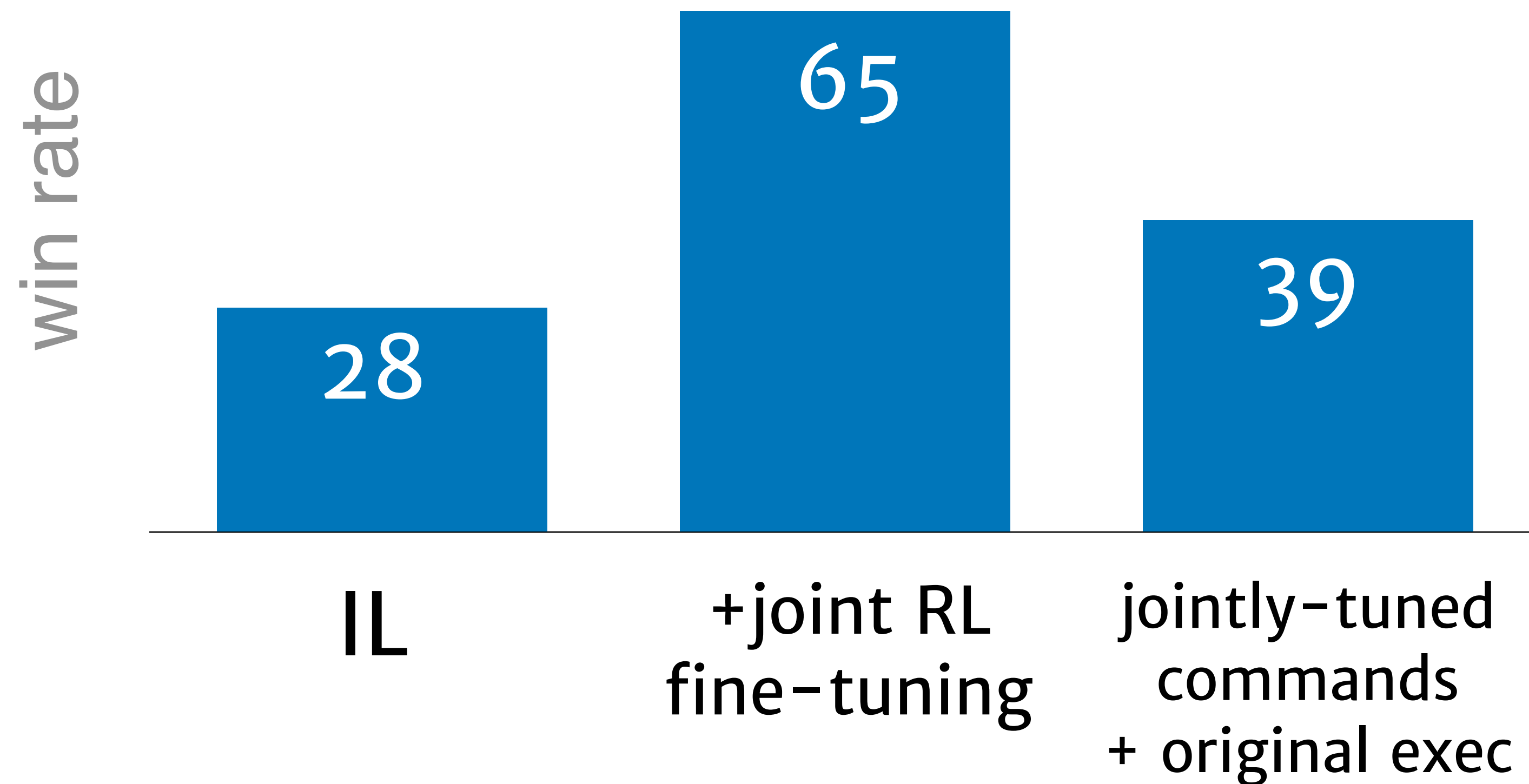
A.P. Jacob



Learning with natural language options



Learning with natural language options



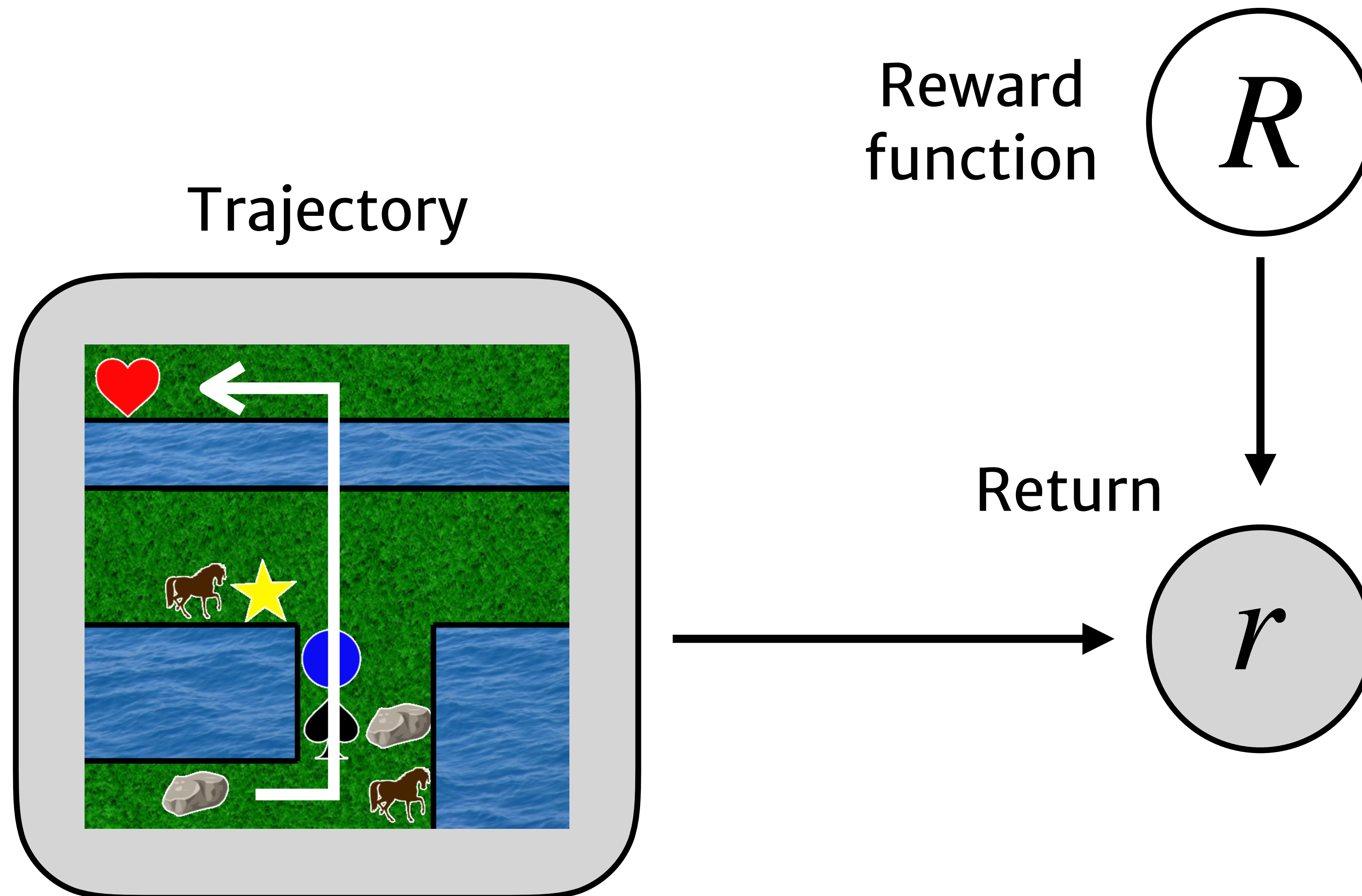
Language as a representation of goals

Language for goal inference

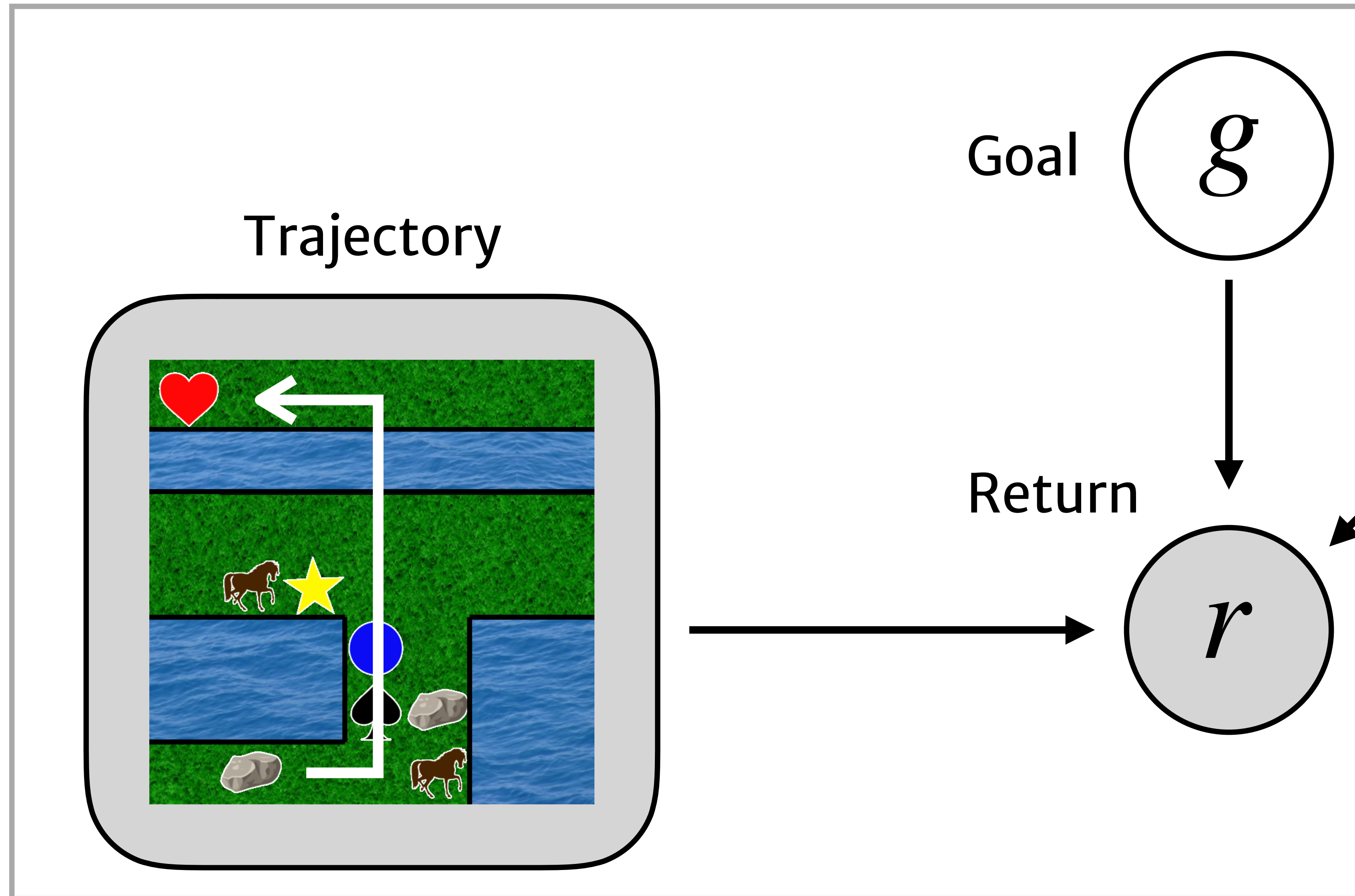


[Hermer-Vazquez, Spelke, Katznelson 1999]

Language for goal inference

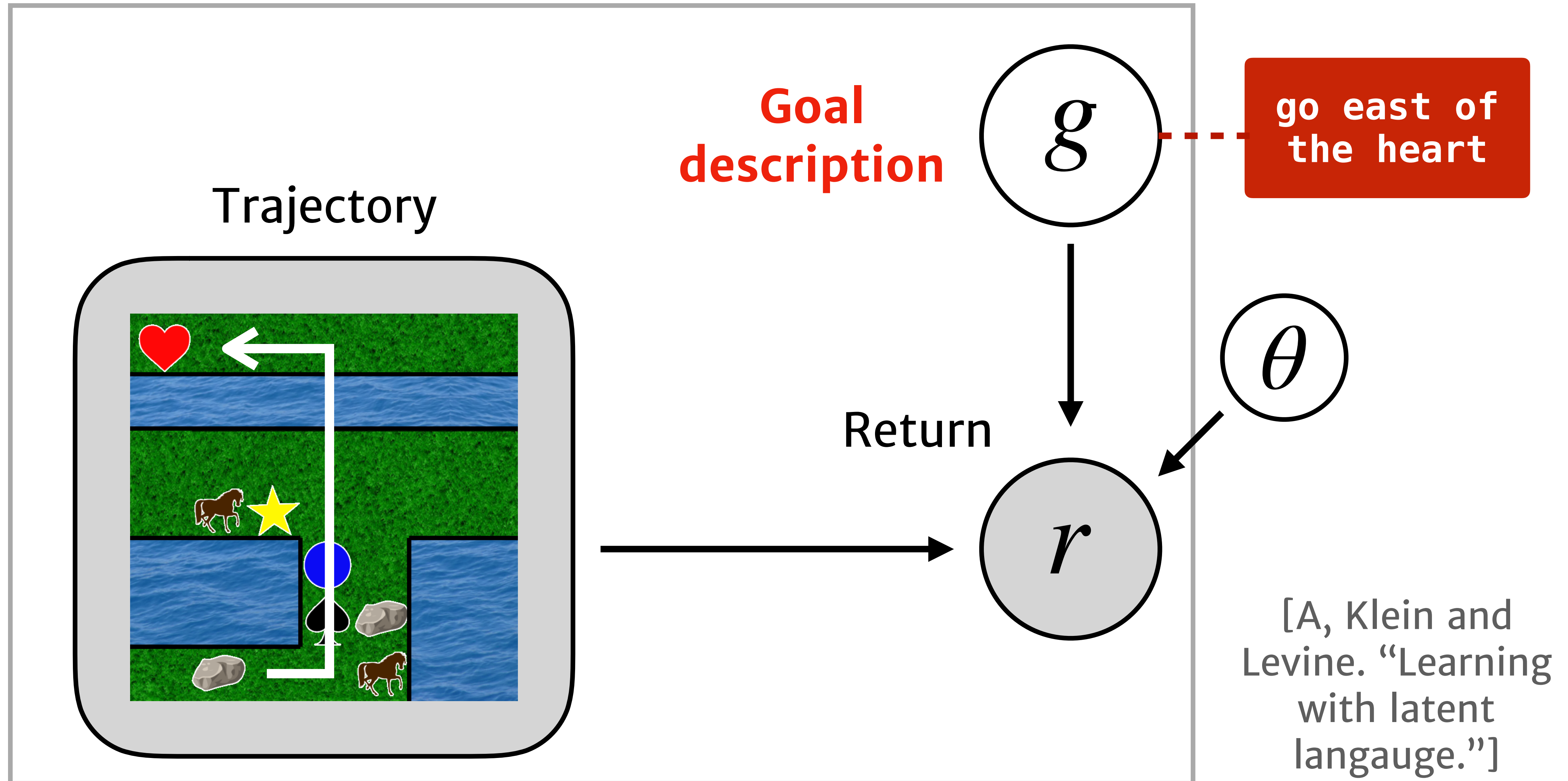


Language for goal inference

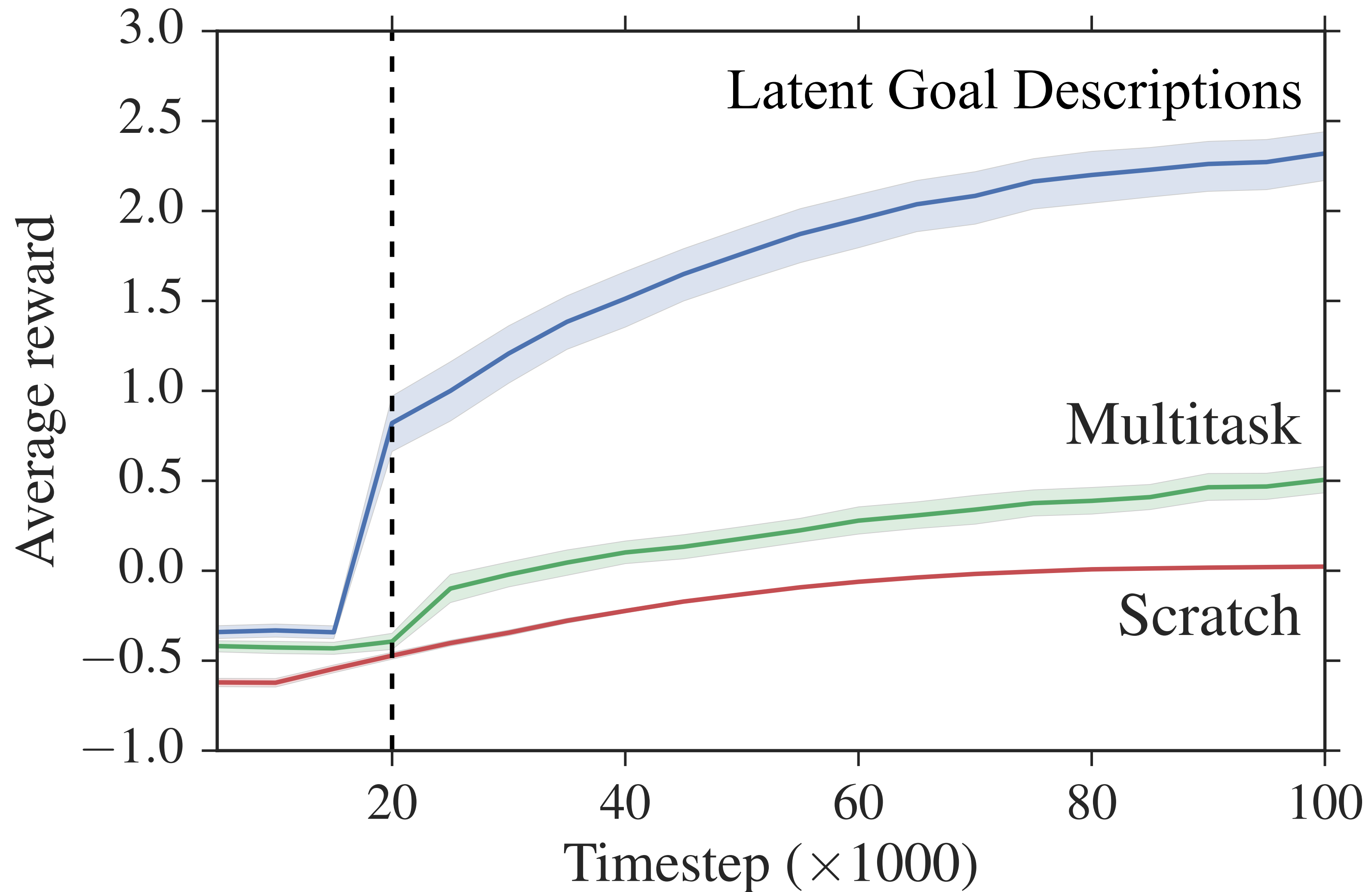


[e.g. Dimitrikakis
& Rothkopf,
“Bayesian
Multitask IRL”]

Language for goal inference



Experimental results



Language for goal inference

examples

emboldens	embo l dec s
kisses	kiss e s
loneliness	lo c el i cess
vein	ve i c
dogtrot	dog t rot

true description

replace all n s
with c

change any n
to a c

pred. description

true output

loocies



loonies

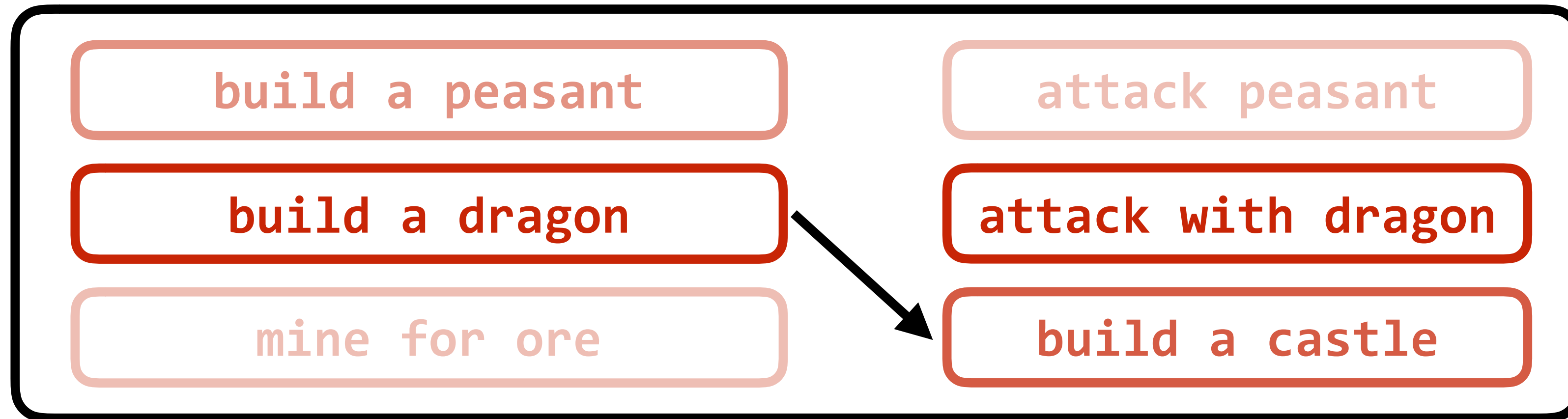


loocies

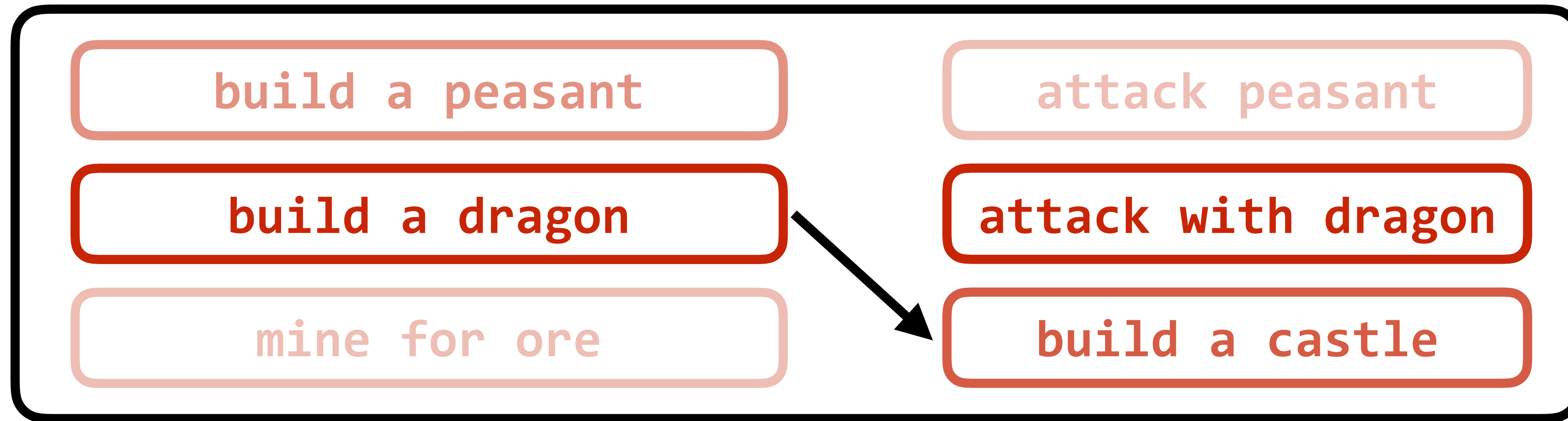
pred. output

Language as a representation of MDPs?

Language for goal inference

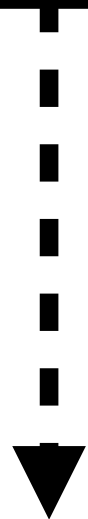


Language for goal inference



$p(\text{string} | \text{string})$

$p(\text{string})$





Language modeling and representation

and

I'll

transformer

cheap

[MASK]

delicious

[SEP]

green

definitely

go

back

[Devlin et al. “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding”]

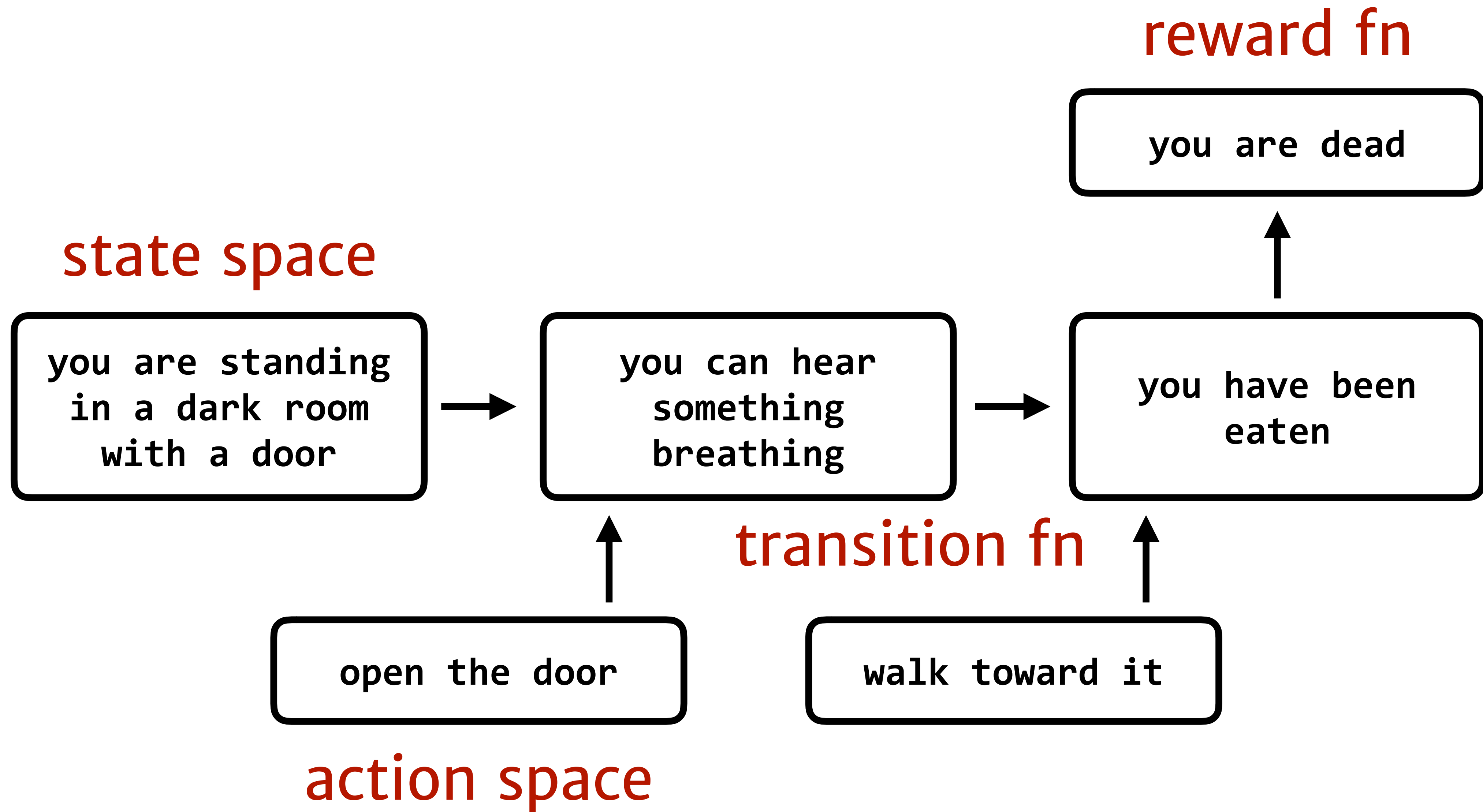



Language modeling and representation

Query	Prediction
The color of a banana is [?].	green
The capital of [?] is Dhaka.	Bangladesh
I can use a [?] to chop a carrot.	knife
I can use a [?] to mince a carrot.	knife
I can use a [?] to scrub a carrot.	brush
Plates are found in the [?] room.	dining
If I drop a glass, it will [?].	explode

[Devlin et al. “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding”]

The string-valued MDP





Text adventure games

Observation: **West of House** You are standing in an open field west of a white house, with a boarded front door. There is a small mailbox here.

Action: **Open mailbox**

Observation: Opening the small mailbox reveals a leaflet.

Action: **Read leaflet**

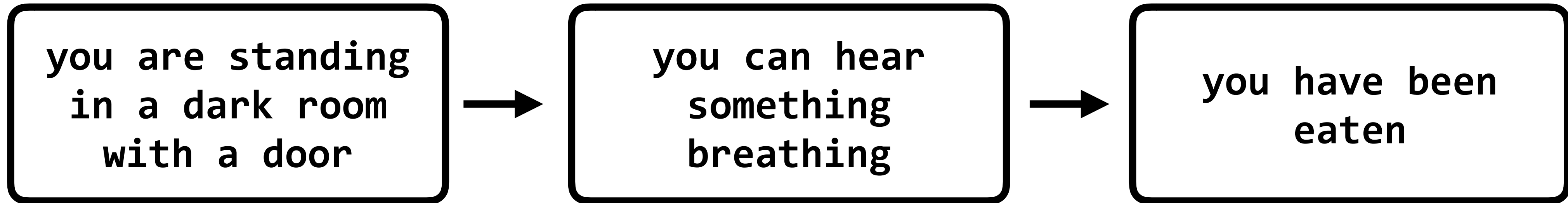
Observation: (Taken) "WELCOME TO ZORK! ZORK is a game of adventure, danger, and low cunning. In it you will explore some of the most amazing territory ever seen by mortals. No computer should be without one!"

Action: **Go north**

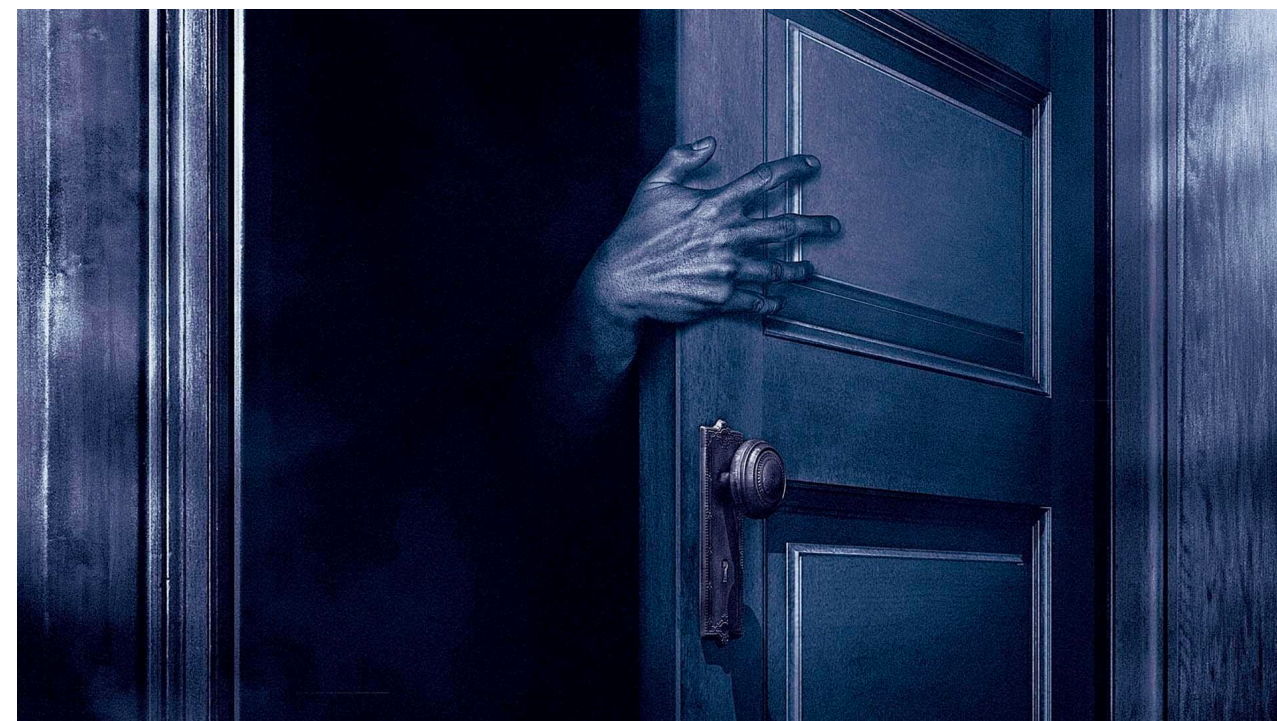
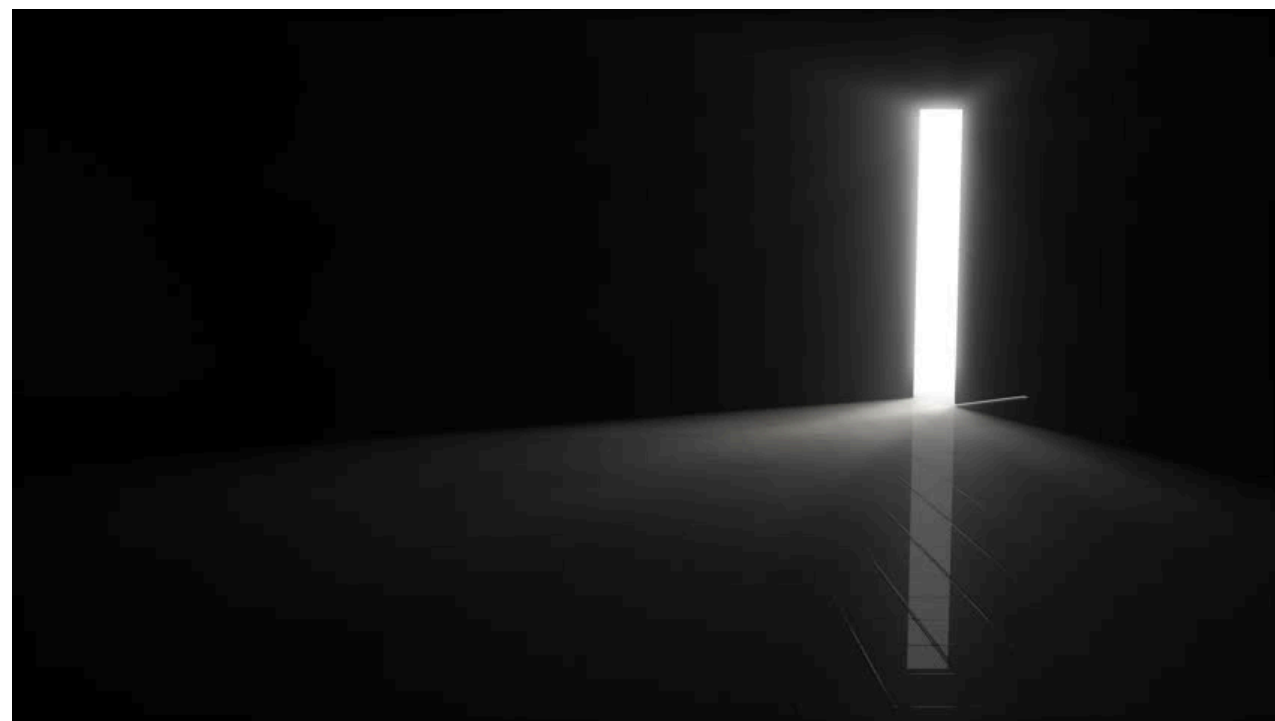
Observation: **North of House** You are facing the north side of a white house. There is no door here, and all the windows are boarded up. To the north a narrow path winds through the trees.

[Ammanabrolu et al. 2020]

From language to the real world



~||



Luketina *et al.*,

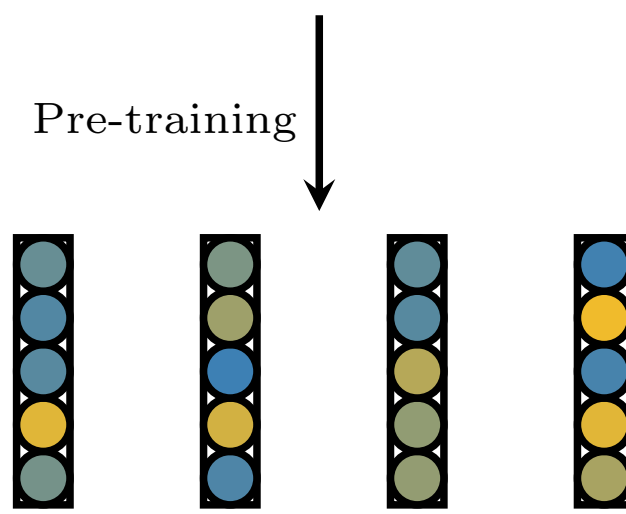
A survey of reinforcement learning informed by natural language

<https://arxiv.org/abs/1906.03926>

Task-independent

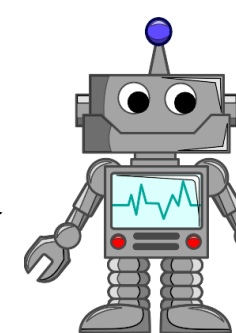
[...] having the correct
[...] known lock and
[...] unless the correct

key can open the lock [...]
key device was discovered [...]
key is inserted [...]



V_{key} V_{skull} V_{ladder} V_{rope}

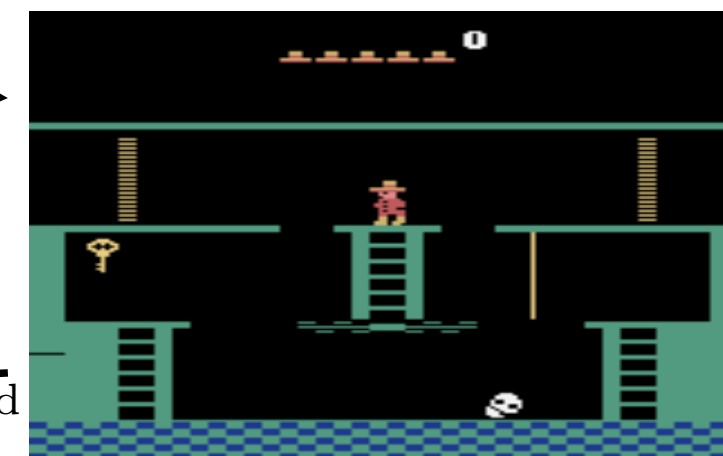
Pre-trained



Agent

Action

State, Reward



Environment

Task-dependent

Language-assisted

Key Opens a door of the same color as the key.

Skull They come in two varieties, rolling skulls and bouncing skulls ... you must jump over rolling skulls and walk under bouncing skulls.

Language-conditional

Go down the ladder and walk right immediately to avoid falling off the conveyor belt, jump to the yellow rope and again to the platform on the right.