Human-aligned Intelligent Robotics

COMP3411/9814: Artificial Intelligence

Lecture Overview

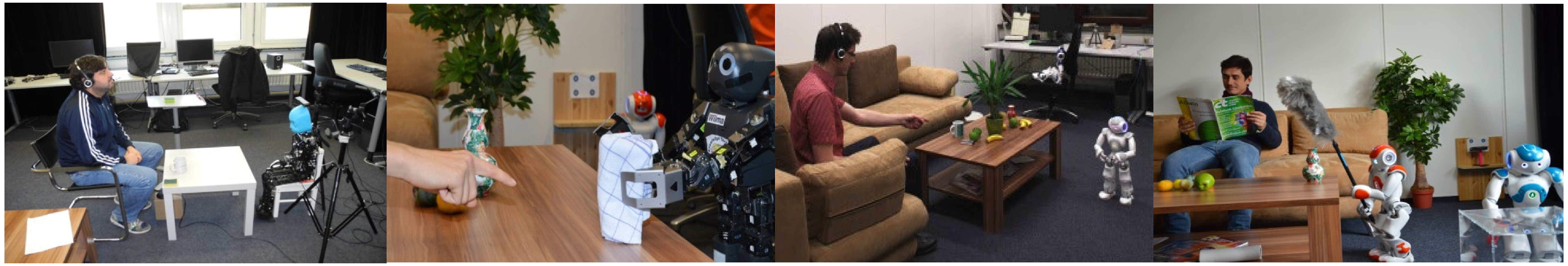
- Introduction
- Human-in-the-loop learning
- Robotic sensory modalities – Multimodal integration
- Contextual affordances
- Explainable robotic systems

Lecture Overview

- **Introduction**
- Human-in-the-loop learning
- Robotic sensory modalities – Multimodal integration
- Contextual affordances
- Explainable robotic systems

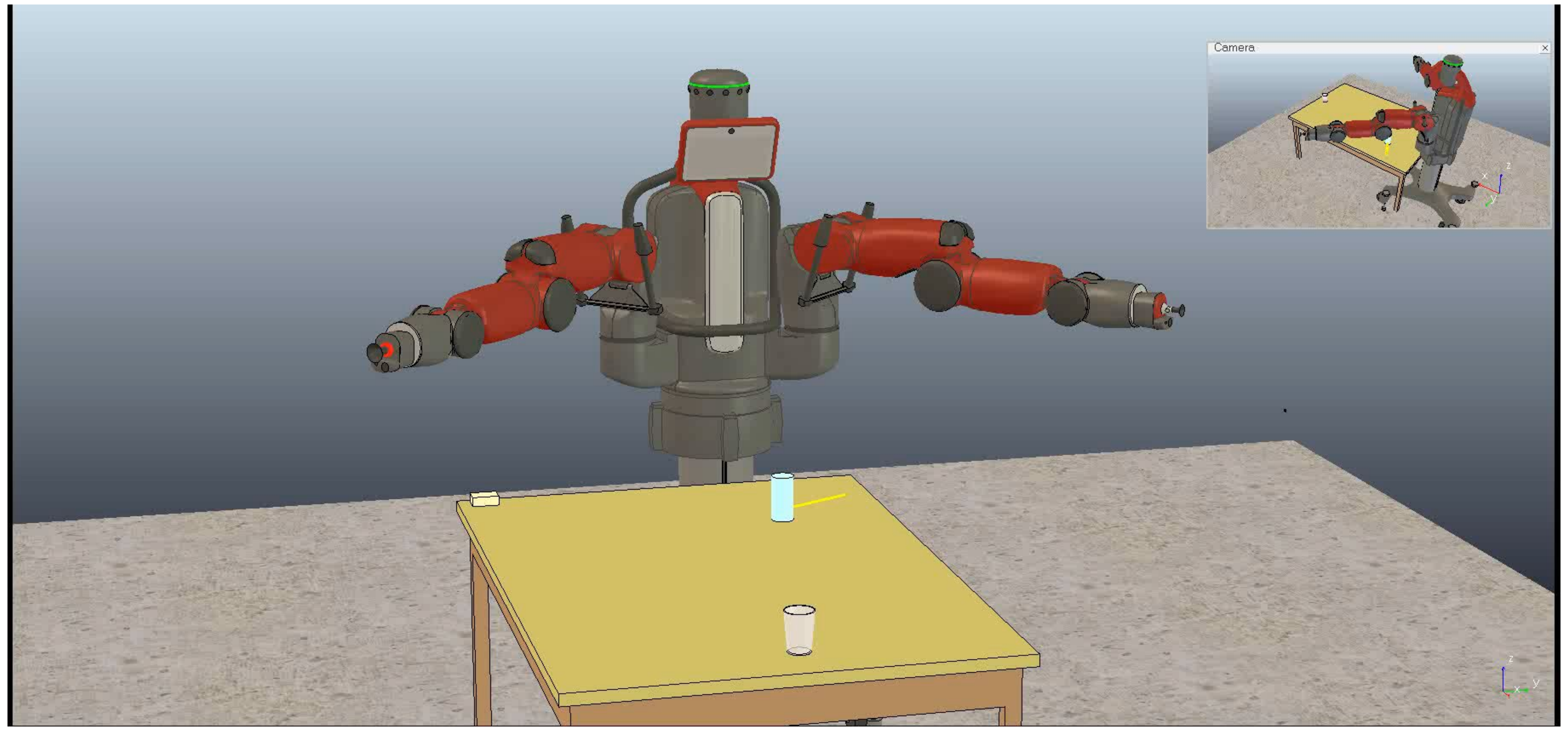
Motivation

- Intelligent assistive robots taken their first steps toward entering domestic scenarios.
- To accomplish complex tasks successfully, can only be addressed if the robot constantly acquires new skills.
- An open challenging issue is the time required by a robot to autonomously learn a new task.



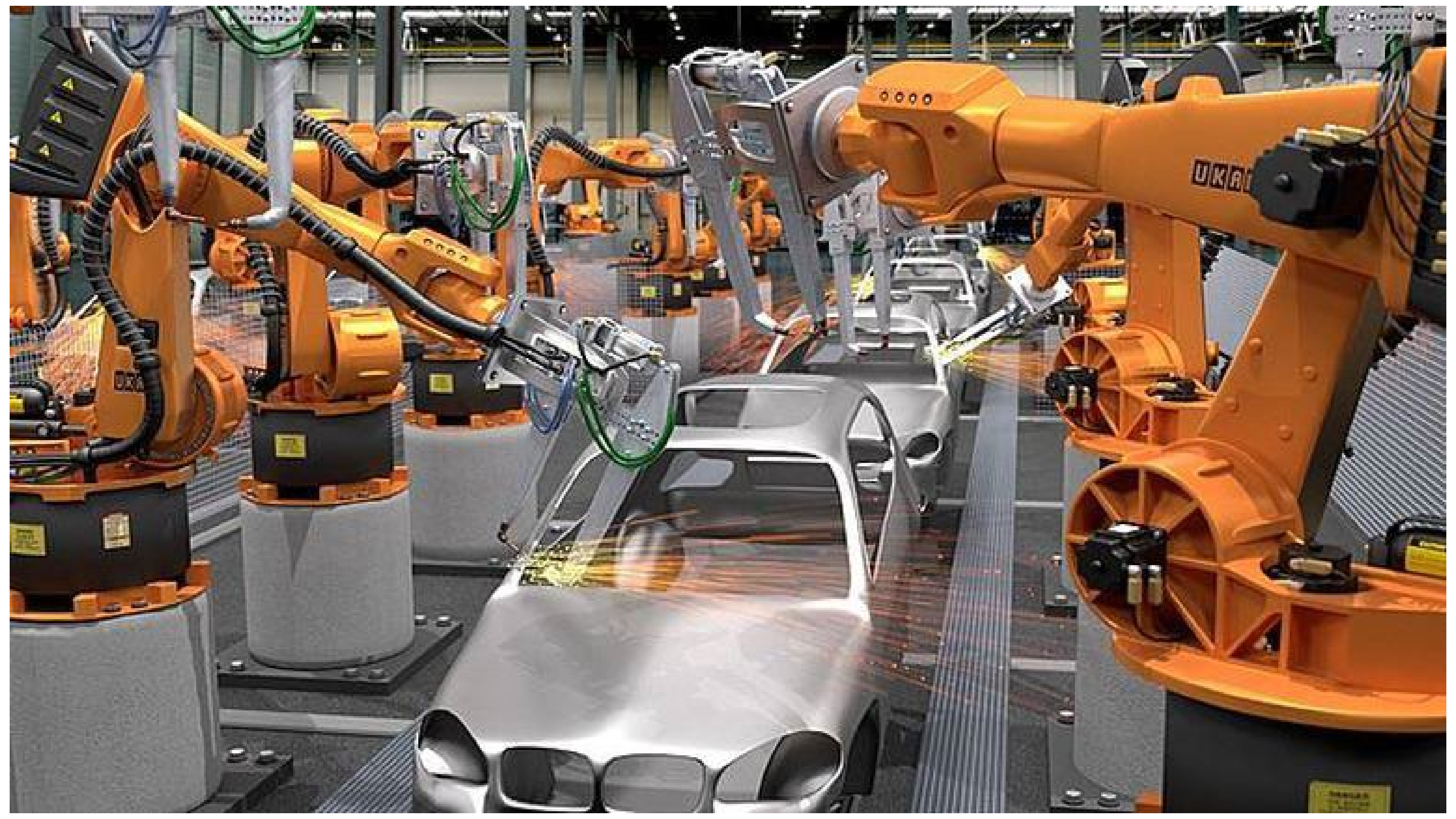
Introduction

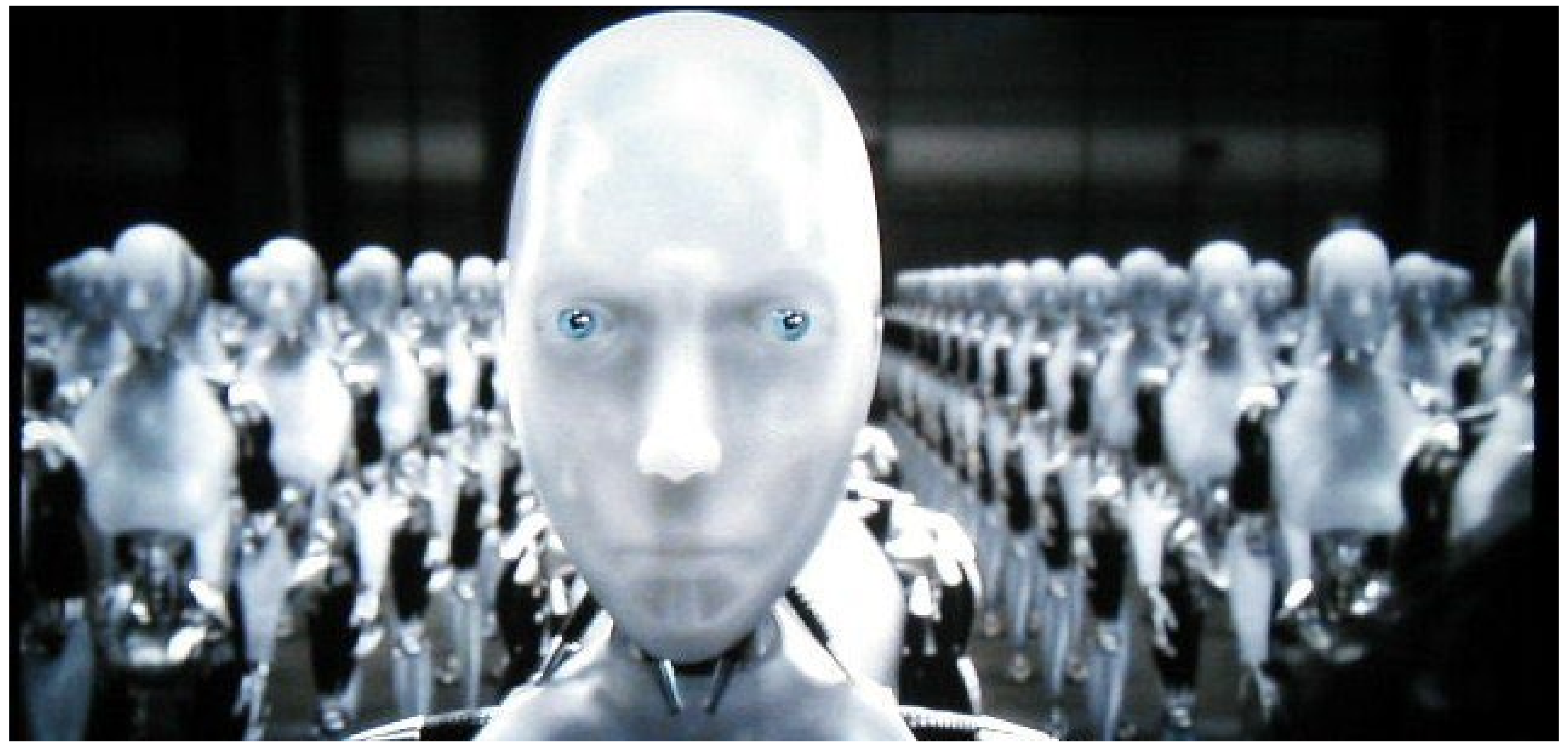
- Robots are able to autonomously learn new tasks. Problem: the time needed for a robot to acquire new skills.
    - Integration of parent-like trainers to scaffold the learning.
    - A parent-like trainer or teacher may speed up the learning process.
    - Advice may not be clear and misunderstood.



Current (mis)understanding of AI

How are currently perceived AI-based systems?



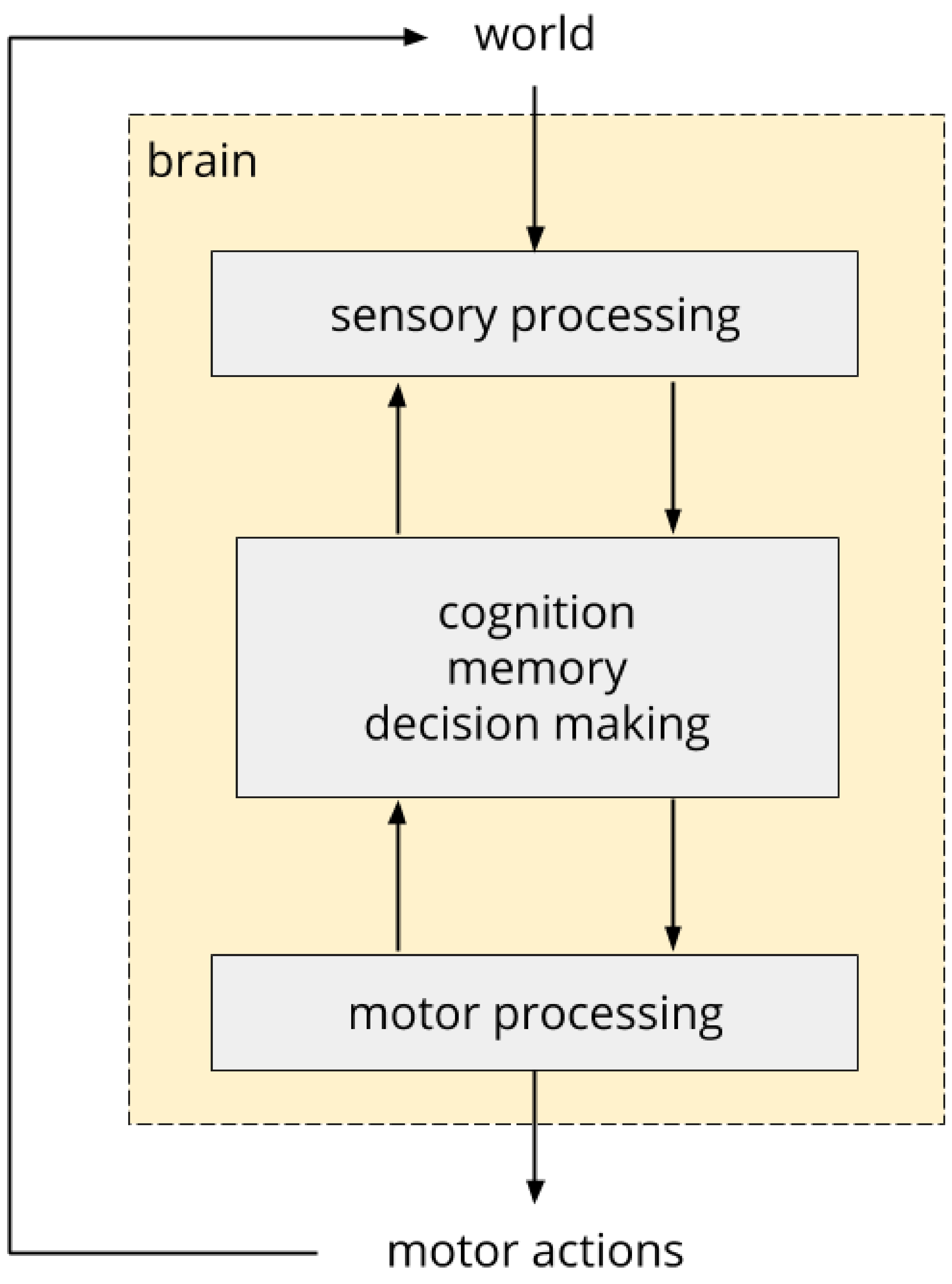


Current (mis)understanding of AI

But in reality, it's still an open problem

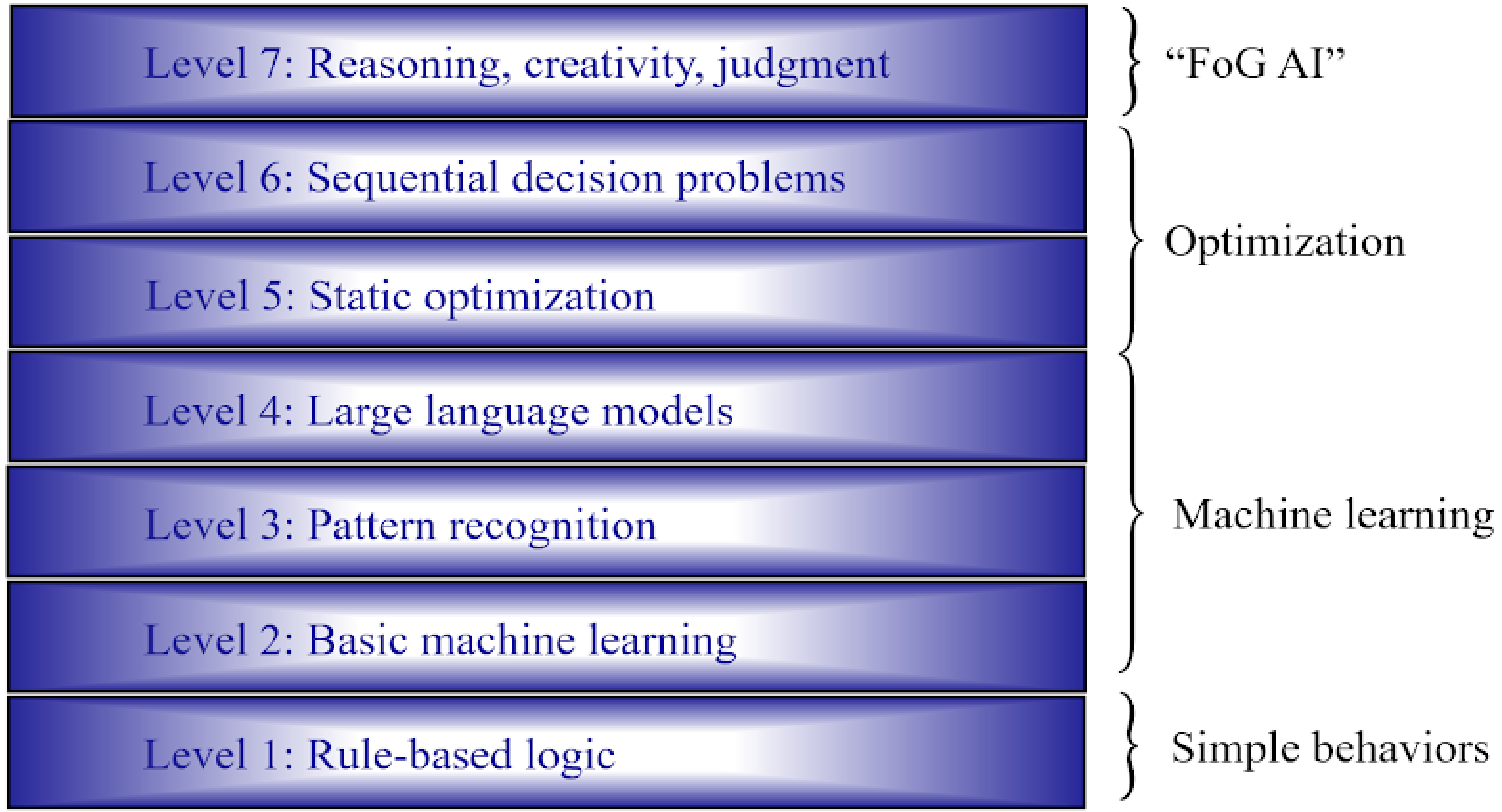
Introduction

- Brain-world Interactive Framework



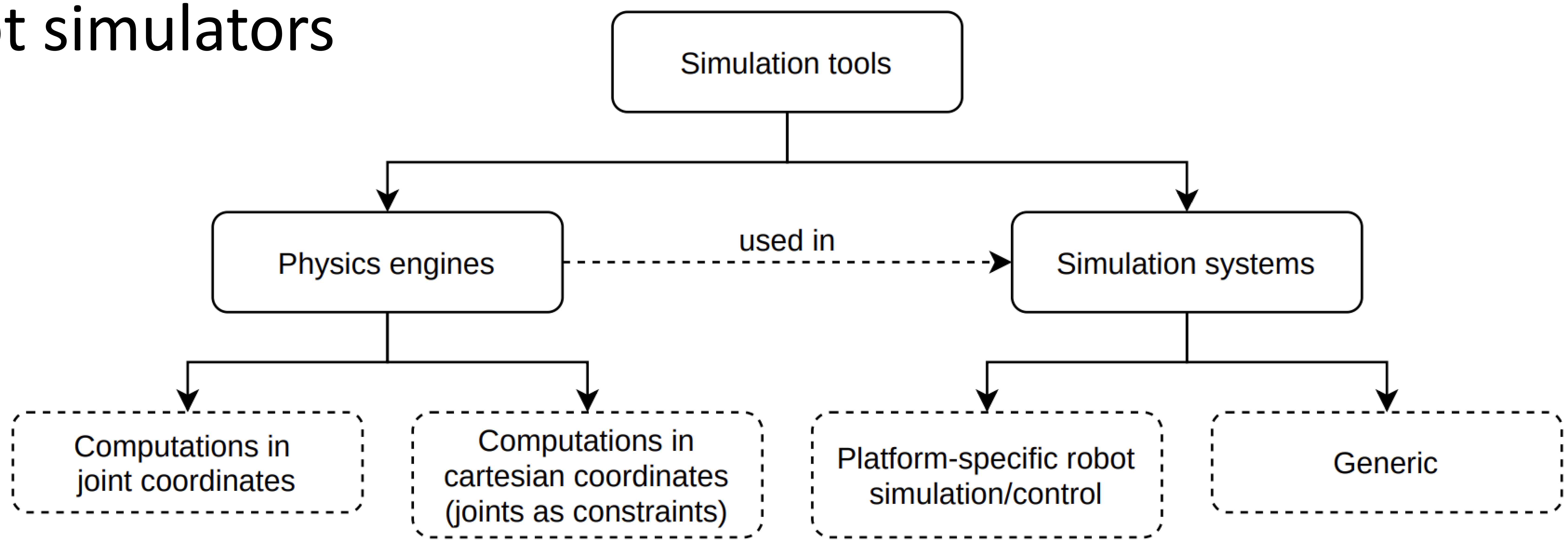
Introduction

- The 7 levels of AI [Powell, 2024].

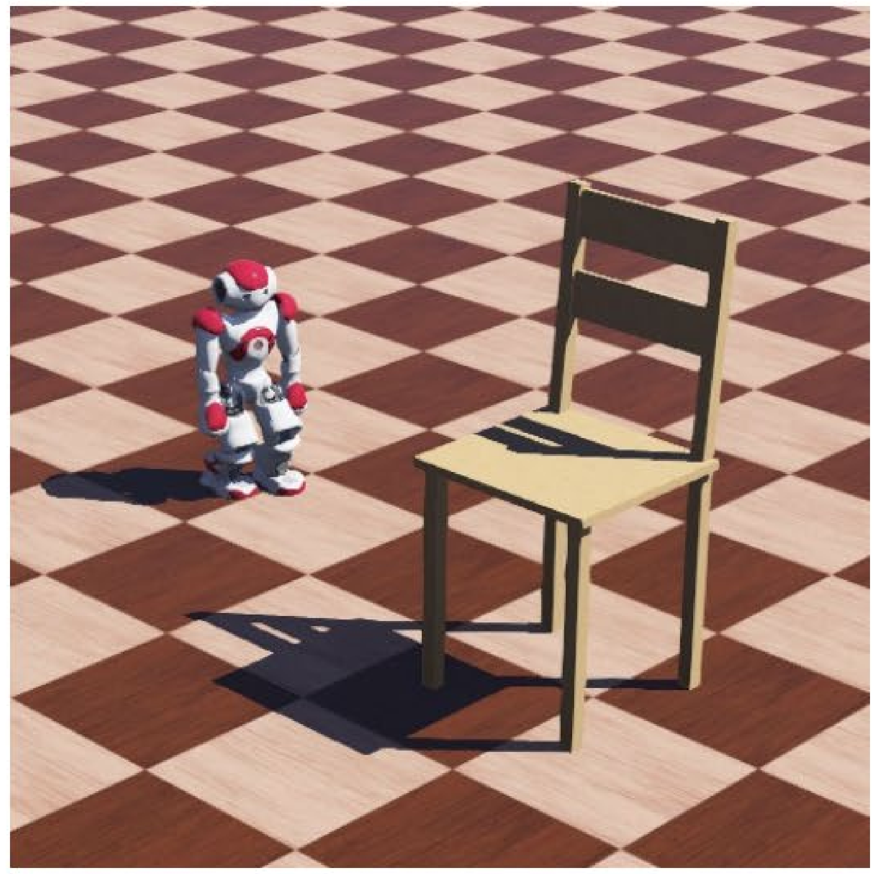


Introduction

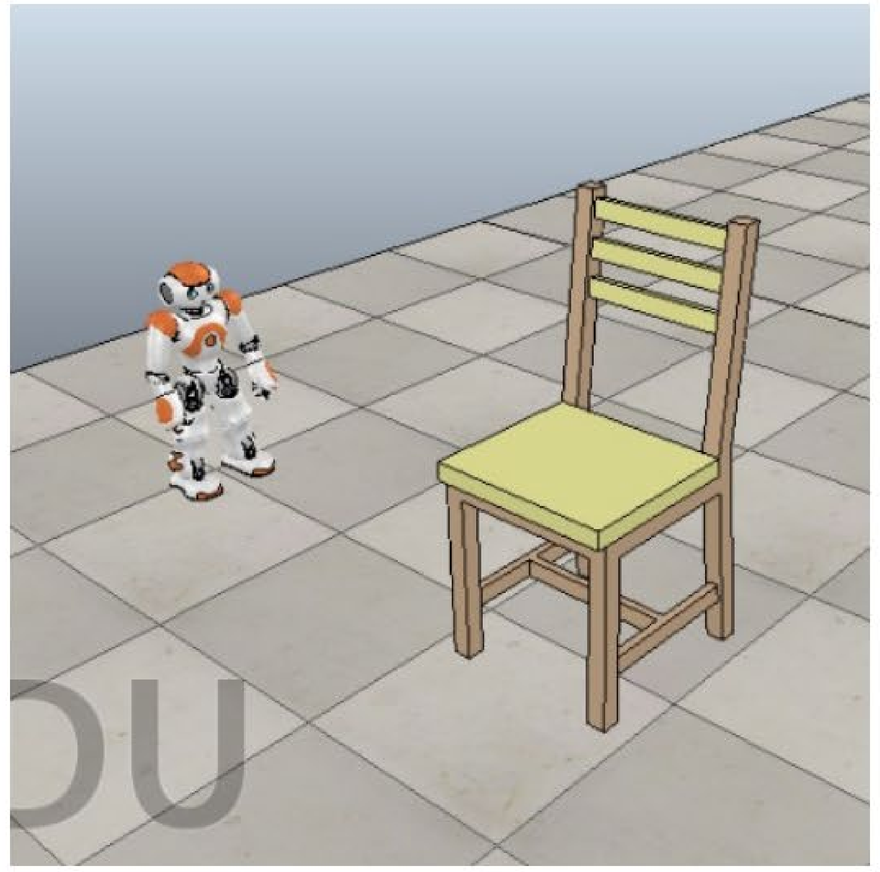
- Robot simulators



(a) Gazebo simulator.



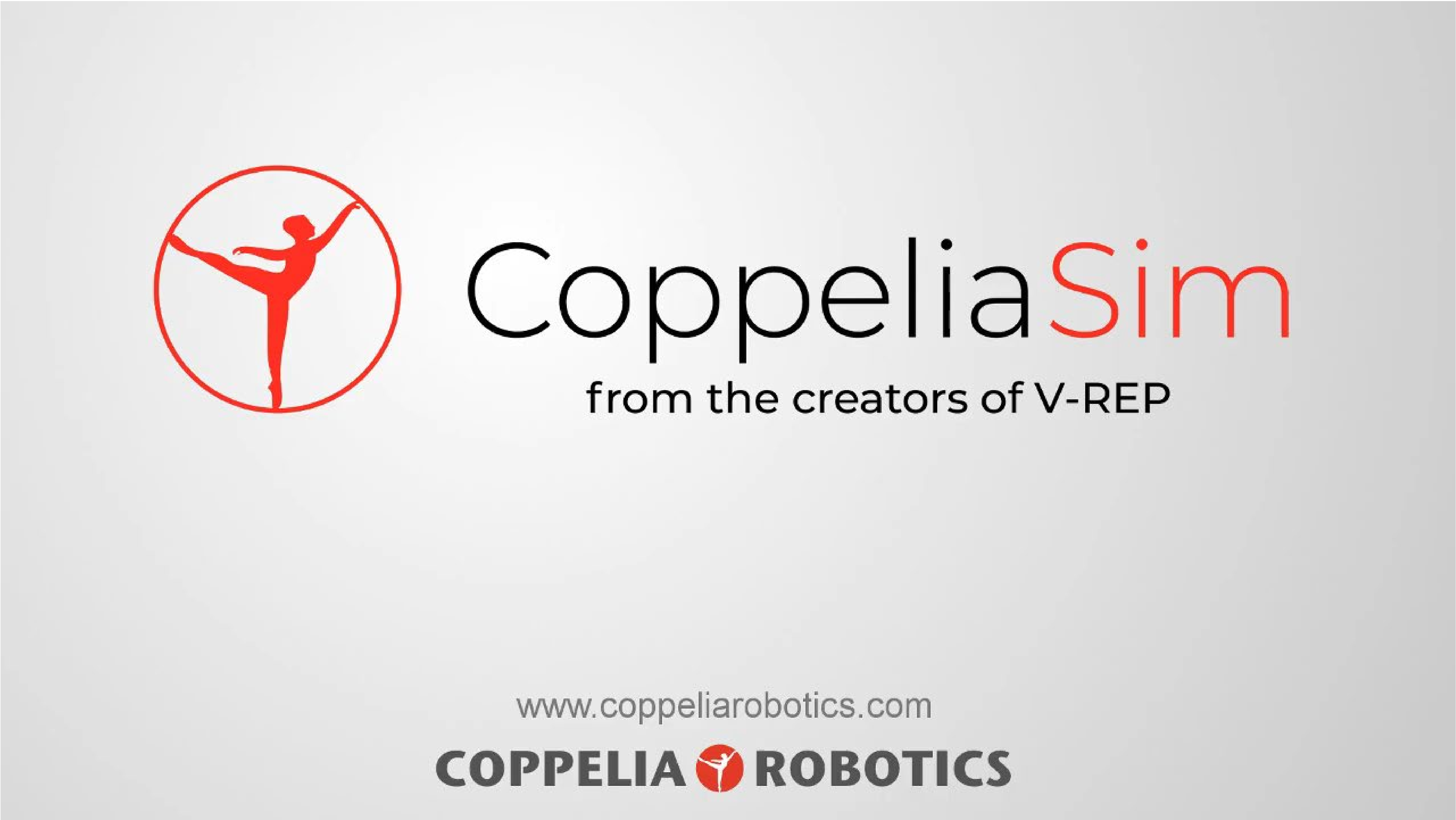
(b) Webots simulator.



(c) V-REP simulator.

Introduction

- Robot simulators
- [Video link]

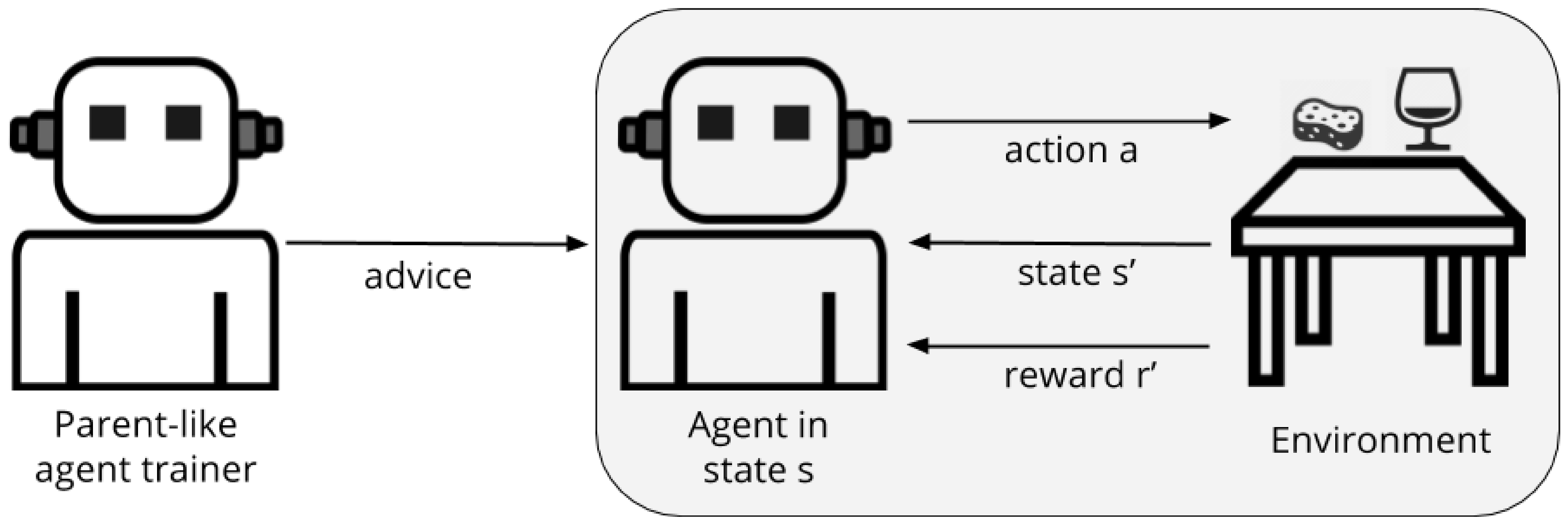


Lecture Overview

- Introduction
- **Human-in-the-loop learning**
- Robotic sensory modalities – Multimodal integration
- Contextual affordances
- Explainable robotic systems

Interactive Feedback*

- Artificial agent or human trainer advising learner-agents.
- Levels of interaction and consistency of feedback.
- Advice might be assumed to be fully observable.

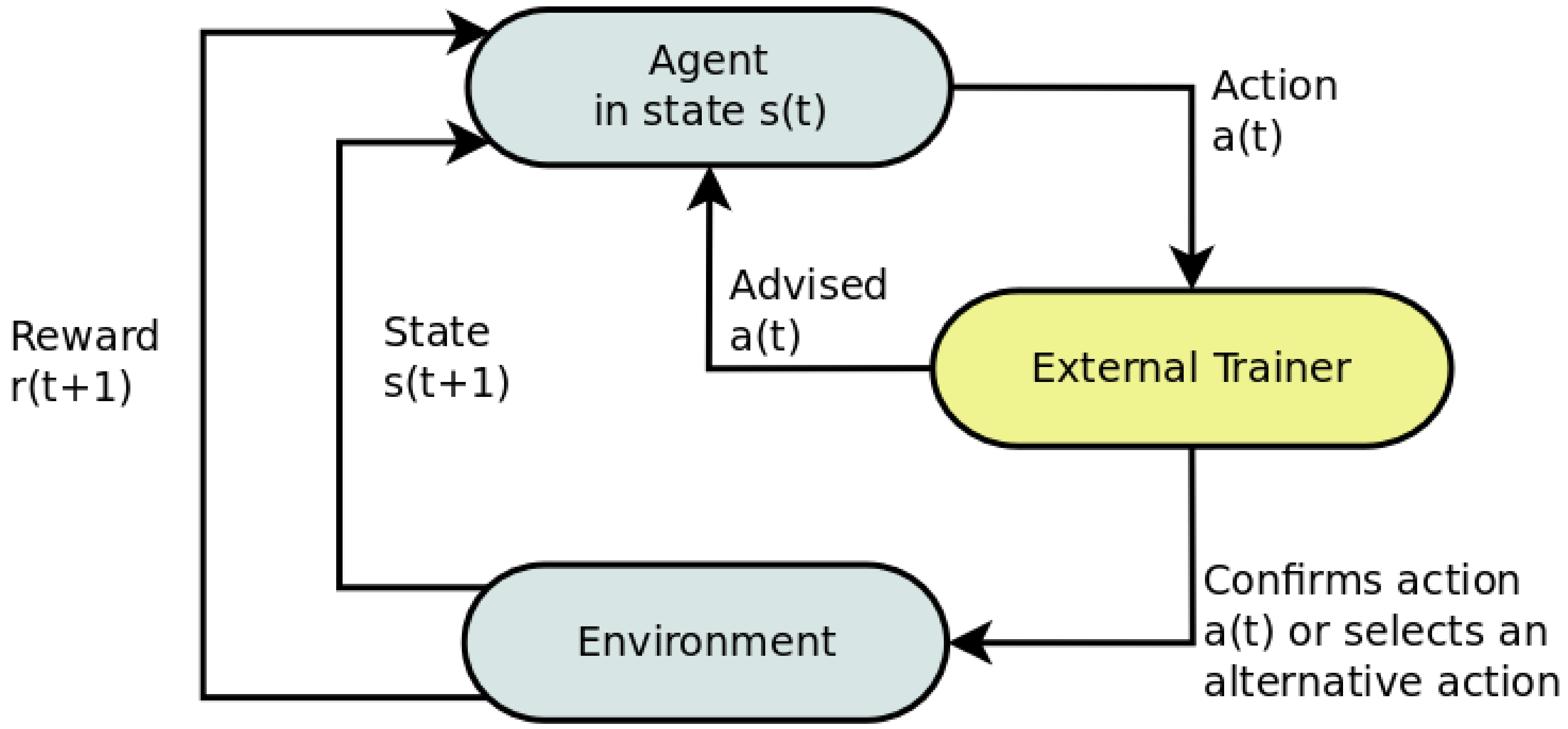


* Cruz, F., Magg, S., Weber, C., and Wermter, S. “Improving reinforcement learning with interactive feedback and affordances”. In *Proceedings of the Fourth Joint IEEE International Conference on Development and Learning and on Epigenetic Robotics (ICDL-EpiRob)*, pp. 125-130, Genoa, Italy, 2014.

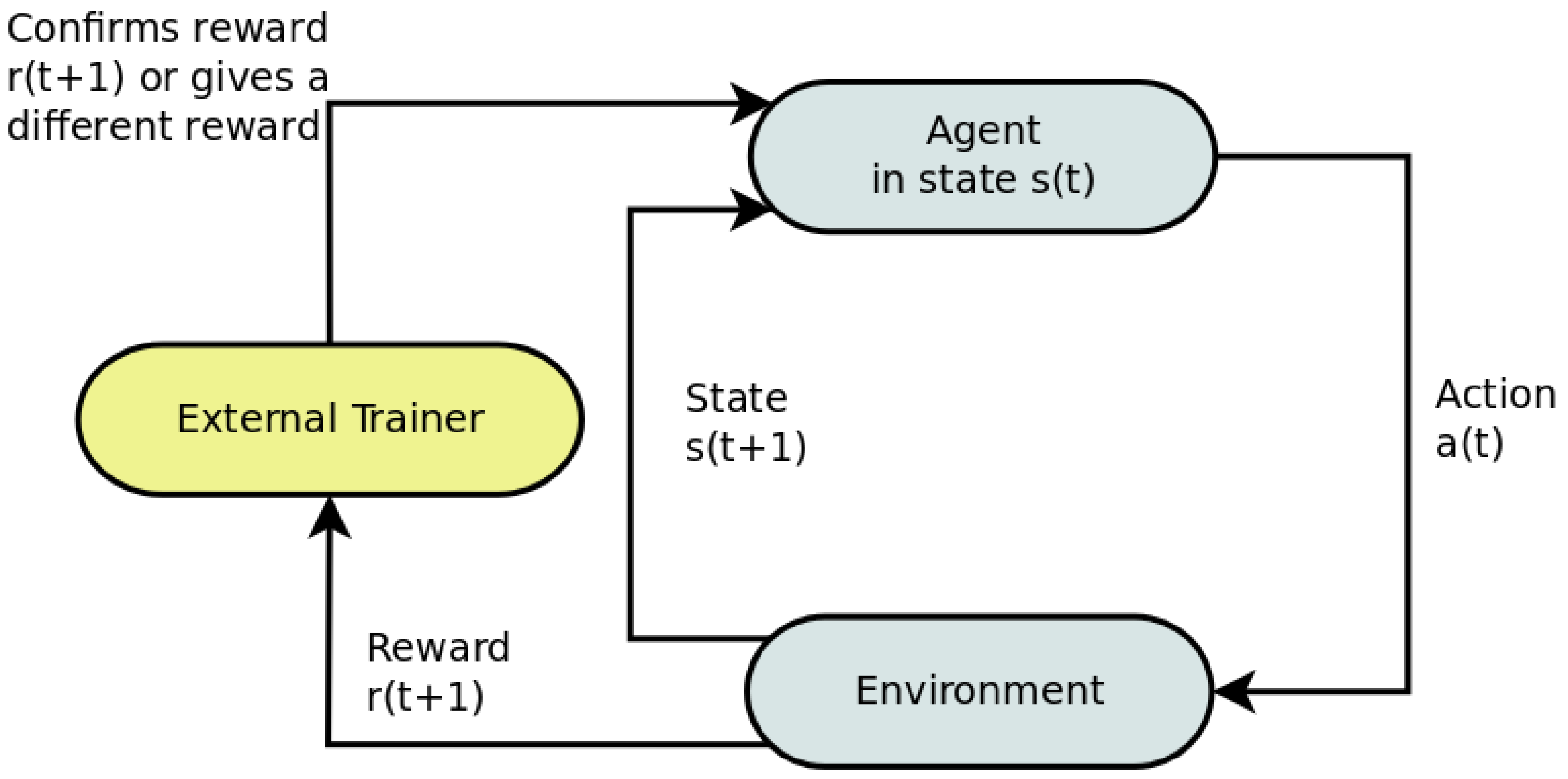
* Cruz, F., Magg, S., Weber, C., and Wermter, S. “Training agents with interactive reinforcement learning and contextual affordances”. *IEEE Transactions on Cognitive and Developmental Systems (TCDS)*, Vol. 8, Nr. 4, pp. 271-284, December 2016

Learning by Feedback

- Policy shaping: the trainer is able to change an action.
- Reward shaping: the trainer is able to modify the proposed reward.



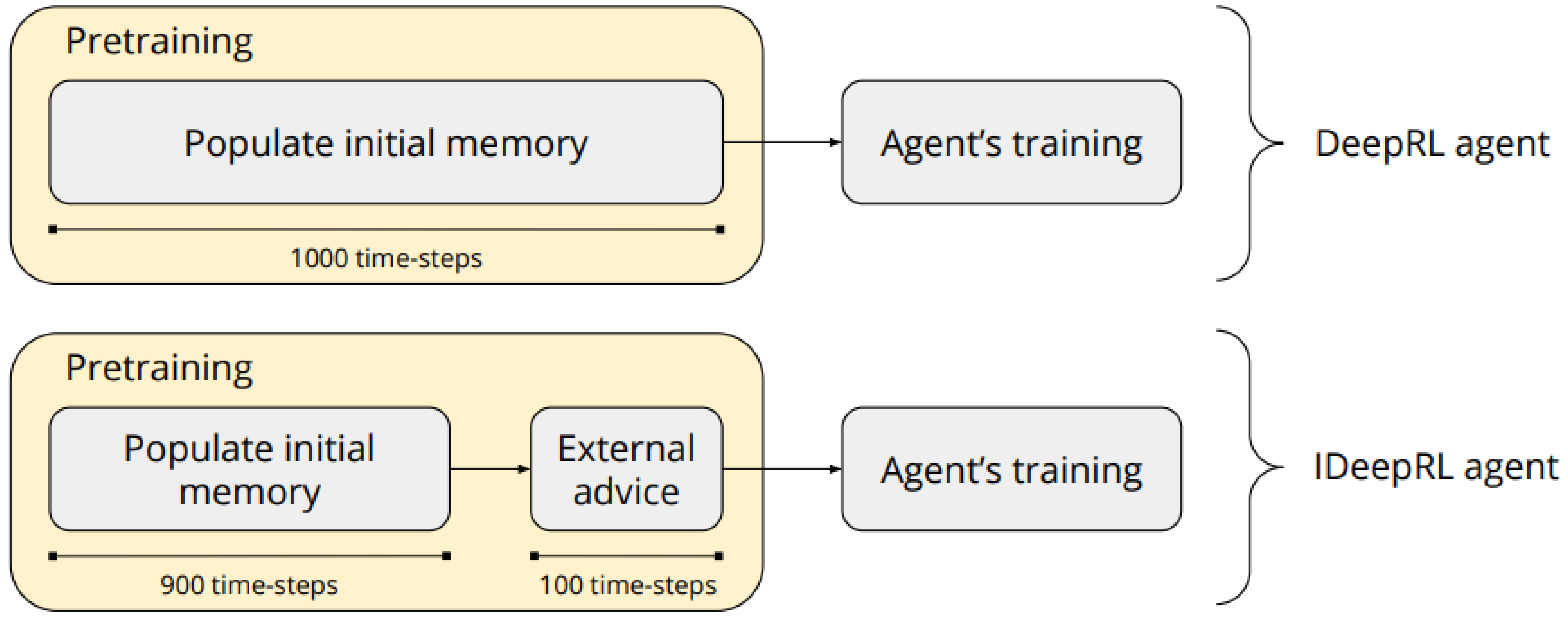
A. Policy shaping

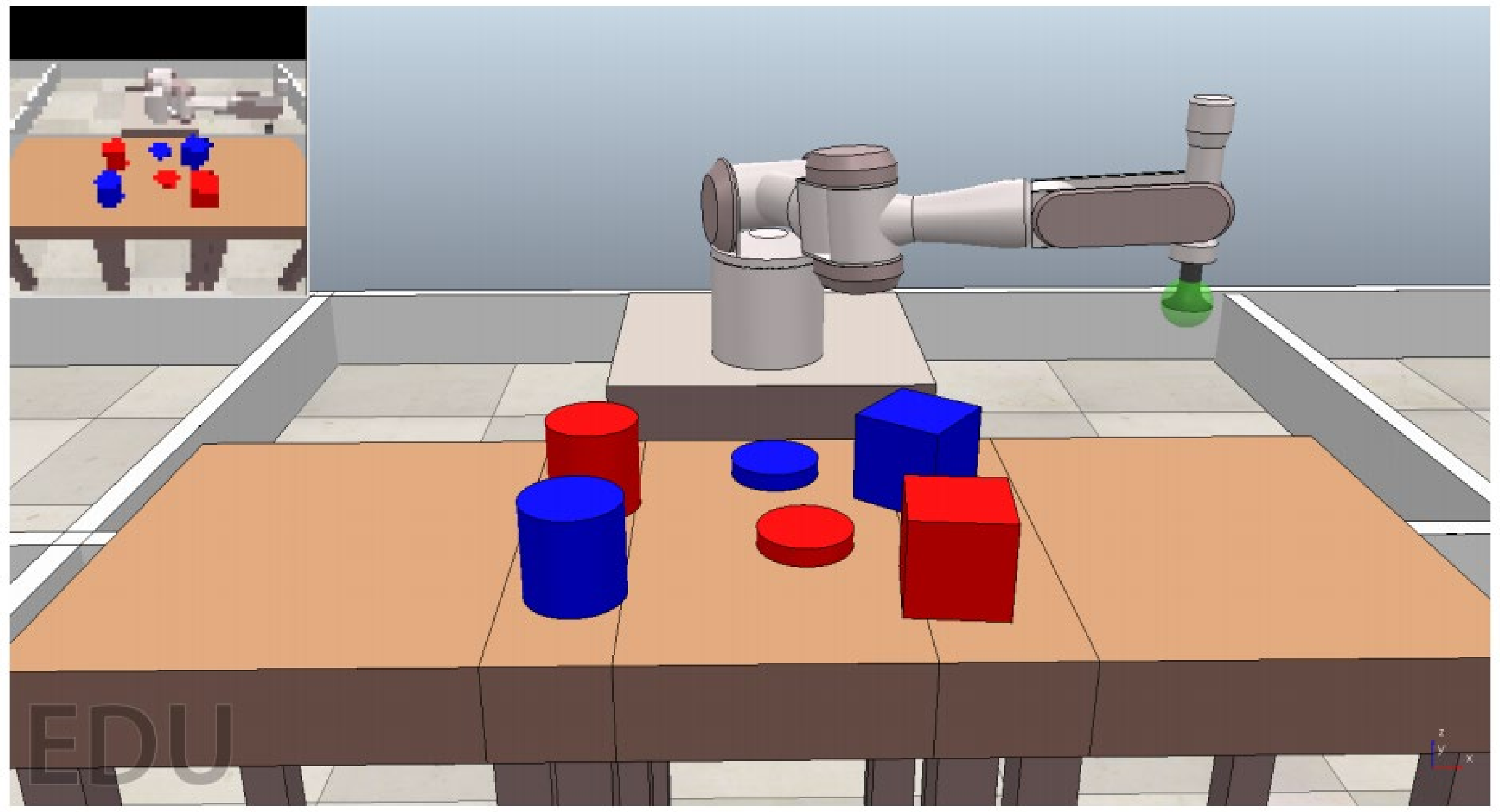


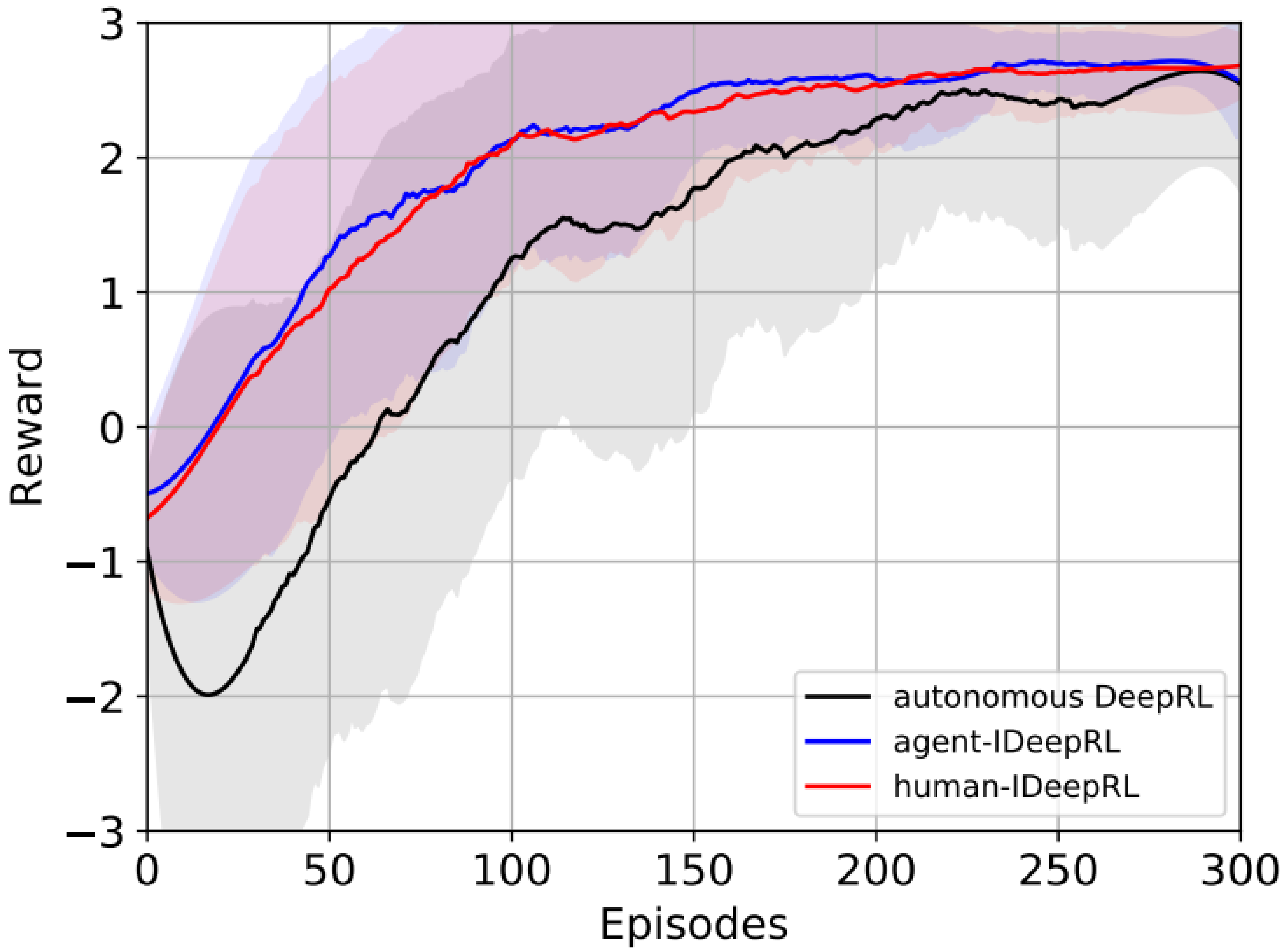
B. Reward shaping

Interactive Reinforcement Learning*

- Convergence speeds up when using advice in a deep RL scenario.

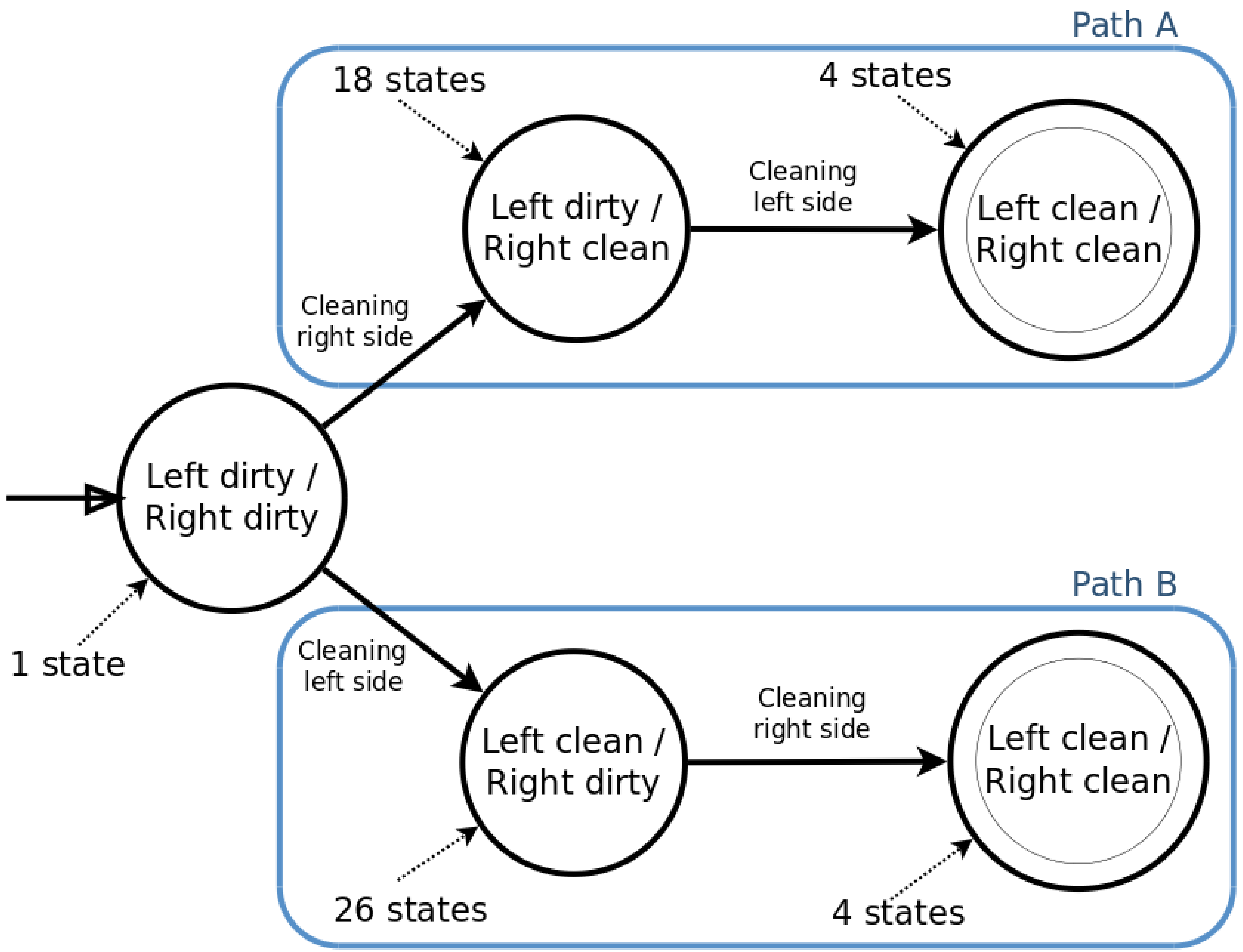






* Moreira, I., Rivas, J., Cruz, F., Dazeley, R., Ayala, A., & Fernandes, B. "Deep Reinforcement Learning with Interactive Feedback in a Human-Robot Environment". *Applied Sciences*, *10*(16), 5574. 2020.

What Makes A Good Teacher*



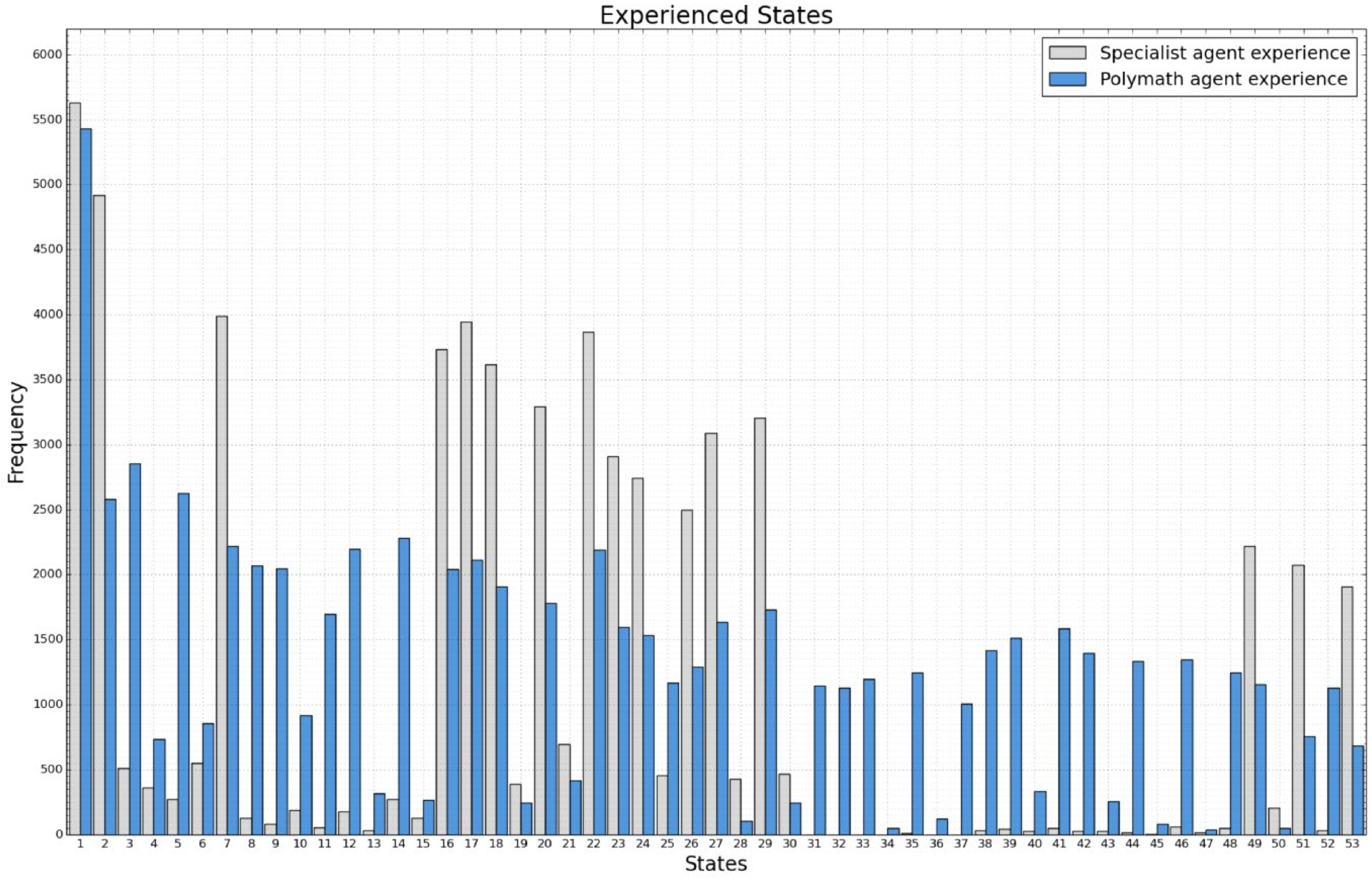
- States transitions distributed on 2 main paths.
- Agents with diverse behaviors.
  - The specialist-A.
  - The specialist-B.
  - The polymath agent.

* Cruz, F., Magg, S., Nagai, Y., and Wermter, S. “Improving interactive reinforcement learning: What makes a good teacher?”. *Connection Science*, Vol. 30, Nr. 3, pp. 306-325, March 2018.

What Makes A Good Teacher

State visit frequency: lower standard deviation in polymath agent.

$$T^* = \underset{i \in Ag}{\operatorname{argmin}} \, \sigma_s^i$$



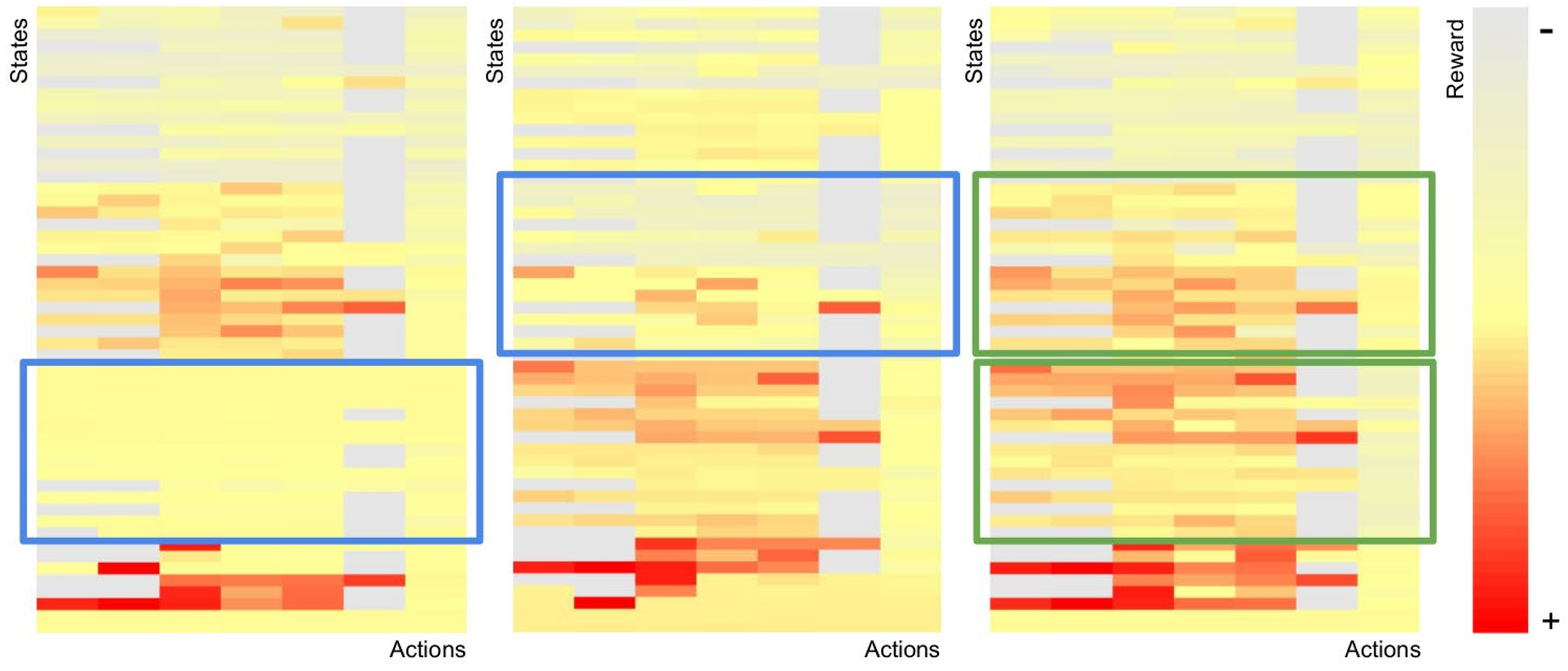
What Makes A Good Teacher

- Choosing and advisor agent.
- Lower stardard deviation in polymath agent.

$$T^* = \underset{i \in Ag}{\text{argmin}}\ \sigma_s^i$$

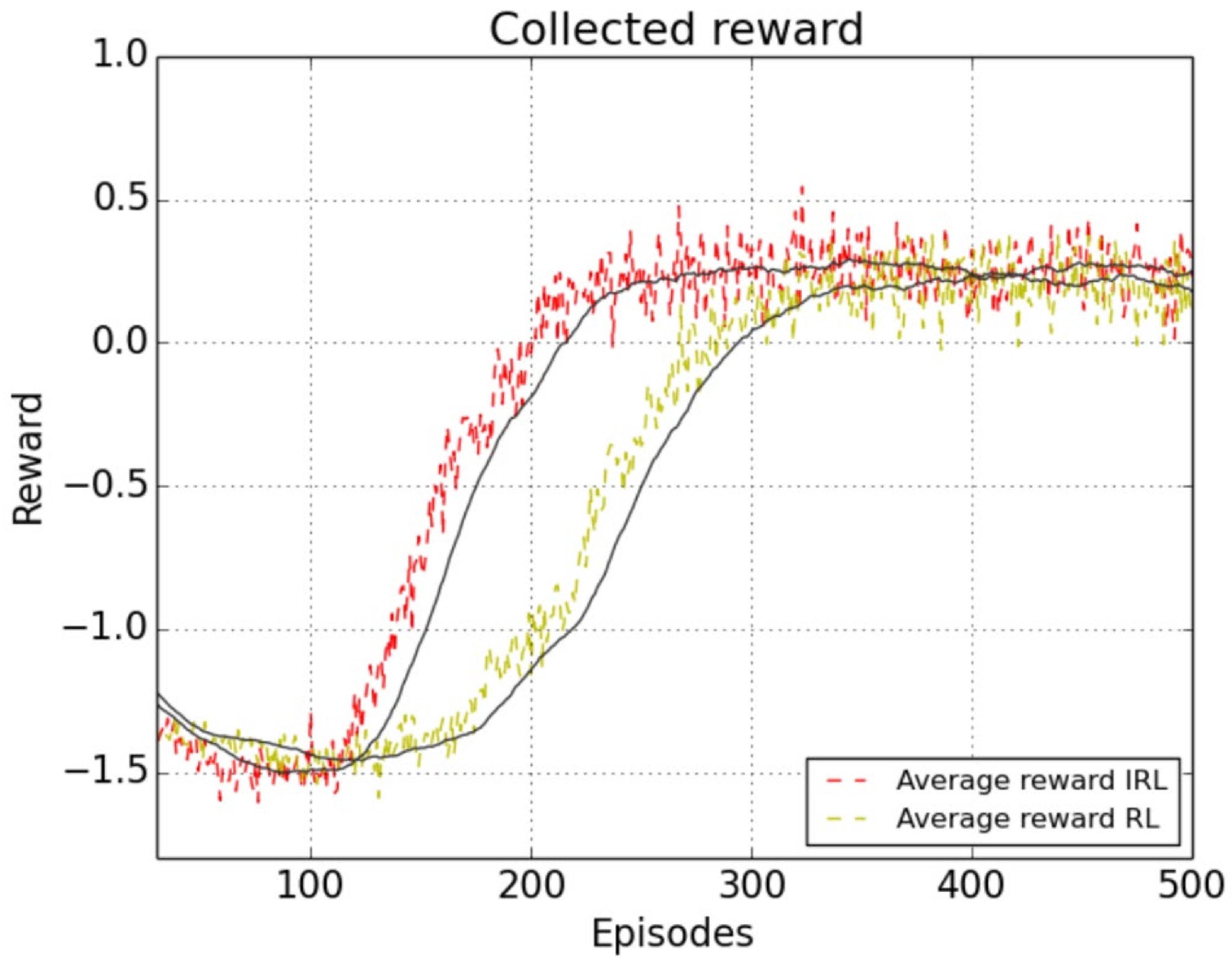
- Different internal representation.

Agent	s	σ_s	r	**R**
Specialist-A agent	1121.21	1570.75	0.11105	333.15
Specialist-B agent	1561.15	1628.70	-0.17839	-535.18
Polymath agent	1307.51	947.96	-0.00427	-12.82

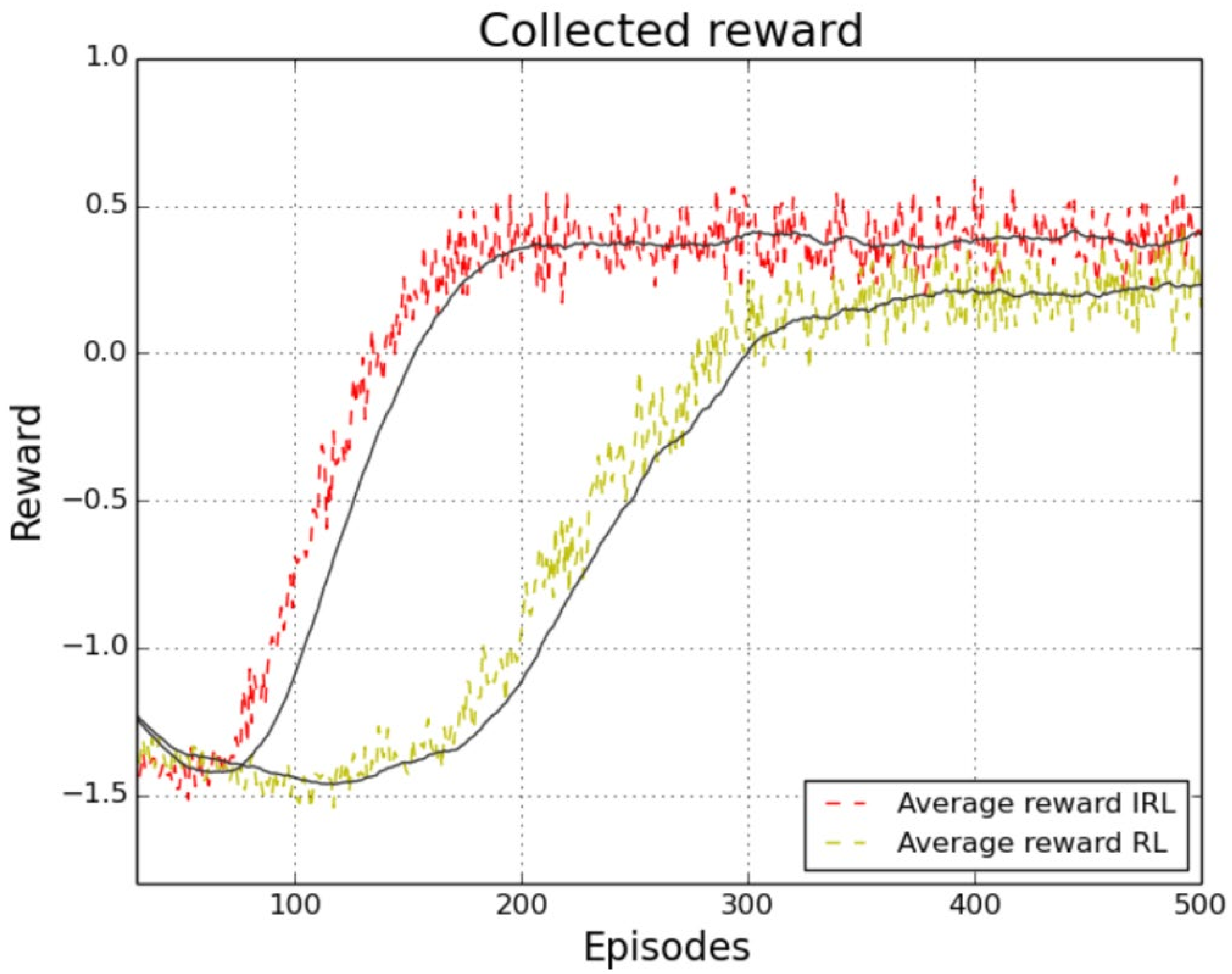


a) The specialist-A agent b) The specialist-B agent c) The polymath agent

What Makes A Good Teacher



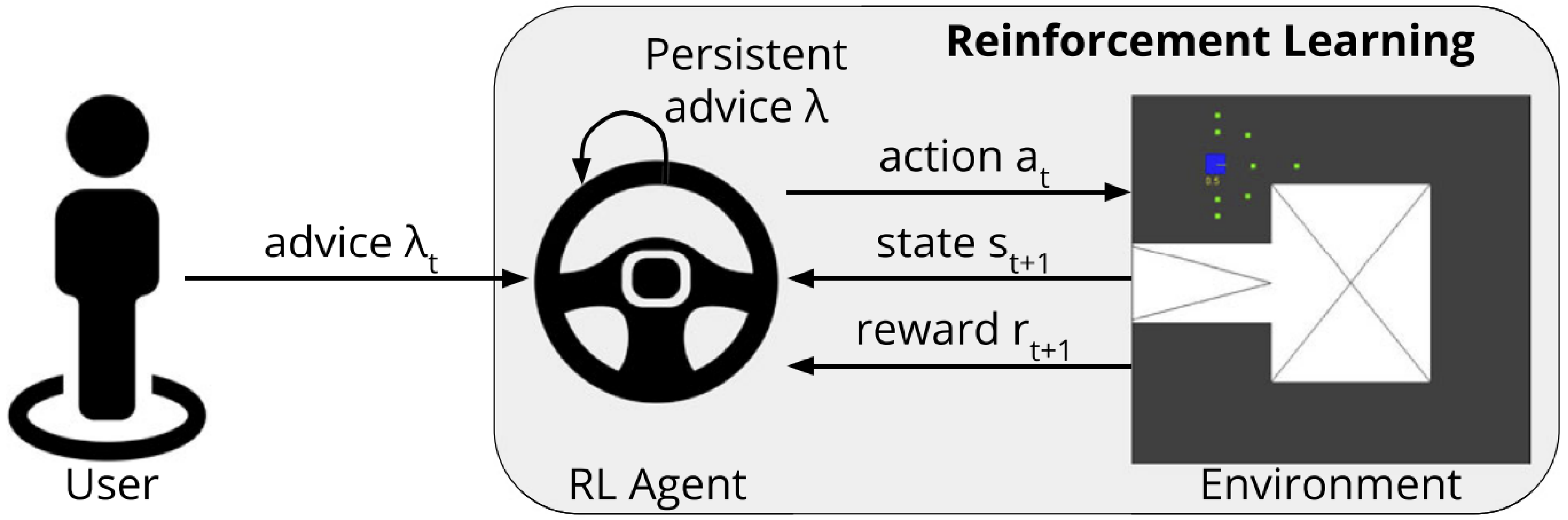
A. Advisor: Specialist-A agent.

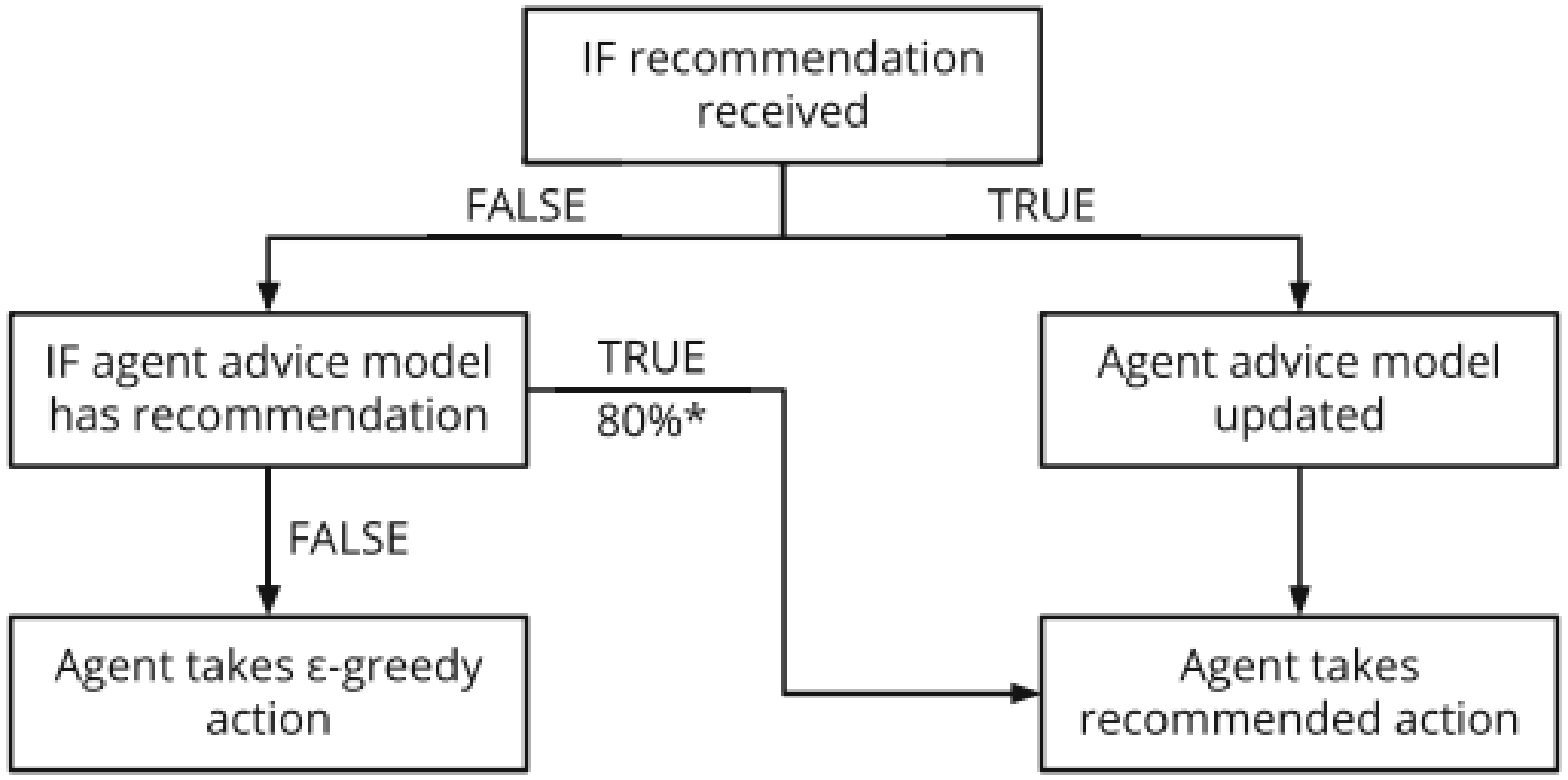


B. Advisor: Polymath agent.

Persistent Rule-based Interactive Reinforcement Learning

- Introduction of persistence to IntRL, a method for information retention and reuse.
  - Probabilistic policy reuse (PPR) is used for discarding or ignoring advice after a period of time.
- A persistent rule-based IntRL approach allows users to provide information in the form of rules, rather than per-state action recommendations.
  - Ripple-down rules (RDR) is an iterative technique for building and maintaining binary decision trees.



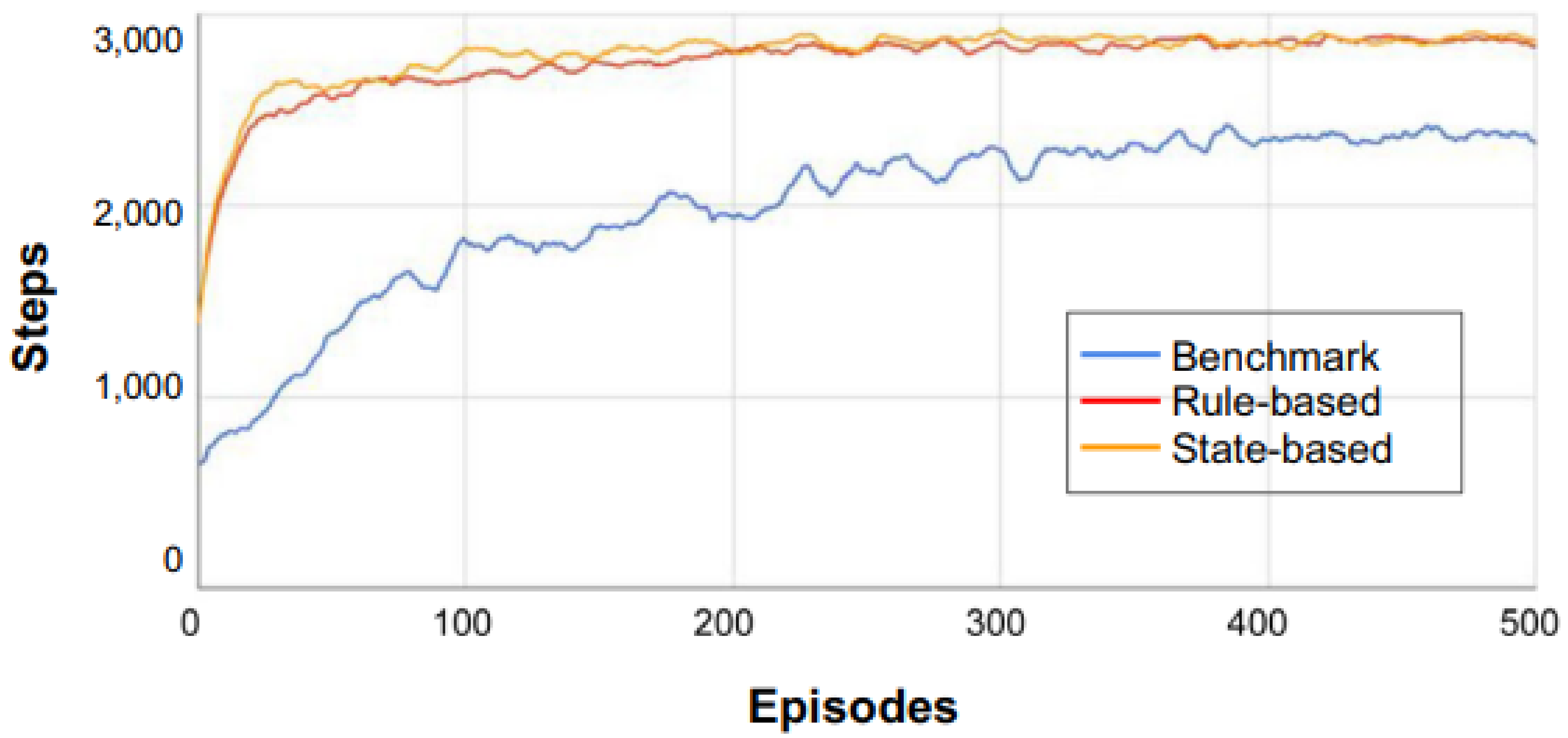


Results – Persistent Rule-based advice

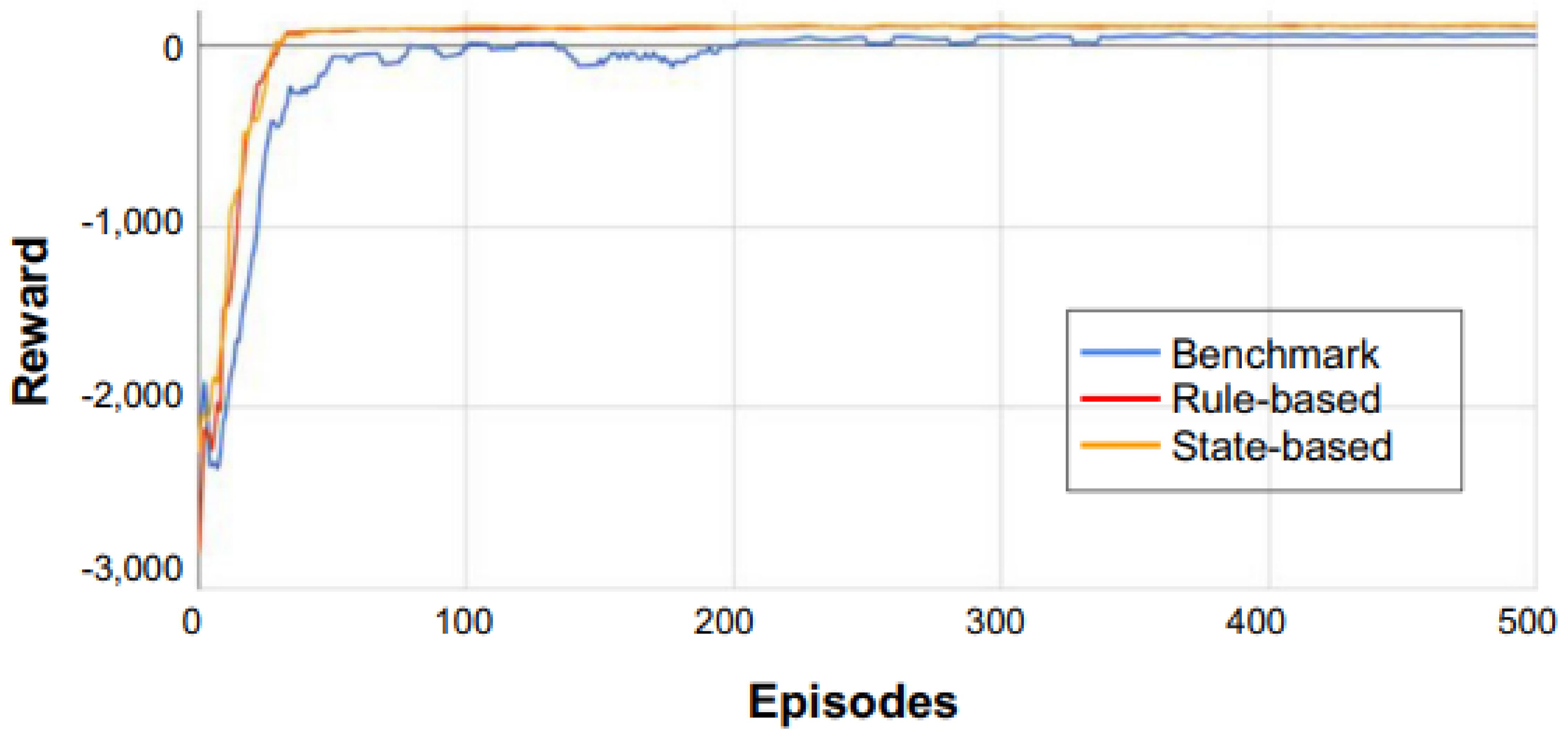
- Simulated self-driving car domain.
- Agents with different level of knowledge and availability.
- Reduced number of interactions.

Table 7 Average number of interactions performed per experiment and the percentage of interactions compared to the steps taken, for each state-based/rule-based agent/user combination in the self-driving car domain

Agent	#Interaction (%)
State-based advice agent	232 (< 0.01%)
Rule-based advice agent	2 (< 0.01%)



(a) Steps per episode.



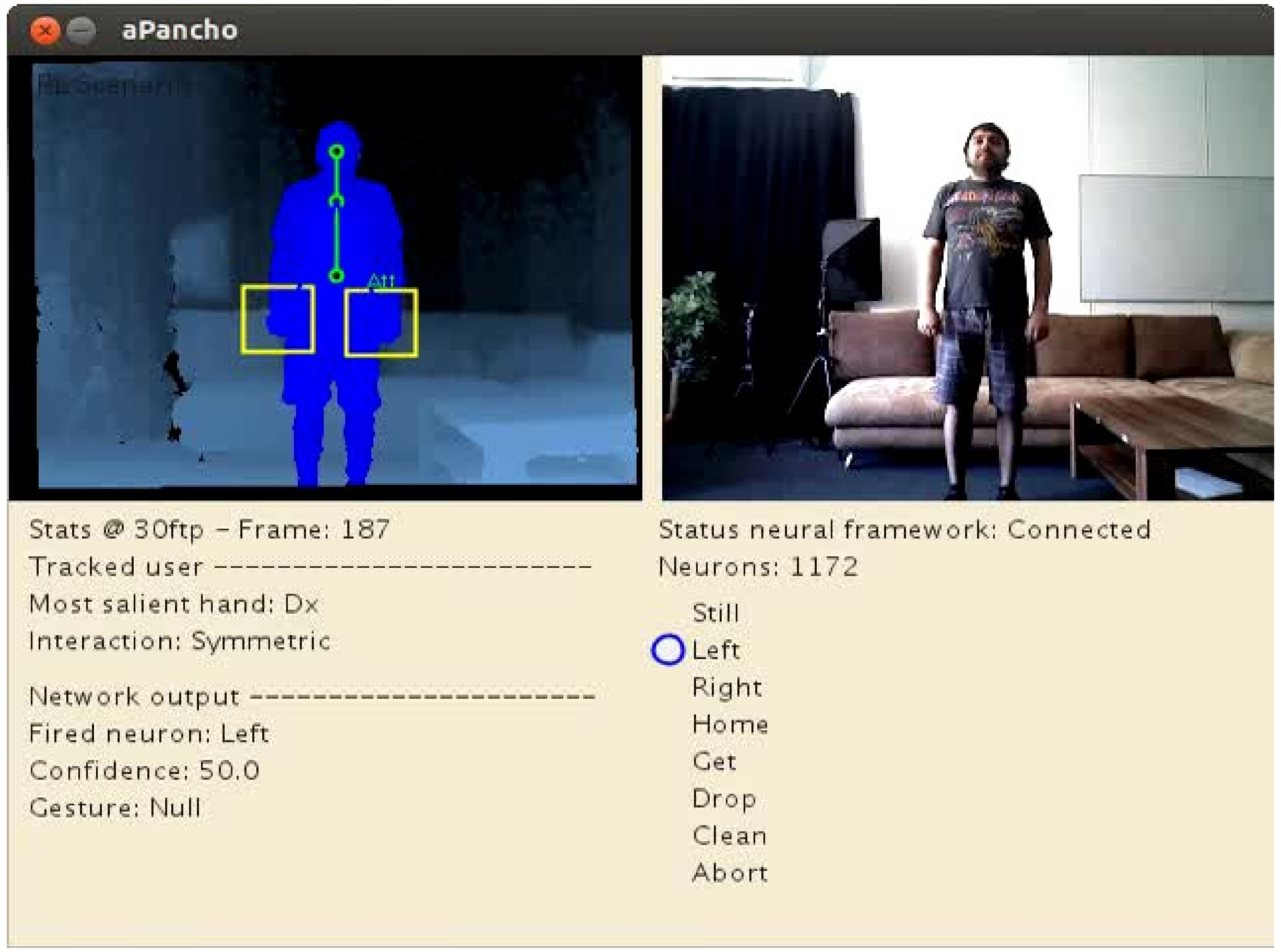
(b) Reward per episode.

Lecture Overview

- Introduction
- Human-in-the-loop learning
- **Robotic sensory modalities – Multimodal integration**
- Contextual affordances
- Explainable robotic systems

Robot Vision

- Using RGB-D sensors [Video link].



Robot Vision

- Skeleton tracking

LLM-based Robot Interaction

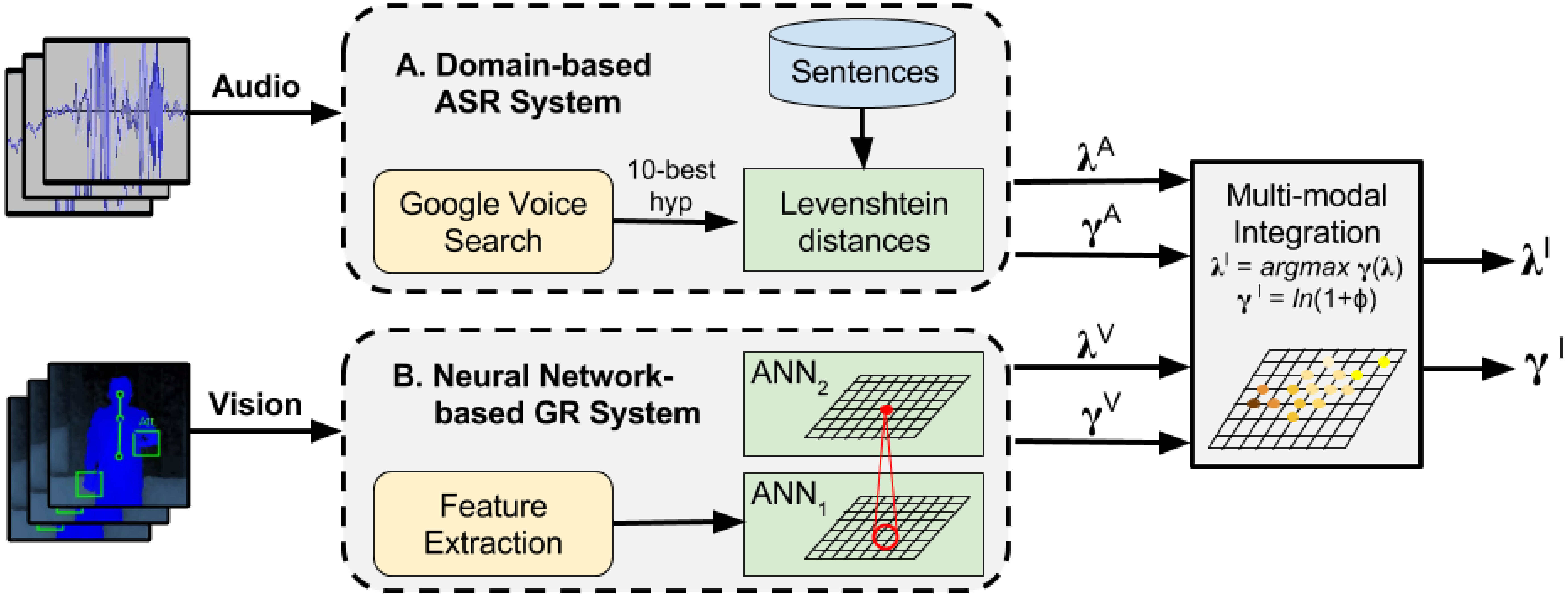
- ChatGPT code generated for drone control [Video link].

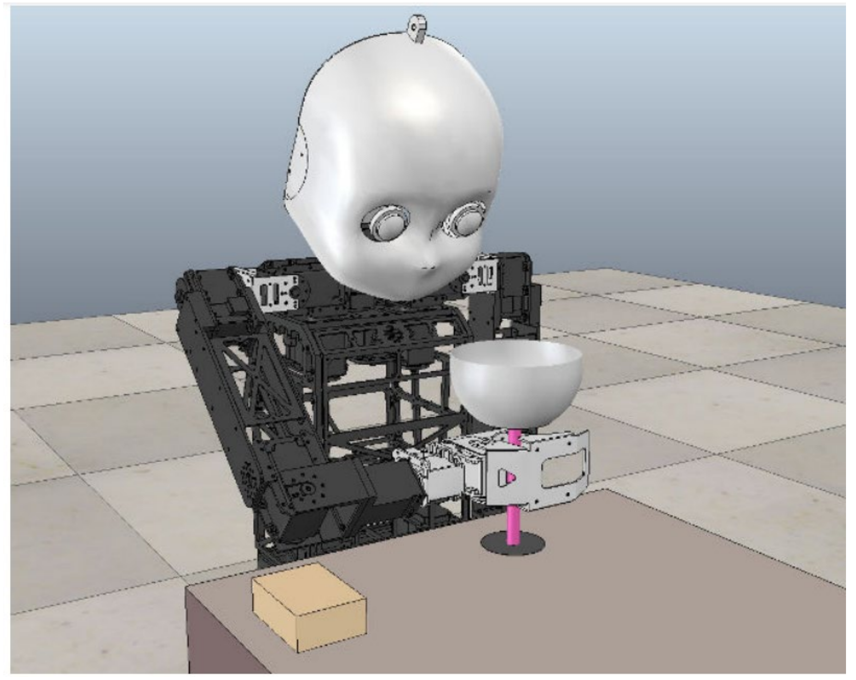
LLM-based Robot Interaction

- LLaMA with ARI social robot companion

Multi-modal Associative Architecture

Uni-modal and multi-modal labels and confidence values.





$$\lambda^A = \operatorname{argmin} \mathcal{L}(h_i, s_j)$$

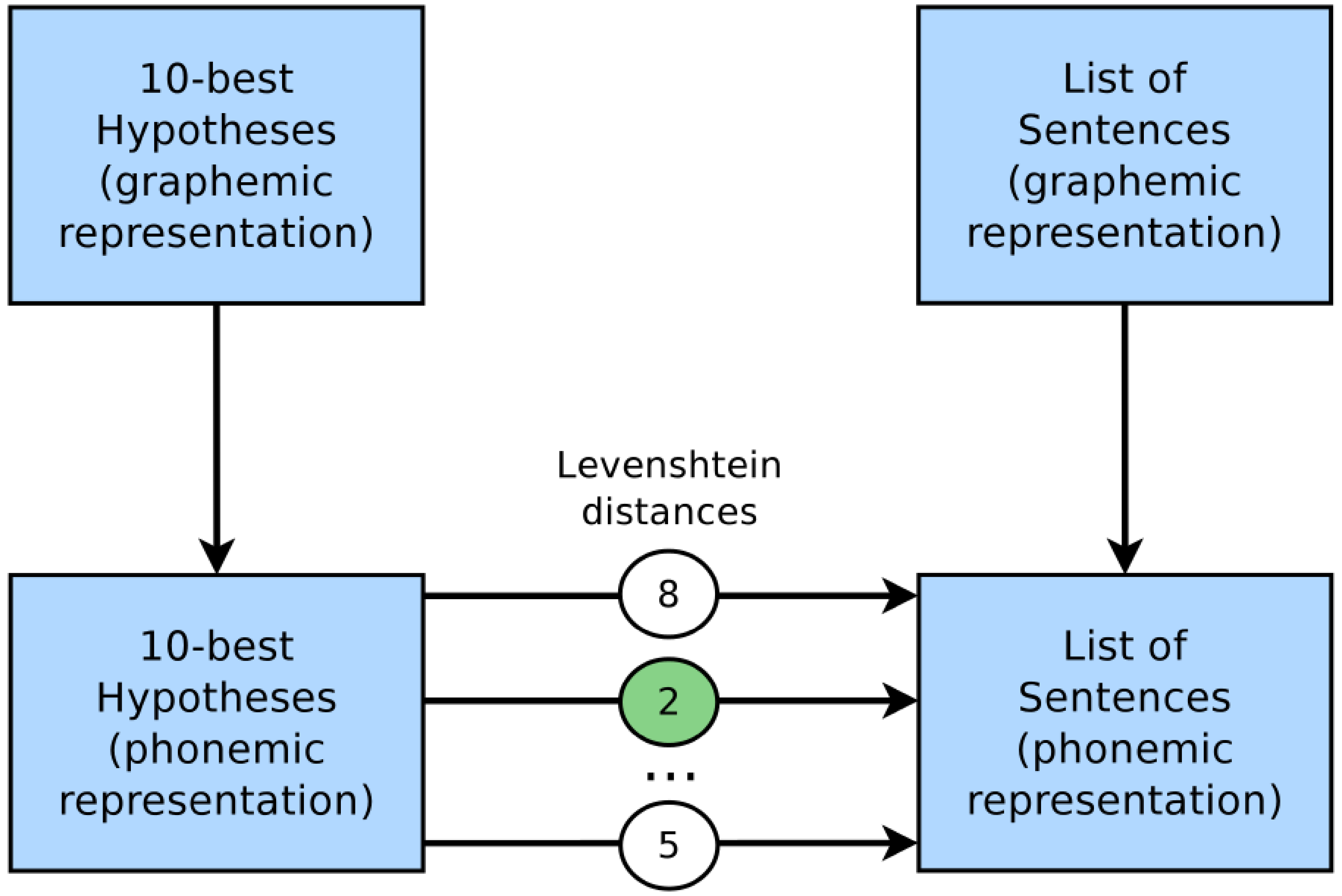
$$\gamma^A = \max(0, 1 - \mathcal{L}(h_i, s_j)/|s_j|)$$

$$\lambda^V = Mo(\Lambda^V)$$

$$\gamma^V = N/|\Lambda^V|$$

Automatic Speech Recognition

- DOCKS, based on Google Voice Search.
- Comparison with Levenshtein distance.
- Domain-specific language model.



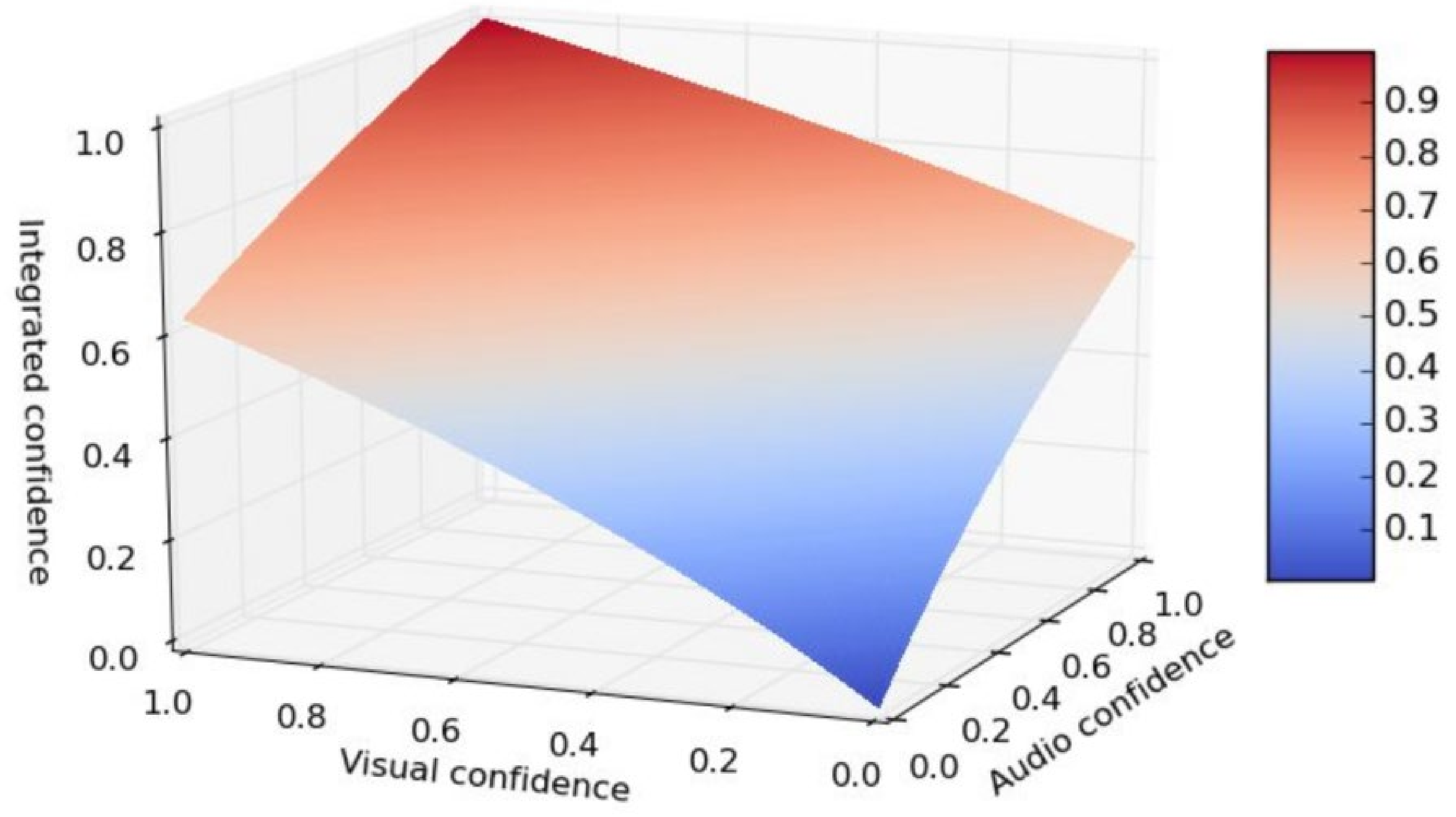
Multi-modal Integration

- Integrated label and confidence value:

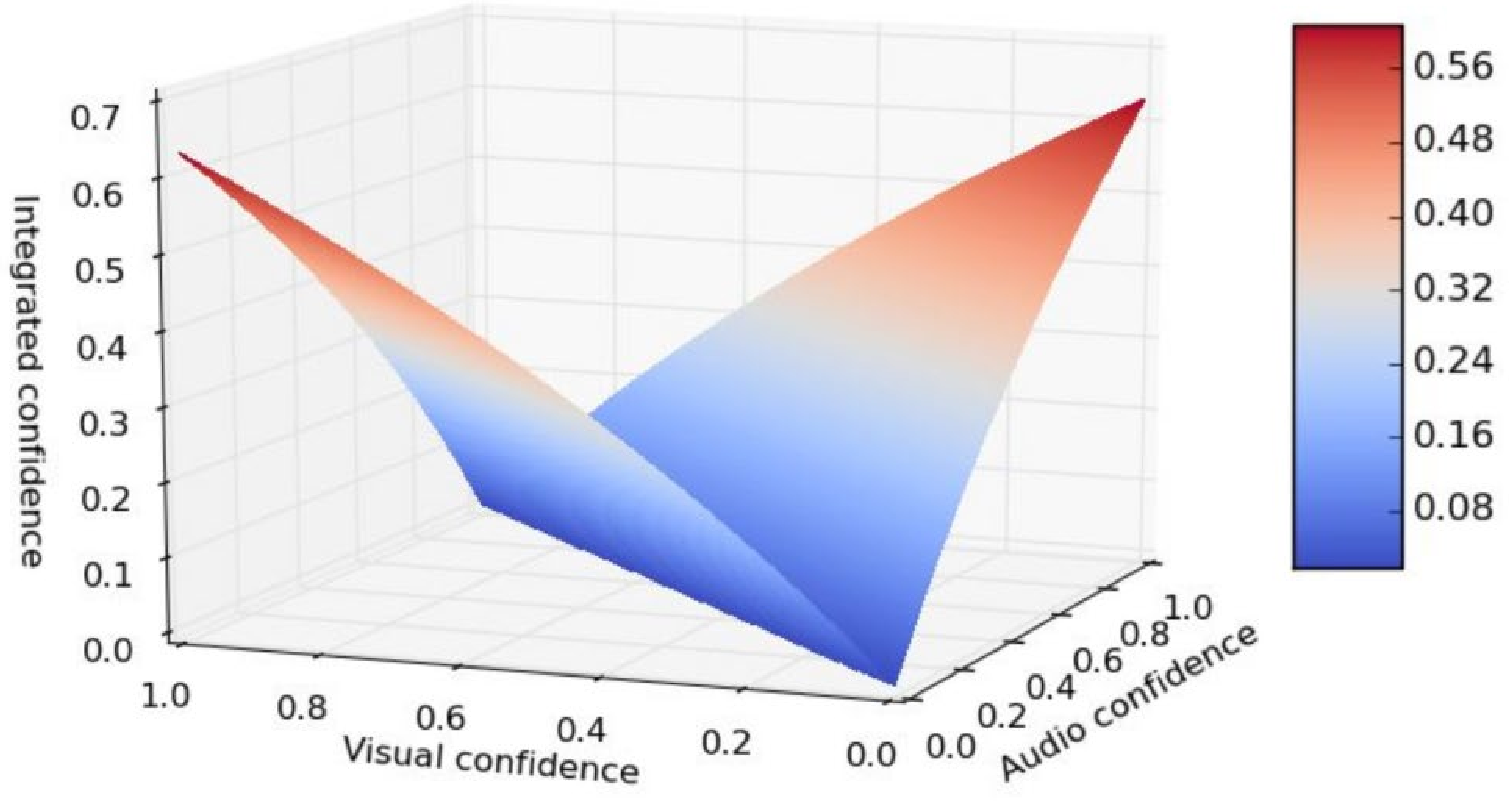
$$\lambda^I = \begin{cases} \lambda^A & \text{if } \gamma^A > \gamma^V \\ \lambda^V & \text{otherwise} \end{cases}$$

$$\phi = \begin{cases} \gamma^A + \gamma^V & \text{if } \lambda^A = \lambda^V \\ |\gamma^A - \gamma^V| & \text{if } \lambda^A \neq \lambda^V \end{cases}$$

$$\gamma^I = \ln(1 + \phi)$$

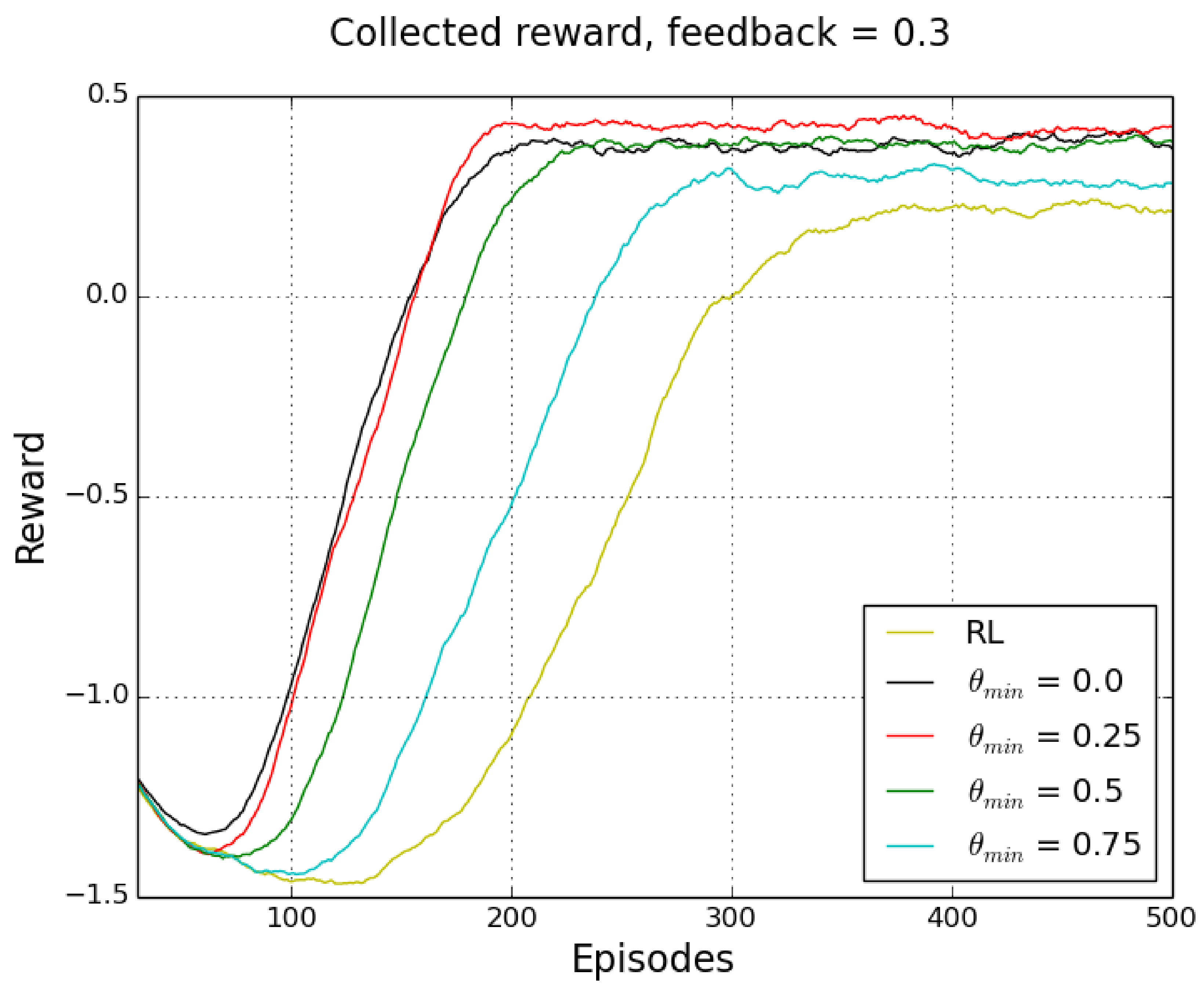


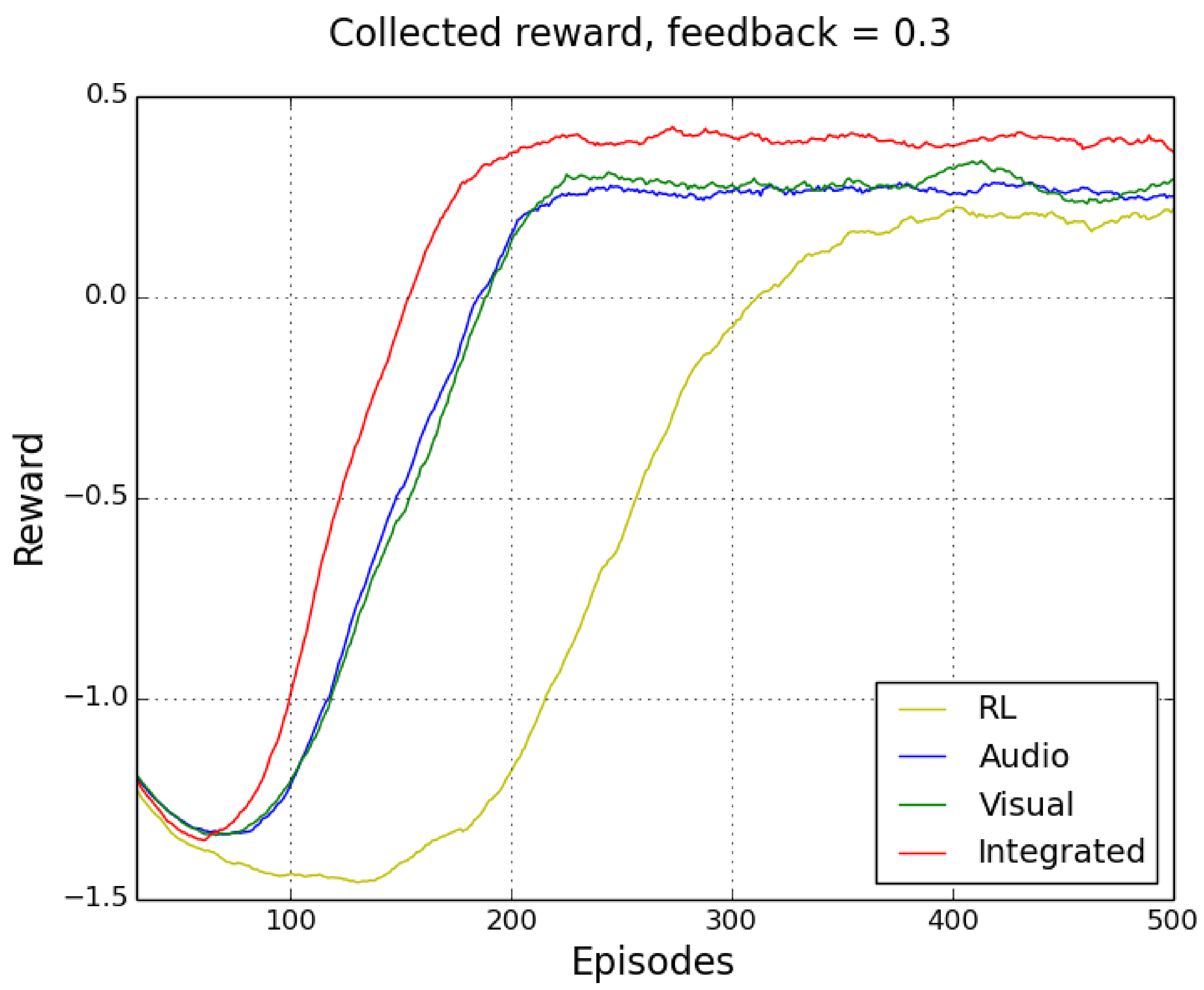
(a) Integrated confidence with equal uni-modal predicted labels



(b) Integrated confidence with different uni-modal predicted labels

IRL Scenario Results





Integrated confidence values over a threshold.

$$\gamma^I > \theta_{min}$$

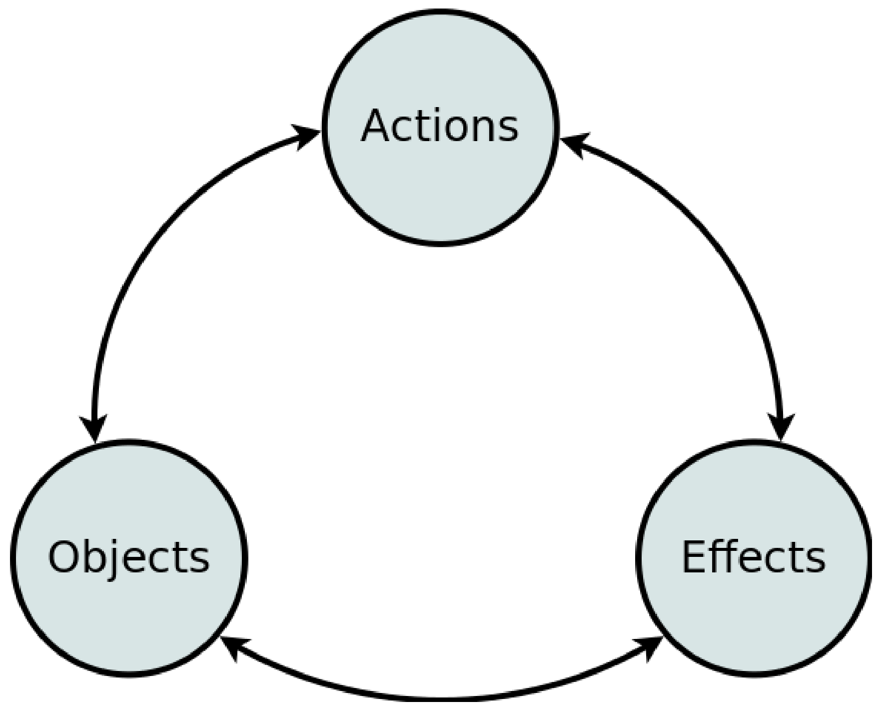
Uni- and multi-modal advice.

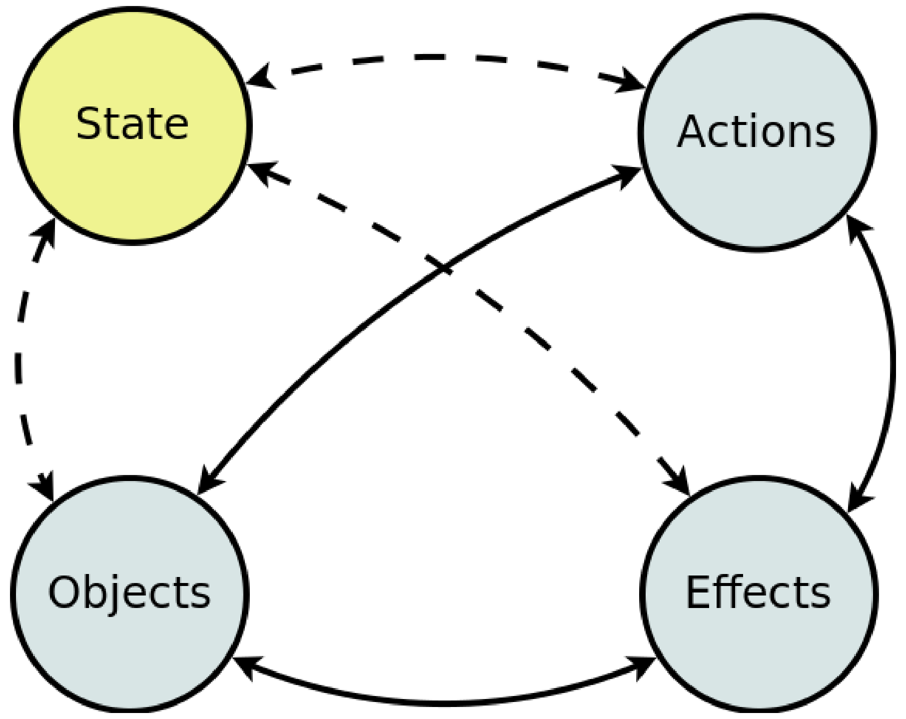
Lecture Overview

- Introduction
- Human-in-the-loop learning
- Robotic sensory modalities – Multimodal integration
- **Contextual affordances**
- Explainable robotic systems

Contextual affordances*

- Affordances represent characteristics of the relation between an agent and an object.
- Used to represent object/action information.
    - *affordance := <object, action, effect>.*
    - *effect = f(object, action).*
- Contextual affordances consider the agent's state.
    - *effect = f(state, object, action).*





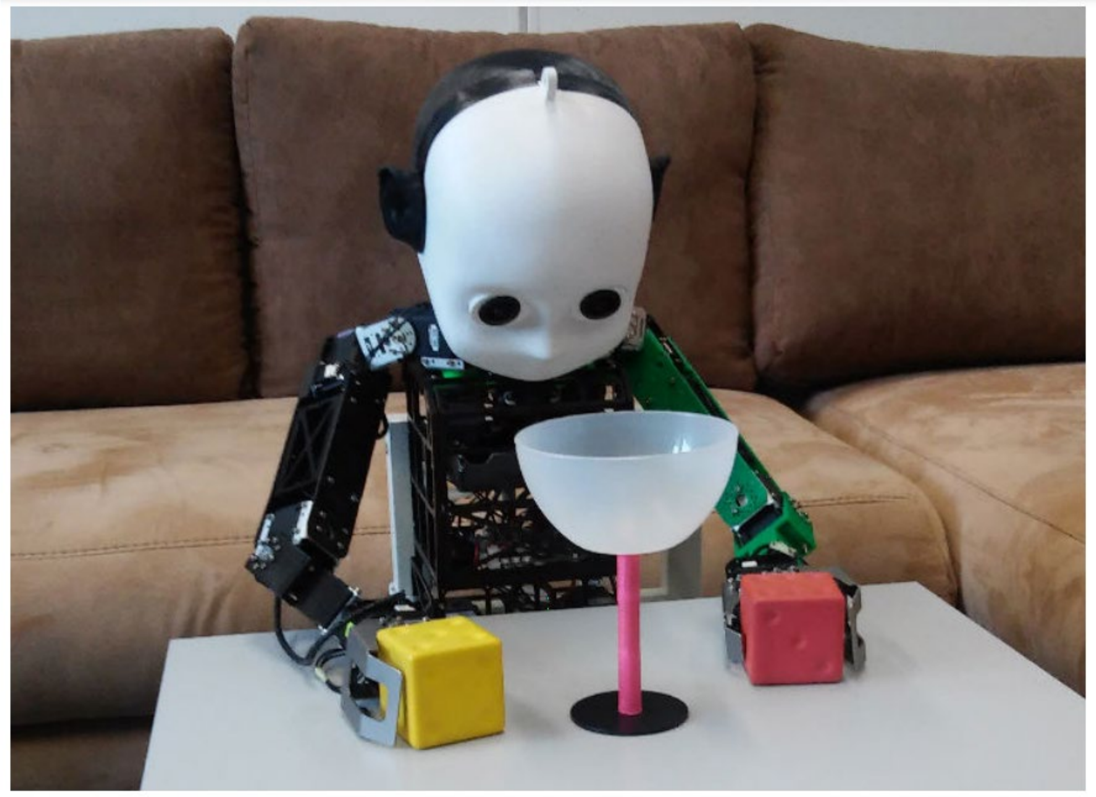
* Cruz, F., Parisi, G. I., and Wermter, S. “Learning contextual affordances with an associative neural architecture”. In *Proceedings of the 24th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN)*, pp. 665-670, Bruges, Belgium, 2016.

Contextual affordances

For instance, given an agent performing the same action a with the same object o, but from a different agent's state $s_1 \neq s_2$: when action a is performed, different effects $e_1 \neq e_2$ could be generated, since the initial states s_1 and s_2 are different. It is unfeasible to establish differences in the final effect when we utilize affordances to represent it, because $e_1 = (a, o)$ and $e_2 = (a, o)$. Hence, to deal with the current states $s_1 \neq s_2$, an agent must distinguish each case and learn them at the same time utilizing contextual affordances defined by $e_1 = (s_1, a, o)$ and $e_2 = (s_2, a, o)$, establishing clear differences between the final effects.

Safety Autonomous Learning*

- Contextual affordances with a neural architecture.



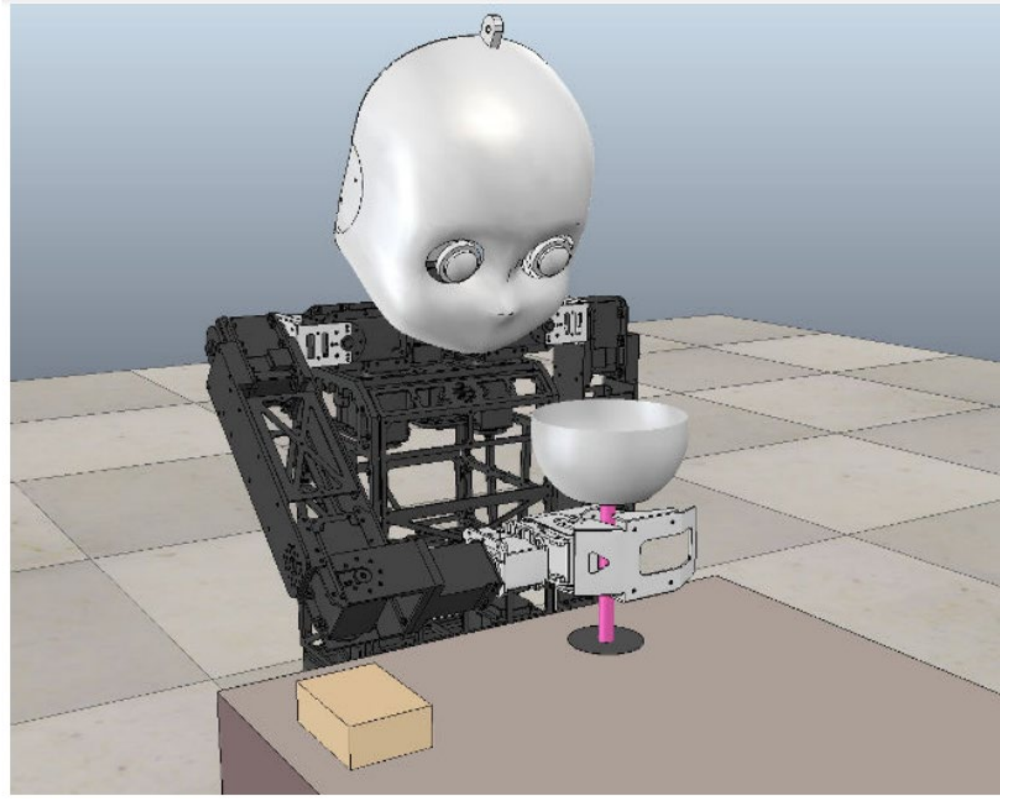
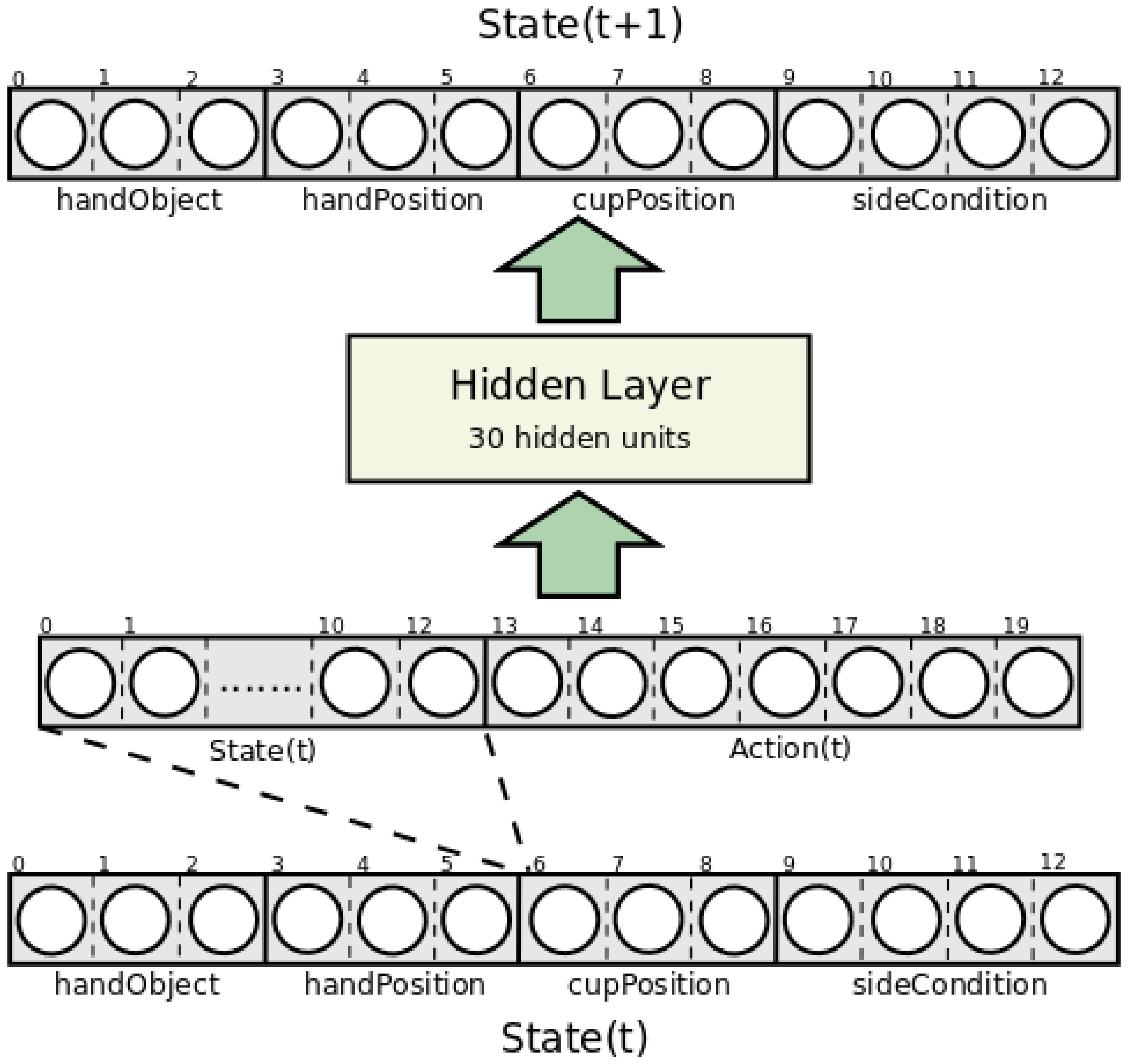


Table 4.1: Representation of training data used for neural network classification.

Data Representation					
Objects		**Locations**		**Side conditions**	
free	[1 0 0]	home	[1 0 0]	dd	[1 0 0 0]
sponge	[0 1 0]	left	[0 1 0]	dc	[0 1 0 0]
goblet	[0 0 1]	right	[0 0 1]	cd	[0 0 1 0]
				cc	[0 0 0 1]

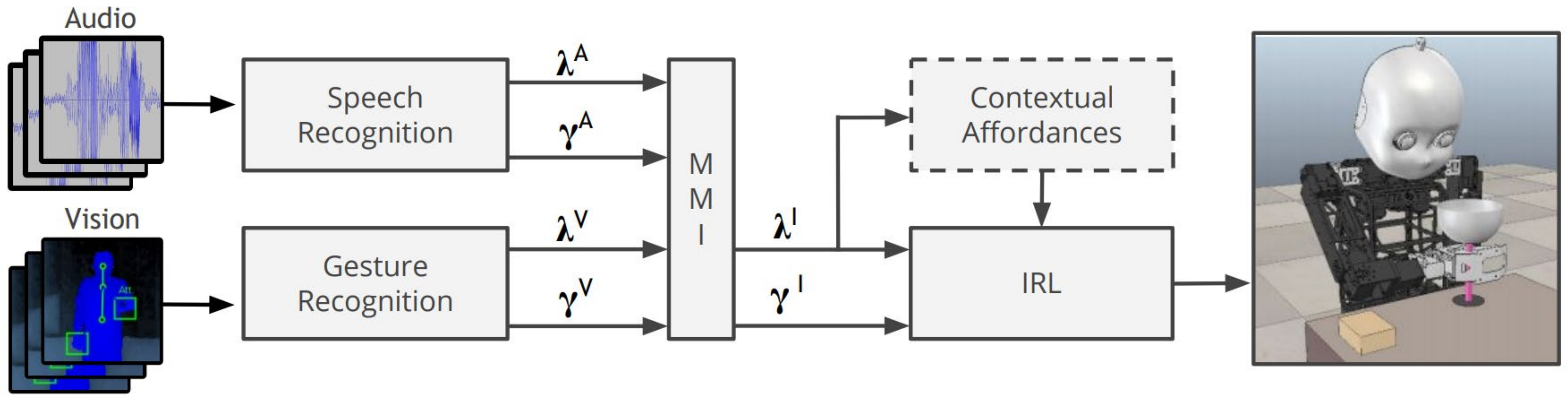
Actions			
grasp	[1 0 0 0 0 0 0]	go right	[0 0 0 0 1 0 0]
place	[0 1 0 0 0 0 0]	clean	[0 0 0 0 0 1 0]
go home	[0 0 1 0 0 0 0]	abort	[0 0 0 0 0 0 1]
go left	[0 0 0 1 0 0 0]		

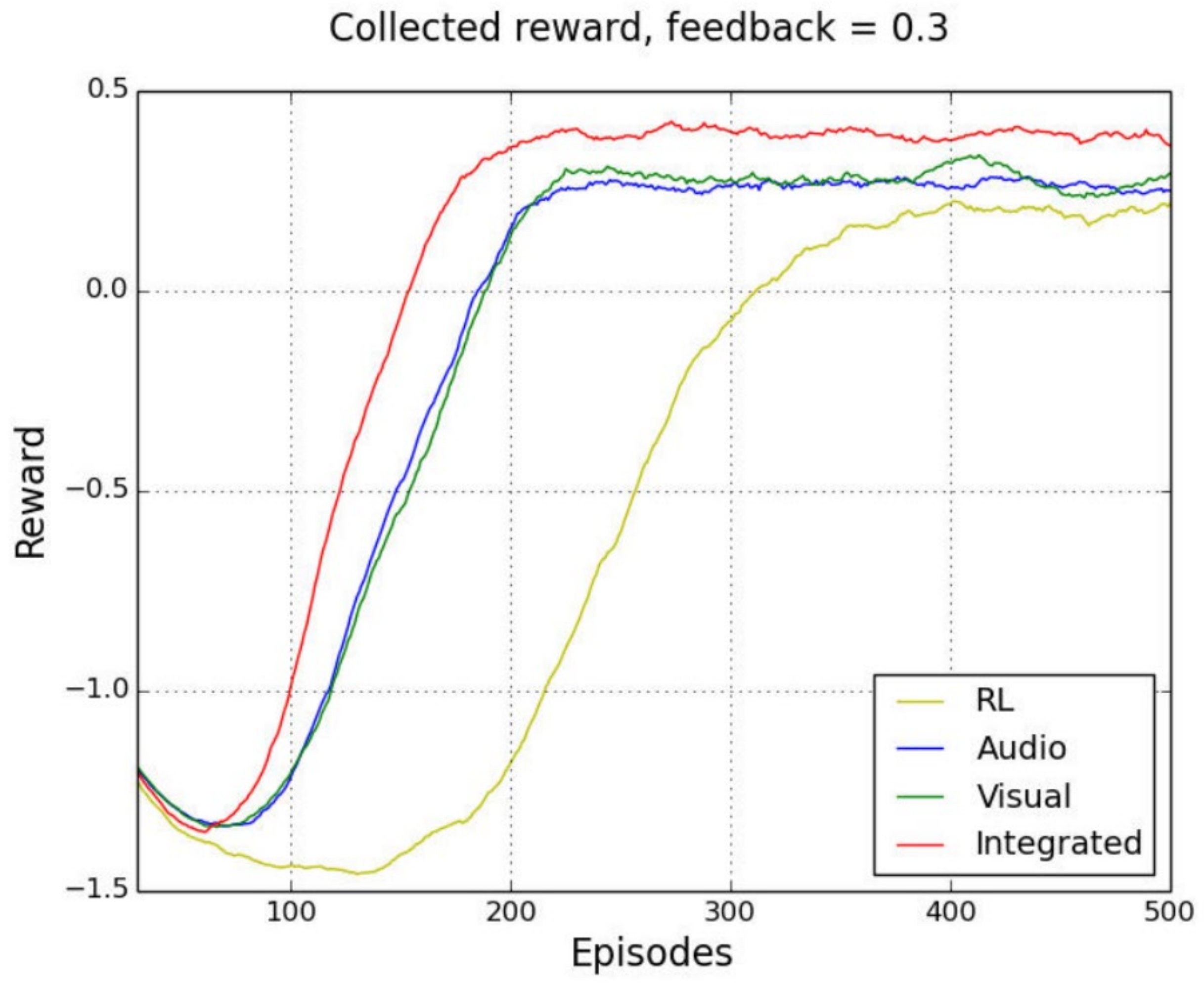


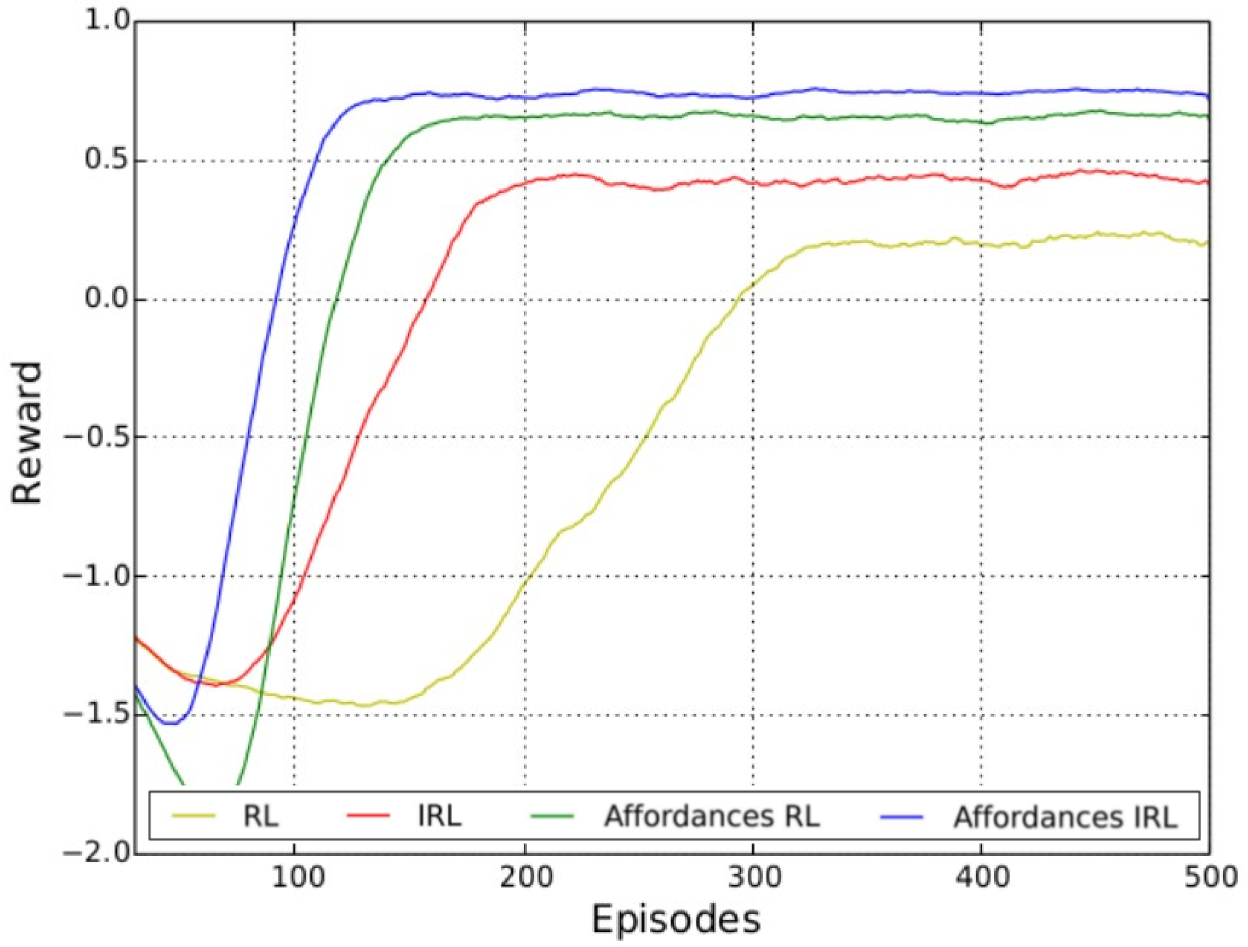
* Cruz, F., Magg, S., Weber, C., & Wermter, S. (2016). Training agents with interactive reinforcement learning and contextual affordances. *IEEE Transactions on Cognitive and Developmental Systems*, *8*(4), 271-284.

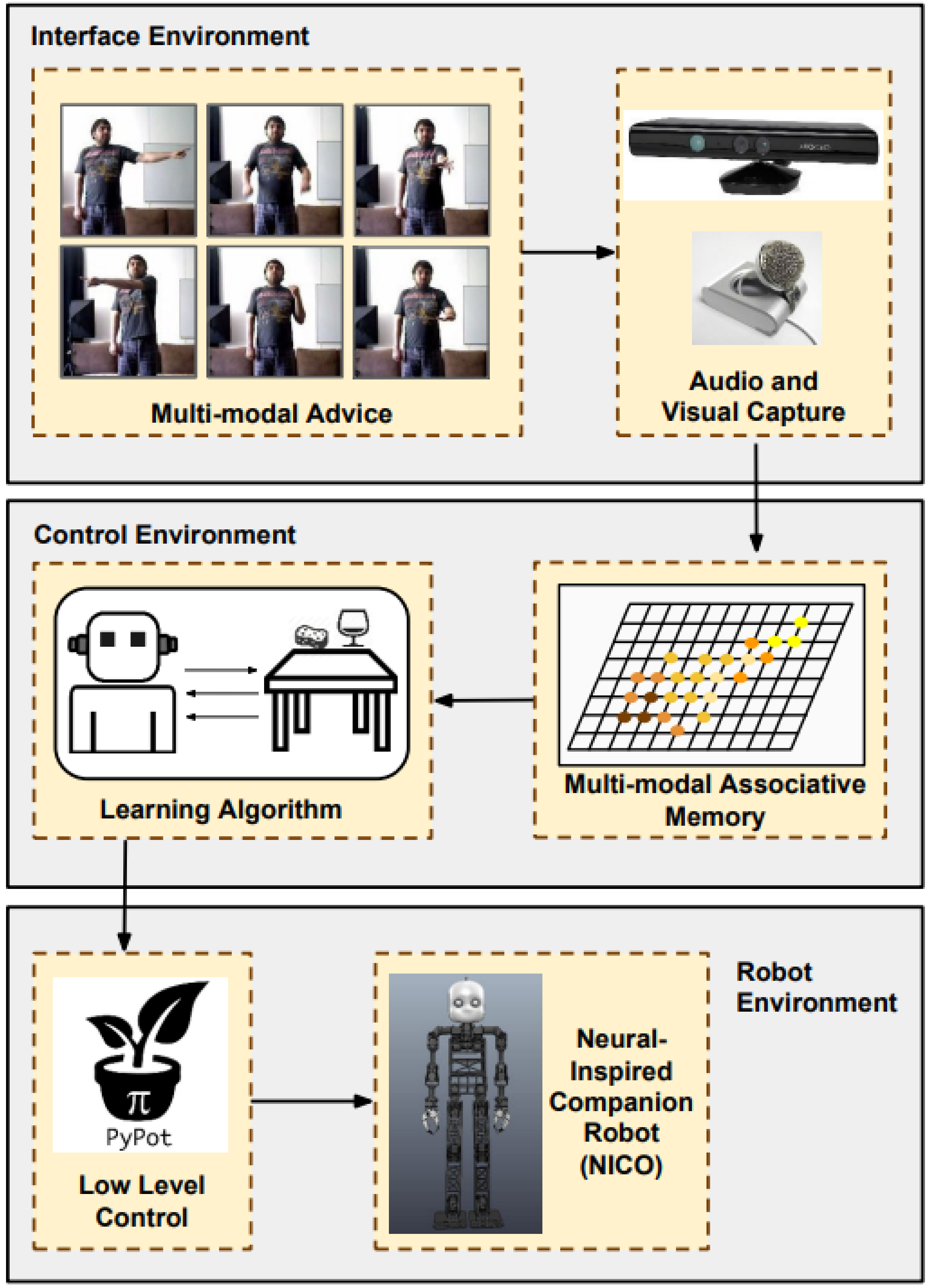
Safety Autonomous Learning*

- Multi-modal integration for interactive feedback.









* Cruz, F., Parisi, G., Twiefel, J., Wermter, W. "Multi-modal Integration of Dynamic Audiovisual Patterns for an Interactive Reinforcement Learning Scenario". In *Proceedings of IEEE/RSJ IROS*, Daejeon, Korea, 2016.

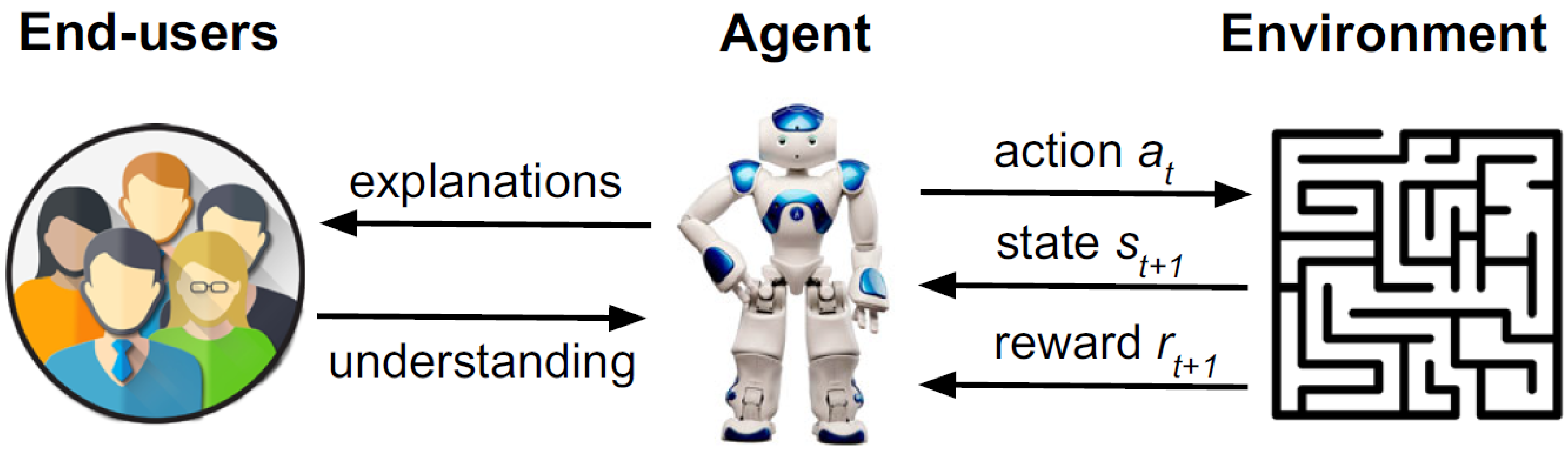
* Cruz, F., Parisi, G., Wermter, S. "Multi-modal Feedback for Affordance-driven Interactive Reinforcement Learning". In *Proceedings of IEEE International Joint Conference on Neural Networks IJCNN*, Rio, Brazil, 2018.

Lecture Overview

- Introduction
- Human-in-the-loop learning
- Robotic sensory modalities – Multimodal integration
- Contextual affordances
- **Explainable robotic systems**

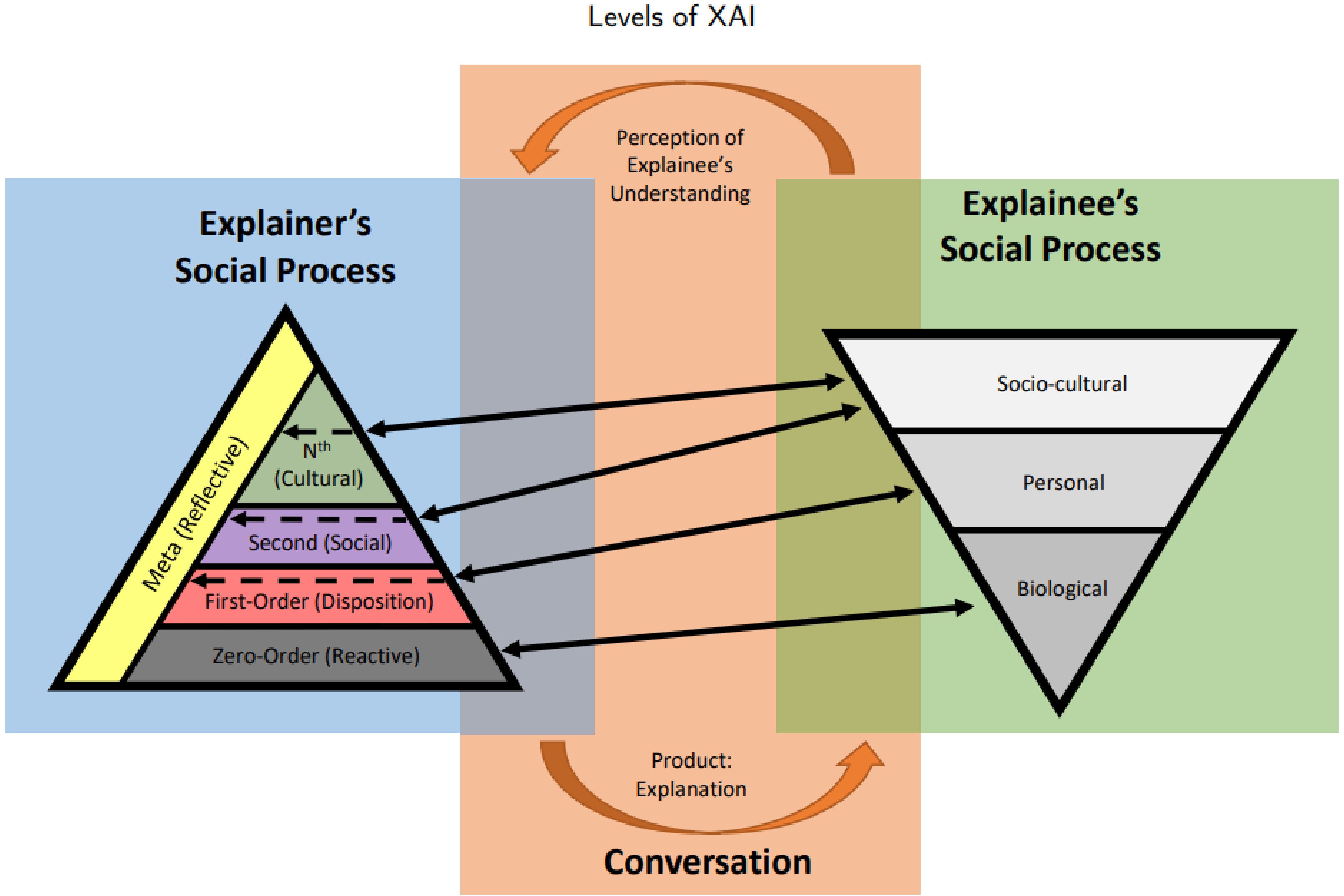
Explainable Robotic Systems

- In human–robot environments, it is crucial that end-users may correctly understand their robotic team-partners.
- A robot can provide featured-based or goal-driven explanations.
- **Not acceptable.** I choose action *left* because it maximizes future collected reward OR I choose action *right* because it is the next one following the optimal policy.
- Using the probability of success is possible to create human-like explanations.



Explainable Artificial Intelligence*

- AI explanations aligned to human communication.



* Dazeley, R., Vamplew, P., Foale, C., Young, C., Aryal, S., & Cruz, F. "Levels of Explainable Artificial Intelligence for Human-aligned Conversational Explanations". *Artificial Intelligence*, 299, 103525. 2021.

Memory-based Method*

- From a non-expert end-user perspective, most relevant questions: 'why?' and 'why not?'. For instance
  - Why did you step forward in the last movement?
  - Why did you not turn to the right in this situation?

- We propose MXRL to compute P_s and N_t using an episodic memory.

- We implement a list of state-action pairs (TList).

[2] Cruz, F., Dazeley, R., Vamplew, P. "Memory-based explainable reinforcement learning". In *Proceedings of the 32nd Australasian Joint Conference on Artificial Intelligence (AI2019)*, pp. 66-67, Adelaide, Australia, 2019.

Memory-based Method

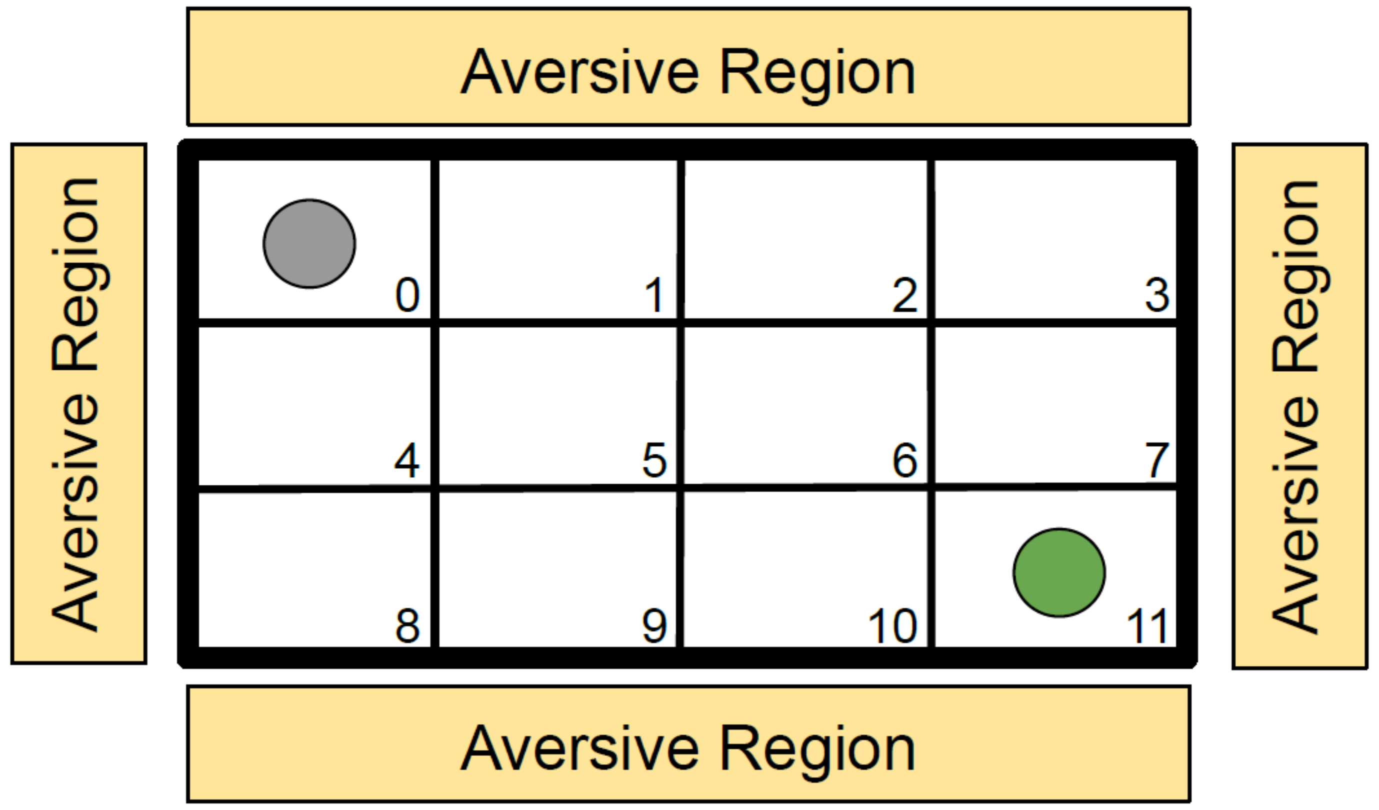
- MXRL algorithm.

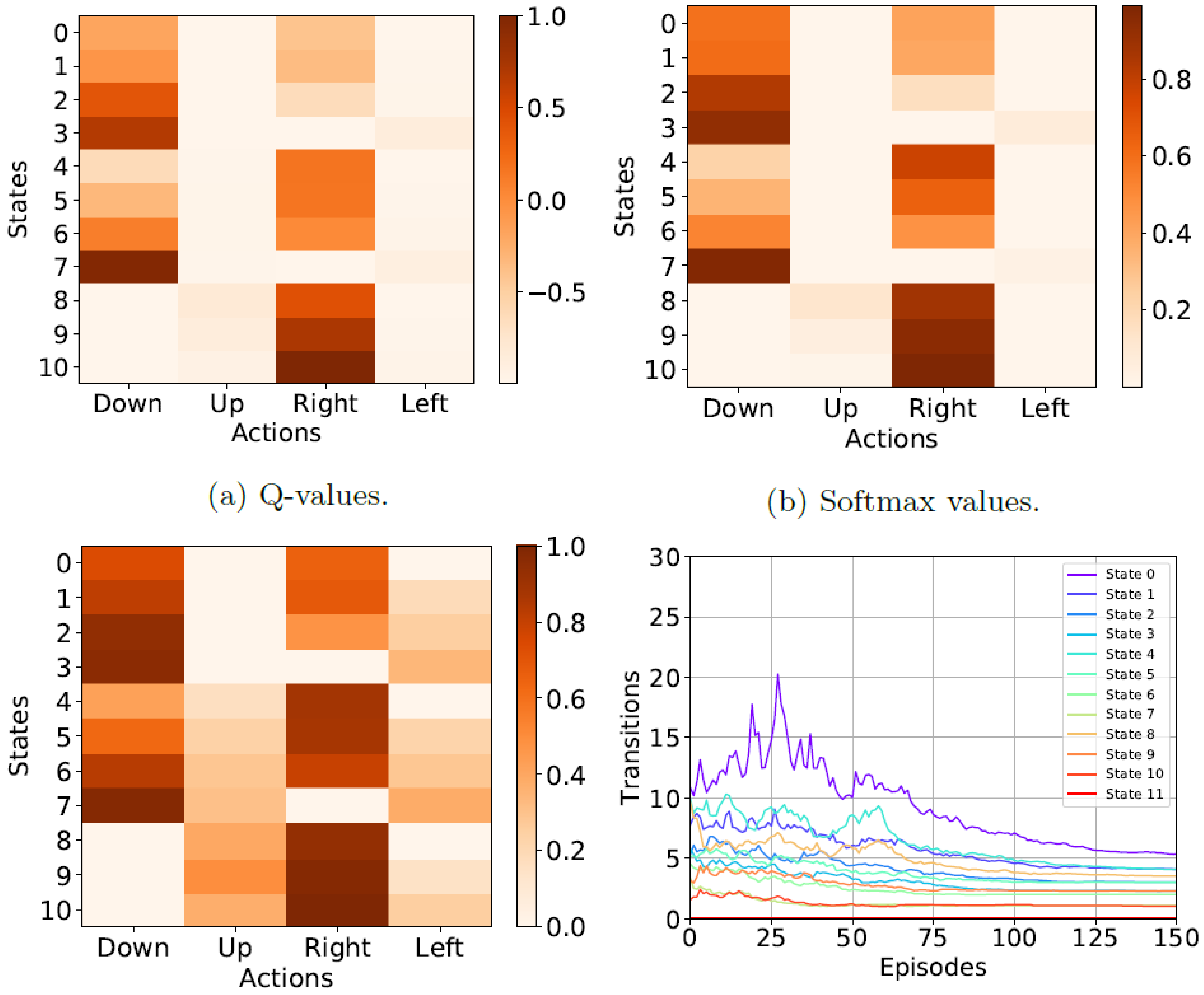
Algorithm 1 Memory-based explainable reinforcement learning approach with the on-policy method SARSA to compute the probability of success and the number of transitions to the goal state.

Initialize $Q(s, a)$, T_t, T_s, P_s, N_t
for each episode **do**
    Initialize $T_{List}[]$
    Choose an action using $a_t \leftarrow \text{SELECTACTION}(s_t)$
    repeat
        Take action a_t
        Save state-action transition T_{List}.add(s, a)
        $T_t[s][a] \leftarrow T_t[s][a] + 1$
        Observe reward r_{t+1} and next state s_{t+1}
        Choose next action a_{t+1} using softmax action selection method
        $Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha[r_{t+1} + \gamma Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t)]$
        $s_t \leftarrow s_{t+1}$; $a_t \leftarrow a_{t+1}$
    until s is terminal (goal or aversive state)
    if s is goal state **then**
        for each s,a $\in T_{List}$ **do**
            $T_s[s][a] \leftarrow T_s[s][a] + 1$
        end for
    end if
    Compute $P_s \leftarrow T_s/T_t$
    Compute N_t for each s $\in T_{List}$ as pos(s, T_{List}) + 1
end for

Memory-based Method

- Experimental results.



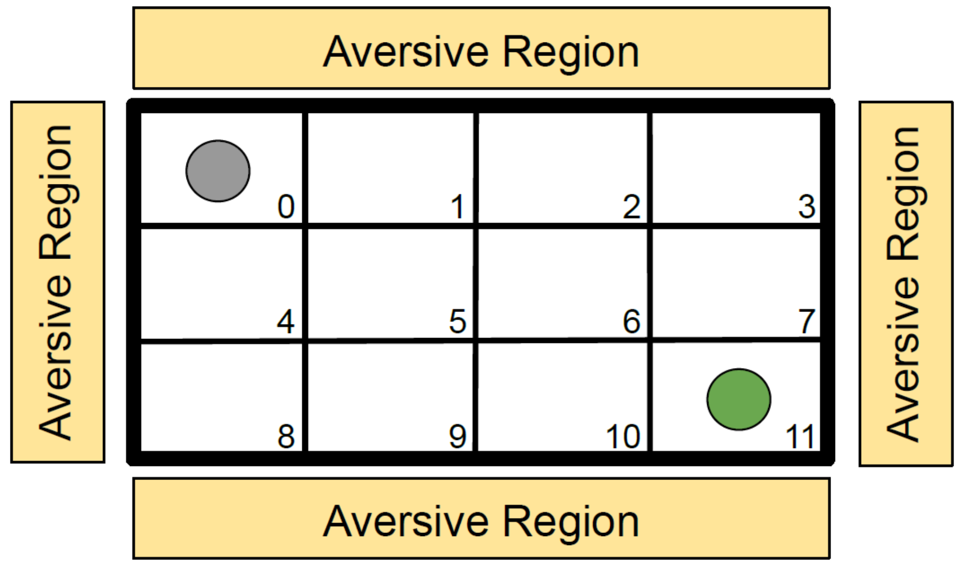


(a) Q-values.

(b) Softmax values.

(c) Probability of success.

(d) Number of transitions.



Memory-based Method

- In this context, one possible question to the artificial agent is:
  - *Why did you choose action down when in state 0?*

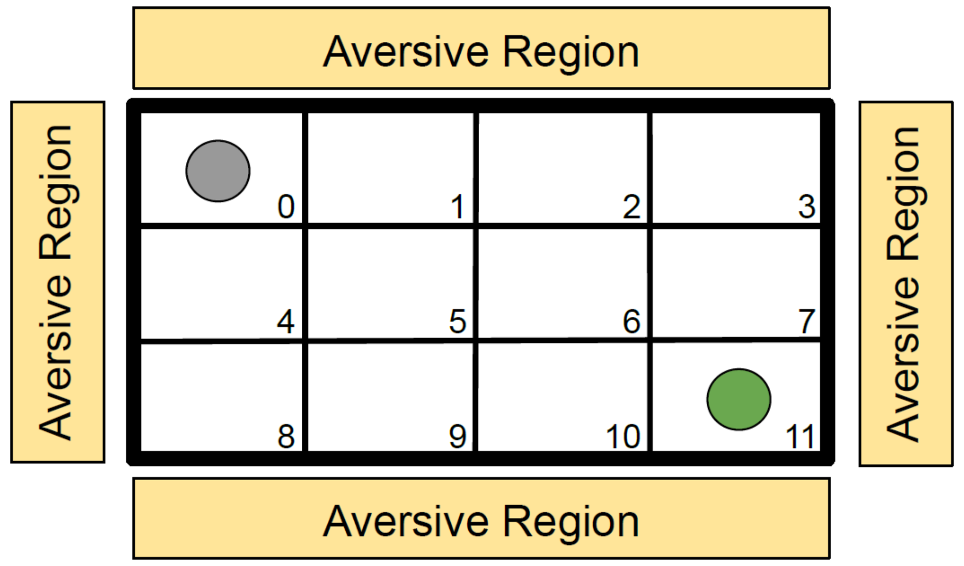
- Using Q-values to explain this is pointless for a non-expert user.

Q(s=0; a=down) = -0.181	Q(s=0; a=up) = -0.998	Q(s=0; a=right) = -0.411	Q(s=0; a=left) = -0.998

- If we use Ps, the agent may answer the end-user: *I chose to go down because that has a 73.6% probability of successfully reaching the goal.*

Ps(s=0; a=down) = 0.736	Ps(s=0; a=up) = 0	Ps(s=0; a=right) = 0.656	Ps(s=0; a=left) = 0

Memory-based Method



- Another possible question to the agent is:
  - *Why did you not choose to go left when in state 0?*

- Using Q-values to explain this is pointless for a non-expert user.

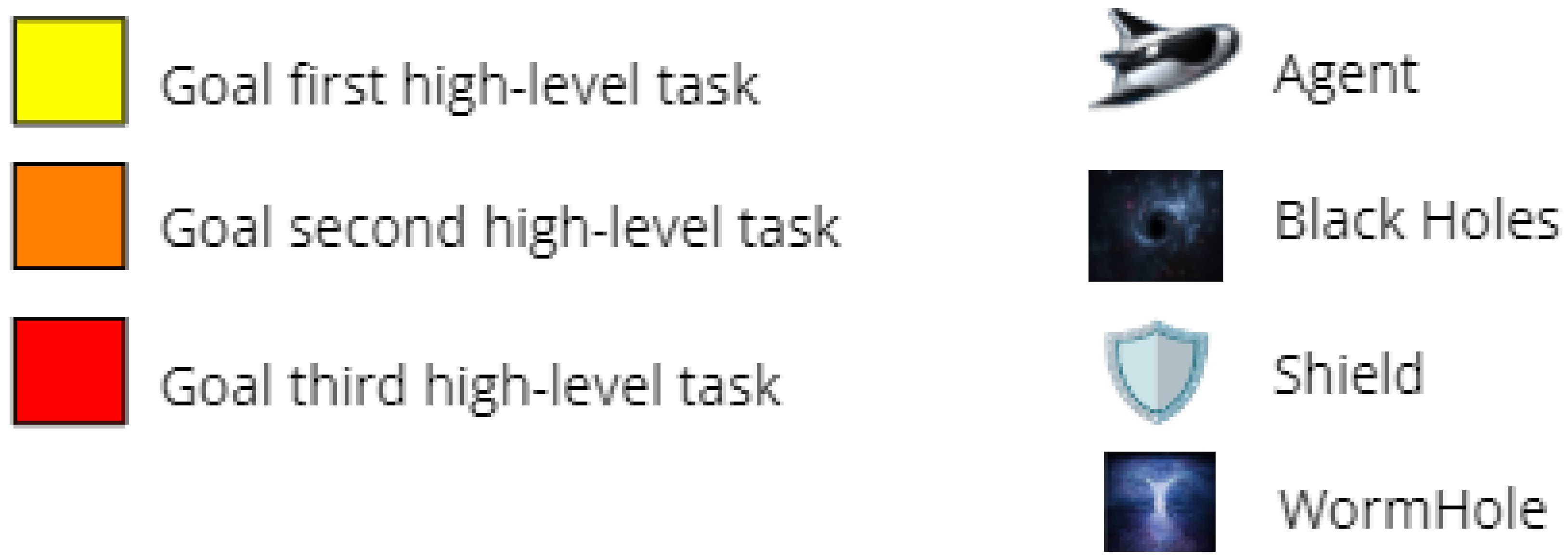
Q(s=0; a=down) = -0.181	Q(s=0; a=up) = -0.998	Q(s=0; a=right) = -0.411	Q(s=0; a=left) = -0.998

- If we use Ps, one possible answer is: *I did not choose left because that has a zero probability of success, whereas by choosing down has a 73.6% probability of success, which was higher than other actions.*

Ps(s=0; a=down) = 0.736	Ps(s=0; a=up) = 0	Ps(s=0; a=right) = 0.656	Ps(s=0; a=left) = 0

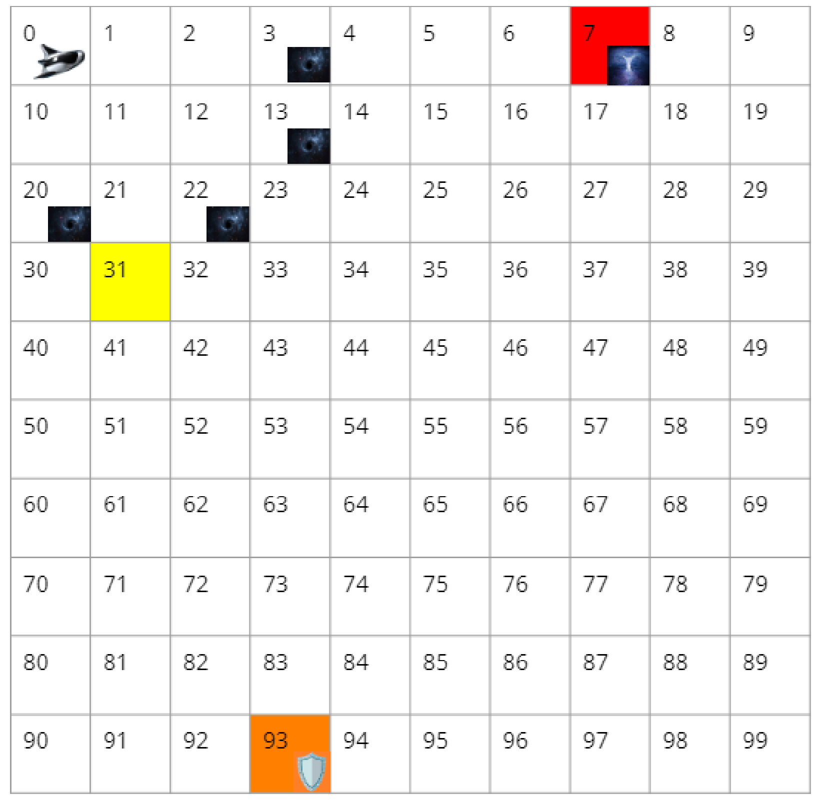
Memory-based in a Hierarchical Scenario*

- **Spaceship problem:**

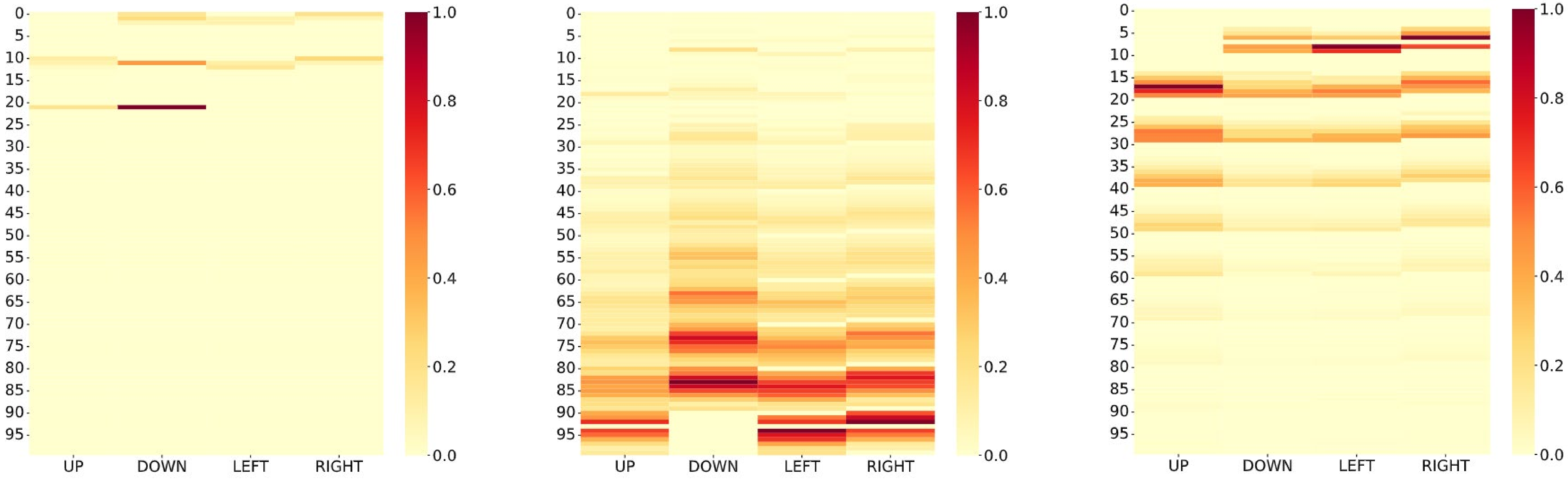


* Muñoz, H., Portugal, E., Ayala A., Fernandes, B., Cruz, F. "Explaining Agent's Decision-making in a Hierarchical Reinforcement Learning Scenario". IEEE 41st International Conference of the Chilean Computer Society (SCCC 2022).

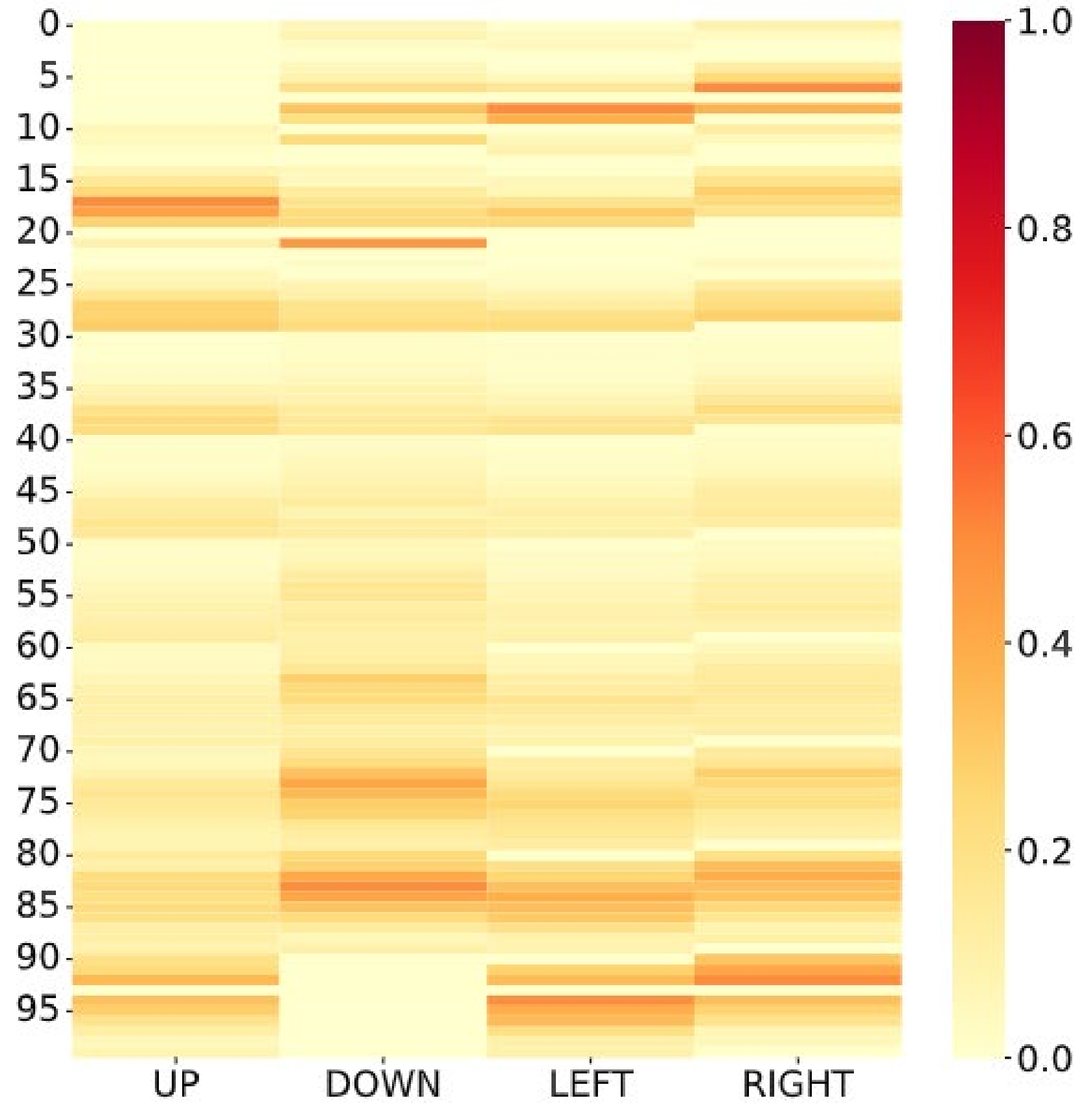
Memory-based Hierarchical Method



- Spaceship problem:



High-level tasks



General task

Learning- and Introspection-based Methods*

- Goal-driven explanations.

Algorithm 2 Explainable reinforcement learning approach to compute the probability of success using the learning-based approach.

Initialize $Q(s,a)$, $\mathbb{P}(s_t,a_t)$
for each episode **do**
  Initialize s_t
  Choose an action a_t from s_t
  repeat
    Take action a_t
    Observe reward r_{t+1} and next state s_{t+1}
    Choose next action a_{t+1} using softmax action selection method
    $Q(s_t,a_t) \leftarrow Q(s_t,a_t) + \alpha[r_{t+1} + \gamma Q(s_{t+1},a_{t+1}) - Q(s_t,a_t)]$
    $\mathbb{P}(s_t,a_t) \leftarrow \mathbb{P}(s_t,a_t) + \alpha[\varphi_{t+1} + \mathbb{P}(s_{t+1},a_{t+1}) - \mathbb{P}(s_t,a_t)]$
    $s_t \leftarrow s_{t+1}$; $a_t \leftarrow a_{t+1}$
  until s_t is terminal (goal or aversive state)
end for

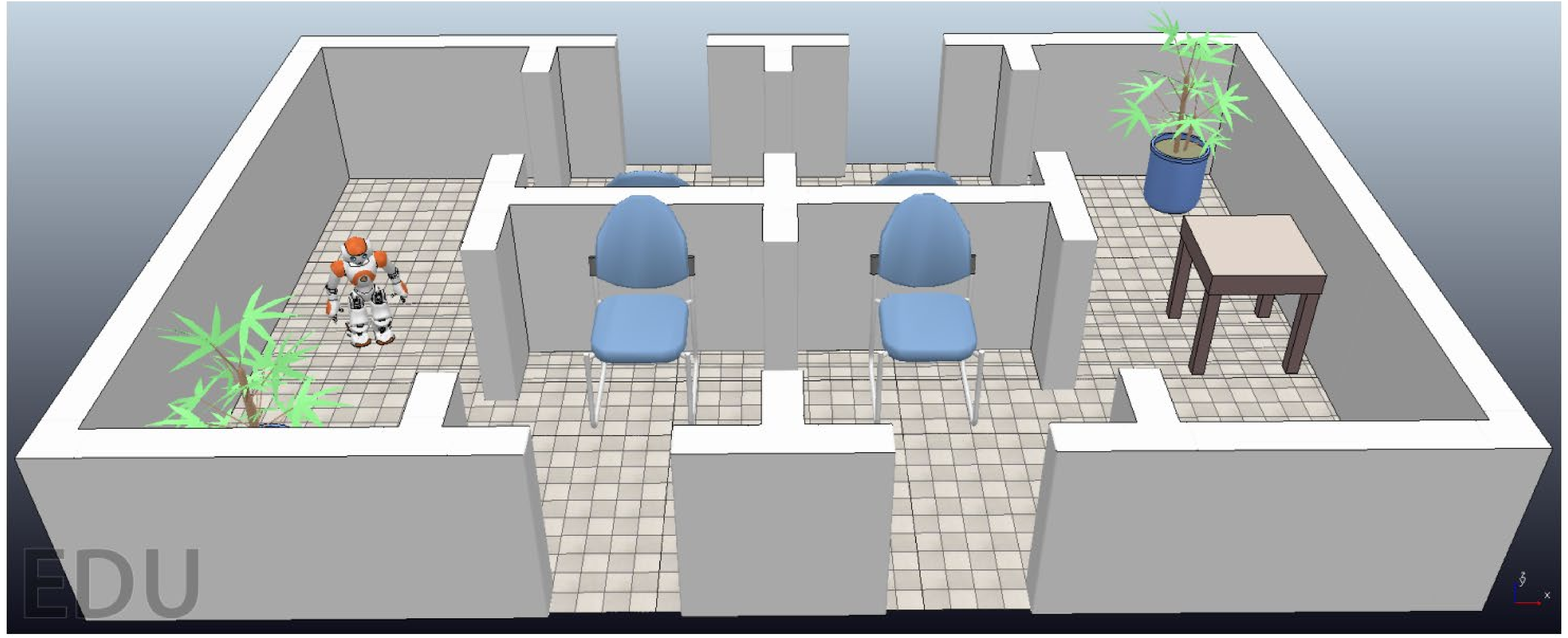
Algorithm 3 Explainable reinforcement learning approach to compute the probability of success using the introspection-based approach.

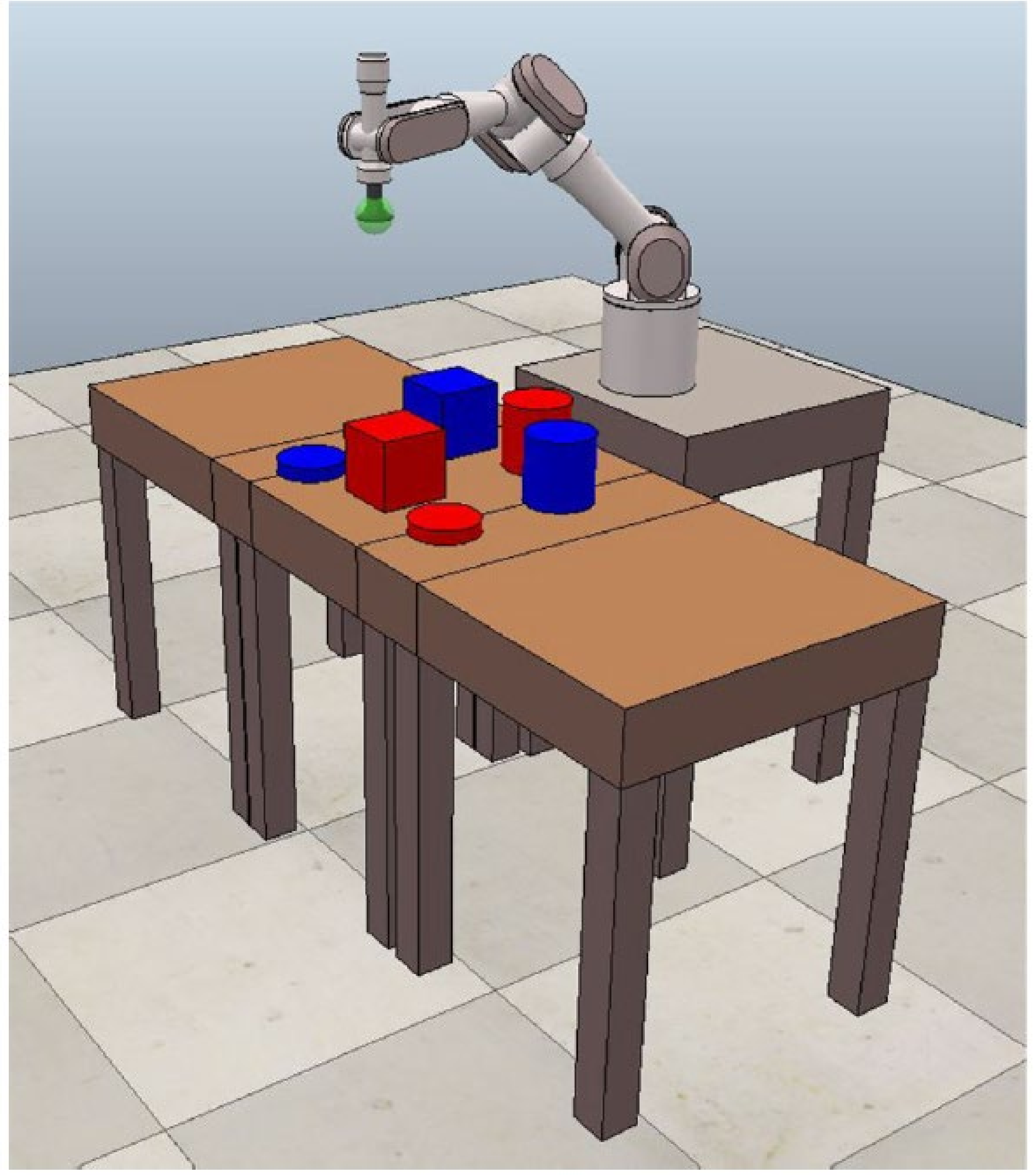
Initialize $Q(s,a)$, $\hat{P}_s$
for each episode **do**
  Initialize s_t
  Choose an action a_t from s_t
  repeat
    Take action a_t
    Observe reward r_{t+1} and next state s_{t+1}
    Choose next action a_{t+1} using softmax action selection method
    $Q(s_t,a_t) \leftarrow Q(s_t,a_t) + \alpha[r_{t+1} + \gamma Q(s_{t+1},a_{t+1}) - Q(s_t,a_t)]$
    $s_t \leftarrow s_{t+1}$; $a_t \leftarrow a_{t+1}$
  until s_t is terminal (goal or aversive state)
  $\hat{P}_s \approx \left[(1-\sigma)\cdot\left(\frac{1}{2}\cdot log_{10}\frac{Q(s_t,a_t)}{R^T}+1\right)\right]_{\hat{P}_s\geq 0}^{\hat{P}_s\leq 1}$
end for

* Cruz, F., Dazeley, R., Vamplew, P., Moreira, I. “Explainable Robotic Systems: Understanding Goal-driven Actions in a Reinforcement Learning Scenario”. *Neural Computing and Applications*. Springer. 2021.

Learning- and Introspection-based Methods

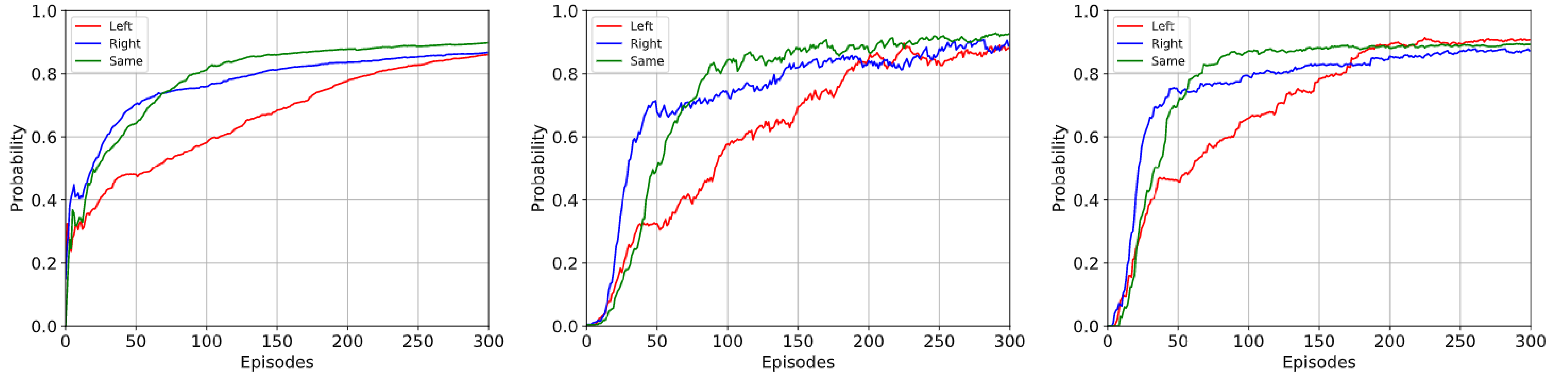
- Deterministic and stochastic navigation task.
- Continuous sorting object task.
- Real-world scenario.



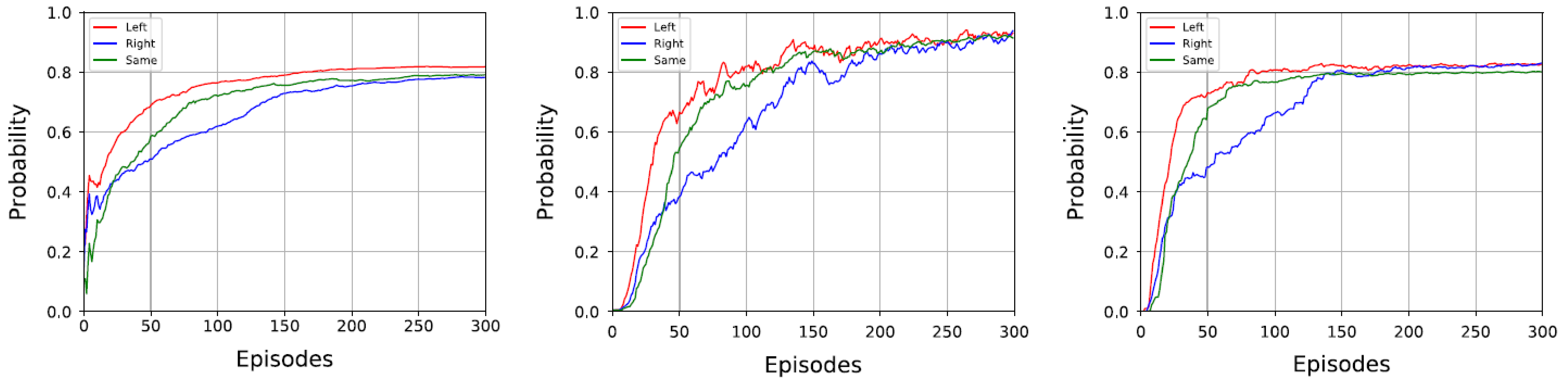


Learning- and Introspection-based Methods

Deterministic and stochastic tasks.

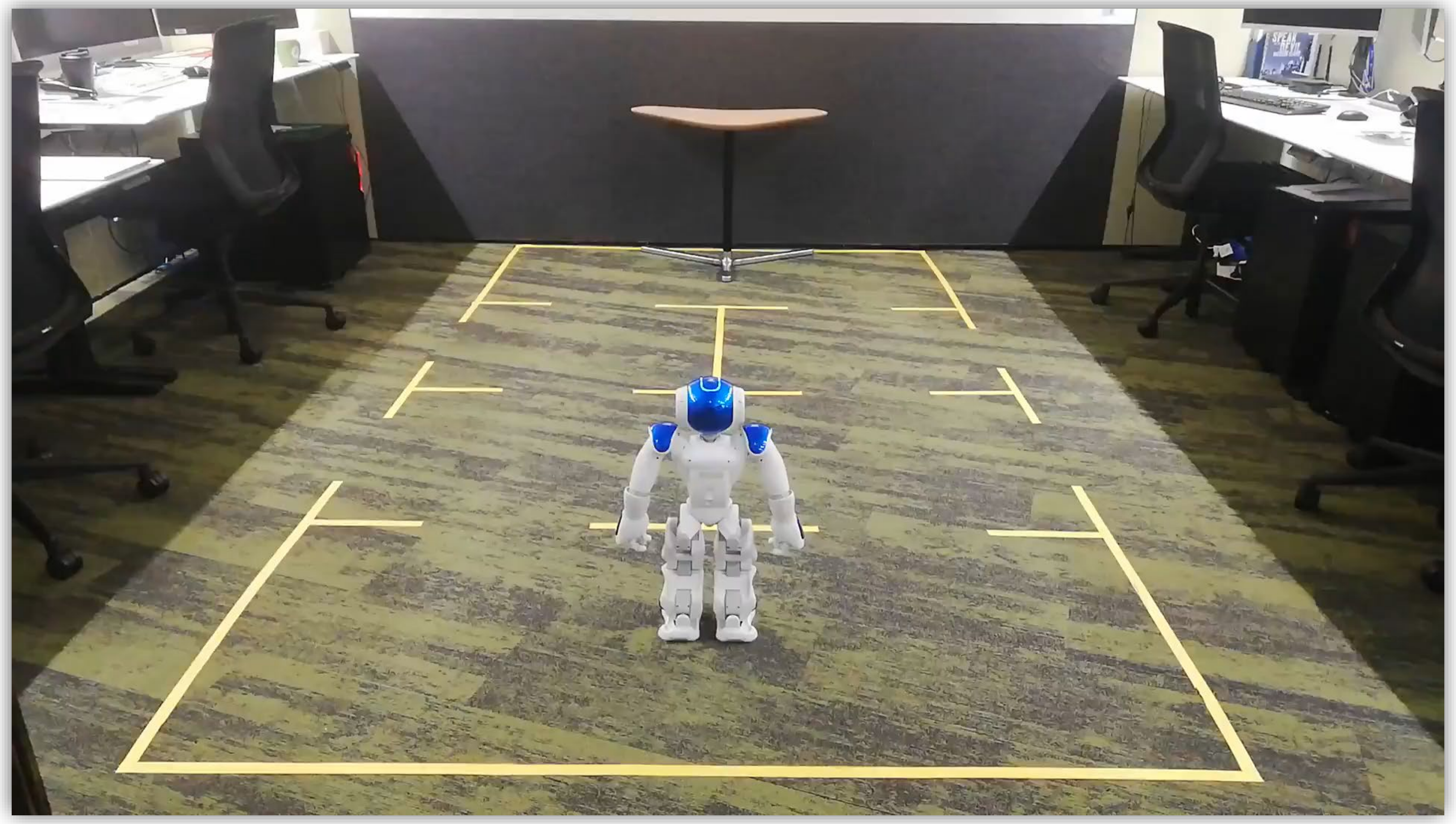


(d) Memory-based approach. (e) Learning-based approach. (f) Introspection-based approach.



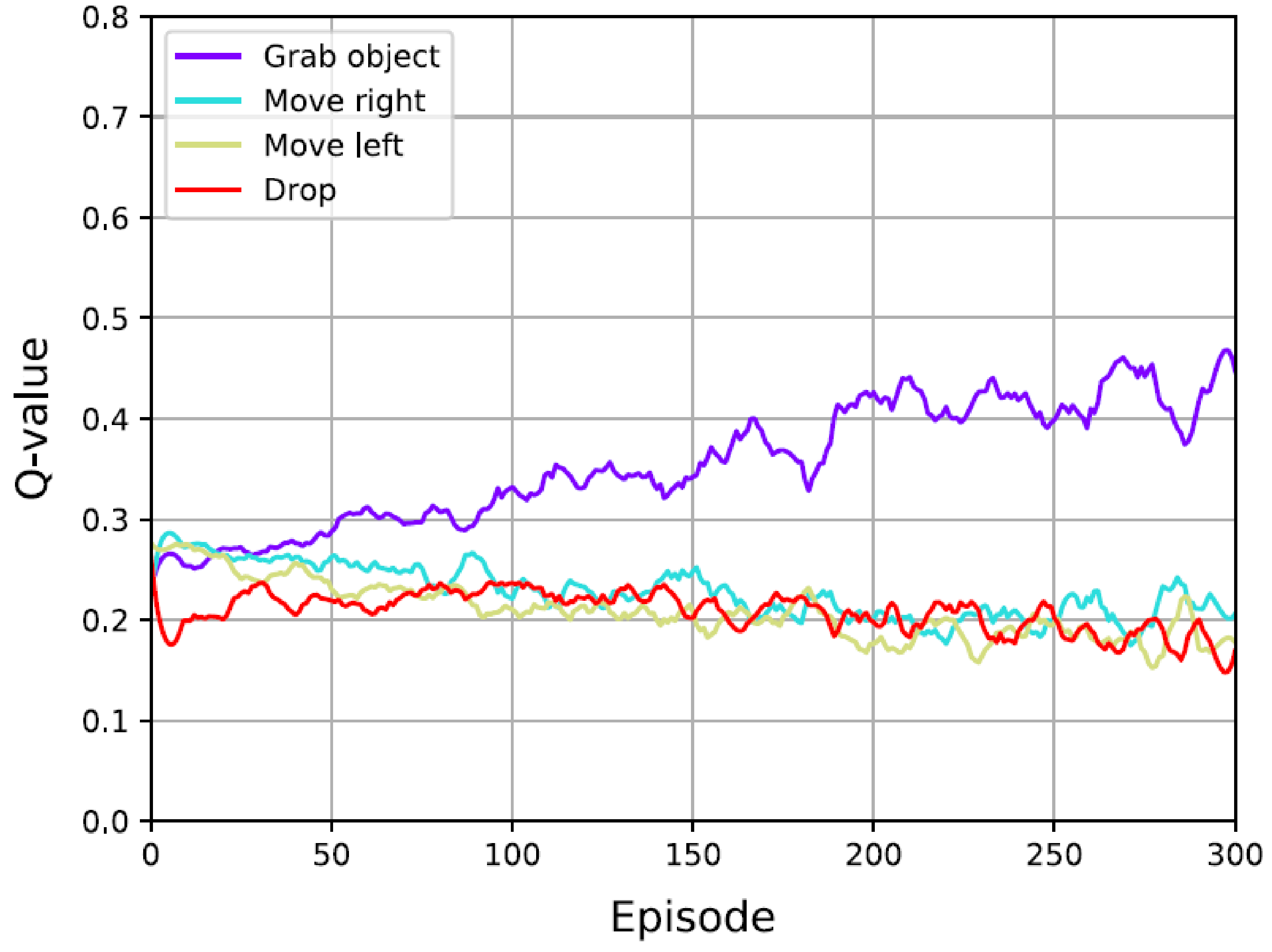
(d) Memory-based approach. **(e)** Learning-based approach. **(f)** Introspection-based approach.

Left Right Same

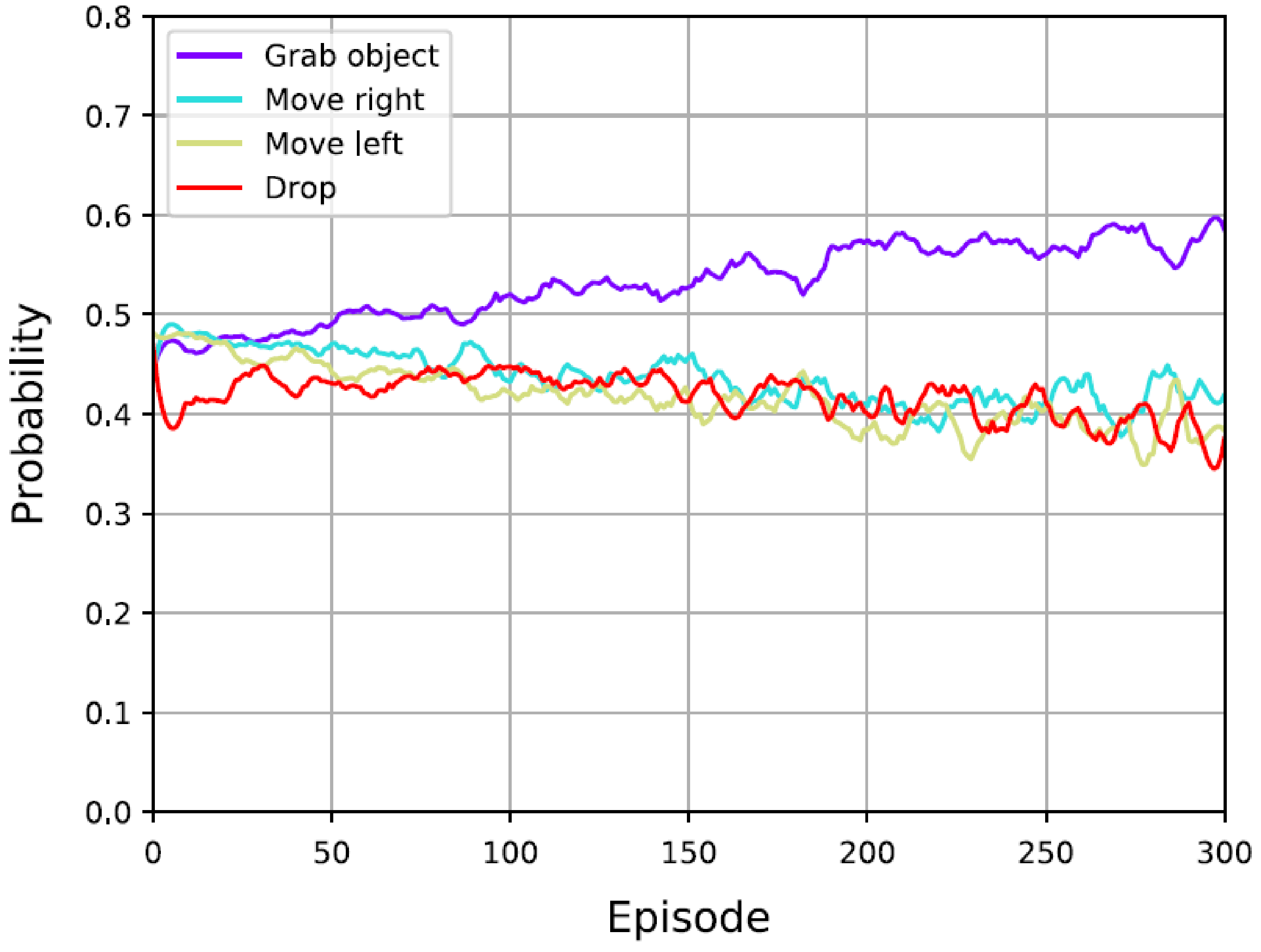


Explanation. I chose to go left because that has a 87.6% probability of reaching the goal successfully

Learning- and Introspection-based Methods

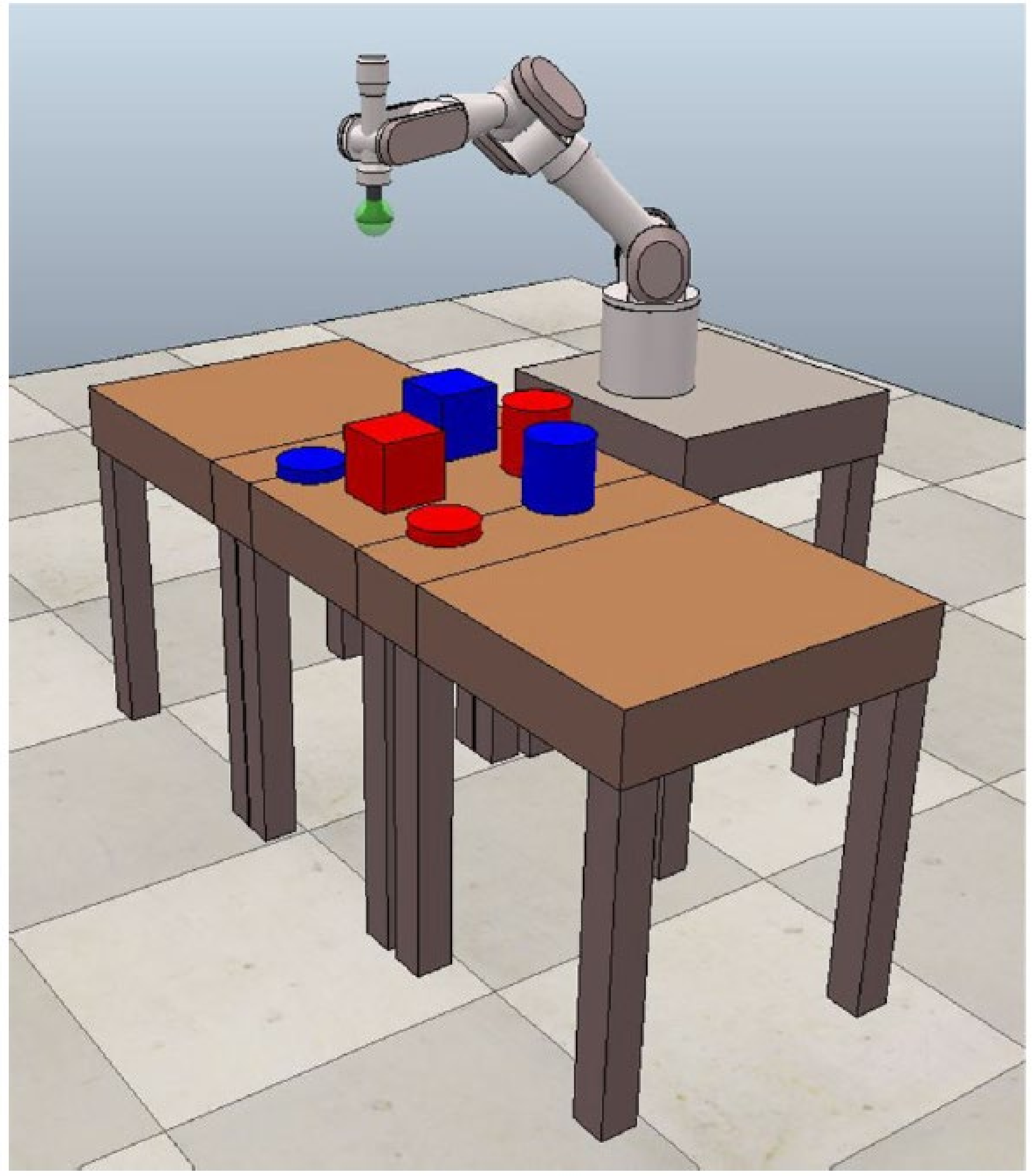


(a) Q-values.



(b) Introspection-based approach.

Grab object Move right Move left Drop

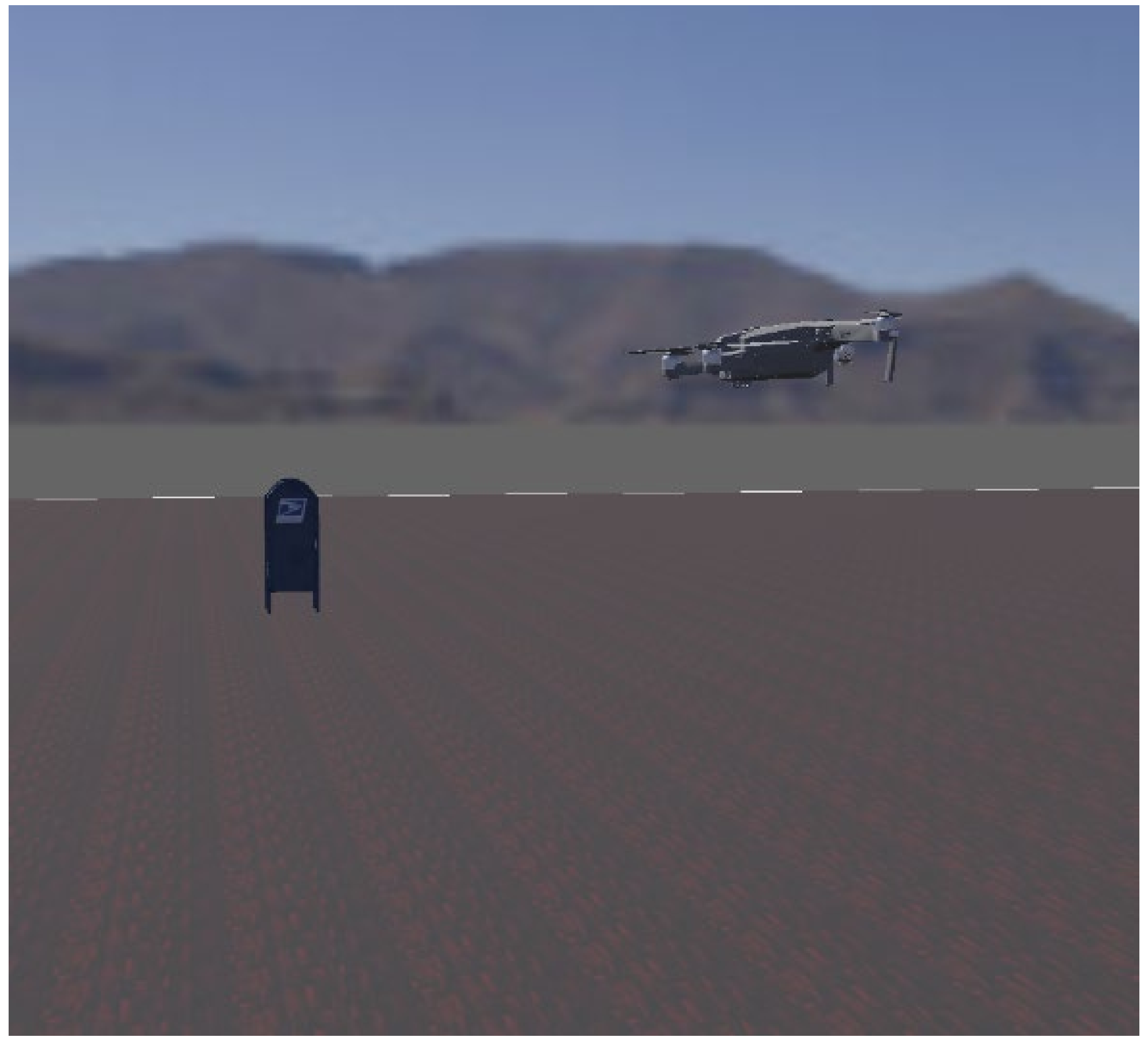


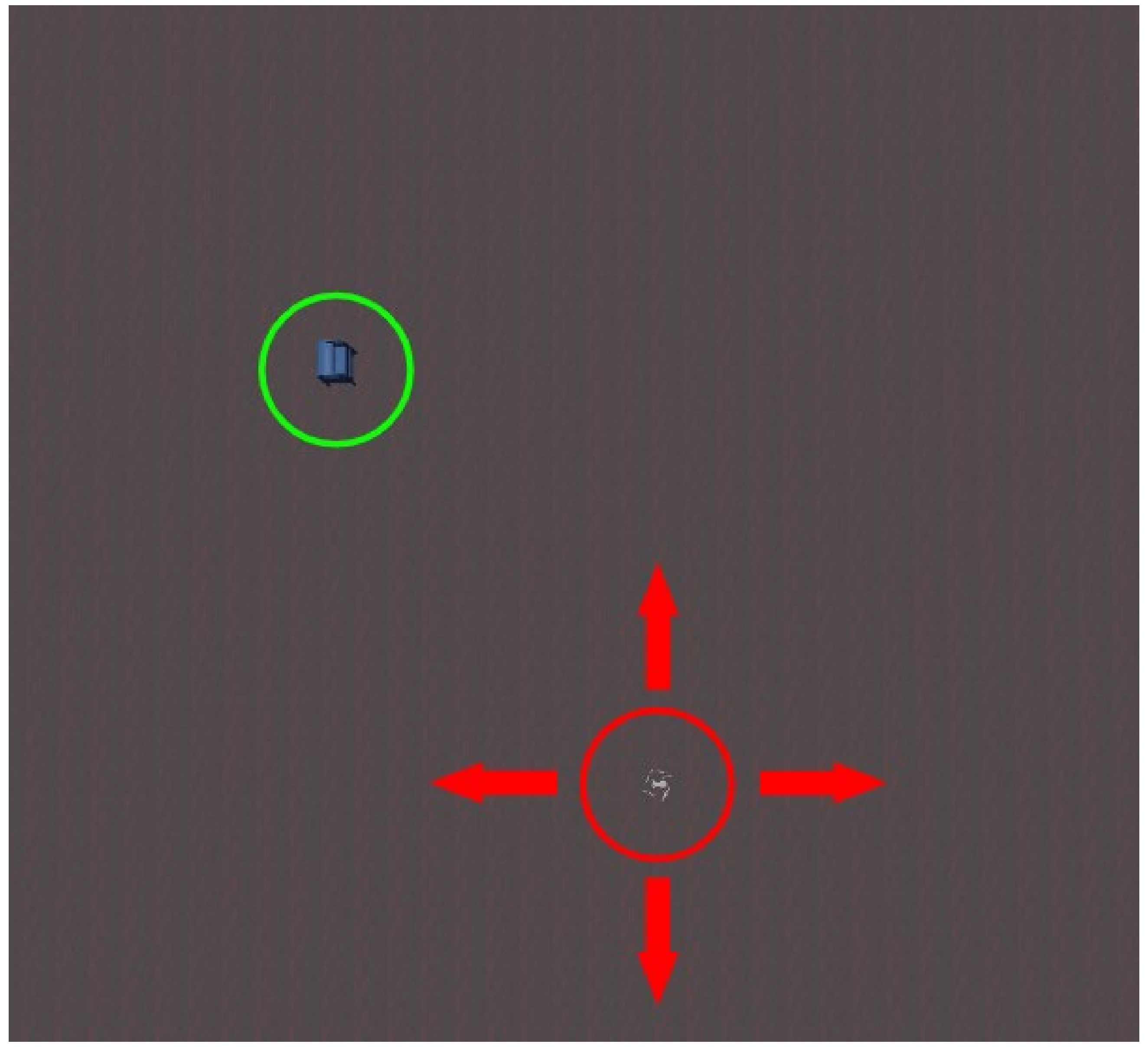
Question. Why the action move right or move left have not been chosen by the agent.

Explanation. I have selected the action grab object because doing so, I have 59% chances of sorting all the objects successfully, while moving left I have only 38% probability of being successful.

Non-episodic and Continuous Domains*

- Drone scenario in Webots.





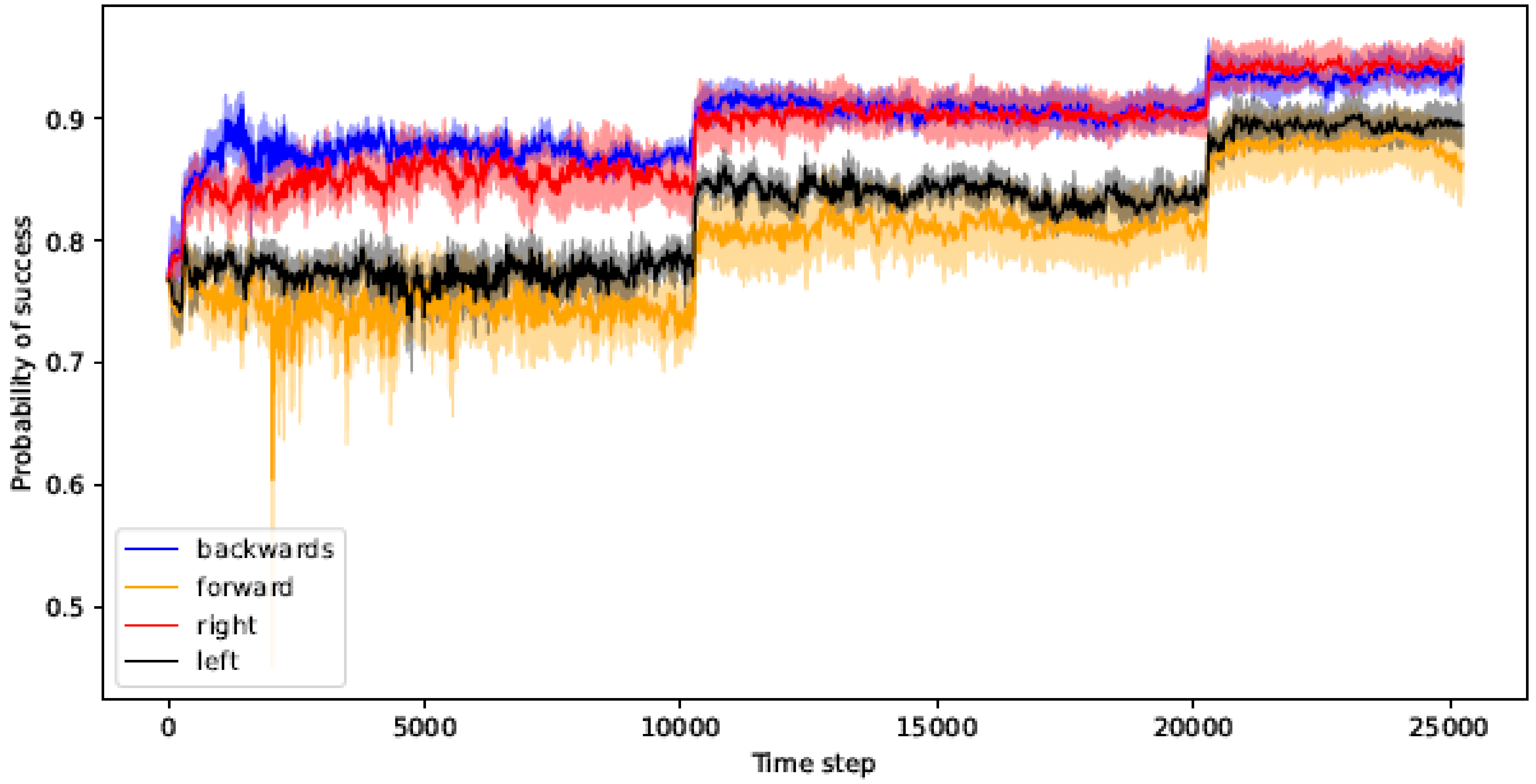


Figure 3: Probabilities of success for every available action in a spot close to the top-left corner.

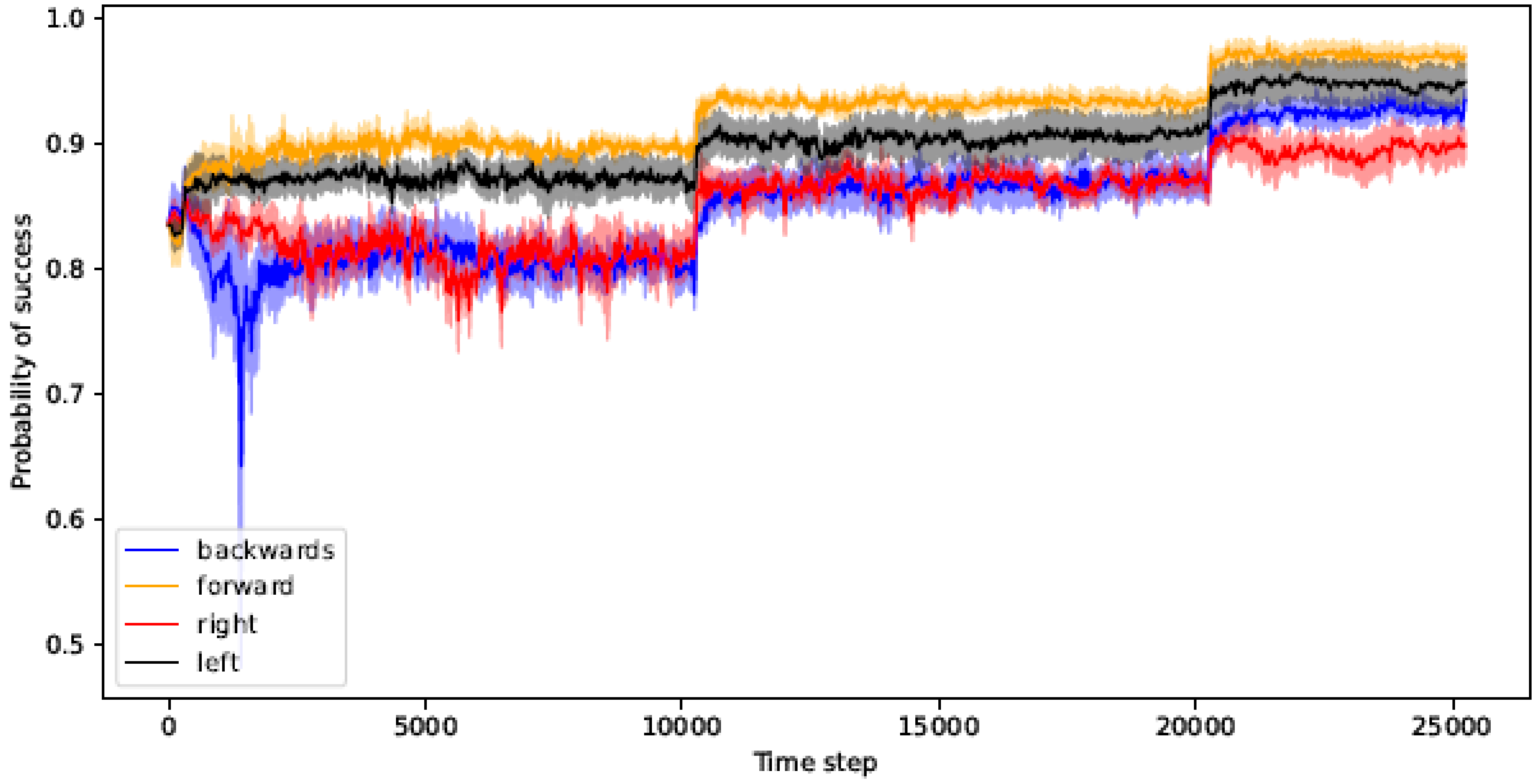
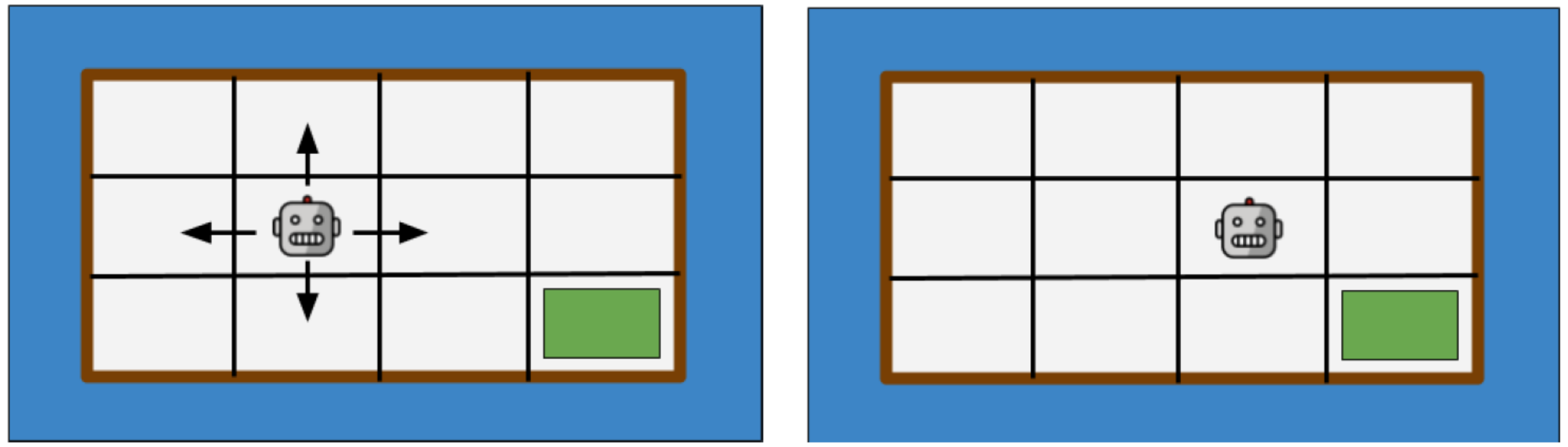


Figure 4: Probabilities of success for every available action in a spot close to the bottom-right corner.

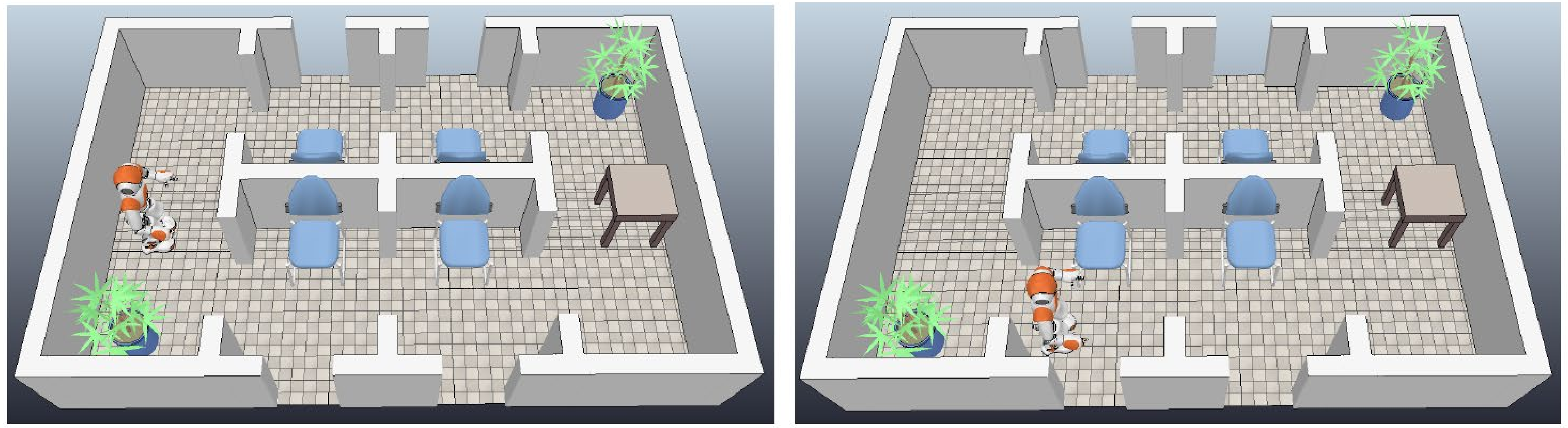
* Schroeter, N., Cruz, F., Wermter, S. "Introspection-based Explainable Reinforcement Learning in Episodic and Non-episodic Scenarios". Australian Conference on Robotics and Automation (ACRA 2022).

Evaluating Goal-driven Explanations by Non-experts End-users*

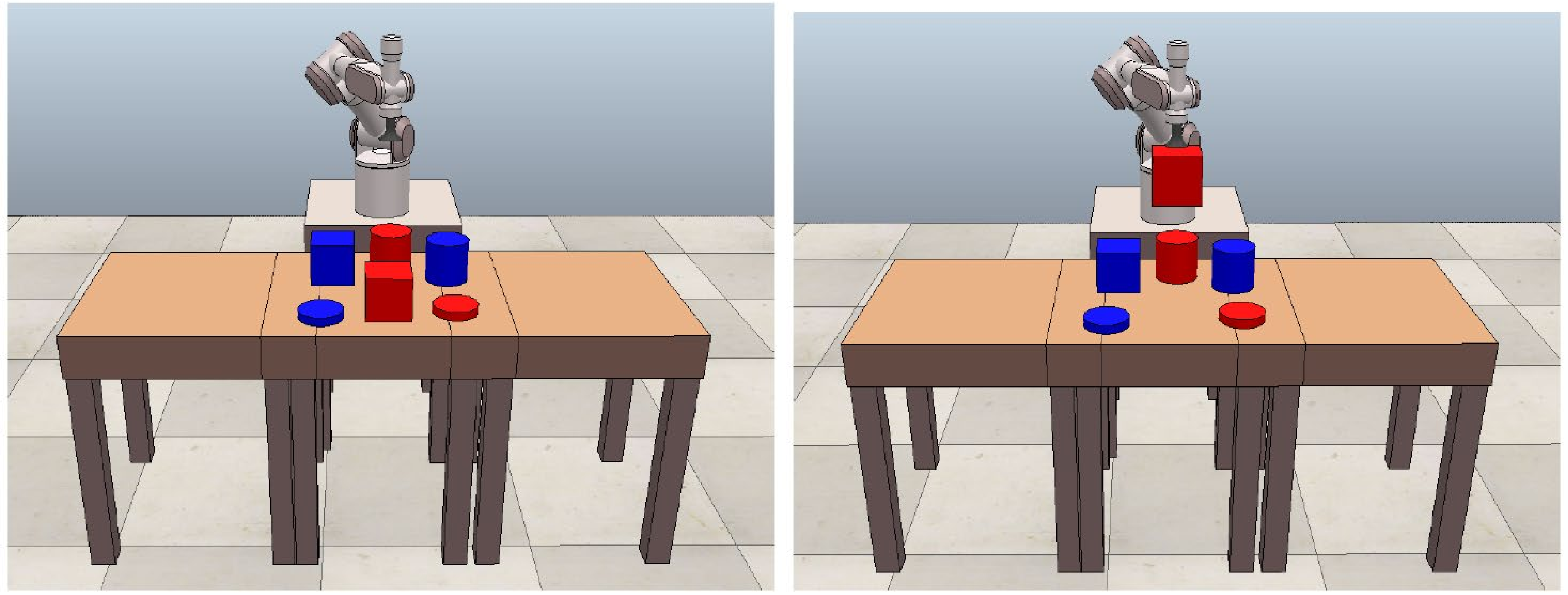
- User study using Amazon Mechanical Turk with 228 participants.



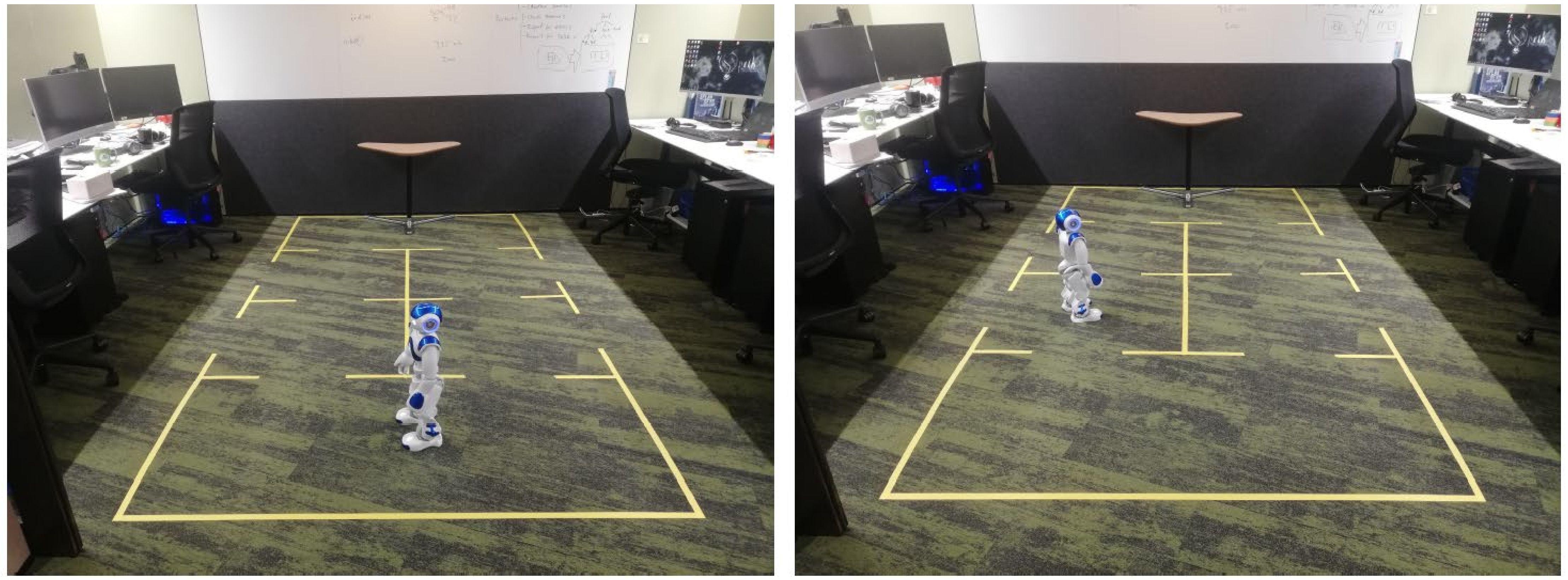
(a) Initial state. (b) State after ‘go east’ action.



(a) Initial state. (b) State after ‘move right’ action.



(a) Initial state. (b) State after ‘grab an object’ action.

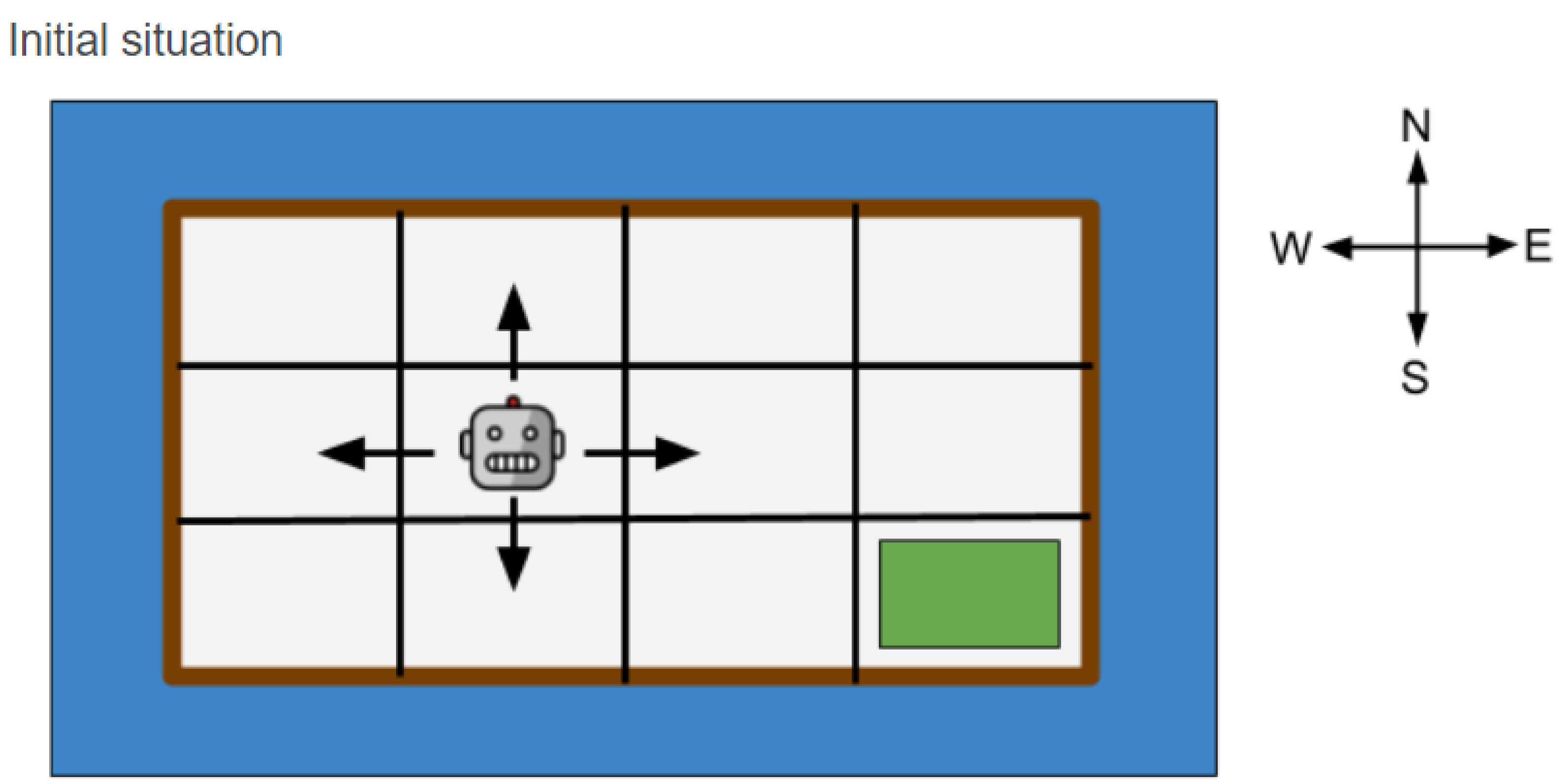


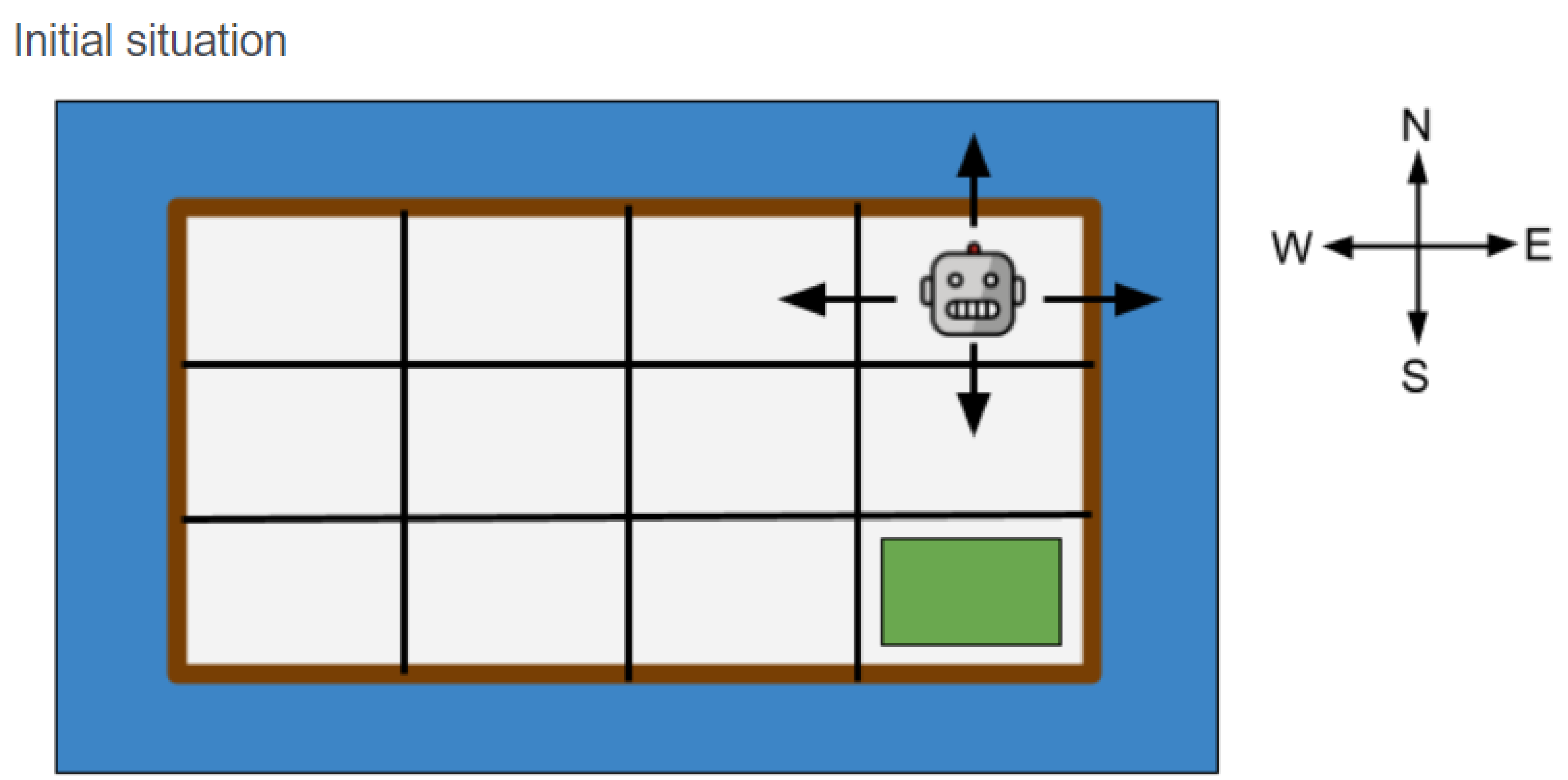
(a) Initial state. (b) State after ‘move to the left’ action

* Cruz, F., Young, C., Dazeley, R., Vamplew, P. "Evaluating Human-like Explanations for Robot Actions in Reinforcement Learning Scenarios". IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Kyoto, Japan, 2022.

Evaluating Goal-driven Explanations by Non-experts End-users

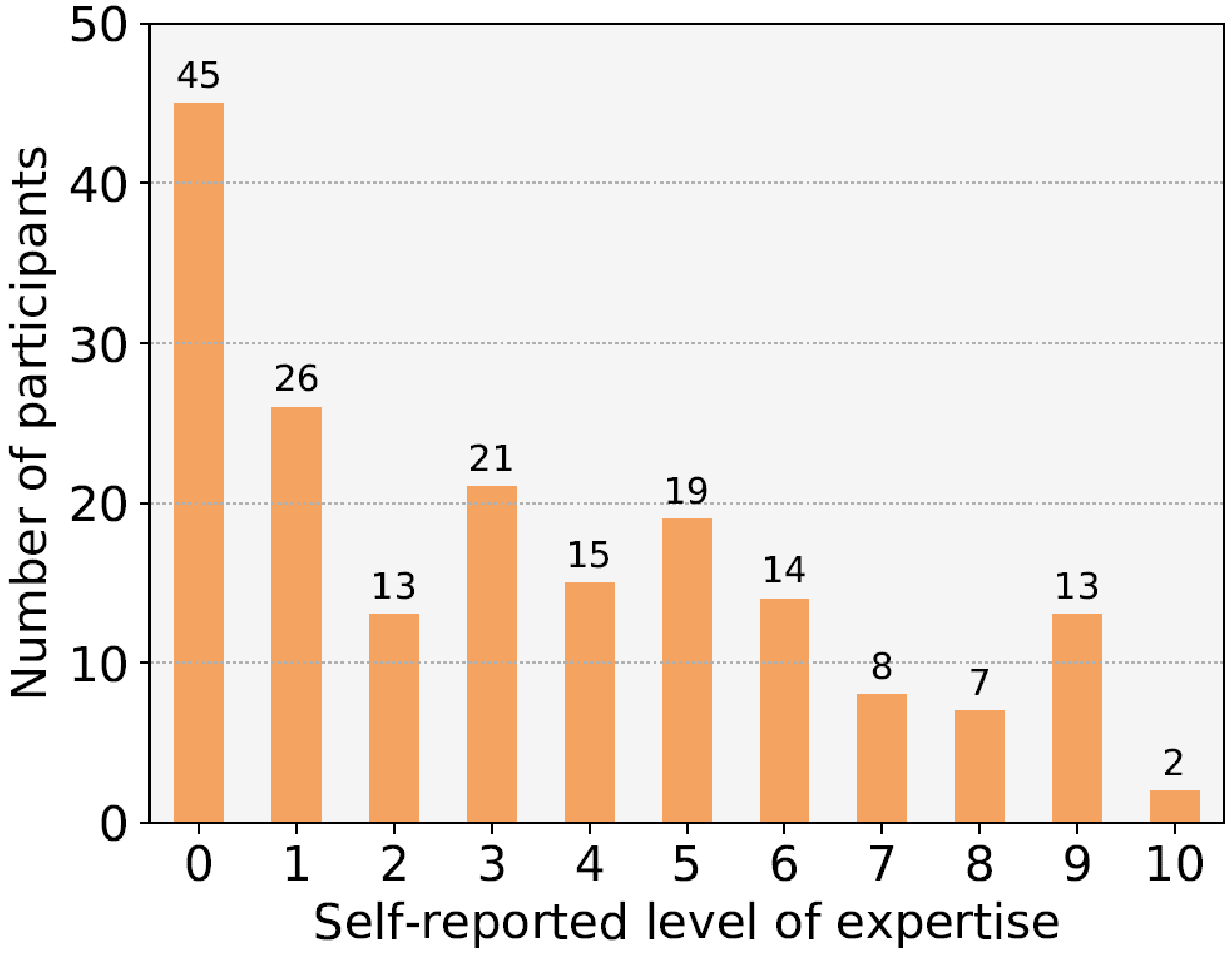
- Technical, human-like and standalone, counterfactual explanations.
  - **[S]** After performing 'go east' from (1,1). Why did you move to the east?
    - **[T]** I moved to the east because it has a Q-value of -0.411
    - **[H]** I moved to the east because it has a 65.6% probability of reaching the green position
  - **[C]** After performing 'go south' from (3,0). Why you did not move to the east?
    - **[T]** I did not move to the east because it has a Q-value of -0.998, while moving south has a Q-value of 0.181
    - **[H]** I did not move to the east because it has 0% probability of reaching the greenposition, instead moving south has 73.6% probability



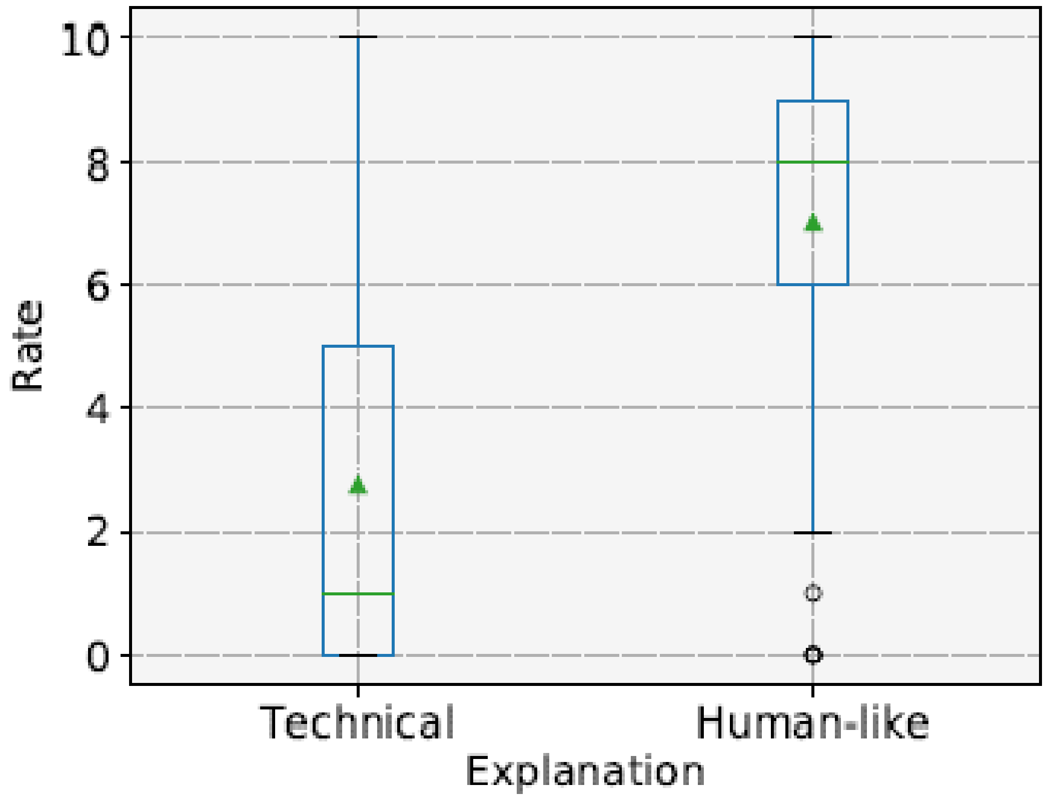


Evaluating Goal-driven Explanations by Non-experts End-users

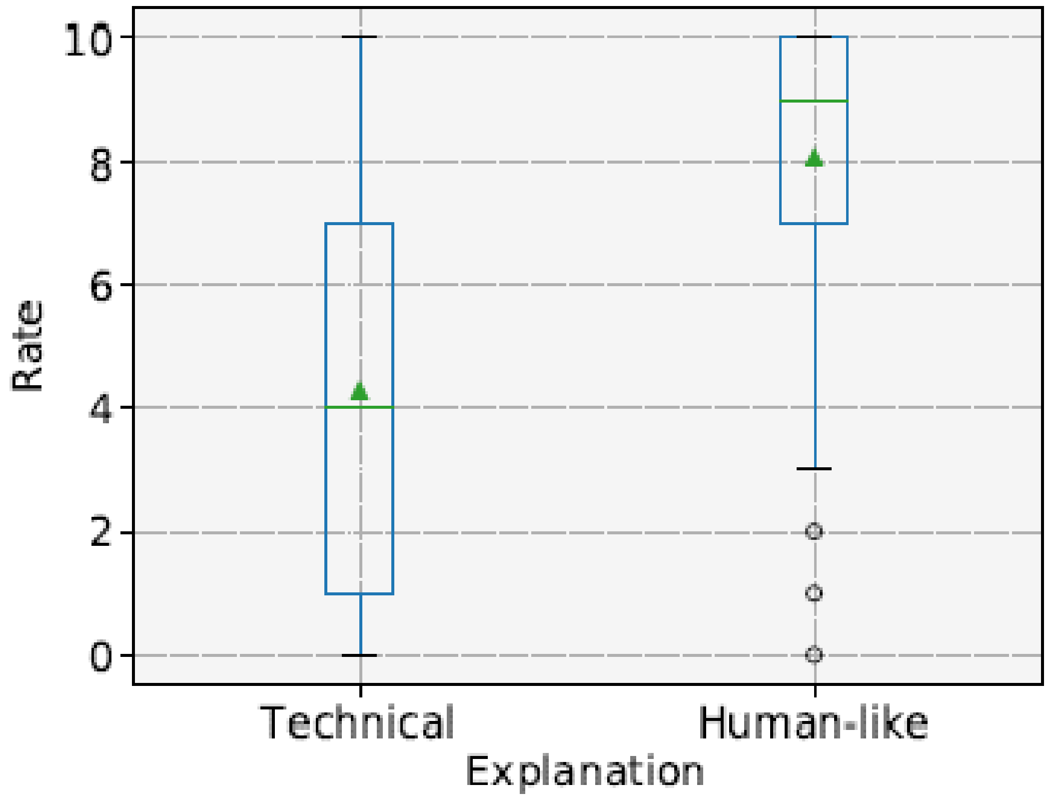
- Most of participants reported no previous expertise in machine learning.



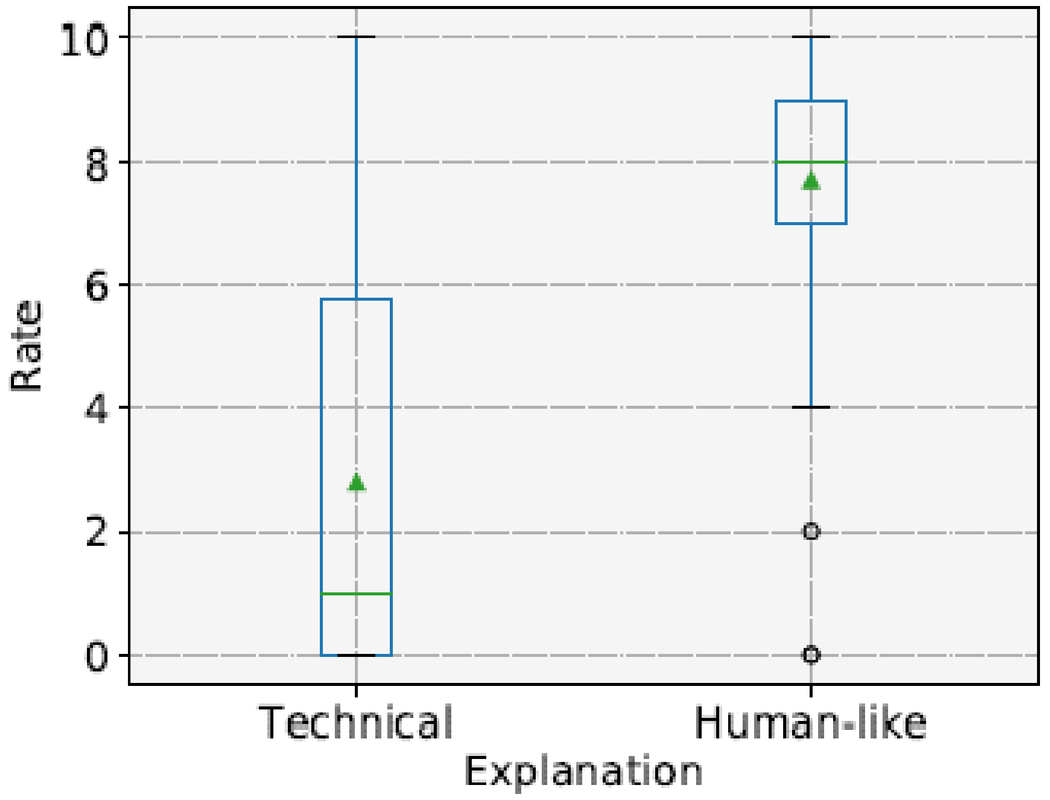
Evaluating Goal-driven Explanations by Non-experts End-users



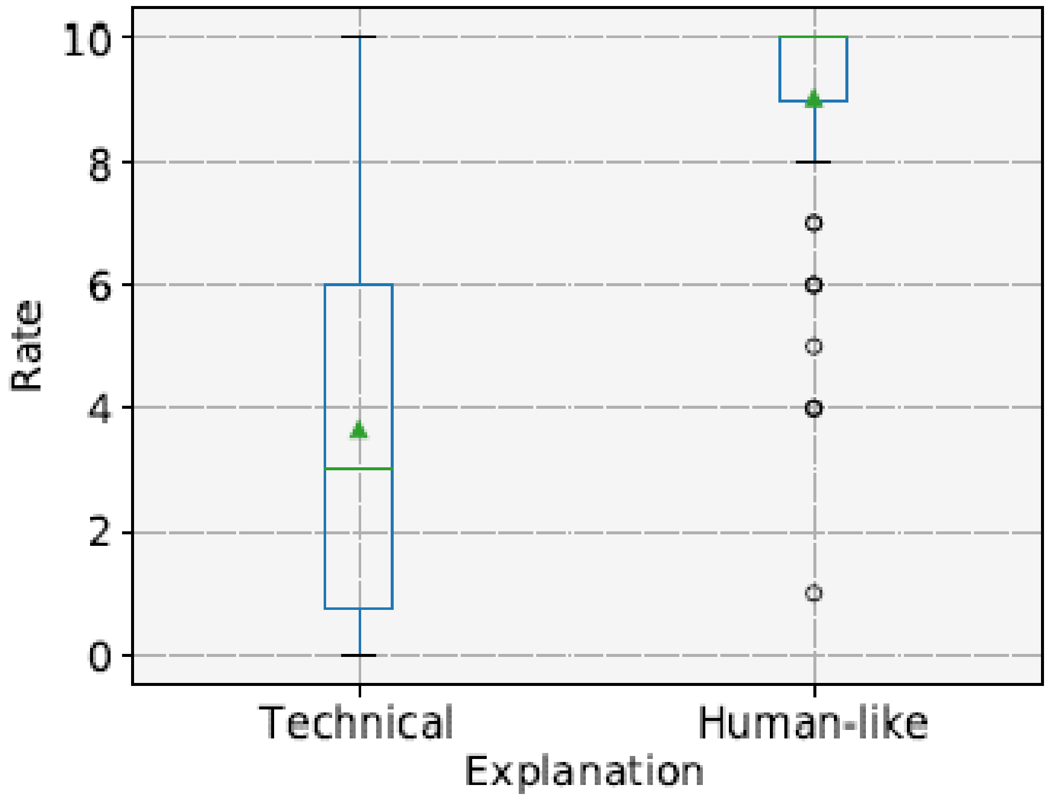
(a) Island scenario – Go east.



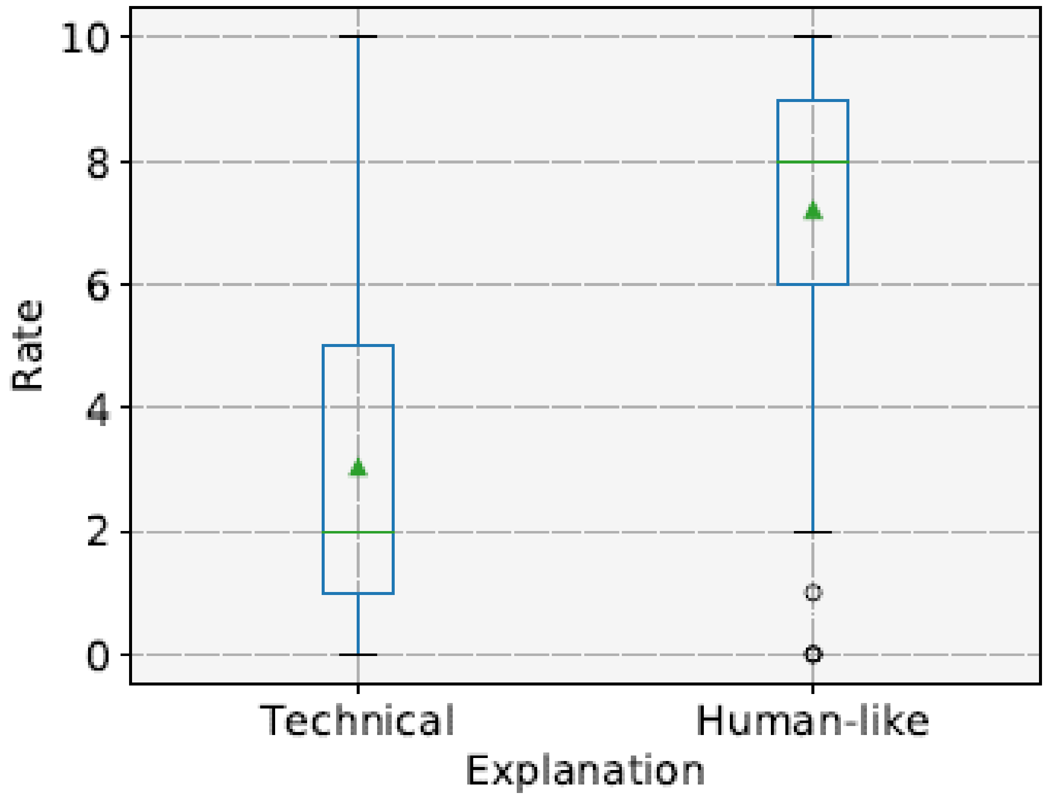
(b) Island scenario – Go south.



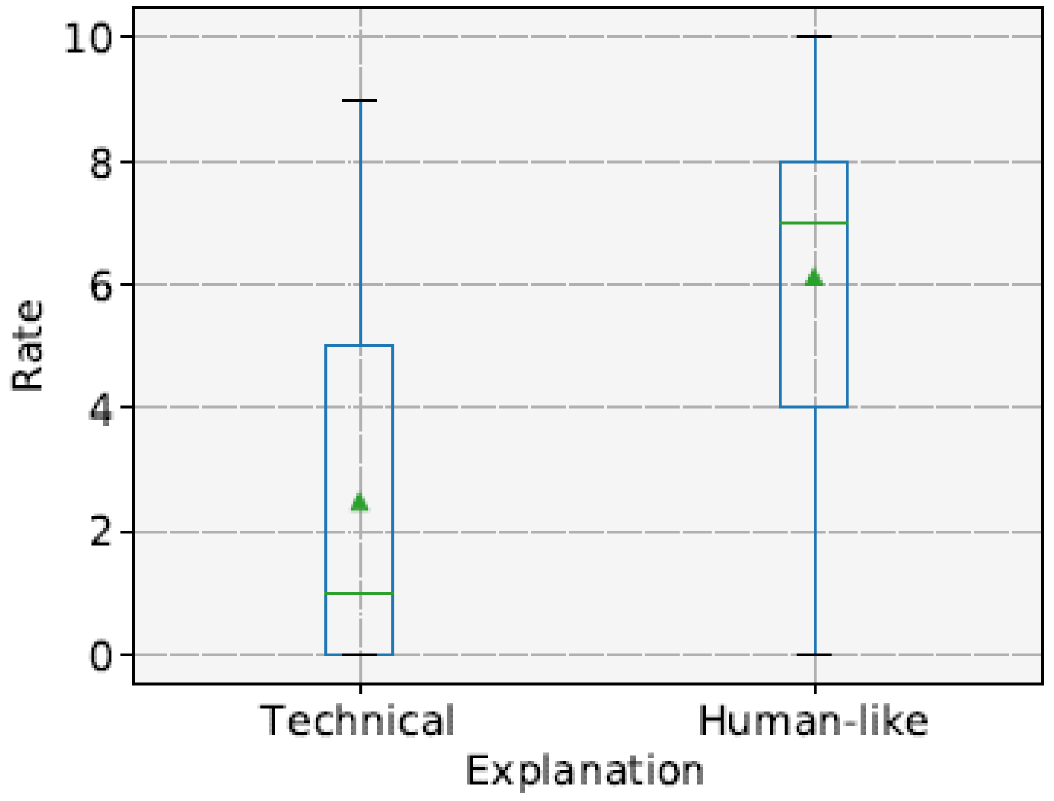
(c) Navigation scenario – Move right.



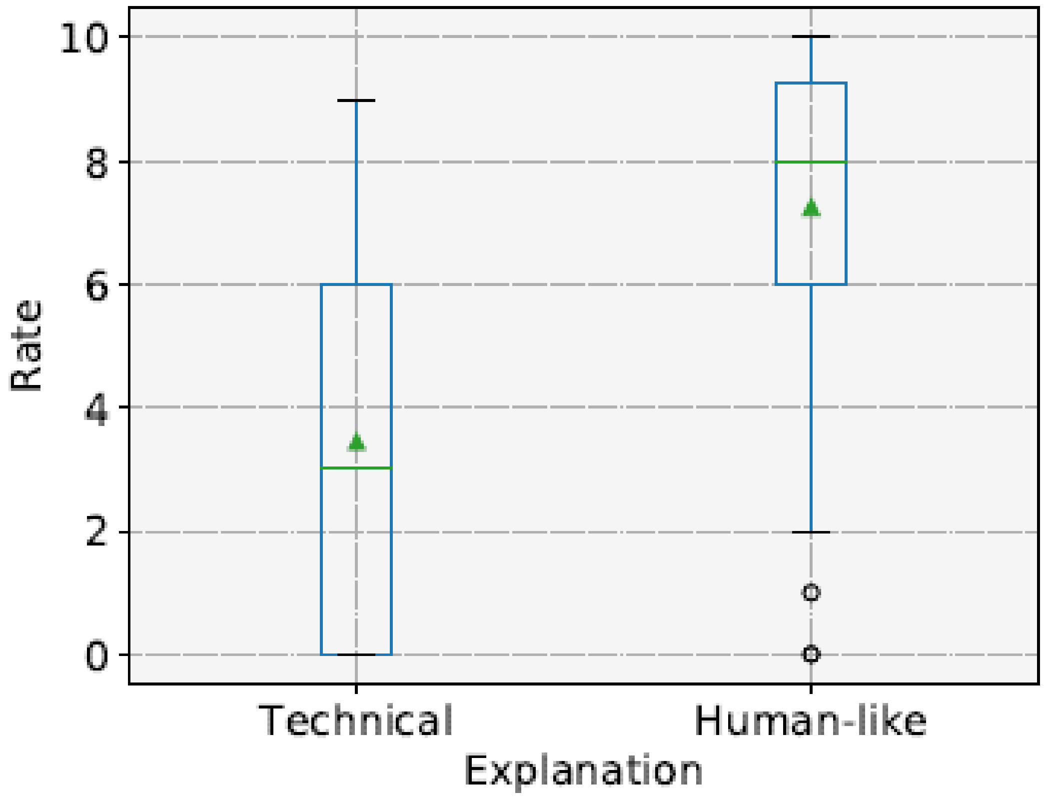
(d) Navigation scenario – Move straight.



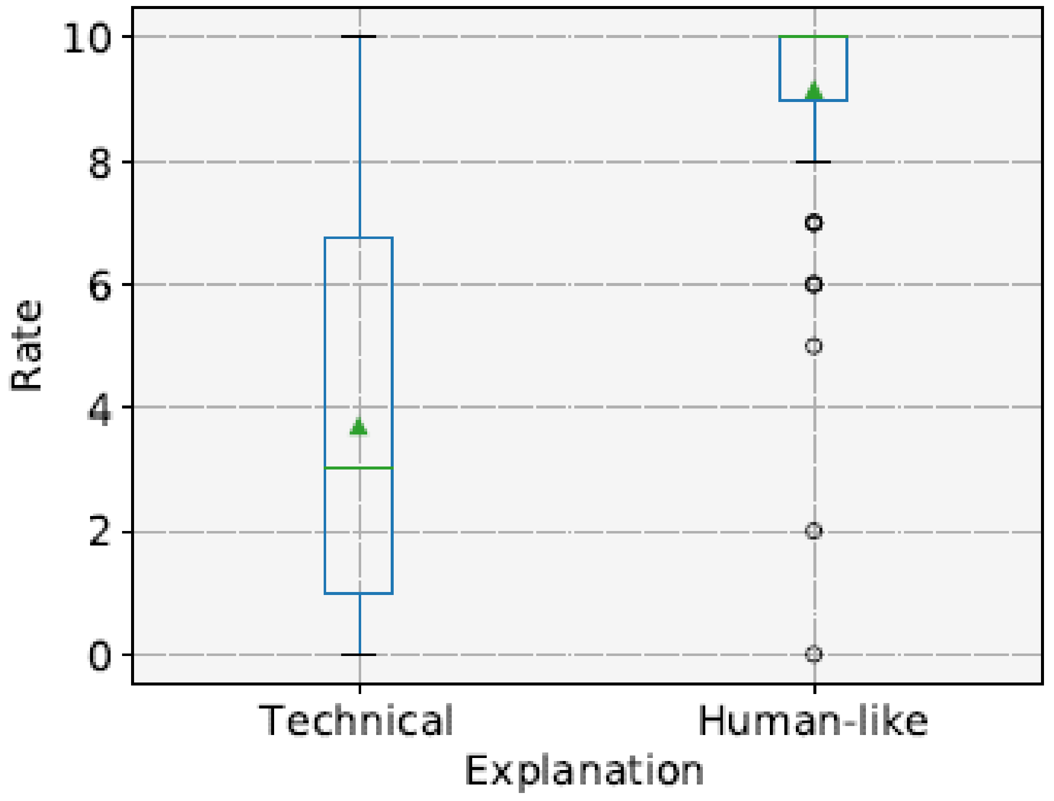
(e) Robot arm – Move to the right.



(f) Robot arm – Grab an object.



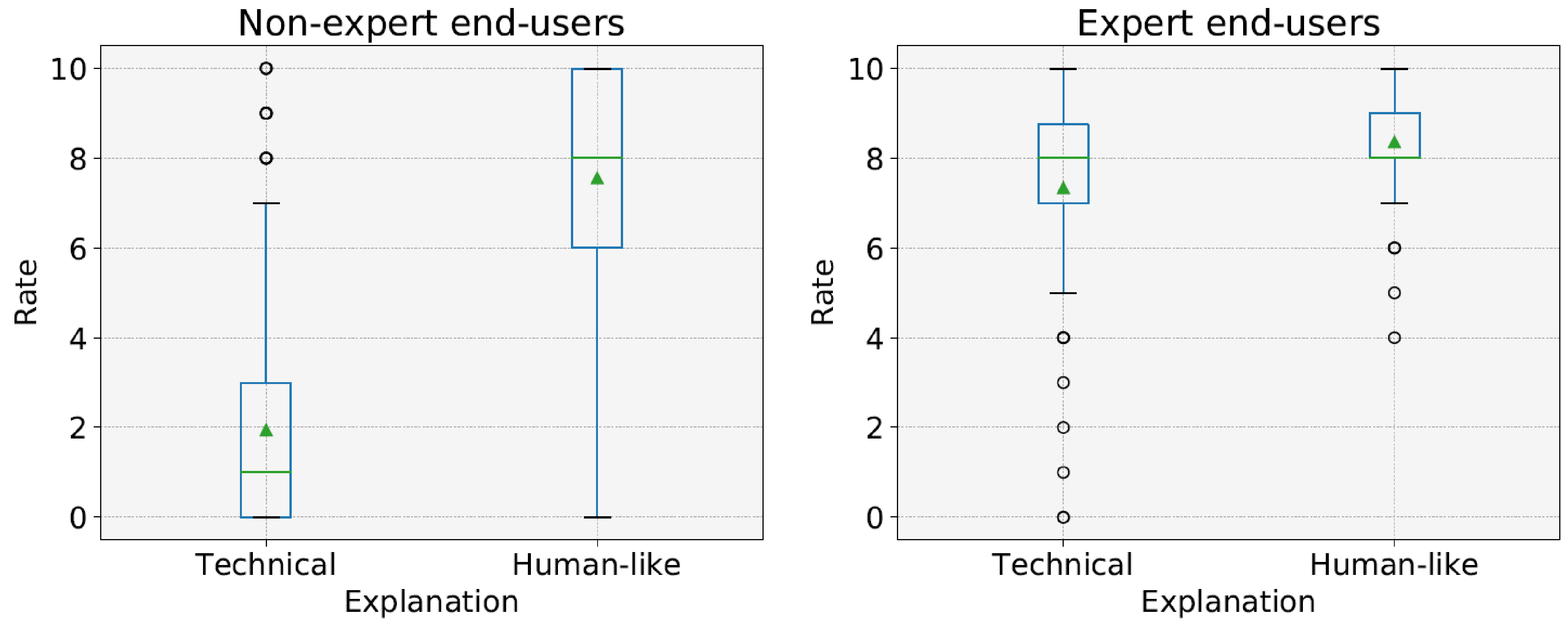
(g) Real-world Nao – Move straight.



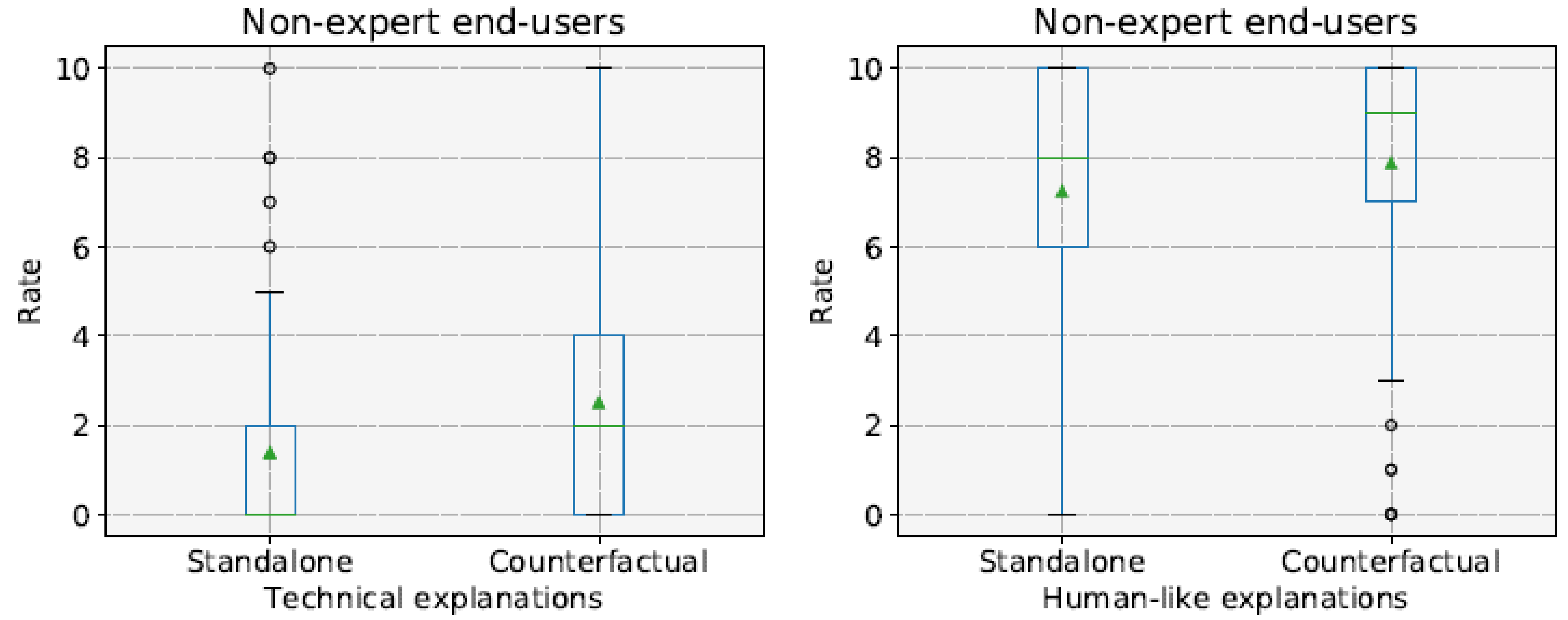
(h) Real-world Nao – Move to the left.

Evaluating Goal-driven Explanations by Non-experts End-users

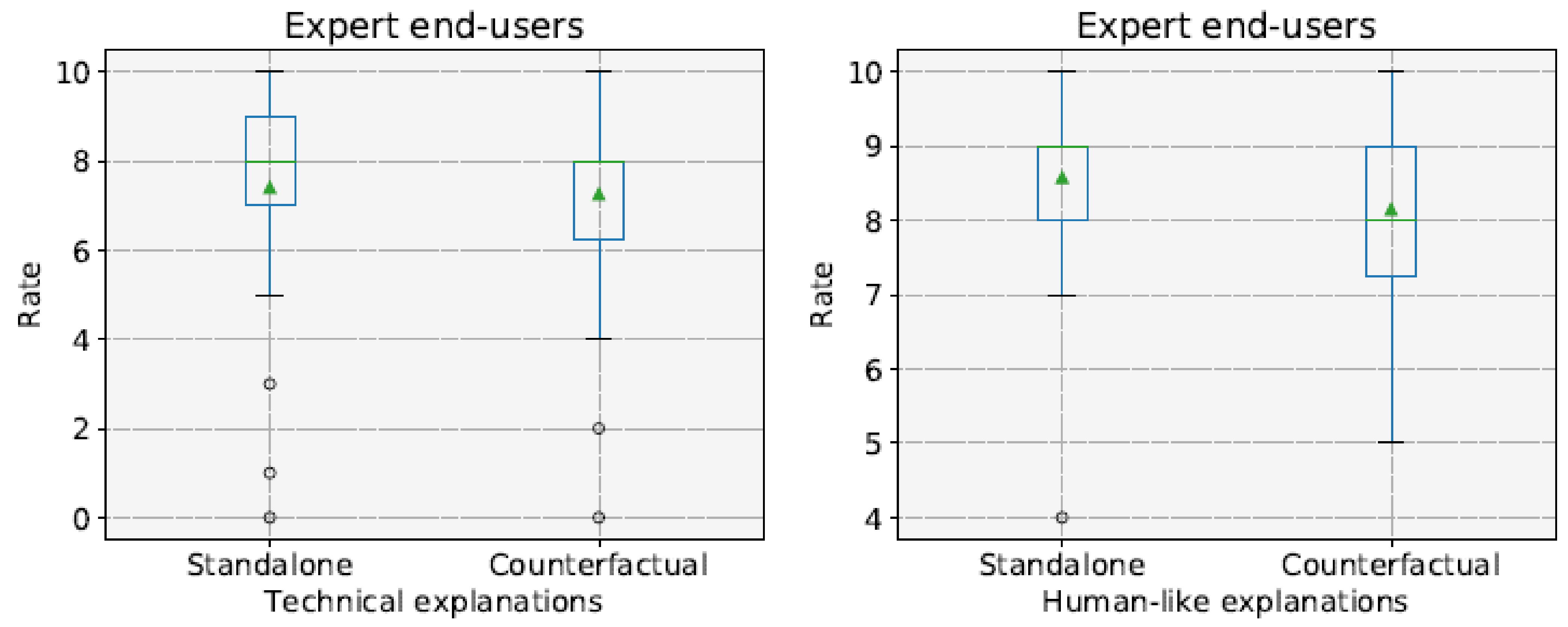
- Expert and non-expert end-users.



(a) Low ML expertise. (b) High ML expertise.



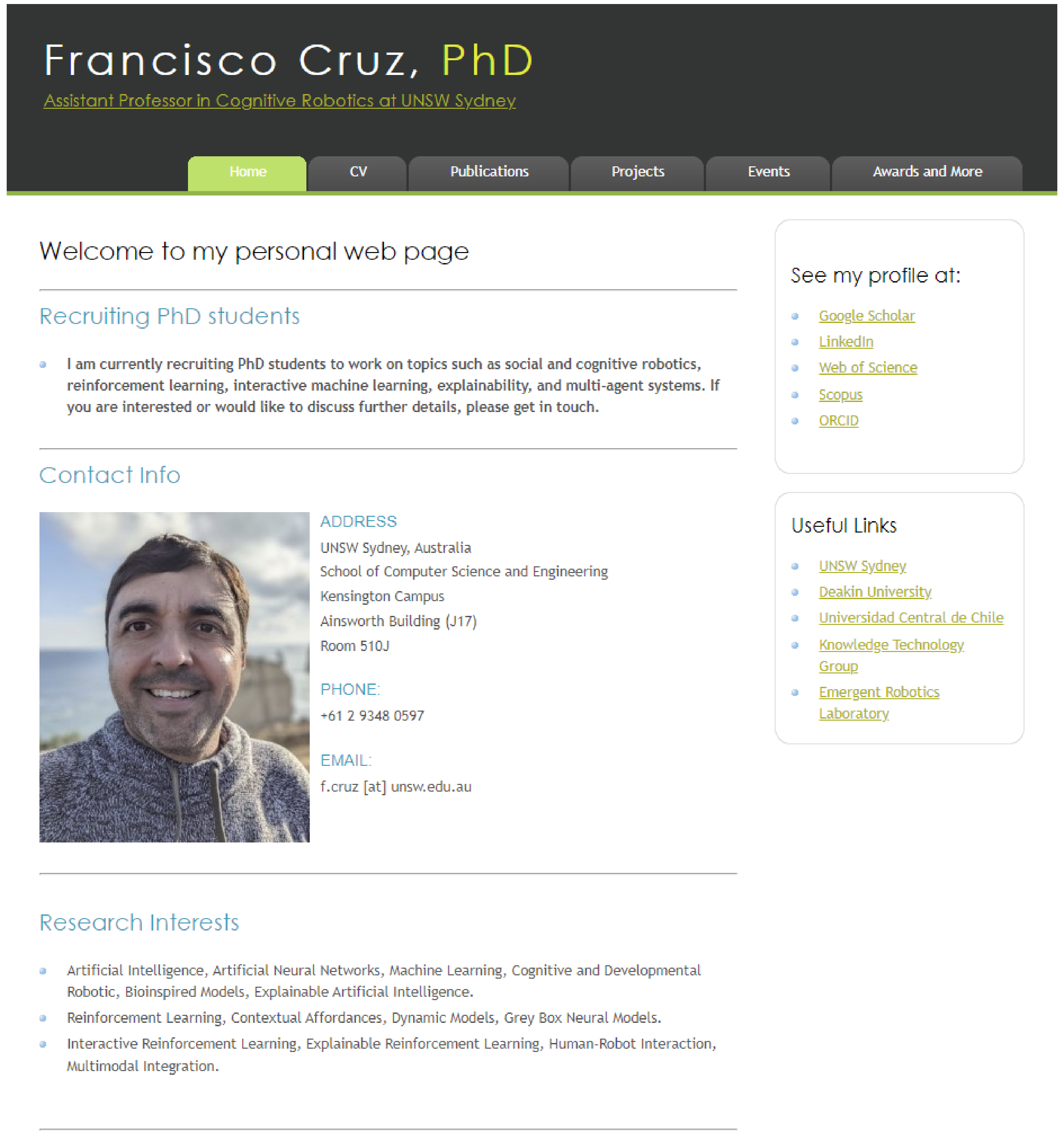
(a) Non-experts – Technical. (b) Non-experts – Human-like.



(c) Experts – Technical. (d) Experts – Human-like.

References

- Visit my personal webpage!
- In Publications Section you will find all the preprints of these papers.

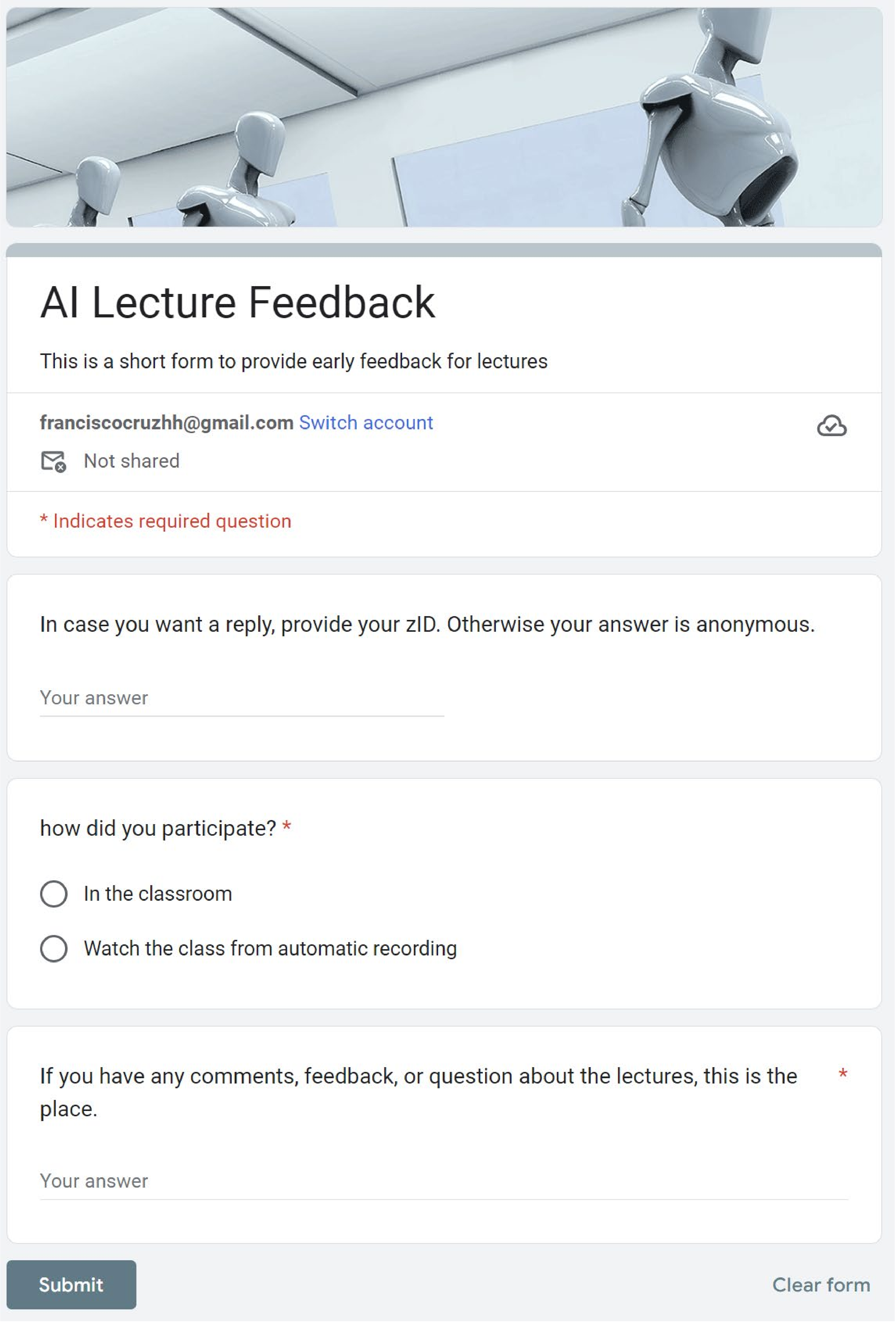


Feedback

- In case you want to provide anonymous feedback on these lectures, please visit:

- https://forms.gle/KBkN744QuffuAZLF8

Muchas gracias!



AI Lecture Feedback

This is a short form to provide early feedback for lectures

franciscocruzhh@gmail.com Switch account

Not shared

* Indicates required question

In case you want a reply, provide your zID. Otherwise your answer is anonymous.

Your answer

how did you participate? *

- In the classroom
- Watch the class from automatic recording

If you have any comments, feedback, or question about the lectures, this is the place. *

Your answer

Submit

Clear form