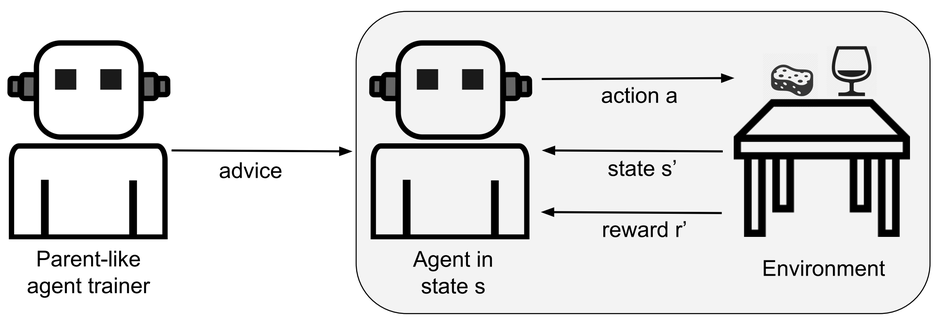
Internal knowledge representation on trainer agents by using interactive reinforcement learning

Francisco Cruz
Escuela de Computación e Informática
Universidad Central de Chile
Email: francisco.cruz@ucentral.cl

1. Motivation

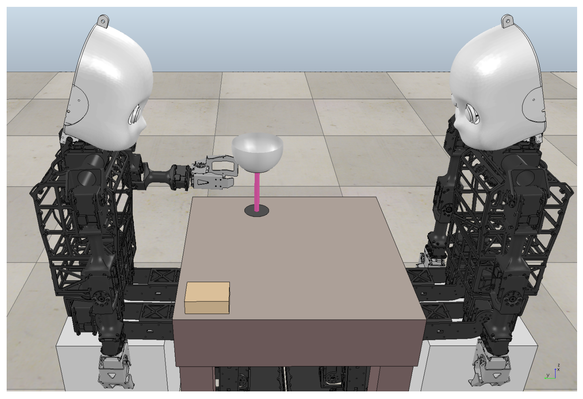
- Interactive reinforcement learning (IRL) has become an important apprenticeship approach to speed up convergence in classic reinforcement learning (RL) problems.
- We study effects of agent-agent interaction in terms of achieved learning when parent-like teachers differ in essence and when learner-agents vary in the way they incorporate the advice.



2. Robotic Scenario

- In a reinforcement learning scenario, a robot learns how to clean a table. We define *objects*, *locations*, and *actions* as follows:

Objects	Locations
sponge	left
goblet	right
	home
Actions	
go <location>	
get	drop
clean	abort



- Each robot state is represented by four variables:

$$s_t =< handPos, handObj, cupPos, sideCond[] > \quad (1)$$

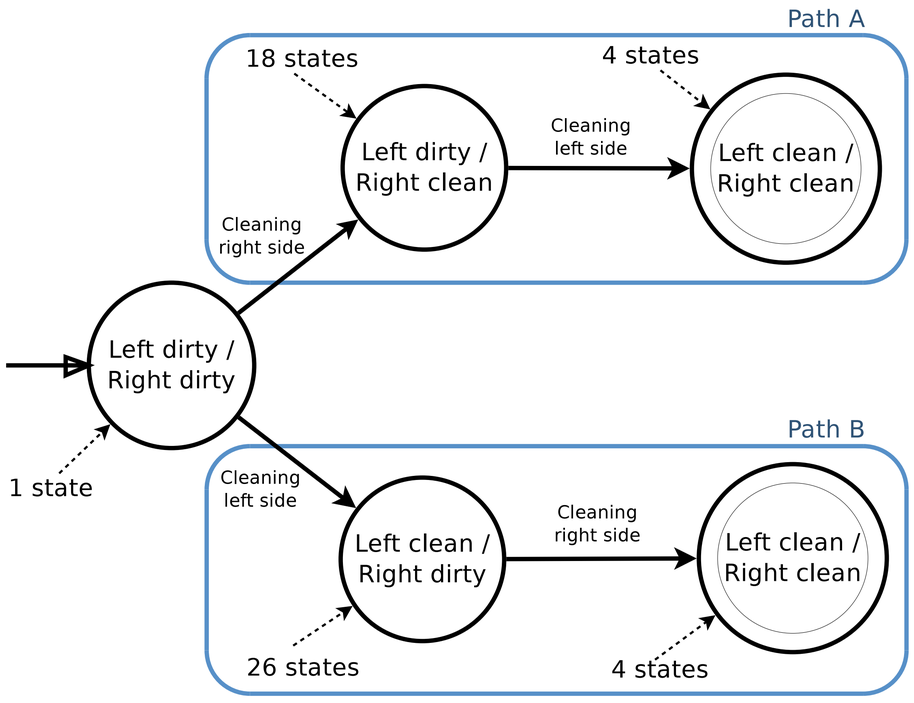
- Transition function:

Table 1. State vector transitions. After performing an action the agent reaches either a new state or a failed condition, if the latter, the agent starts another training episode from the initial state s_0.

Action	State vector update
Get	**if** handPos = *home* **&&** handObj = *cup* **then** FAILED **if** handPos = *cupPos* **&&** handObj = *sponge* **then** FAILED **if** handPos = *home* **then** handObj = *sponge* **if** handPos = cupPos **then** handObj = *cup*
Drop	**if** handPos = *home* **&&** handObj = *cup* **then** FAILED **if** handPos != *home* **&&** handObj = *sponge* **then** FAILED **otherwise** handObj = *free*
Go < pos>*	handPos = pos **if** handObj = *cup* **then** cupPos = pos
Clean	**if** handPos = cupPos **then** FAILED **if** handPos = *home* **then** FAILED **if** handObj = *sponge* **then** sideCond[handPos] = *clean*
Abort	handPos = *home* handObj = *free* cupPos = random(pos) sideCond = [dirty]*|pos|

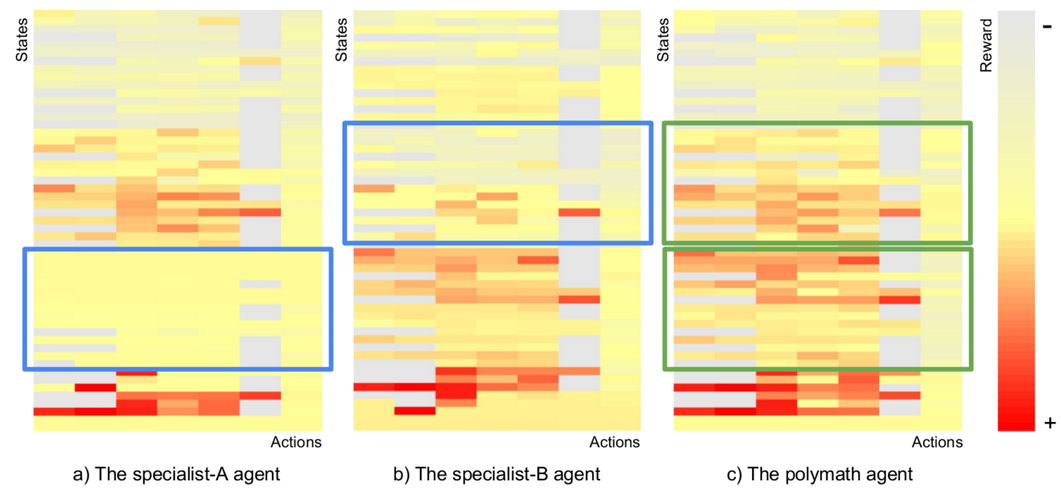
* < *pos* > may be any defined location, therefore three actions are represented by this transition, i.e. go left, go right, and go home.

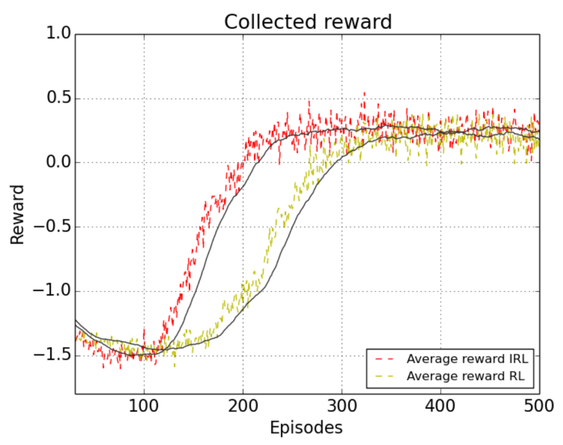
- Summarized state machine:



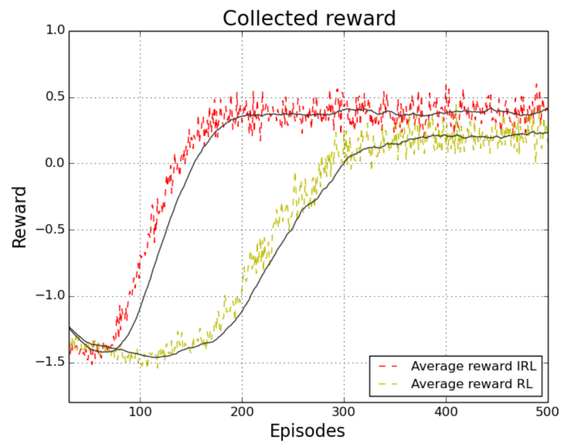
3. What Makes a Good Teacher

- Agents with diverse behaviors: specialist-A, specialist-B, and polymath agent.
- Lower standard deviation in polymath agent $T^* = \text{argmin}\ \sigma_s^i$.
- Different internal representation of Q-values.

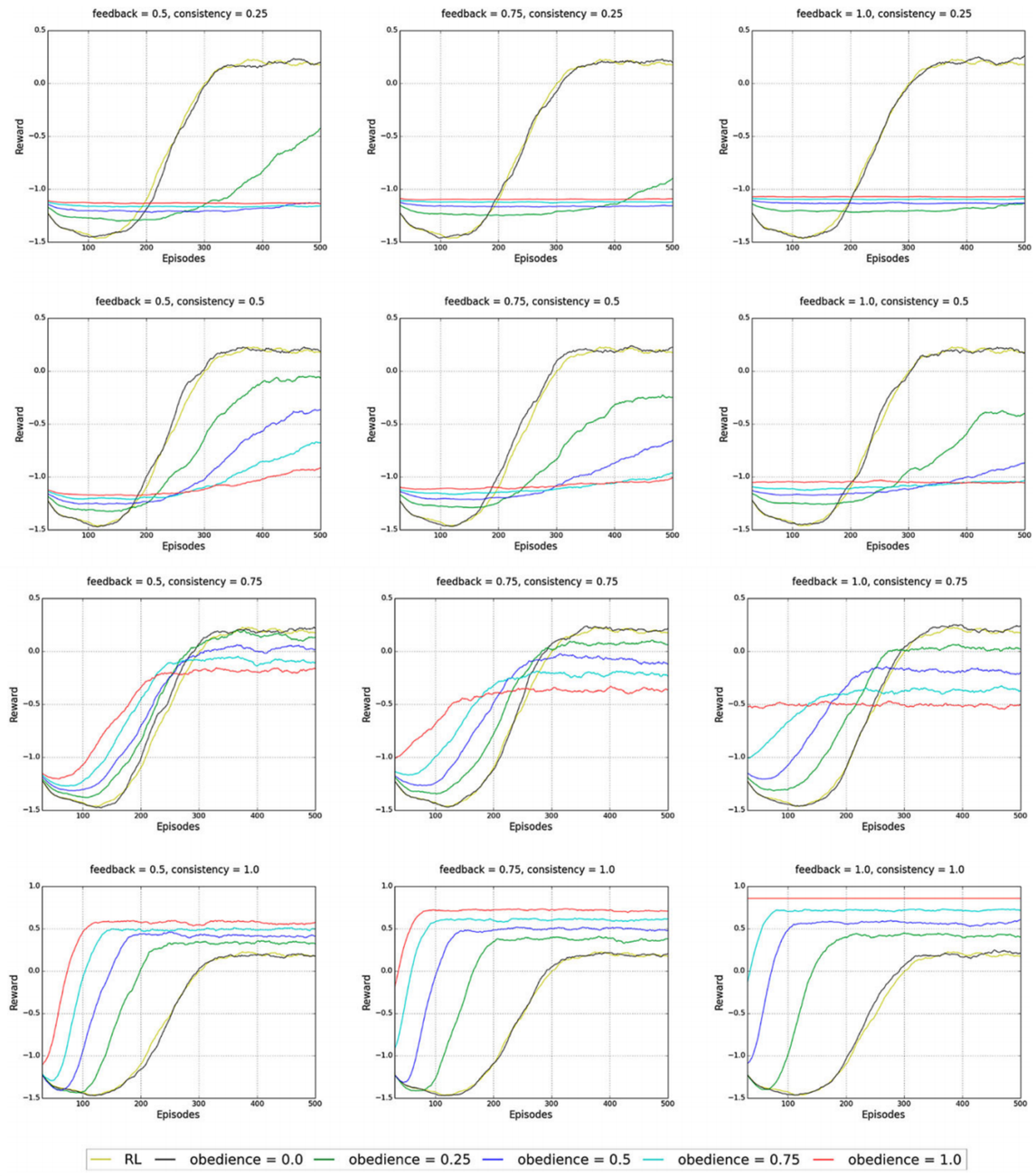




Advisor: Specialist-A agent.



Advisor: Polymath agent.



4. Conclusions

- Interactive feedback provides advantages over RL, but parent-like trainers need to give good feedback.
- Agents collecting more reward are not necessarily good trainers. Agents with better distribution of knowledge are preferred candidates.
- Polymath trainer-agent properly advises in more situations.

Reference

Francisco Cruz, Sven Magg, Yukie Nagai, and Stefan Wermter. "Improving interactive reinforcement learning: What makes a good teacher?". Journal Connection Science, Vol. 30, Nr. 3, pp. 306-325, March 2018. Open Access.