2050 追逐早上七八点钟的太阳 云栖

如何利用人类语言帮助训练人工智能

Yuhuai Wu (吴宇怀)

University of Toronto

UNIVERSITY OF TORONTO

VECTOR INSTITUTE

2018.5.27

ACTRCE: Augmenting Experience via Teachers' Advice

Harris Chan

Sanja Filder

Jimmy Ba

Challenges in Reinforcement Learning

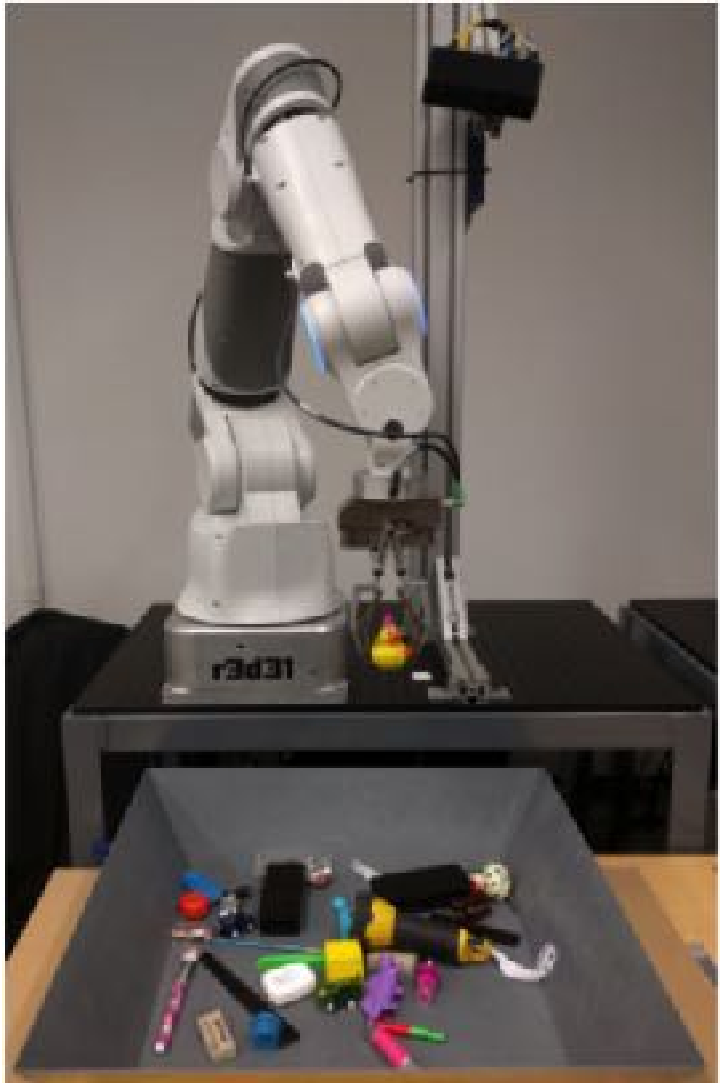
Sample efficiency

- ACKTR (actor) – Wu et al., 2017 (NIPS)

Exploration Problem

- ACTRCE (actress) – Wu et al., 2018

For example…



Sparse Rewards – “mostly nothing”

- **Sparse Reward:** reward of 1 given if the task is completed successful, otherwise 0
- **Slow/difficult to learn from**

Potential solutions?

Design a dense reward function.

e.g., Euclidean distance to the goal

However! We do not like this! Because…

What's the problem with dense reward function?

1. It will lead to biased learning (stuck in a local optimum).

What's the problem with dense reward function?

which is even dangerous!

What's the matter with dense reward function?

2. It is rather complicated and requires a significant engineering effort.

For example, a seemingly simple task of stacking Lego blocks, Popov et al. needed 5 complicated reward terms with different importance weights.

Sparse Reward function

- **Advantages:**
  - Don’t need to hand engineer the reward shaping / domain knowledge
  - Avoid biased learning

失败乃成功之母

Hindsight Experience Replay (HER)

Relabel the goal to utilize failure experience!

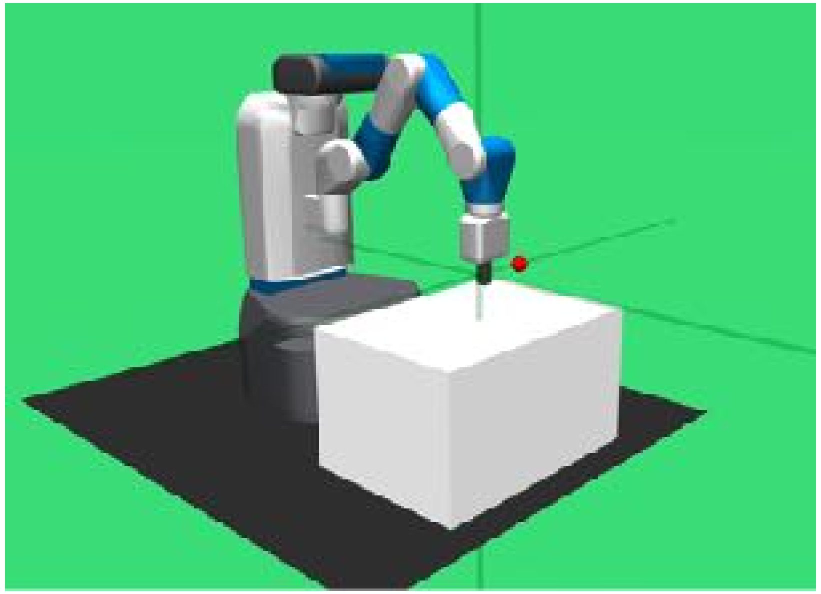
Goal-oriented MDP

A goal is chosen at every episode and stay fixed.

The policy, and the reward function depends on the current goal.

Hindsight Experience Replay (HER)

Reach object at (3,1)



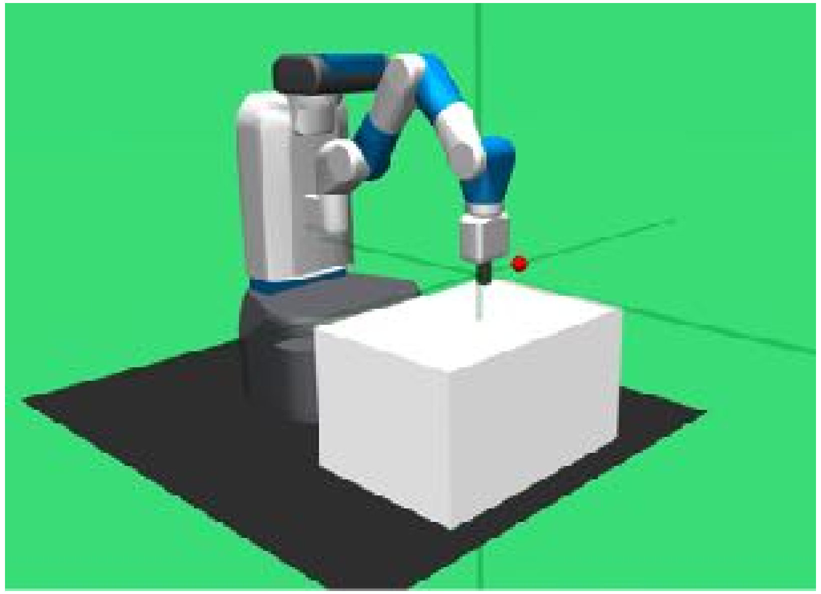
Reached (2,4)

Reward 0

Hindsight Experience Replay (HER)

Reach object at (2,4)

Reward 0

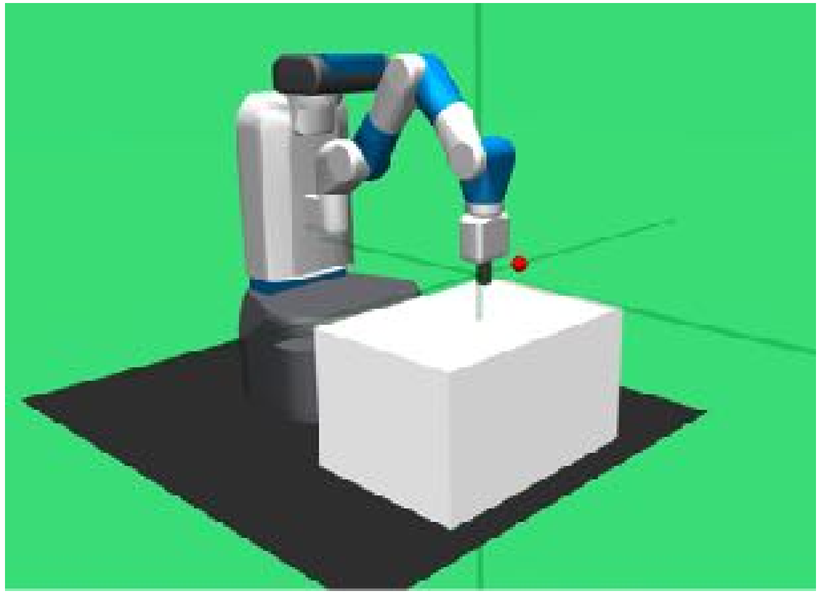


Reached (2,4)

Hindsight Experience Replay (HER)

Reach object at (2,4)

Reward 1

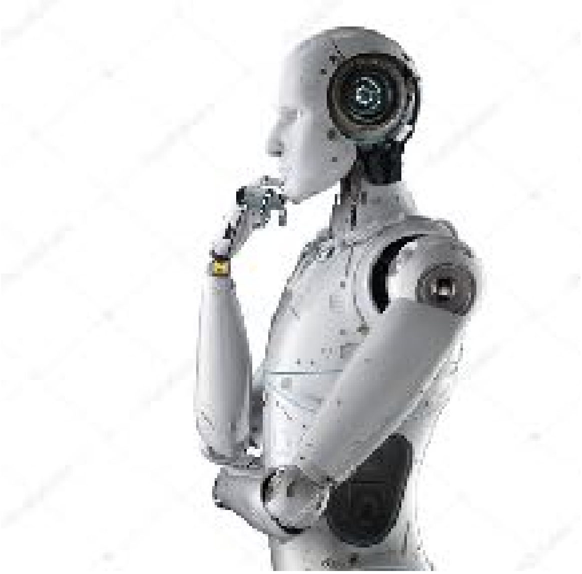


Reached (2,4)

A Crucial Assumption Behind HER

For every state, there exists a goal that is achieved in this state.

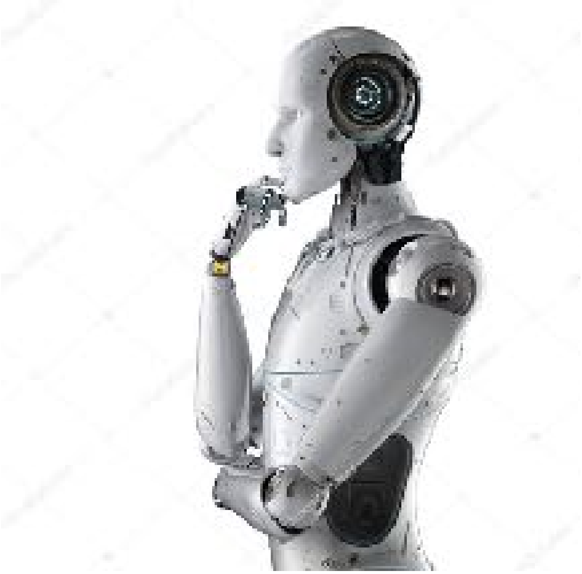
我总可以重新幻想我的目标!



A Crucial Assumption Behind HER

A trivial example: goal space = state space

到哪儿就算哪儿是目标!



A Crucial Assumption Behind HER

Such goal representation will create a lot of redundancy in general.

For example, all the following can be thought of representing the same goal:

Driving straight; Avoiding colliding

Goal 1

Goal 2

Goal 3

Question: How do we represent the goal in general?

Question: What’s a good representation?

Question: What's a good representation?

1. Universal
2. Compact & abstract.

Using language as goal representation!

Two important attributes of language:

1. Universal
2. Compact & abstract.

就是它了!

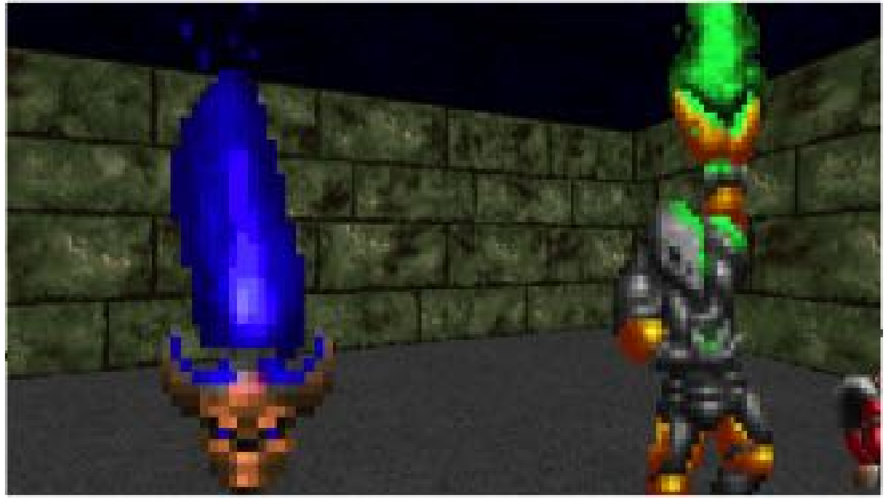
ACTRCE!

- Combining HER framework with language representation.
- Demonstrating two great attributes of language.

ACTRCE!

Reach the armor!

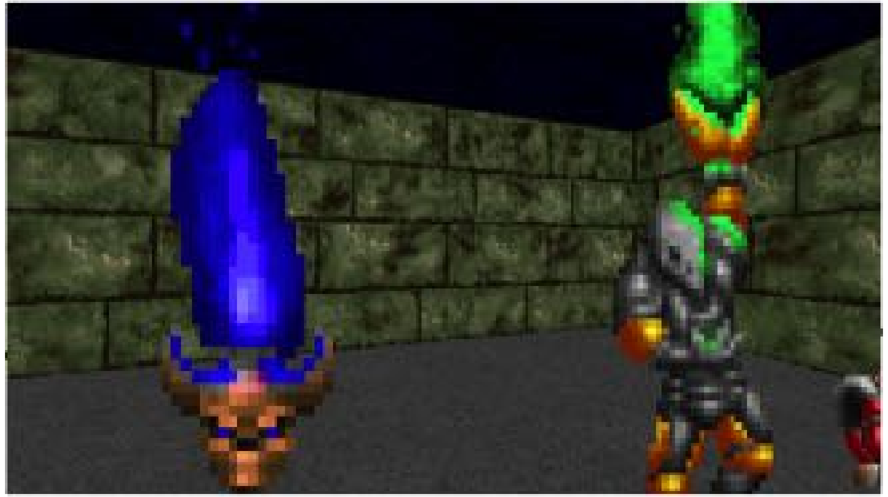
Reward 0



Reached the blue torch

ACTRCE!

Reach the blue torch!

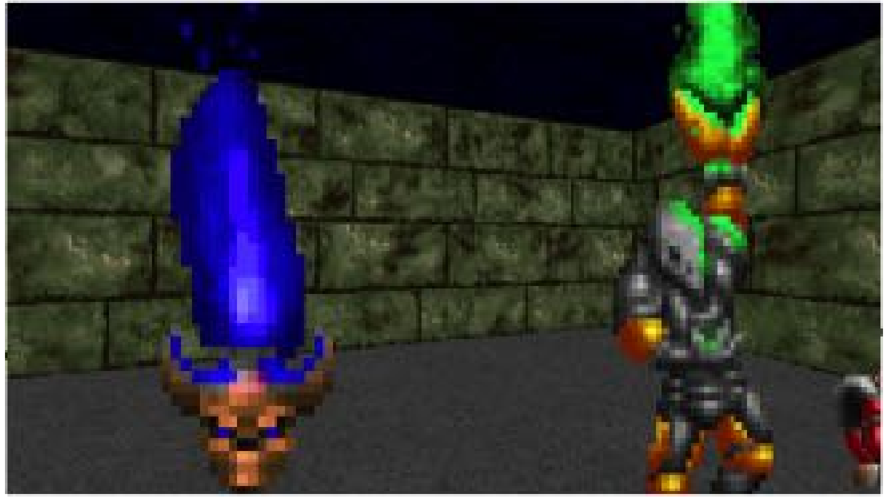


Reached the blue torch

Reward 0

ACTRCE!

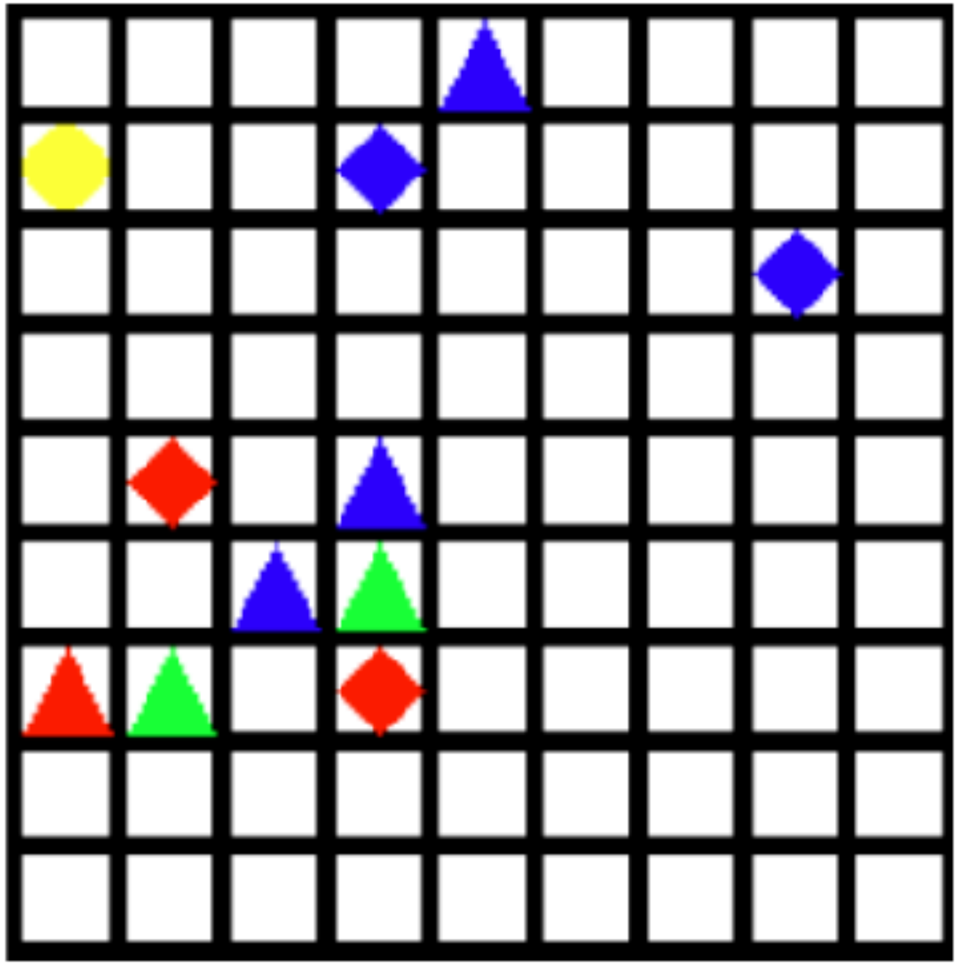
Reach the blue torch!



Reached the blue torch

Reward 1

KrazyGrid World 2D env



Triangles: treasure

Squares: lavas

KrazyGrid World 2D env

Functionality: Goal, Lava, Normal, and Agent.
Colour attribute: Red, Blue, Green.

Desired goal： Reach _ treasure.

Other goals: Reach _ lava. Avoid any goal. Avoid any lava.

Optimistic Teacher

When a desired goal is achieved, I'll describe what has been achieved as advice to the agent.

Knowledgeable Teacher

I'll always describe what has been achieved as advice to the agent.

Discouraging Teacher

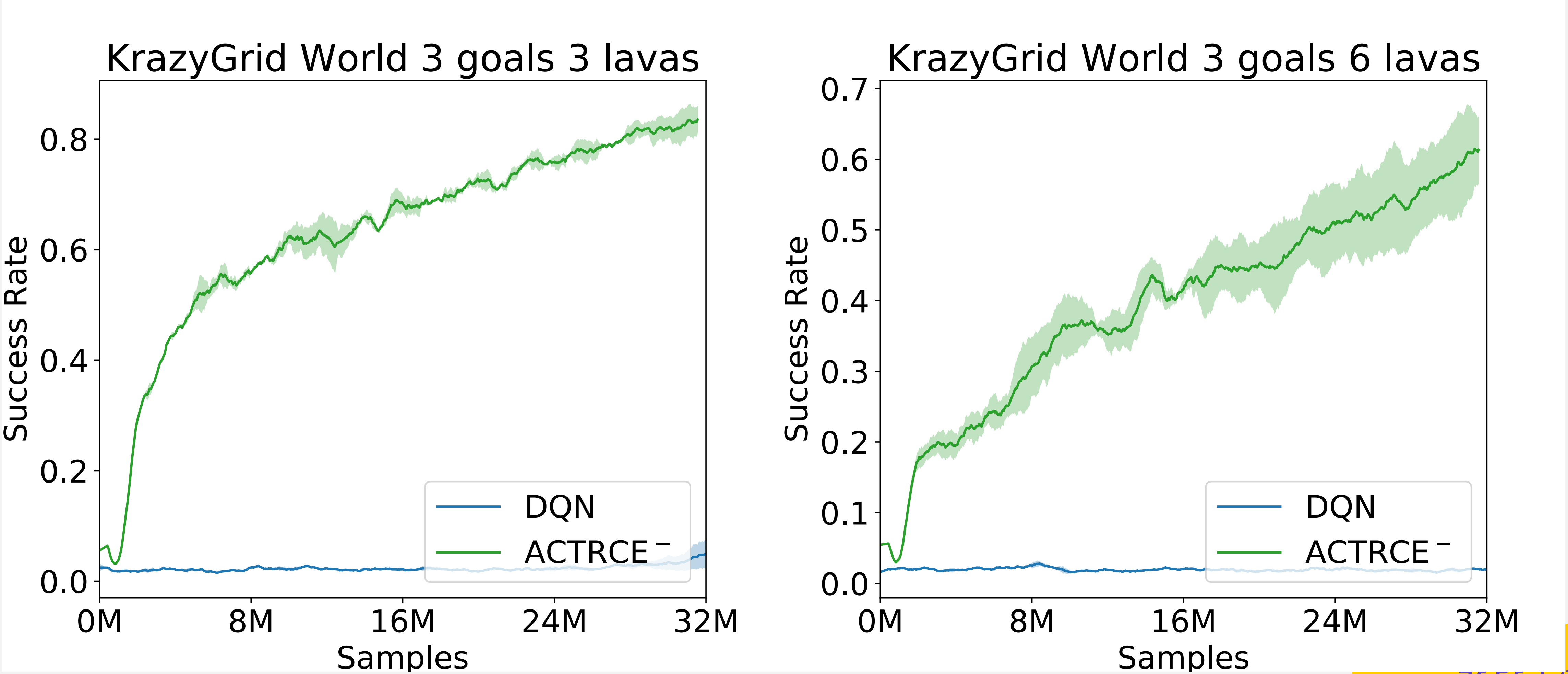
I’ll describe an unachieved desired goal as advice to the agent.

2050大会

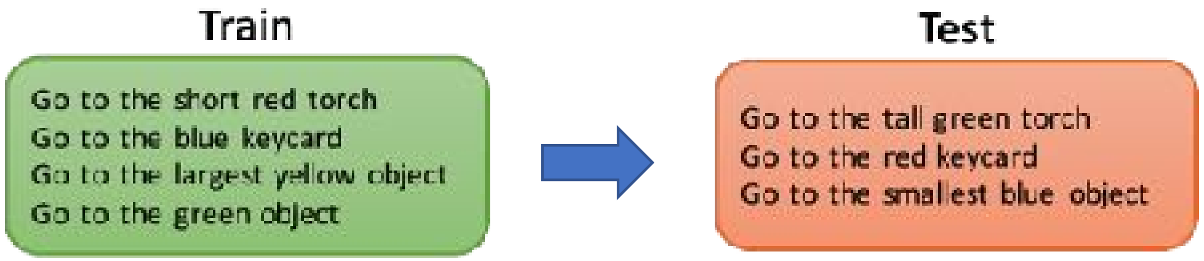
Comparison to baseline

ACTRCE- : Optimistic teachers + Discouraging teachers

KrazyGrid World Results

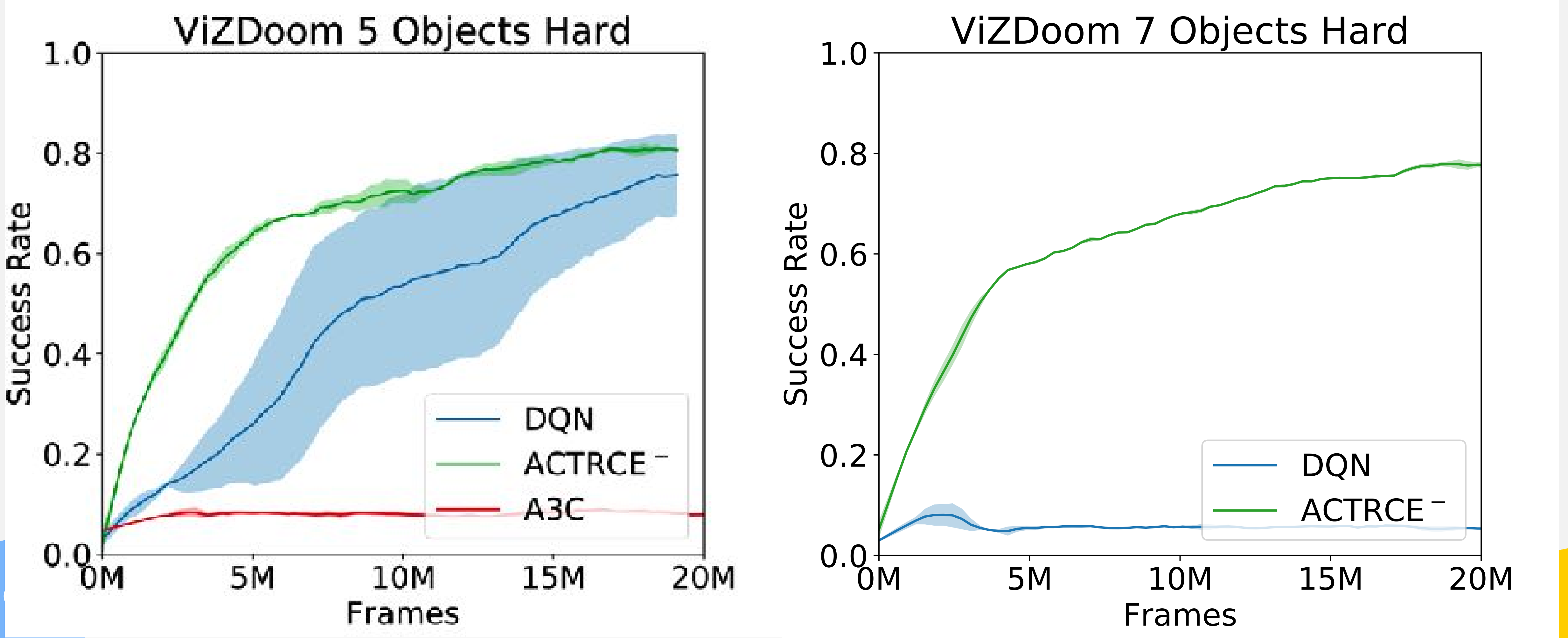


Doom 3D language environment (Chaplot et al., 2017)



State: 3 x 300 x 168 RGB Image
Action: [TurnLeft, TurnRight, MoveForward]
Reward: 1.0 if correct object, -0.2 for incorrect, 0.0 otherwise
Training Instructions: 55 instructions
Testing Instructions: 15 instructions

Doom Results



Doom visualization

Language is abstract:

——allowing generalization

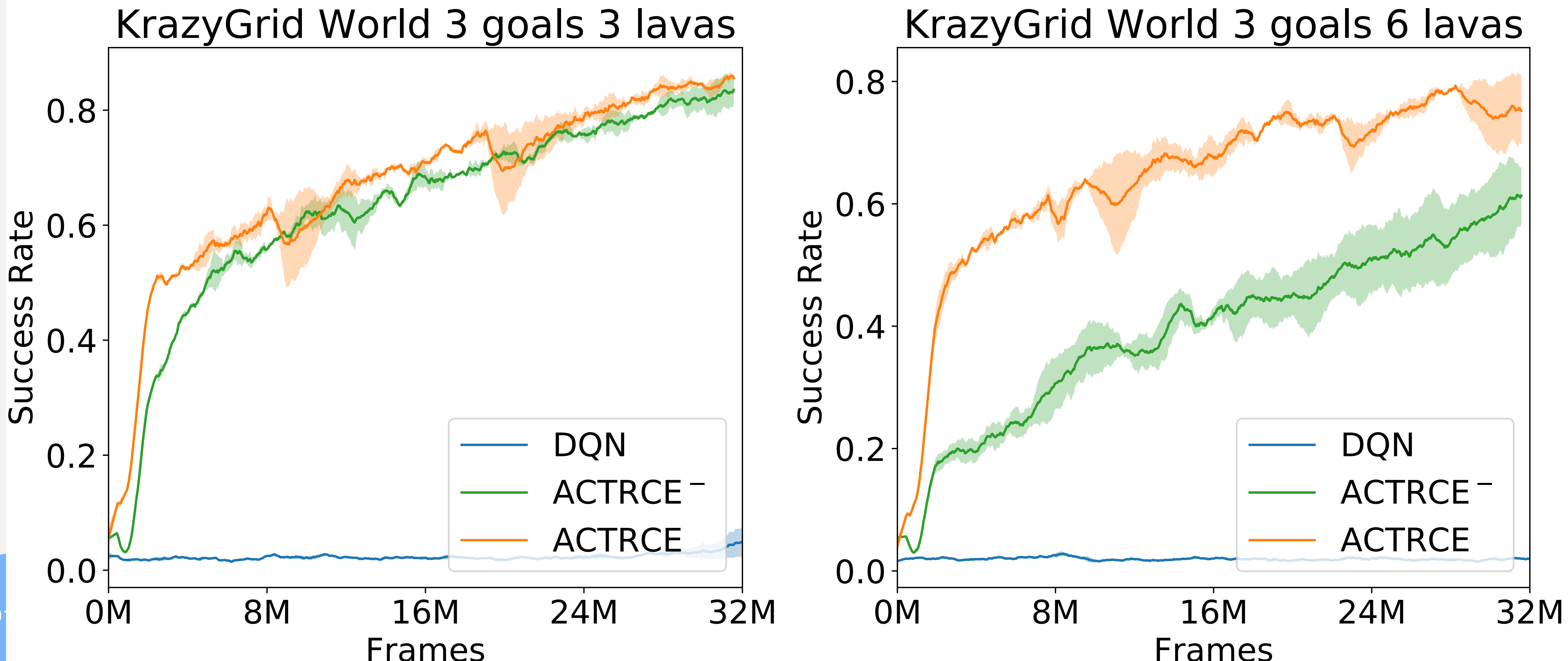
Observation: More language helps!

Increasing the language set

Option 1: use Knowledgeable Teachers.

ACTRCE : Knowledgeable teachers +
Discouraging teachers

ACTRCE vs ACTRCE-



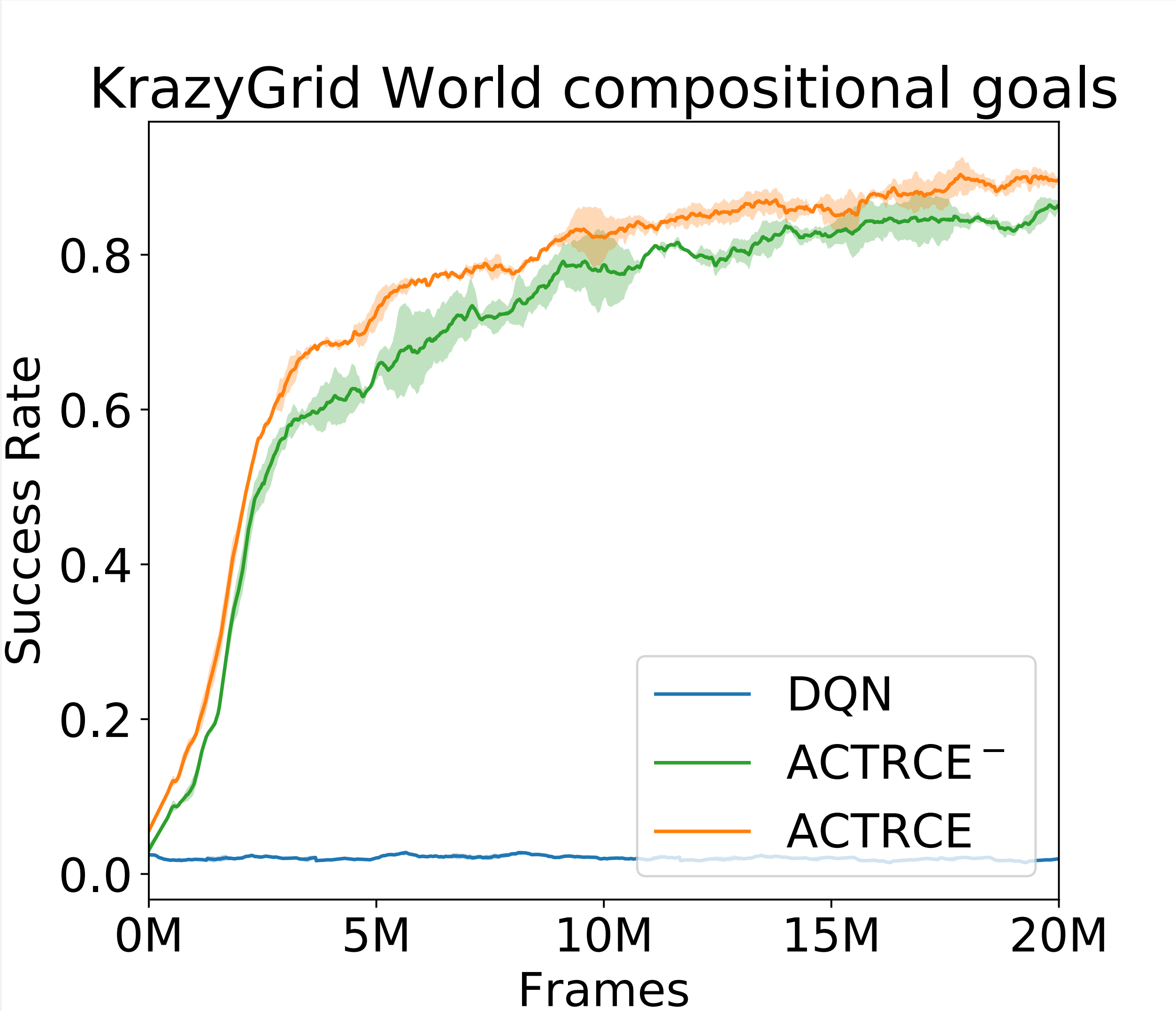
Increasing the language set

Option 2: Increasing goal space
by considering compositions of tasks.

Desired goal: Reach _ treasure and/or Reach _ treasure

Other goals: Reach _ lava and/or Reach _ lava

Compositional tasks

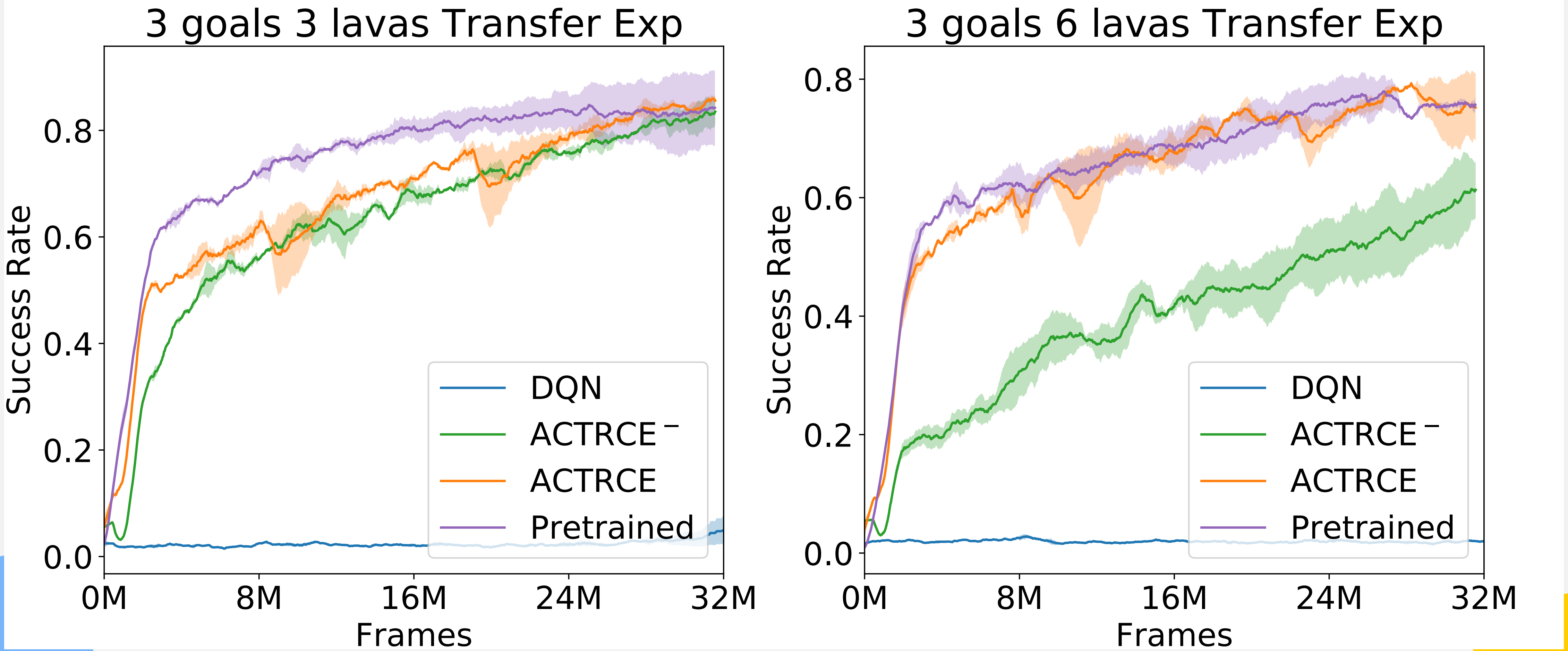


Why more language helps?
— Transfer learning!

Pessimistic teacher: Only gives advice when an undesired goal is achieved.

Validating experiment: Pretrain with pessimistic teacher. Train with ACTRCE-. Compare.

Transfer learning works!



Concluding Remarks

It is very difficult to build a high-fidelity simulated environment — not in the near future.

However, there is a beautiful world inside language corpus! — Great resources for world representation.

2050

追逐早上七八点钟的太阳

云栖

THANKS