

[RL22] Q&A Session on Exploration in Deep RL

21.07.2022

Christopher Mutschler

Course Information

Some remarks on the exam

- The exam will take place on **August, 10th** (14:00 – 15:30 ← this is still preliminary!)
- Preliminary room: H11 (Felix-Klein-Building, Cauerstraße 11)
- Please take with you your student ID card
- No material allowed during the exam (only pen, we will provide paper)
- Next week we can do a Q&A on the exam – please prepare yourself if needed
 - we will not provide a mock exam or example questions

Some remarks on the evaluation

- Please do give us feedback via the course evaluation – we will discuss the results next week!

In-Presence meeting next week

- We will meet in H13 (Radon), Felix-Klein-Building (Cauerstraße 11)

Let's play Kahoot!

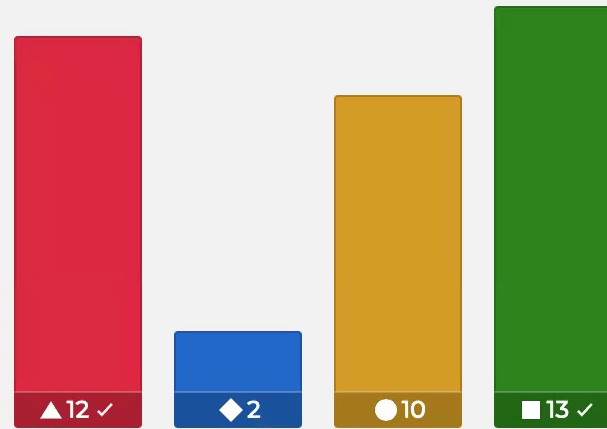
Kahoot!

Let's play Kahoot!

The image shows a screenshot of the Kahoot! website homepage. The browser address bar shows 'kahoot.com'. The navigation menu includes 'Kahoot!', 'News', 'School', 'Work', 'Home', 'Study', 'Academy', 'AccessPass', 'Contact sales', 'Explore content', 'Play', 'Sign up', 'Log in', and 'EN'. The 'Play' button is circled in red. Below the navigation are four promotional banners:

- Make learning awesome!**
Kahoot! delivers engaging learning to billions.
[Sign up for free!](#)
- Make your team superstar presenters**
Set your whole team up to deliver awesome presentations with Kahoot! 360 Spirit, our best plan from only \$16 per month.
[Learn more >](#)
[Buy now](#)
- NEW! Create a branded experience with Kahoot! themes**
Boost audience engagement by customizing your kahoots for your work setting.
[Choose Kahoot! 360 Pro Max](#)
- Meet Kahoot! Kids!**
Spark your child's curiosity for learning with our new playful app experience.
[Get started today](#)

What makes exploration hard?



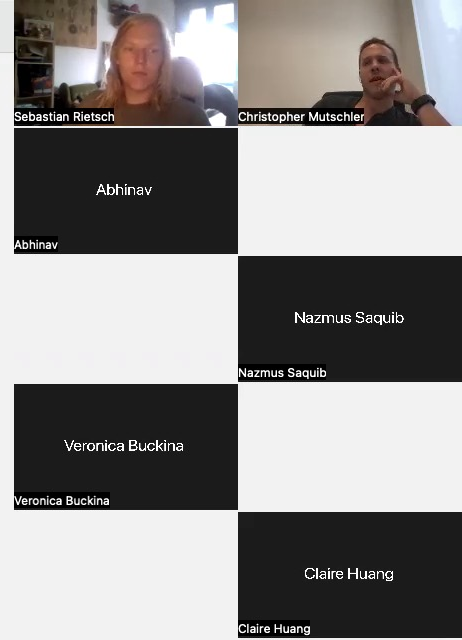
Show media

▲ Sparse rewards ✓

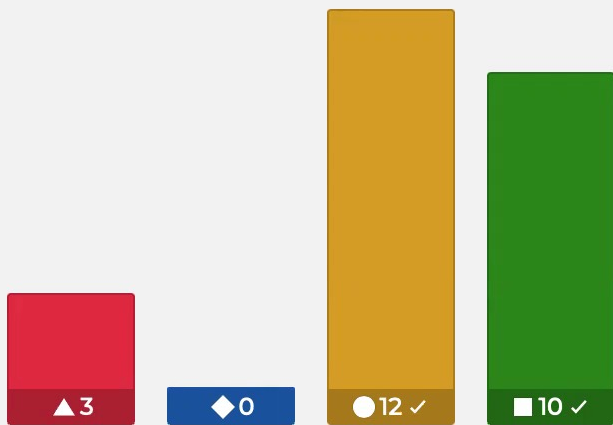
◆ Value function approximation ✗

● Uncontrollable dynamics ✗

■ Stochasticity in the environment ✓



The intrinsic reward



Show media

▲ motivates the agent to maximize rewards received from the environment ✕

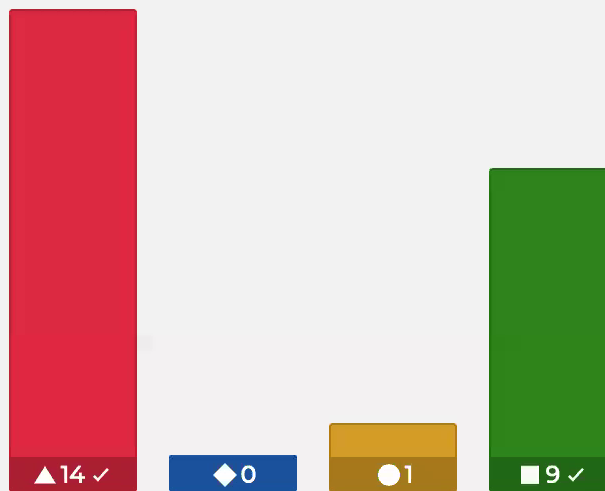
◆ is a measure of the agent's computational power ✕

● motivates the agent to discover novel states ✓

■ helps to improve the agent's knowledge about the environment ✓

Sebastian Rietsch
Christopher Mutschler
Abhinav
Nazmus Saquib
Veronica Buckina
Claire Huang

What are environments with hard exploration?



Show media

▲ Montezuma's Revenge ✓

◆ Cartpole ✗

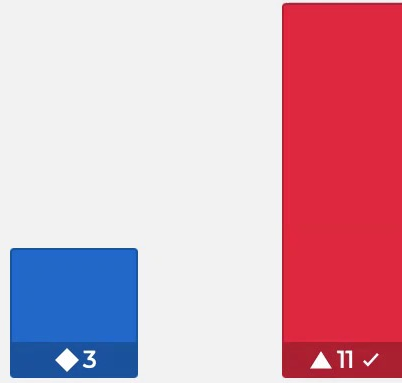
● Pong ✗

■ Seaquest ✓

Sebastian Rietsch
Christopher Mutschler
Abhinav
Nazmus Saquib
Veronica Buckina
Claire Huang

Forward Dynamics Prediction approximates the environment dynamics. High errors tu
intrinsic rewards.

Sebastian Rietsch
Christopher Mutschler
Abhinav
Nazmus Saquib
Veronica Buckina
Claire Huang

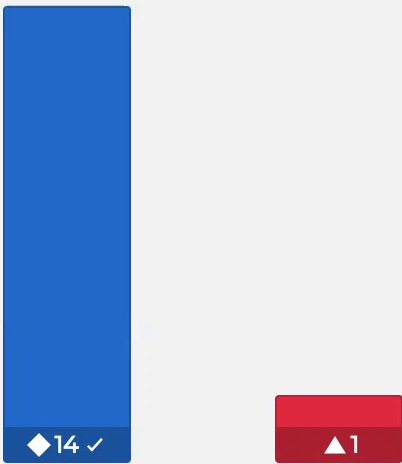


Show media

◆ True ✕

▲ False ✓

The intrinsic reward usually decreases over the course of training



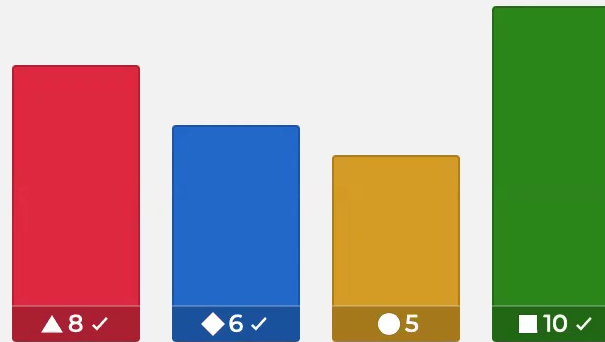
Show media

◆ True ✓

▲ False ✗

A grid of video call windows. The top row shows two windows with video feeds and names: Sebastian Rietsch and Christopher Mutschler. Below them are four windows with names: Abhinav, Nazmus Saquib, Veronica Buckina, and Claire Huang. The bottom row shows two windows with names: Abhinav and Claire Huang.

Intrinsic Curiosity Module: what elements does it learn?



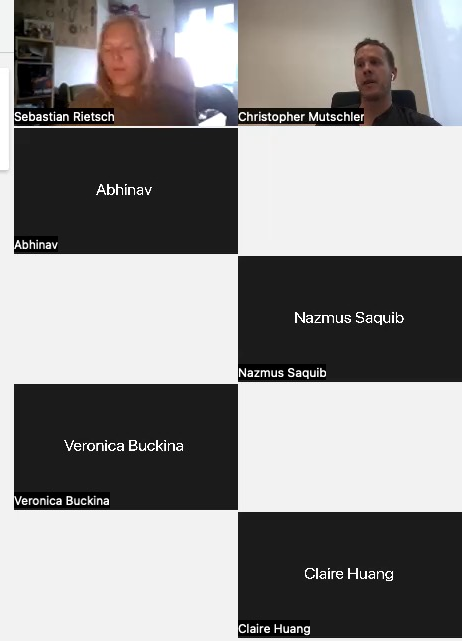
Show media

▲ forward dynamics model ✓

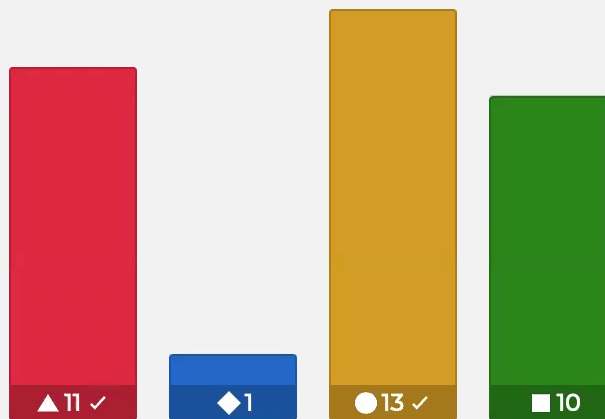
◆ an efficient feature representation of the states ✓

● an efficient exploration policy ✗

■ inverse dynamics model ✓



Limitations of forward models: what are **(unwanted)** reasons for prediction errors (although we have the data!)



Show media

▲ Model capacity too low ✓

◆ Dynamics are linear ✗

● Environment is too stochastic ✓

■ Failure of generalizing from previously seen examples ✗

Sebastian Rietsch
Christopher Mutschler
Abhinav
Nazmus Saquib
Veronica Buckina
Claire Huang