

# [RL22] Q&A Session on Model-based RL #2

---

07.07.2022

Christopher Mutschler

Let's play Kahoot! again...

**Kahoot!**

# Let's play Kahoot!

The image shows a screenshot of the Kahoot! website homepage. At the top, there is a navigation bar with the Kahoot! logo on the left and several menu items: News (with a red notification badge), School, Work, Home, Study, Academy, and AccessPass. On the right side of the navigation bar, there are buttons for 'Contact sales', 'Explore content', 'Play' (circled in red), 'Sign up', 'Log in', and a language selector 'EN'. Below the navigation bar, there are four main promotional cards. The first card on the top left is purple and white, titled 'Make learning awesome!' and features a photo of two young women laughing. The second card on the top right is blue and white, titled 'Make your team superstar presenters' and features a photo of a meeting with a video call overlay. The third card on the bottom left is white and blue, titled 'NEW! Create a branded experience with Kahoot! themes' and features a photo of a Kahoot! game interface with a score of 55427 and names like Rik, Mal, Mille, and Martin. The fourth card on the bottom right is purple and blue, titled 'Meet Kahoot! Kids!' and features a colorful illustration of a rocket and stars. Each card has a green call-to-action button.

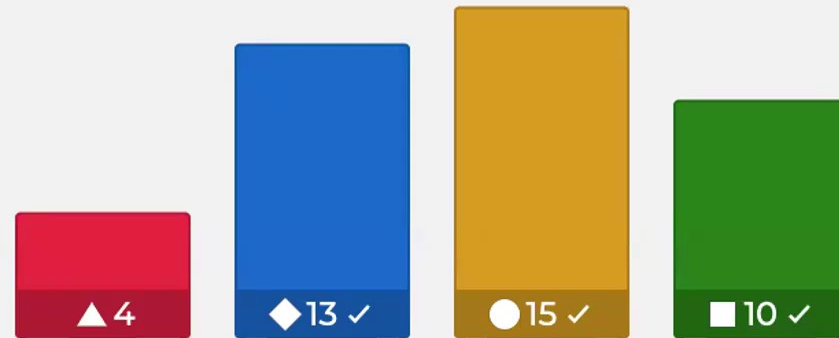
**Make learning awesome!**  
Kahoot! delivers engaging learning to billions.  
[Sign up for free!](#)

**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](#)

# Trajectory Optimization with Derivatives



Show media

▲ LSTMs help with the poor conditioning induced by the recursive gradients



◆ Samples the initial action sequence randomly



● Optimizes the action sequence via gradient descent



■ Collocation makes downstream state changes less sensitive to early changes



Christopher Mutschler

Sebastian Rietsch

Nicolas Kolbenschlag

Nicolas Kolbenschlag

Veronica Buckina

Veronica Buckina

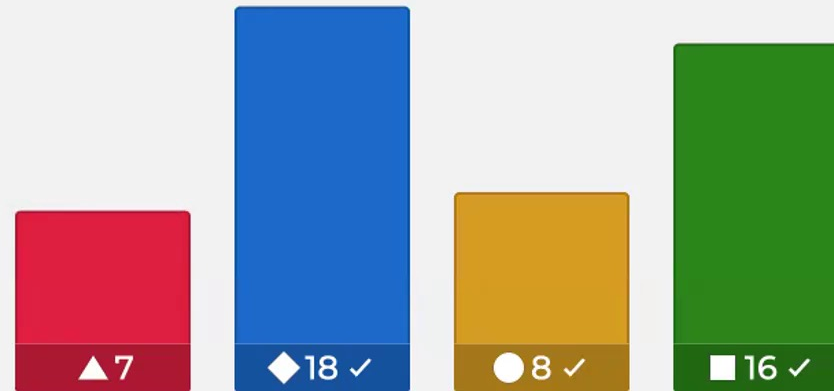
FAU084414

FAU084414

Takuma Nishimura

Takuma Nishimura

# Cross Entropy Maximization



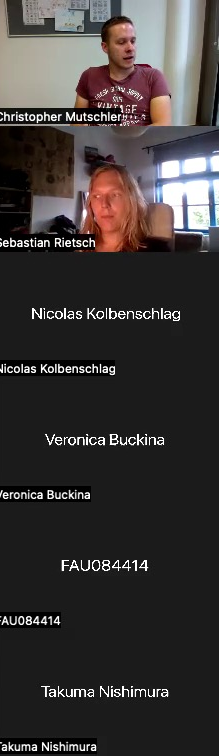
Show media

▲ Uses gradients to optimize a number of action sequences ❌

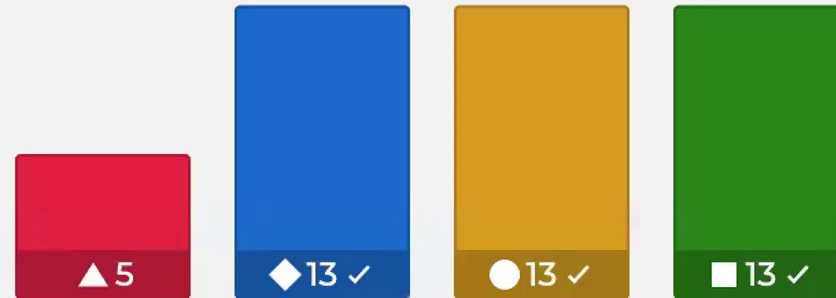
◆ Fits a Gaussian to sample policy parameters at every iteration ✓

● Is able to find the global optimum ✓

■ Is easy to implement and embarrassingly easy to parallelize ✓



# Uncertainty of models in RL



Show media

▲ Aleatoric uncertainty can be reduced close to 0 by training with more data ✘

◆ Epidemic uncertainty often leads to extrapolation errors ✔

● Epidemic uncertainty changes over the course of learning ✔

■ Aleatoric uncertainty captures the world's inherent stochasticity ✔

A vertical video call interface on the right side of the screen. It shows a list of participants with their names and small video thumbnails. The participants listed are Christopher Mutschler, Sebastian Rietsch, Nicolas Kolbenschiag, Veronica Buckina, FAU084414, and Takuma Nishimura.



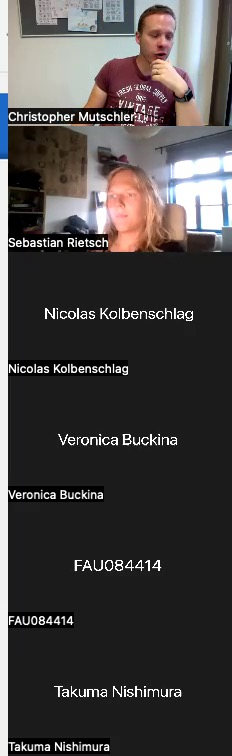
**Model Predictive Control applies a sequence of actions at once (with a maximum length of the planning horizon)**



Show media

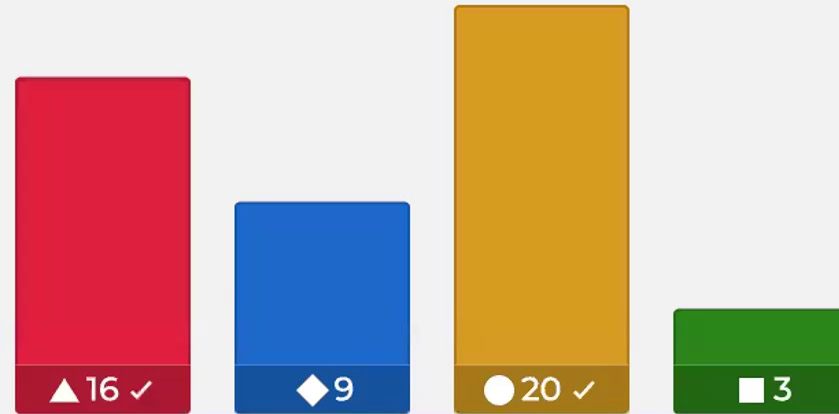
◆ True ✕

▲ False ✓





# Uncertainty Estimation: what models can we use?



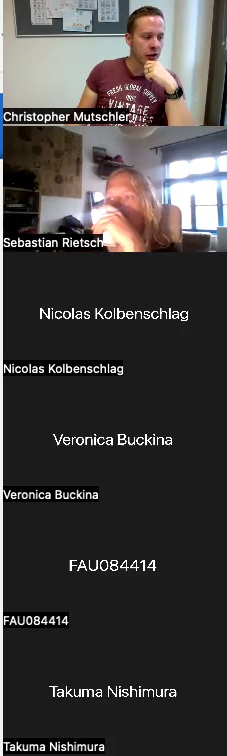
Show media

▲ Model Ensembles ✓

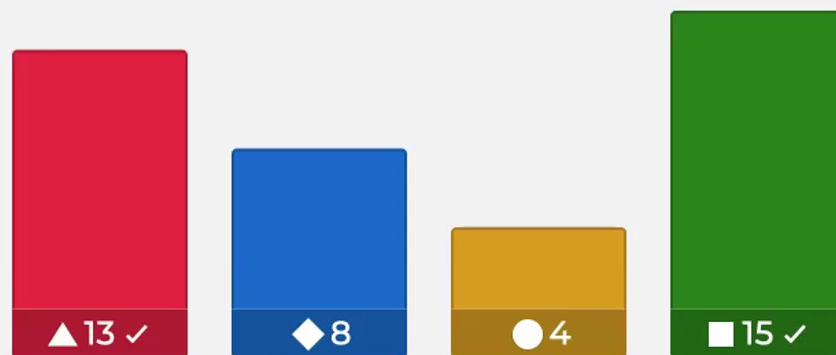
◆ Recurrent Neural Networks ✗

● Gaussian Processes ✓

■ Support Vector Machines ✗



# Uncertainty Propagation: what methods can we use?



Show media

▲ Moment Matching ✓

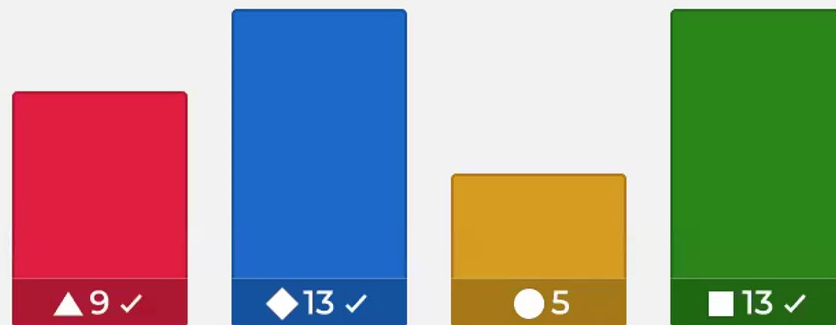
◆ Stochastic Gradient Descent ✗

● Time-Difference Learning ✗

■ Trajectory Sampling ✓

Christopher Mutschler  
Sebastian Rietsch  
Nicolas Kolbenschiag  
Veronica Buckina  
FAU084414  
Takuma Nishimura

# PILCO



Show media

▲ Is robust to small model errors ✓

◆ Is very sample efficient ✓

● Does not require smooth dynamics ✗

■ Is computationally expensive w.r.t. the available training data ✓

Christopher Mutschler

Sebastian Rietsch

Nicolas Kolbenschlag

Nicolas Kolbenschlag

Veronica Buckina

Veronica Buckina

FAU084414

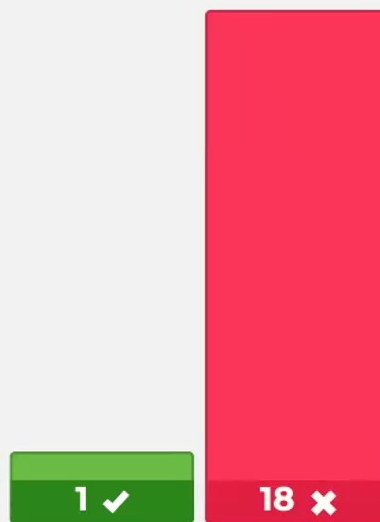
FAU084414

Takuma Nishimura

Takuma Nishimura



# MBRL: Sample Efficiency (top: requires many samples, bottom: requires little samples)



Show media

- Augmented Random Search ✓
- ◆ Policy-based RL ✓
- ▲ Value-based RL ✓
- Model-based RL ✓

Christopher Mutschler

Sebastian Rietsch

Nicolas Kolbenschiag

Nicolas Kolbenschiag

Veronica Buckina

Veronica Buckina

FAU084414

FAU084414

Takuma Nishimura

Takuma Nishimura