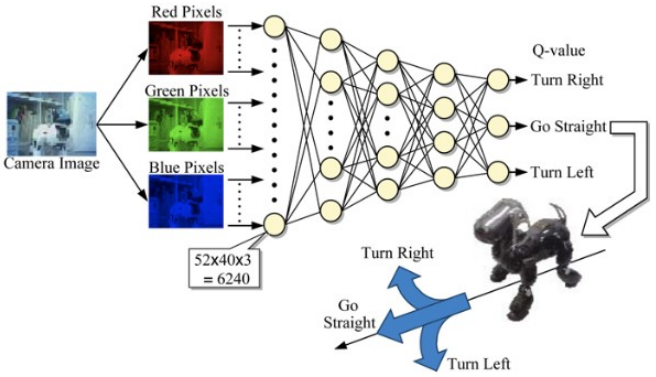
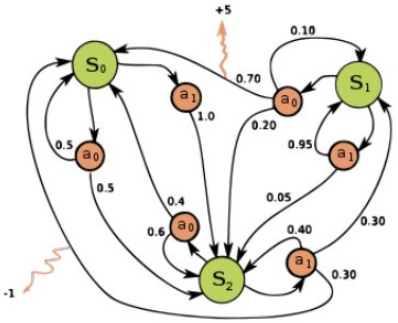
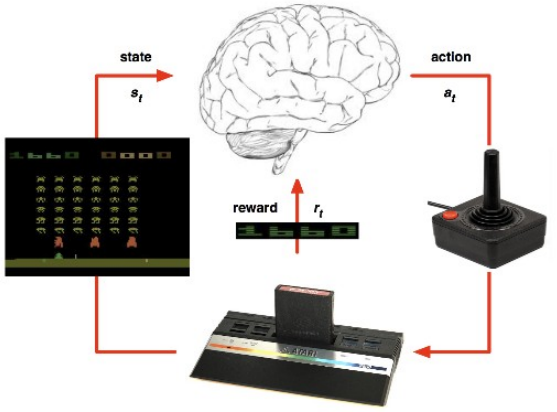
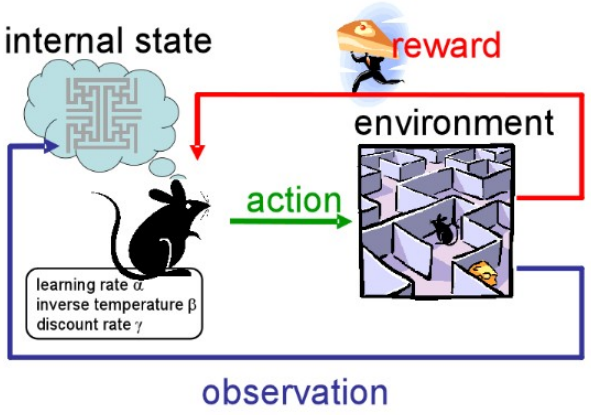
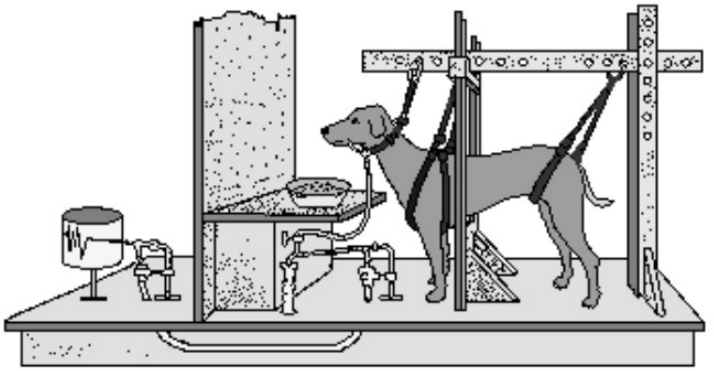
COMP 138: Reinforcement Learning





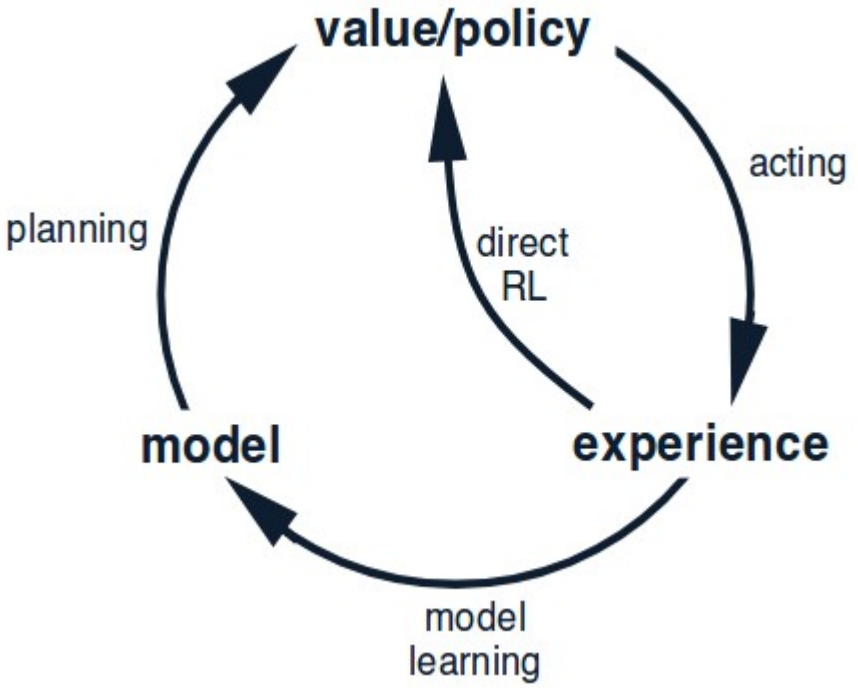






Instructor: Jivko Sinapov

Today


value/policy
acting
planning
direct
RL
model
experience
model
learning

Announcements

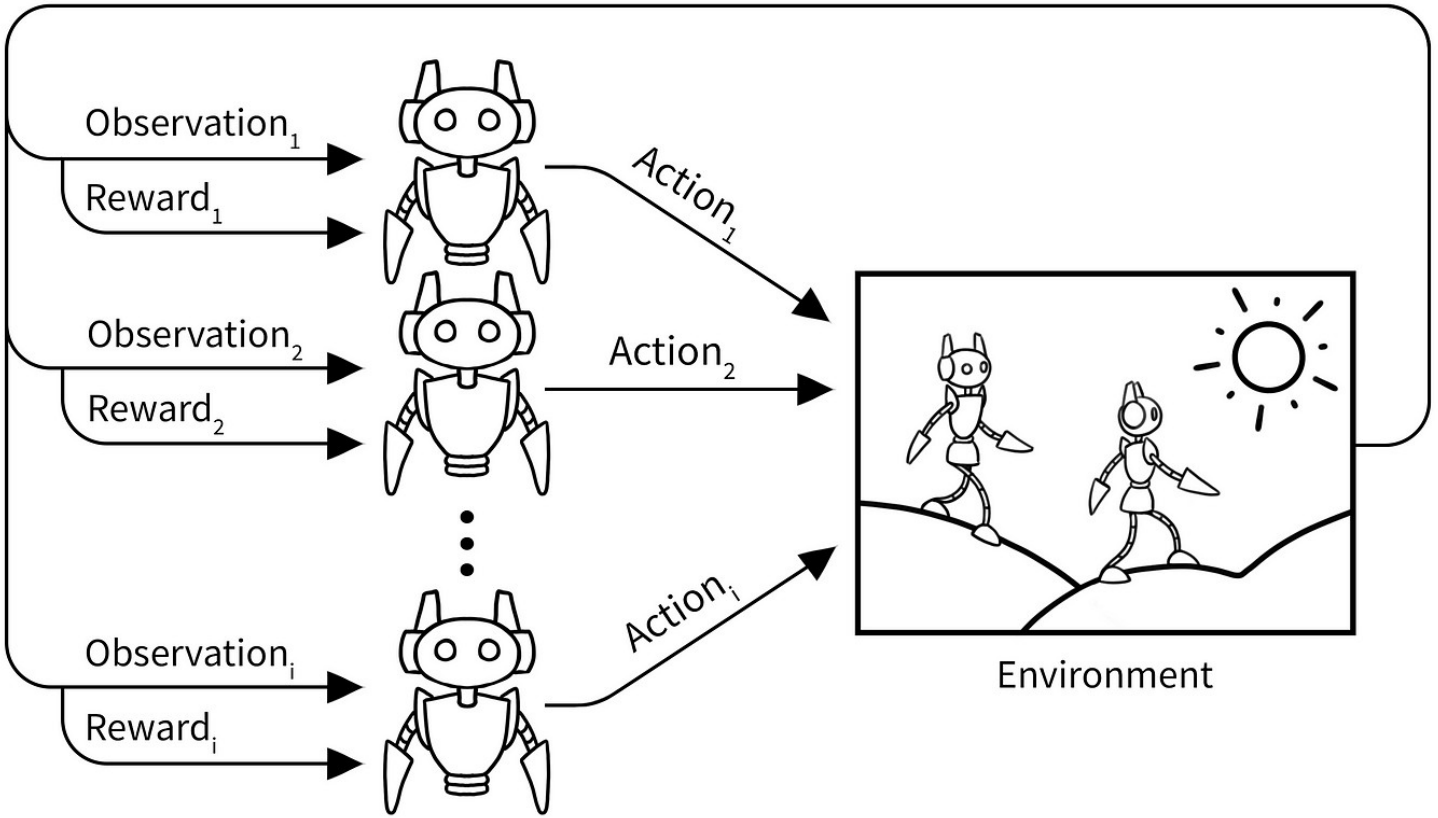
Upcoming Project Due Dates

- Team Formation – Oct 17
- Project Proposal – Oct 31st

Policy Shaping with Human Feedback

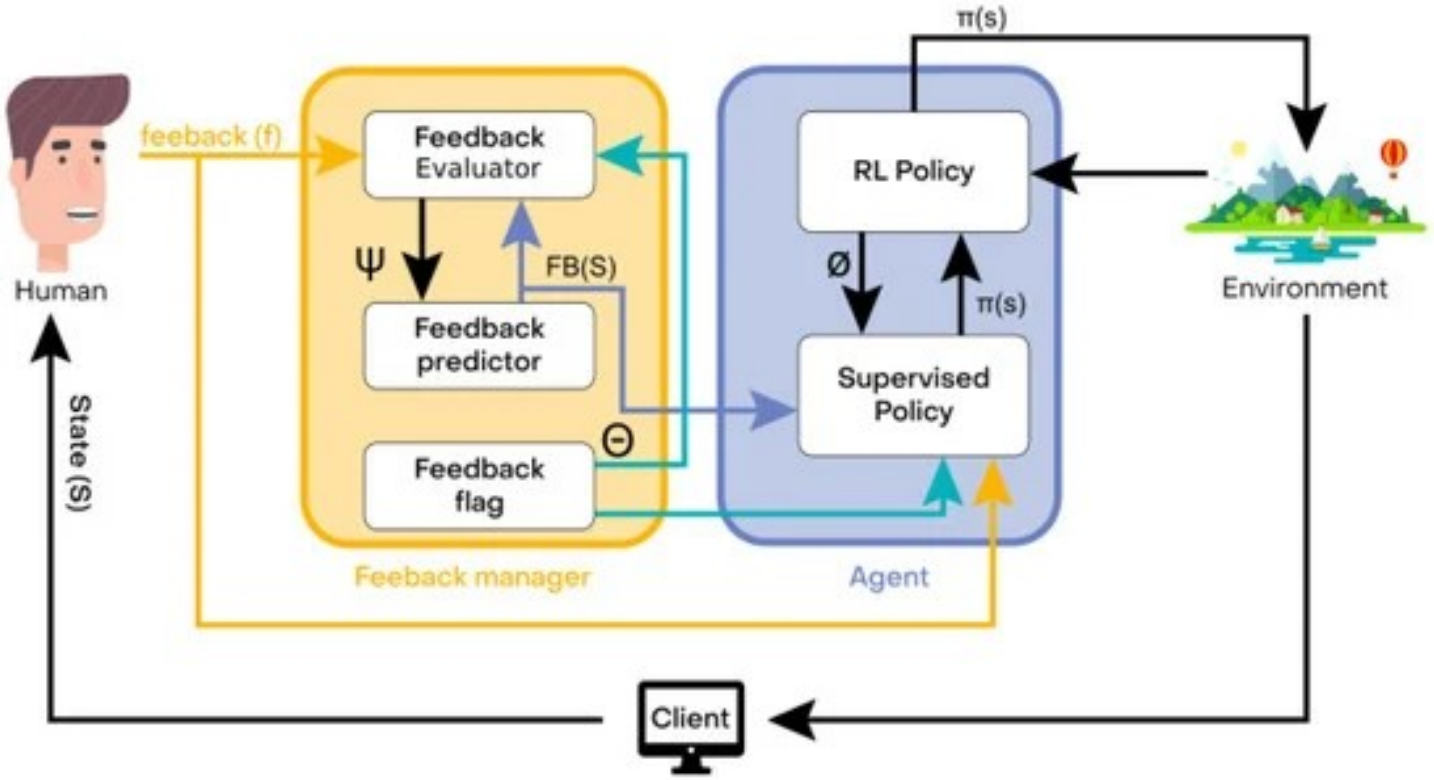
Project Topics

Multi-Agent RL



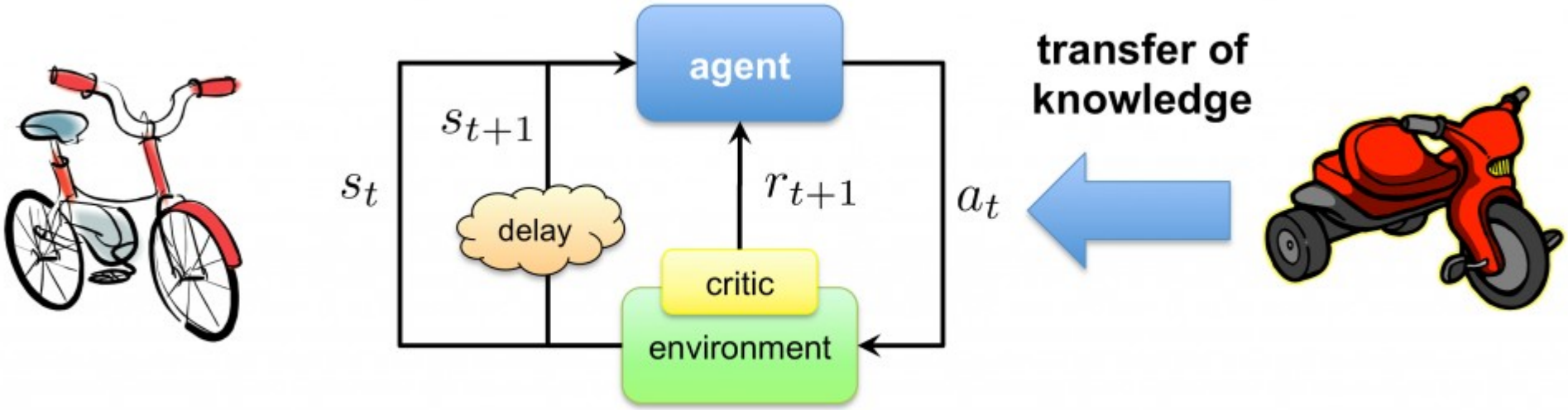
https://towardsdatascience.com/multi-agent-deep-reinforcement-learning-in-15-lines-of-code-using-pettingzoo-e0b963c0820b

RL with Human Feedback



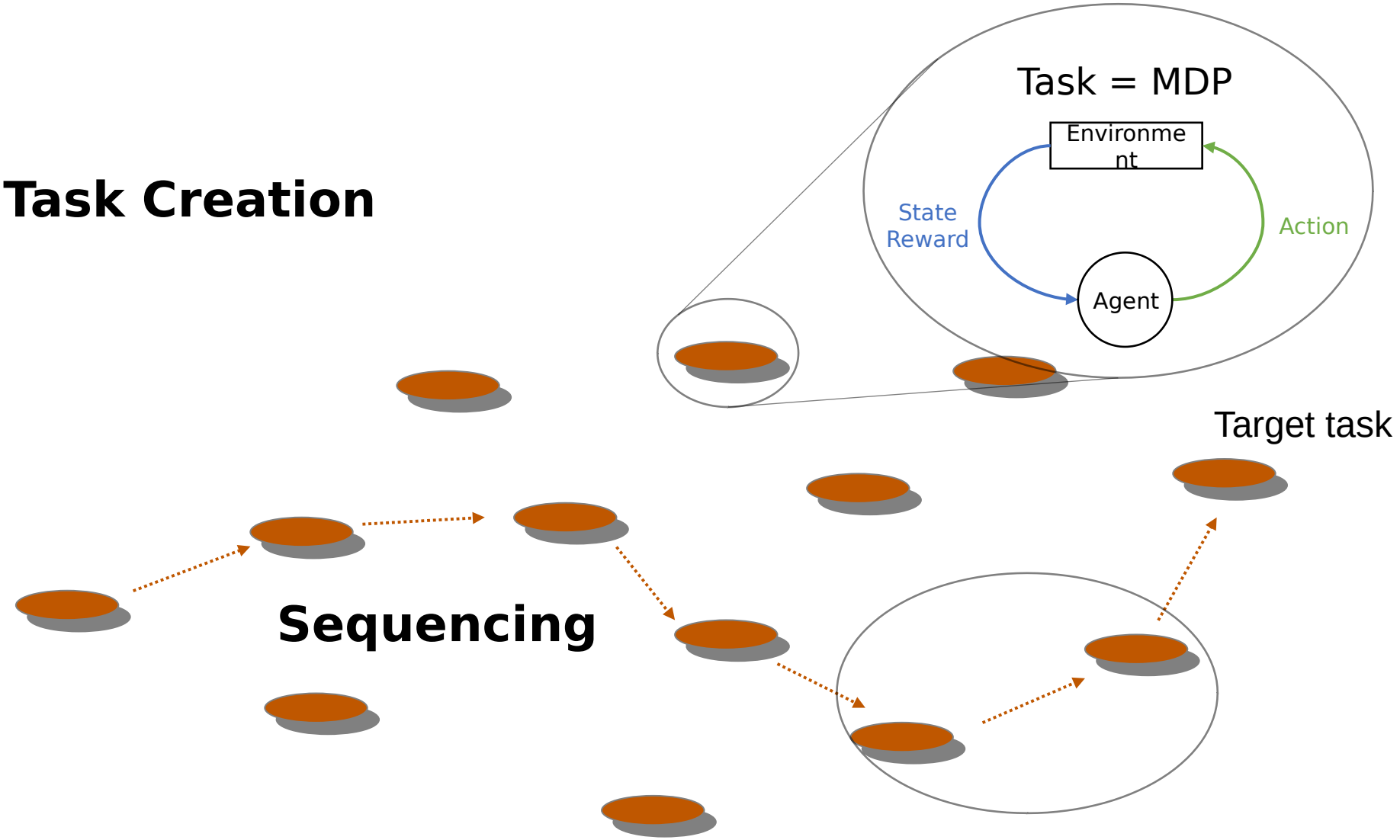
https://www.mdpi.com/2076-3417/11/7/3068

Transfer Learning



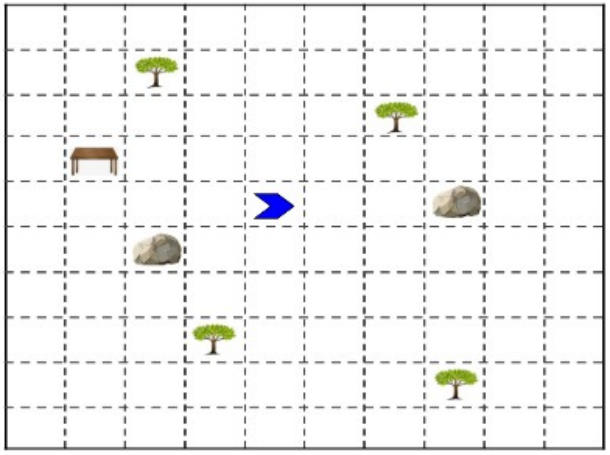
https://project.inria.fr/ExTra-Learn/an-other-news/

Curriculum Learning

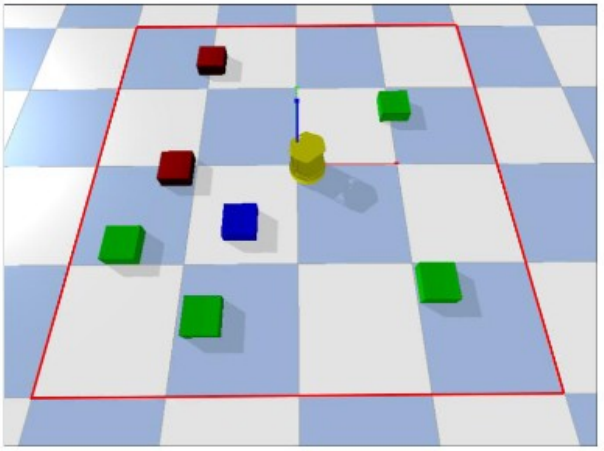


Narvekar, S., Peng, B., Leonetti, M., Sinapov, J., Taylor, M. E., & Stone, P. (2020). Curriculum learning for reinforcement learning domains: A framework and survey. The Journal of Machine Learning Research, 21(1), 7382-7431.

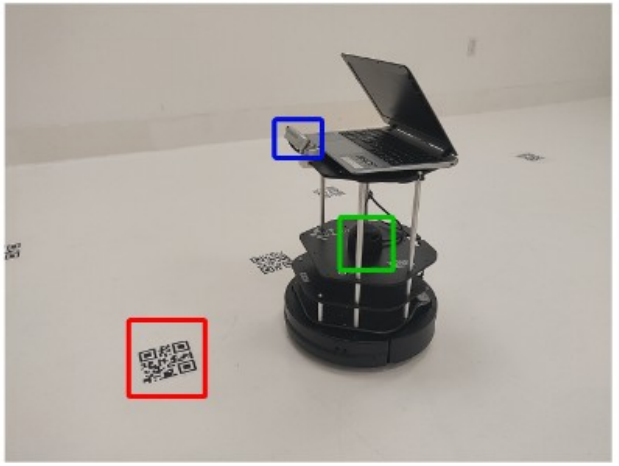
Low-Fidelity to High-Fidelity Transfer



(a) Target task in Low Fidelity Environment



(b) Target task in High Fidelity Environment



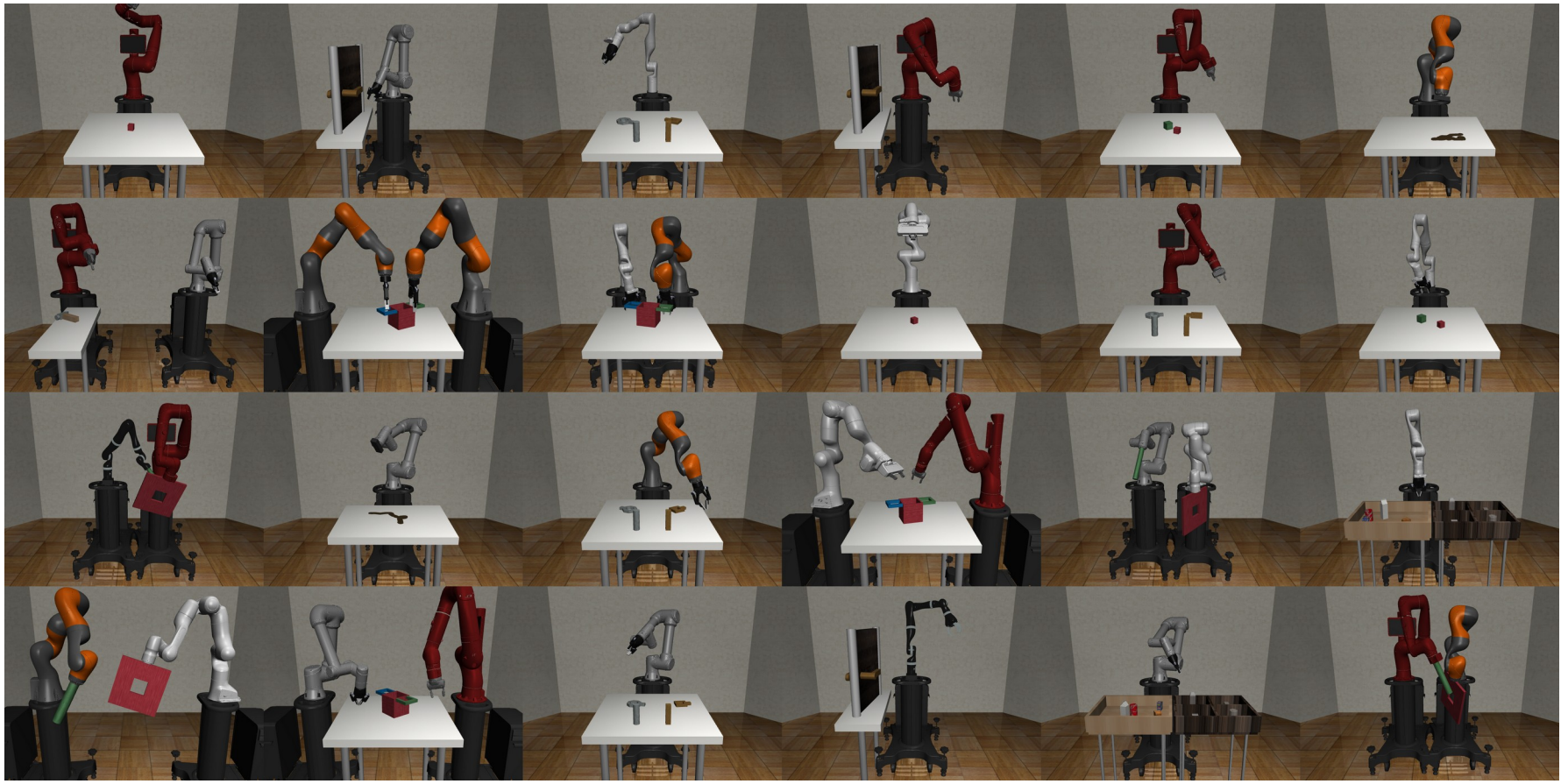
(c) Target task in Physical Environment, using a camera (blue) to interact with fiducials (red). LIDAR (green) is also visible.

Shukla, Y., Thierauf C., Hosseini R., Tatiya G., and Sinapov J. (2022)
ACuTE: Automatic Curriculum Transfer from Simple to Complex Environments
In Proceedings of International Conference on Autonomous Agents and Multiagent Systems (AAMAS), Online, 2022.

RL in Robotics Control

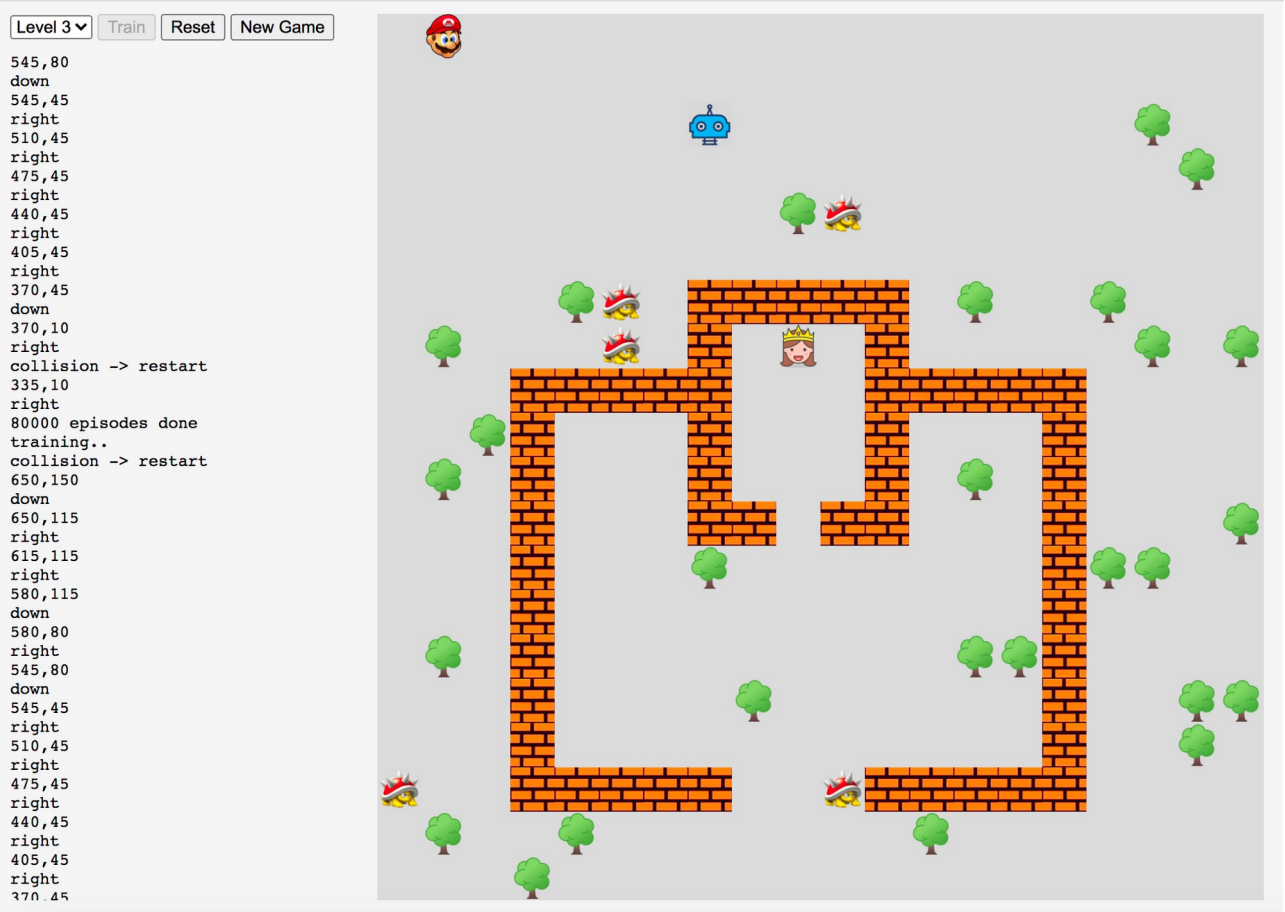
https://www.youtube.com/watch?v=gn4nRCC9TwQ

RL in Robotics Control



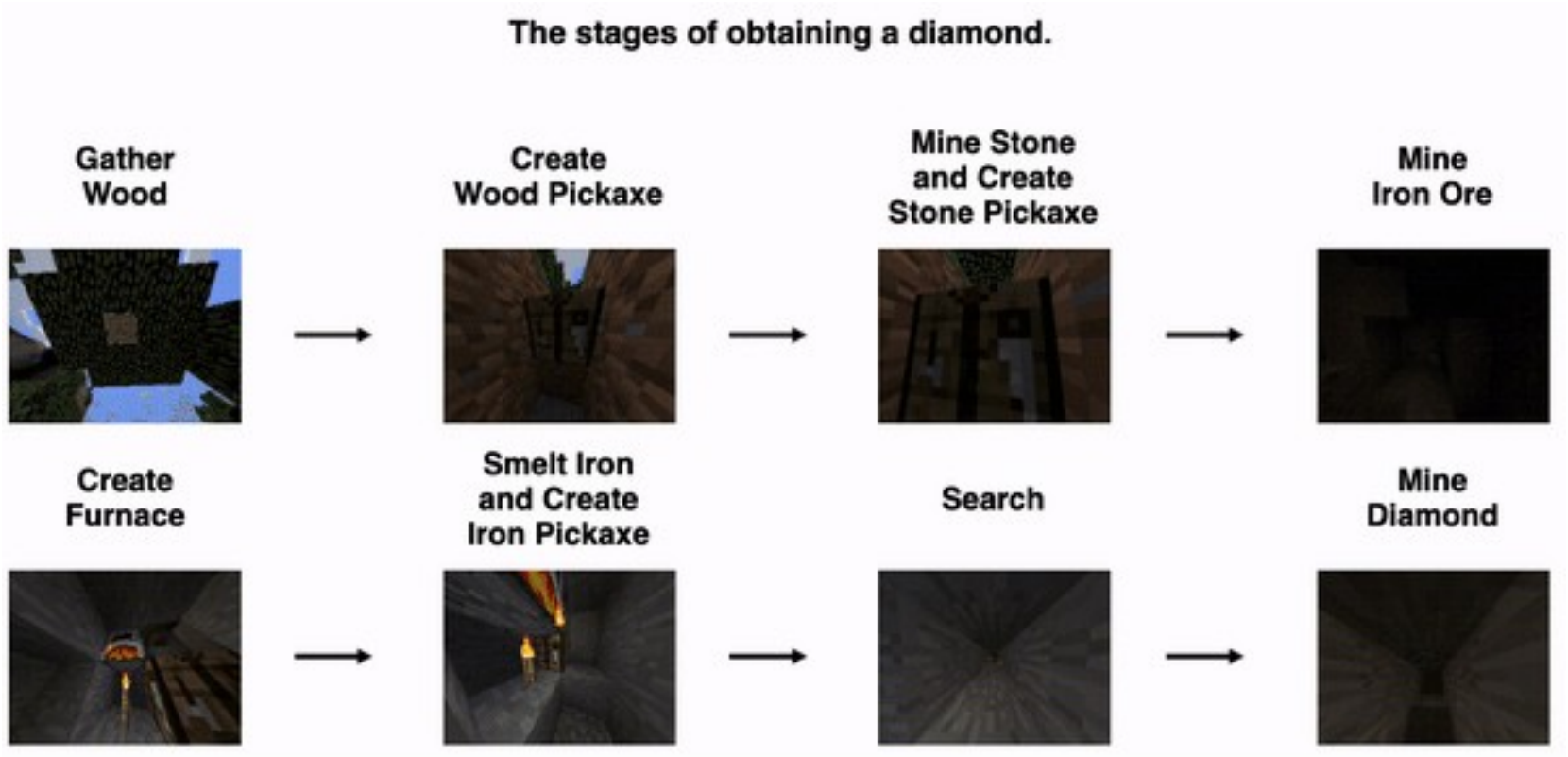
https://robosuite.ai/docs/overview.html

RL Environments



https://github.com/topics/reinforcement-learning-environments?l=javascript

RL Challenges: MineRL



https://www.aicrowd.com/challenges/neurips-2021-minerl-diamond-competition

Overview of 8.1 and 8.2

How do we make Dyna-Q handle stochastic environments?

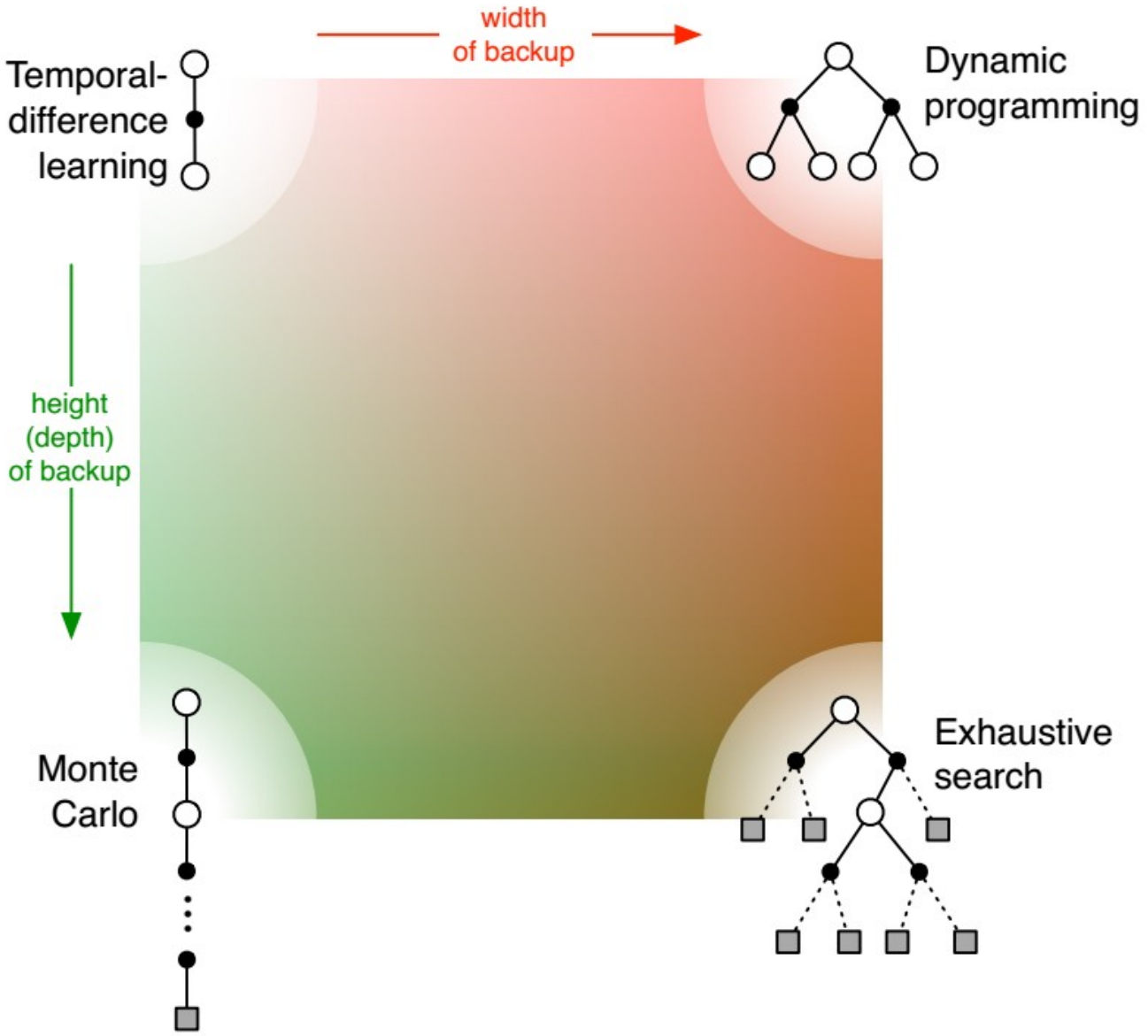
- Small group activity
- Re-write the pseudocode and produce an algorithm which handles stochastic environments
- Now, modify the algorithm you wrote to handle “gradual” non-stationarity (or if it already does, discuss why)

Moderated Discussion

Overview of 8.1 and 8.2

- Exercise 8.2 Why did the Dyna agent with exploration bonus, Dyna-Q+, perform better in the first phase as well as in the second phase of the blocking and shortcut experiments?
- Exercise 8.3 Careful inspection of Figure 8.6 reveals that the difference between Dyna-Q+ and Dyna-Q narrowed slightly over the first part of the experiment. What is the reason for this?

Unified View



[source: Sutton]

Planning and Learning

- Model vs. Model-Free RL
- Types of Models:
  - Distributional
  - Sample
- Q-planning and Dyna-Q

THE END