THE UNIVERSITY OF BRITISH COLUMBIA

Topics in AI (CPSC 532S):
Multimodal Learning with Vision, Language and Sound

Lecture 18: Deep Reinforcement Learning

Types of **Learning**

Supervised training

— Learning from the teacher

— Training data includes desired output

Unsupervised training

— Training data does not include desired output

Reinforcement learning

— Learning to act under evaluative feedback (rewards)

* slide from Dhruv Batra

What is **Reinforcement Learning**

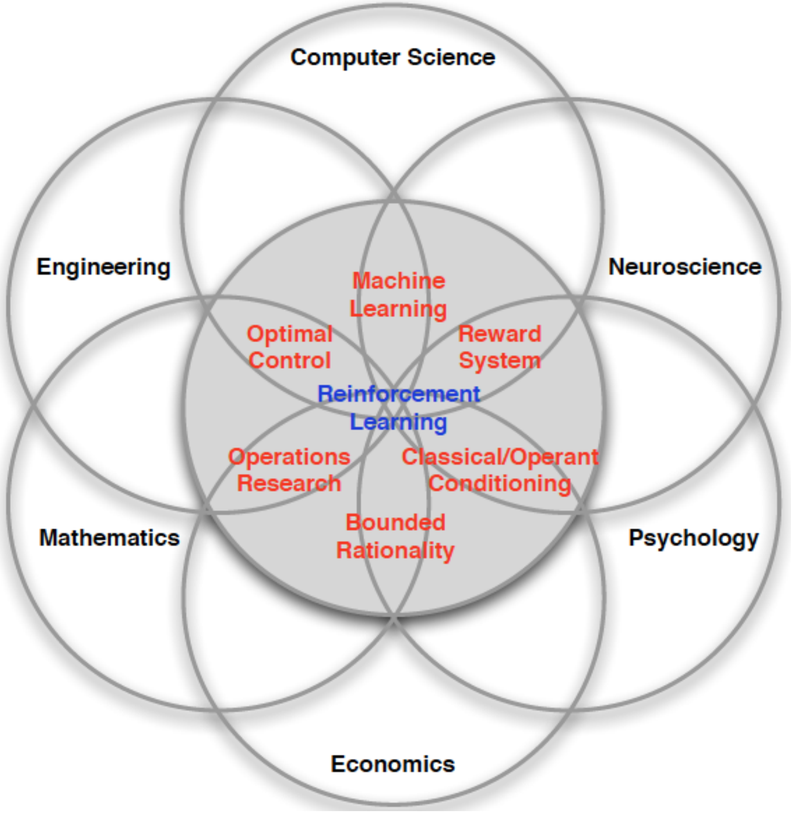
Agent-oriented learning — learning by interacting with an environment to achieve a goal

— More realistic and ambitious than other kinds of machine learning

Learning **by trial and error**, with only delayed evaluative feedback (reward)

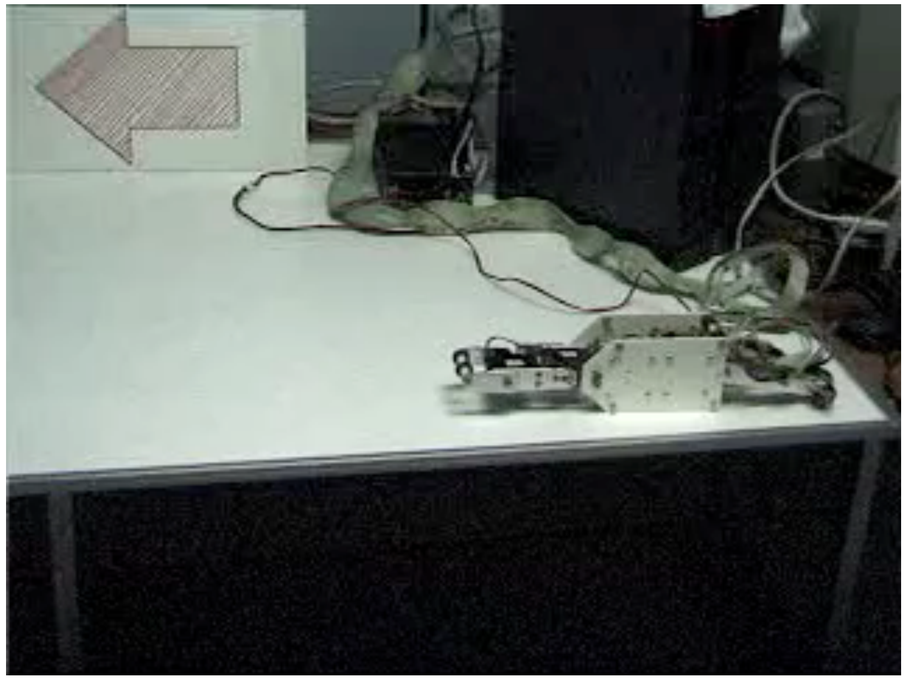
— The kind go machine learning most like natural learning

— Learning that can tell for itself when it is right or wrong

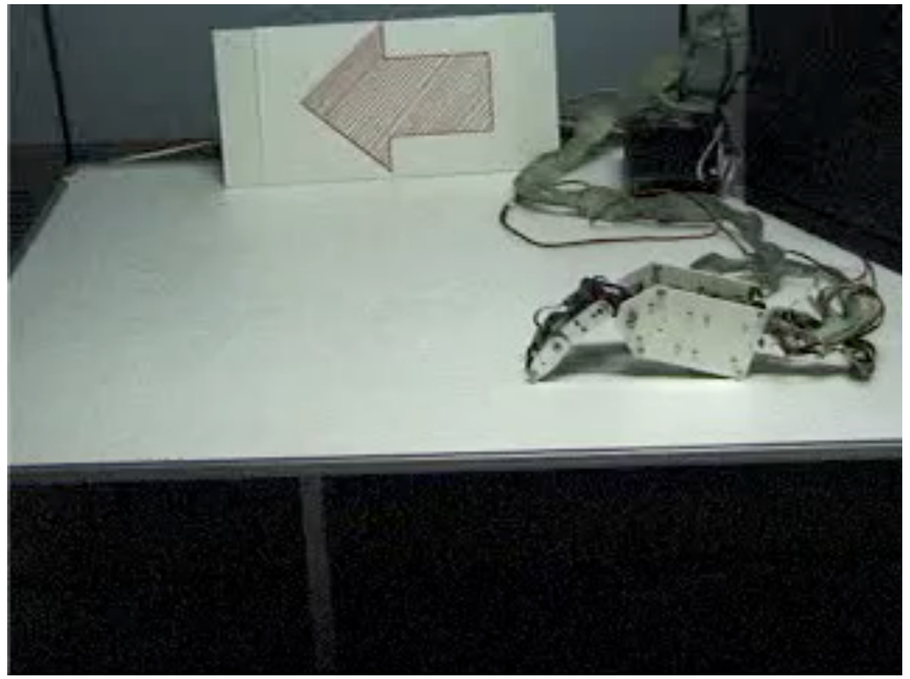


* slide from David Silver

Example: Hajime Kimura's RL Robot



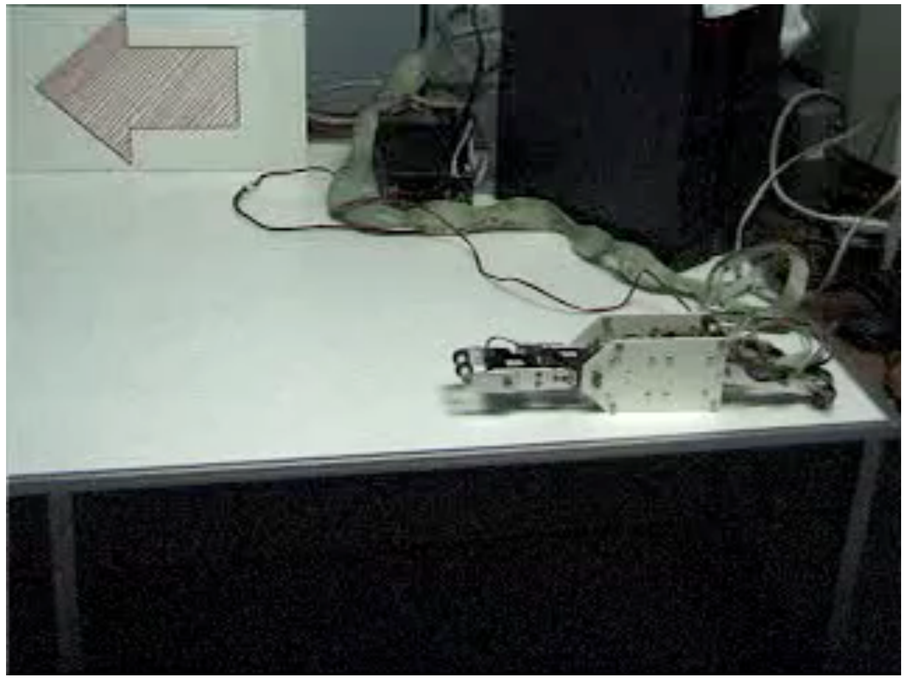
Before



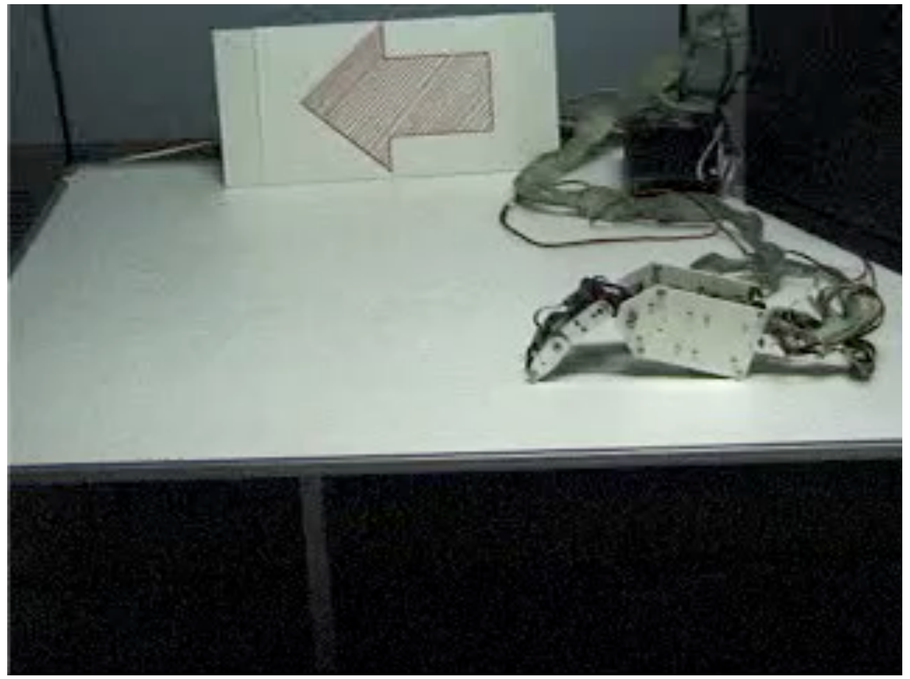
After

* slide from Rich Sutton

Example: Hajime Kimura's RL Robot



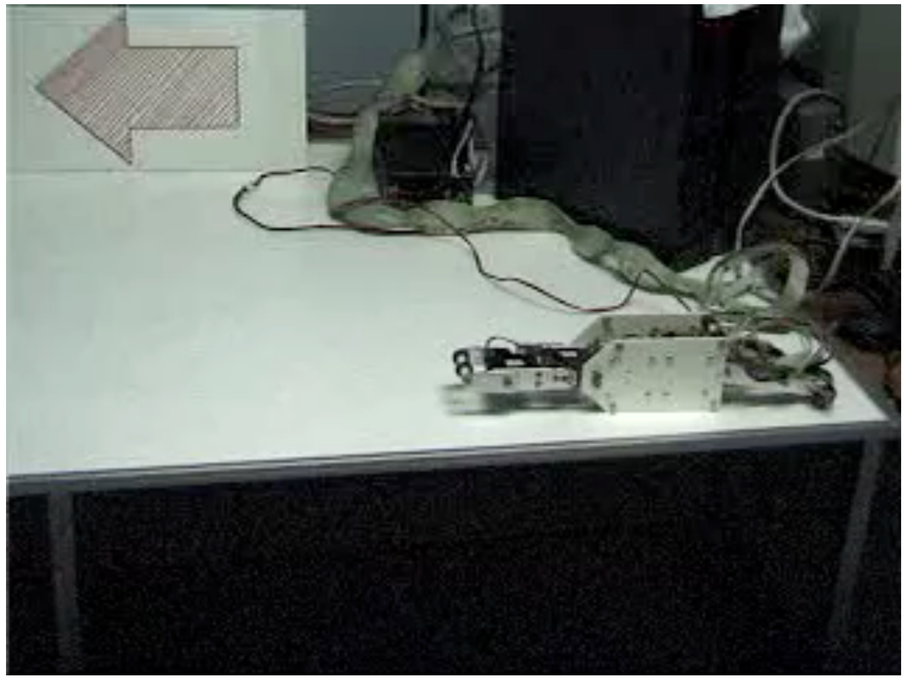
Before



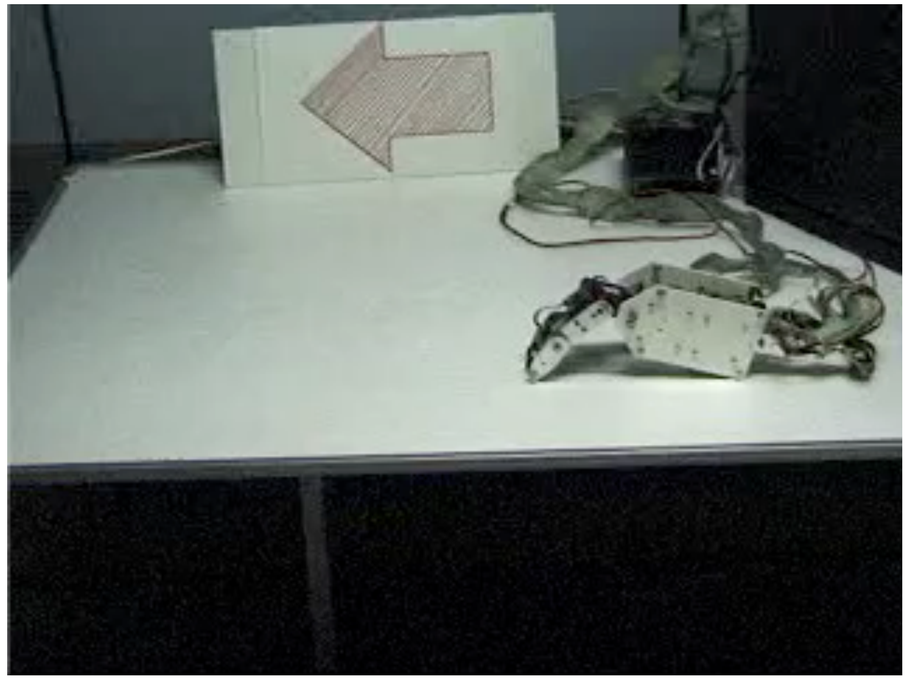
After

* slide from Rich Sutton

Example: Hajime Kimura's RL Robot



Before



After

* slide from Rich Sutton

Human Objectives

“I think it is just the product of a few principles that will be considered very simple in hindsight, so simple that even kids will be able to understand and build intelligent, continually learning, more and more general problem solvers.”

High Level Objectives: Maximize Happiness, Don’t Die

What would be an emergent behavior would evolve

if we have these high level objectives?

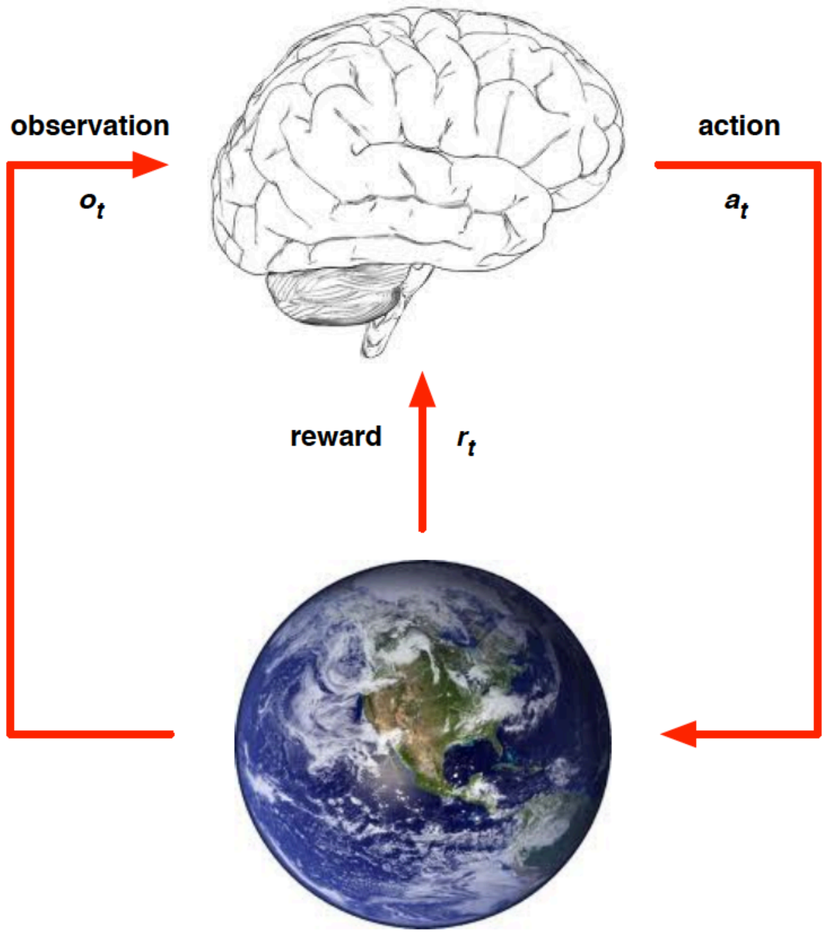
Jurgen Schmidhuber

Challenges of RL

- Evaluative feedback (reward)
- Sequentiality, delayed consequences
- Need for trial and error, to explore as well as exploit
- Non-stationarity
- The fleeting nature of time and online data

* slide from Rich Sutton

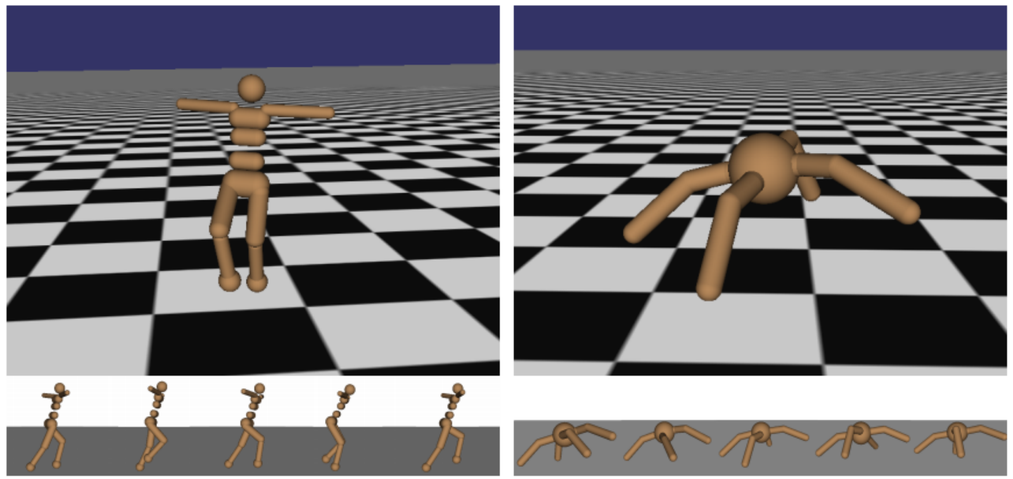
How does **RL** work?



- At each step t the agent:
  - Executes action a_t
  - Receives observation o_t
  - Receives scalar reward r_t
- The environment:
  - Receives action a_t
  - Emits observation o_{t+1}
  - Emits scalar reward r_{t+1}

* slide from David Silver

Robot Locomotion



Objective: Make the robot move forward

State: Angle and position of the joints
Action: Torques applied on joints
Reward: 1 at each time step upright + forward movement

* slide from Fei-Dei Li, Justin Johnson, Serena Yeung, **cs231n Stanford**

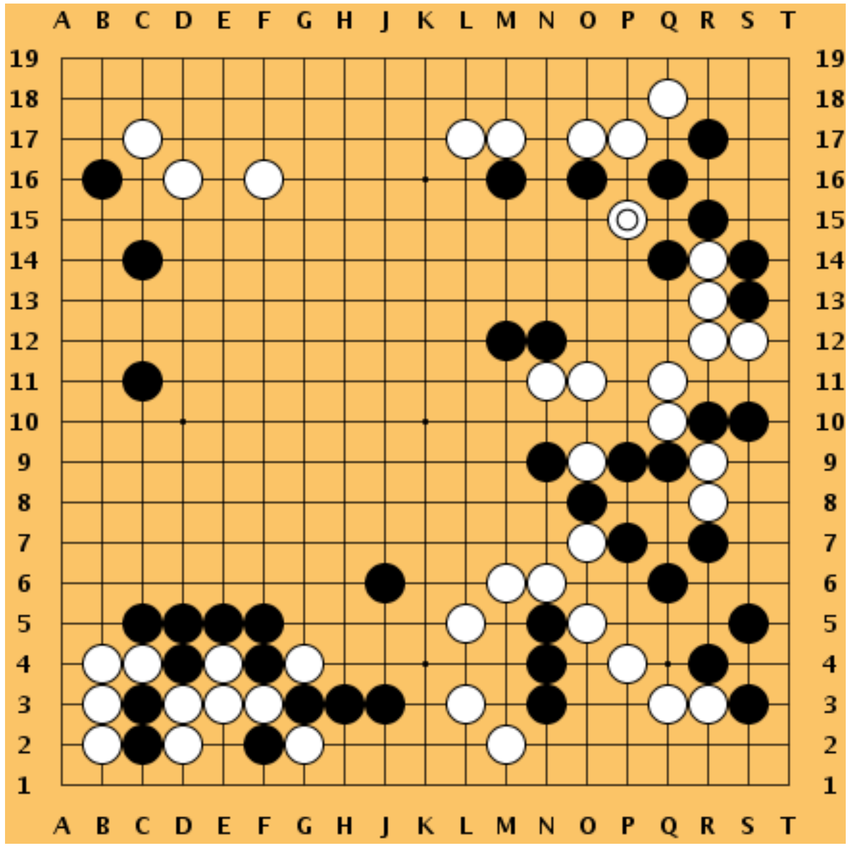
Atari Games

Objective: Complete the game with the highest score

State: Raw pixel inputs of the game state
Action: Game controls e.g. Left, Right, Up, Down
Reward: Score increase/decrease at each time step

* slide from Fei-Dei Li, Justin Johnson, Serena Yeung, **cs231n Stanford**

Go Game (AlphaGo)



Objective: Win the game!

State: Position of all pieces
Action: Where to put the next piece down
Reward: 1 if win at the end of the game, 0 otherwise

* slide from Fei-Dei Li, Justin Johnson, Serena Yeung, **cs231n Stanford**