CMPUT 397

Reinforcement Learning

Fall 2020

Instructor: Martha White
University of Alberta

Some background

- This course **used** to be taught as CMPUT 366
- It was time to make a Reinforcement Learning course
  - The UofA is a world-leader in RL
  - The approaches in RL will be useful in science & industry
- We made a MOOC, to make the topic more accessible to the world

This Course

- We will use the RL MOOC for this course (all lectures)
- In-class time will be spent on
  - Group discussions
  - Worksheets and short answer questions
  - Some free-form question and answer sessions
  - Some fun sessions with guest lectures and demos

What is Reinforcement Learning?

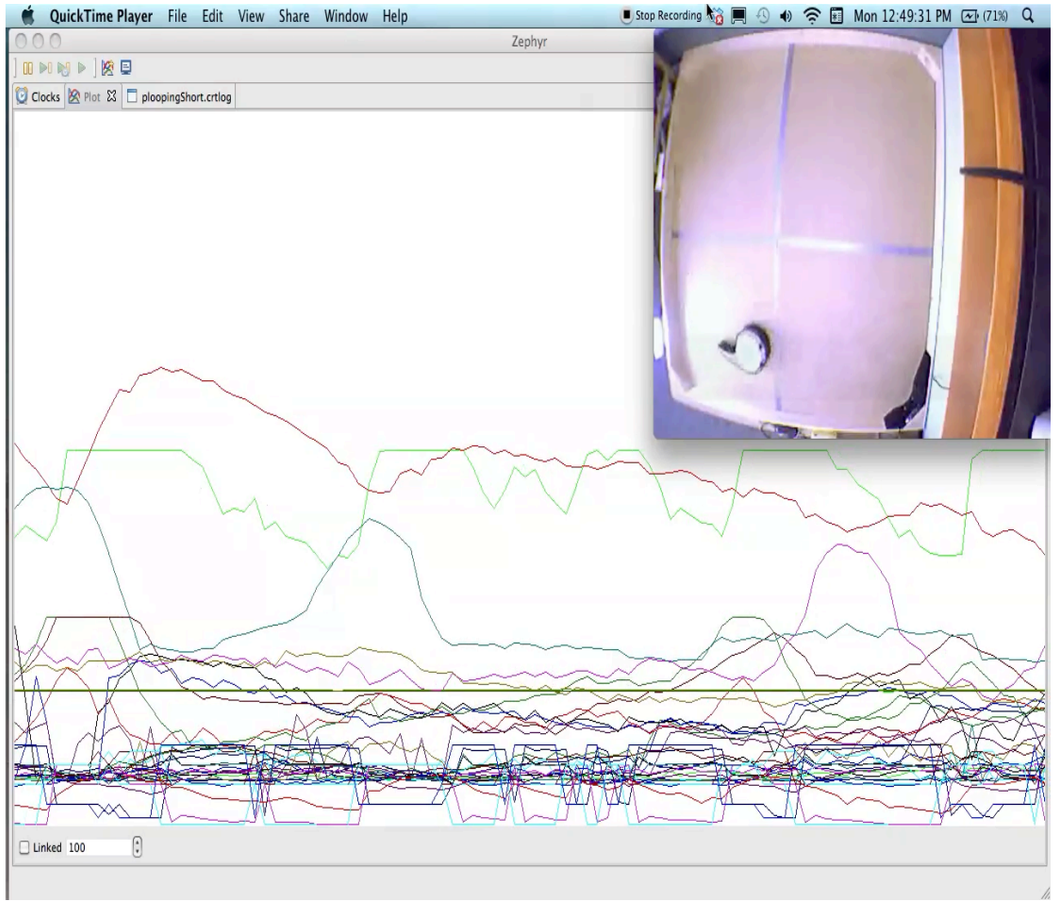
- **A statistical approach to AI**

- An agent constantly interacting with the world around it, learning to make decisions that result in better outcomes

  - Learning is statistical, because it extracts patterns from the experience it has already gathered

  - A key choice in RL is that goodness of outcomes is measured by a scalar reward signal

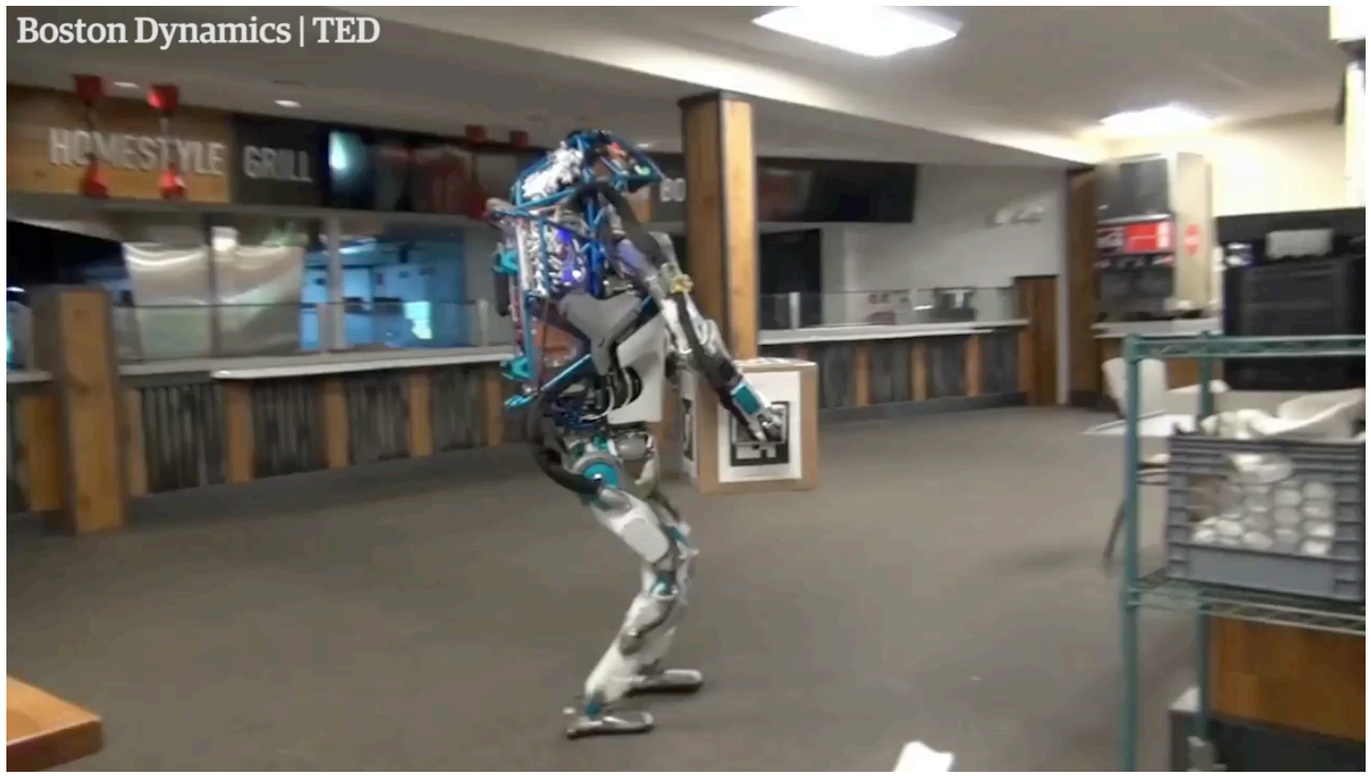
An Example of an RL System: The Critterbot



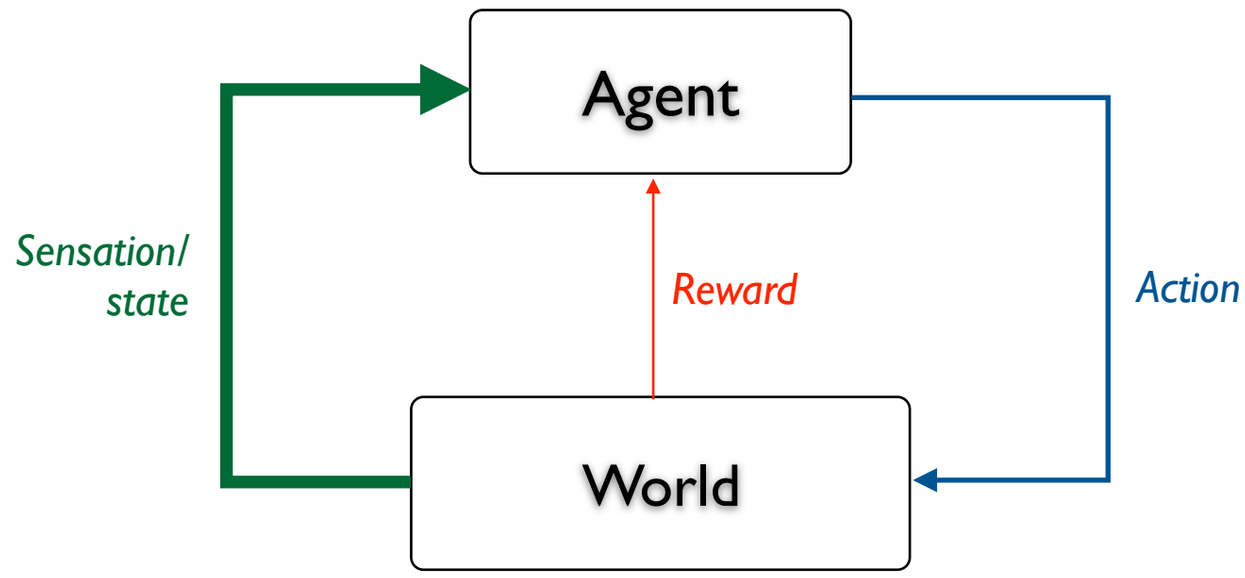
What alternatives are there to RL?

- Traditional approach is control engineering
  - obtain model from first principles
- Experts can engineer impressive systems, using
  - rules
  - assumed models of the world
- These approaches remain important, but do not scale as well with data and can be brittle

Impressive but non-adaptive robot

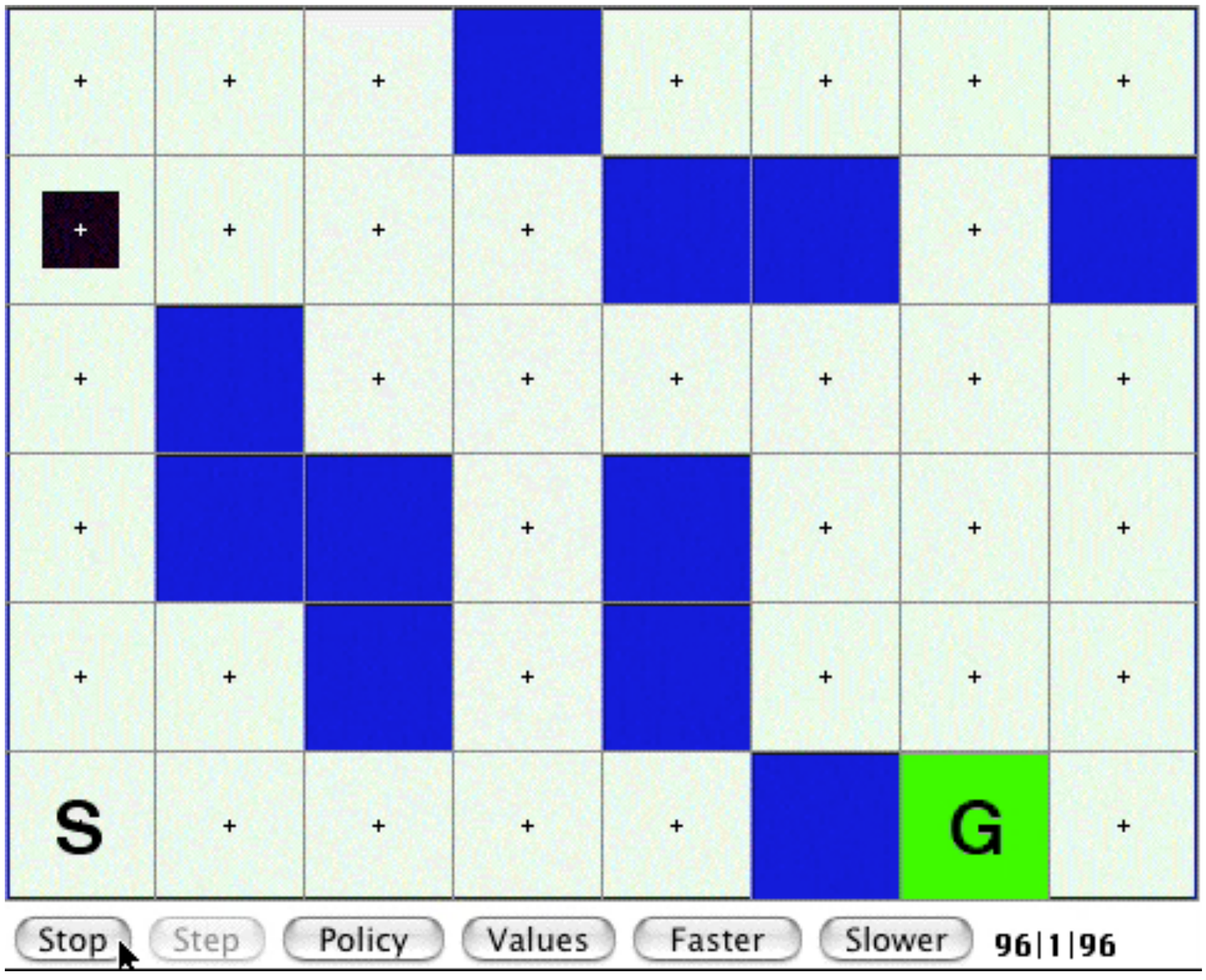


Reinforcement learning is ***more*** *autonomous learning*



- Learning that requires less input from people
- AI that can learn for itself, during its normal operation (adaptive)

GridWorld Example



Why is it important to learn about RL?

- Reinforcement Learning will become a standard tool in industry, to improve decision-making
    - Machine learning is arguably already a standard tool
- It is not that easy to use RL, and requires a good understanding of the algorithms
    - and which ones are appropriate to use where
- You have an opportunity to leverage this expertise

Why is it important to learn about RL?

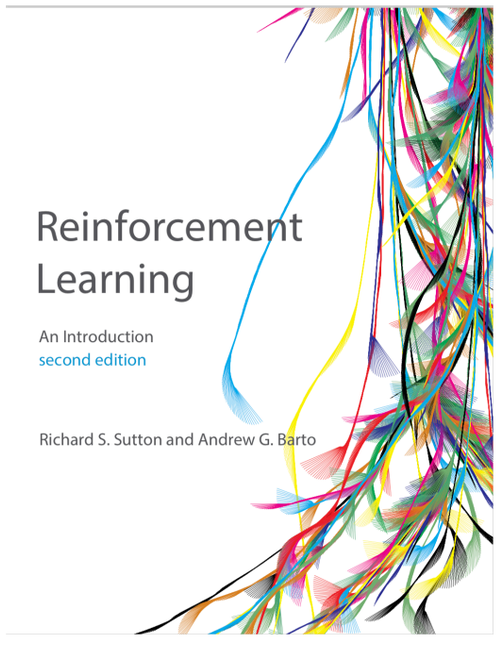
- And of course its just a very elegant approach to AI

- And getting agents to learn is just fun

Course Information

- Github pages with updated schedule and syllabus
- Coursera RL Mooc
- Course eClass page
  - some official information, and place to submit work
- Getting Started and FAQ on eClass

Textbook



- Readings will be from: Reinforcement Learning: An Introduction, by R Sutton and A Barto, MIT Press.
    - available freely online
    - you can order online - be wary of fake sellers

Registering for RL on Coursera

- We have our own private session, on Coursera
- Please register today —> the first item is due on Tuesday!
- See the announcement and the Getting Started document on eClass

Evaluation

- Assignments/Quizzes (completed in Coursera) – 30%
- Project - 10%
- Participation - 10%
- Midterm – 20%
- Final – 30%

Weekly Quizzes and Assignments

- Each week is a different module, with an associated Practice Item and Graded Item
    - usually a Practice Quiz and a Graded Notebook (programming)
- In preparation for class, on your own you need to:
    - Watch the lectures online (at most 1 hour of time)
    - Do the assigned reading for that module (at most 1 hour)
    - Complete the quizzes/assignments (about 3-4 hours)

Weekly Quizzes and Assignments

- In preparation for class, on your own you need to:
    - Watch the lectures online (at most 1 hour of time)
    - Do the assigned reading for that module (at most 1 hour)
    - Complete the quizzes/assignments (about 3-4 hours)
- **You must complete the ungraded component by EOD Sunday and the graded component by EOD Friday**

Typical Flow

- Watch videos and complete Practice Quiz by 11:59 am on Sunday for a given module M (e.g., C1M2)

- Monday-Friday we do in-class sessions discussing the content for that module

    - outside of class you are presumably starting to watch videos for next week's module N, so you can complete the practice quiz for the upcoming Sunday

- You submit the graded assignment for the module M by 11:59 am on Friday.

Deadlines for Quizzes and Assignments

- All deadlines are listed on the github schedule
- We start discussing Mini-Course 1 Module 1 (Sequential decision-making) in-class next week
    - You should start watching the videos now
- For only this first week, we have extended the deadline for the Practice Quiz to Tuesday at midnight
- The Graded Notebook is still due on Friday, Sept 11

Project

- Most will complete the capstone project (Course 4 of the RL Mooc)
- OR it might be possible to join a project with a graduate student, in the RL graduate course
    - less clear-cut and more difficult (research is hard)
    - please email me if you want to do this, but do know that it could be time consuming and risky

Evaluation

- Assignments/Quizzes (Graded Items on Coursera) – 30%
- Project (Capstone on Coursera) - 10%
- Participation (Practice Items + slido questions, to be prepared to participate in class) - 10%
- Midterm – 20%
- Final – 30%

Slido

- Each week, before EOD Sunday, identify a question you have about the module or would like me to address

- Go to Slido and
    - post a question, and/or
    - upvote an existing question
- Half of your participation mark is completing the Slido question submission and/or upvote

Let’s test Slido today

- Go to: https://app.sli.do/

- Use event number: U457

In-class Sessions

- Monday: Q&A Session, and I address some of your questions from Slido, we start a worksheet
- Wednesday or Friday we work on a worksheet
- For Fun Session weeks: worksheet on Wed, Fun session on Friday
- For Discussion weeks: worksheet on Friday, Discussion led by a graduate student on Wed

Fun Session

- Mainly, to have a bit of fun and change up the routine
- Typically these will be guest lectures
    - e.g., Rich presenting about animal learning
- We might do a trivia session

Discussion Session

- These will occur three times
- You will join a small group of about 4-5
- A graduate student (from the RL Grad course) will direct the discussion and help you identify points of confusion
- The primary goal is to help you coherently express what you don’t know and what you need help with
- The grad student will collate the key questions, and I will discuss some of them in class

Worksheet sessions

- We will post a worksheet with 2-4 questions to do in class
  - help reinforce the material
  - get you to answer more open-ended questions than the quizzes on the platform
    - give you a better idea what long-form answers might look like on an exam

- We will randomly assign you to a break-out room in zoom, with a TA

- If you want to have your own group, we encourage one member to use their personal zoom meeting room

Disclaimer: We might adjust as we go along

- CMPUT 397, as a flipped class, is still quite new
- The virtual experience might change things too
- We are planning for discussion sessions, for example, but logistically this might end up changing
- Additionally, if you find something about the structure is not working, or wish we could do something else in class, we might be able to do it! So feel free to tell me.

Breakout Session

- You will be invited to join a Breakout Room, led by a TA
- Go around the circle and **answer one of** the questions
  - Q1: What you hope to get out of this course?
  - Q2: Do you think you will use RL in your future career?
  - Q3: Why are you interested in RL?

No Lab

- There is No Lab

- In-class time is already hands-on

Grades are not based on a normal curve

- Grades are relative to your fellow classmates, but I do not fail the bottom X% (thus, not a normal curve)

- We will provide letter grade boundaries at the end of the course

- Letter grades are provided by a clustering of percentages

  - This allows for adjusting due to yearly differences

- This is a third year course, so grades are typically skewed a bit higher

Prerequisites

- Some comfort or interest in thinking abstractly and with mathematics
- Elementary statistics, probability theory
    - conditional expectations of random variables
        - there will be two class sessions devoted to a tutorial review of basic probability and statistics
- Basic linear algebra: vectors, vector equations, gradients
- Programming skills (Python)
    - If Python is a problem, we have posted some tutorials

Instruction Team

- Prof: Martha White
- TAs (grad students doing research in RL)
    - Matthew
    - Banafsheh
    - Dhawal
    - Sina
    - Abhishek

Contacting us

- Use Student Question forum on eClass
  - Start here if you can. Others have your question too!
- Use course email to email TAs: grp-fall20-cmput397-lec-a1@ualberta.ca
- For personal issues (e.g., missing your exam), please email Martha
- See FAQ for more details

Office Hours

- I have office hours on Wednesday, 2-4 (after class)
    - Let’s test out slido!
- TAs have their office hours listed on eClass
- We will try to accommodate different timezones
    - Note: all in-class sessions will be recorded too to accommodate different timezones

Let’s test Slido today

- Go to: https://app.sli.do/
- Use event number: U457
- Poll 1: Do my proposed office hours work for you?

Collaboration

- Working together to solve the problems is encouraged
- But you must write-up your answers individually
- You must acknowledge all the people you talked with in solving the problems

What is Plagiarism

- Taking things from others and passing it off as your own work without credit

Test time: are these ok?

- Writing down answers to assignments in a group?
- Getting a tutor to help write your code?
- Letting your friend look at your code or assignment question?
- Searching for and using assignment solutions from the internet?
- Not indicating on your assignment who you talked with?
- Discussing ideas without writing anything down?

Policies on Integrity

- Cheating is reported to university whereupon it is out of our hands

- Possible consequences:
  - A mark of 0 for assignment
  - A mark of 0 for the course
  - A permanent note on student record
  - Suspension / Expulsion from university

Academic Integrity

- The University of Alberta is committed to the highest standards of academic integrity and honesty. Students are expected to be familiar with these standards regarding academic honesty and to uphold the policies of the University in this respect. Students are particularly urged to familiarize themselves with the provisions of the Code of Student Behavior (online at www.ualberta.ca/secretariat/appeals.htm) and avoid any behavior which could potentially result in suspicions of cheating, plagiarism, misrepresentation of facts and/or participation in an offence. Academic dishonesty is a serious offence and can result in suspension or expulsion from the University.

Course Overview

- Main Topics:
  - Learning (by trial and error)
  - Planning (search, reason, thought, cognition)
    - Prediction (evaluation functions, knowledge)
    - Control (action selection, decision making)
- Recurring issues:
  - Demystifying the illusion of intelligence

Birds-eye view

- Mini-Course 1: Fundamentals of RL
    - Bandits and the Model-based setting, where someone gives you how the world works
- Mini-Course 2: Sampled-based Learning Methods
    - Learning only through trial-and-error interaction
- Mini-Course 3: Prediction and Control with Function Approximation
    - Extending all the stuff before to the setting where we have to approximate the functions/models (e.g., using neural networks)
- Putting it all together (your project) is Mini-Course 4

High-level view

- Bandits and Online learning (Ch2, C1M1):
  - formalizing a problem and discussing solution methods
  - A miniature version of the entire course
- Markov Decision Processes and Value Functions (Ch3, C1M2 and C1M3):
  - Our formalization of reinforcement learning and AI…no solution methods here
  - Students usually get impatient here

High-level view (2)

- MDP solution method, given a model:
  - Dynamic programming - (Ch 4, C1M4)
- MDP solution methods, if you can only learn from interaction
  - Monte Carlo (MC) - (Ch 5, C2M1)
  - Temporal difference learning (strengths of both DP and MC) - (Ch 6, C2M2,C2M3)
- Planning with learned models - (Ch 8, C2M4)

High-level view (3)

- Everything up to and including Chapter 8 is tabular solution methods:
  - The foundation of modern RL
- In Chapters 9, 10, 13 cover approximate solution methods:
  - Function approximation (including Neural Nets)
- The foundations established in chapter 3-8 will largely transfer to the function approximation case

Demo of Bandits

- https://www.coursera.org/learn/fundamentals-of-reinforcement-learning/ungradedWidget/44Z9R/lets-play-a-game

- https://www.coursera.org/learn/fundamentals-of-reinforcement-learning/ungradedWidget/jEYTO/whats-underneath