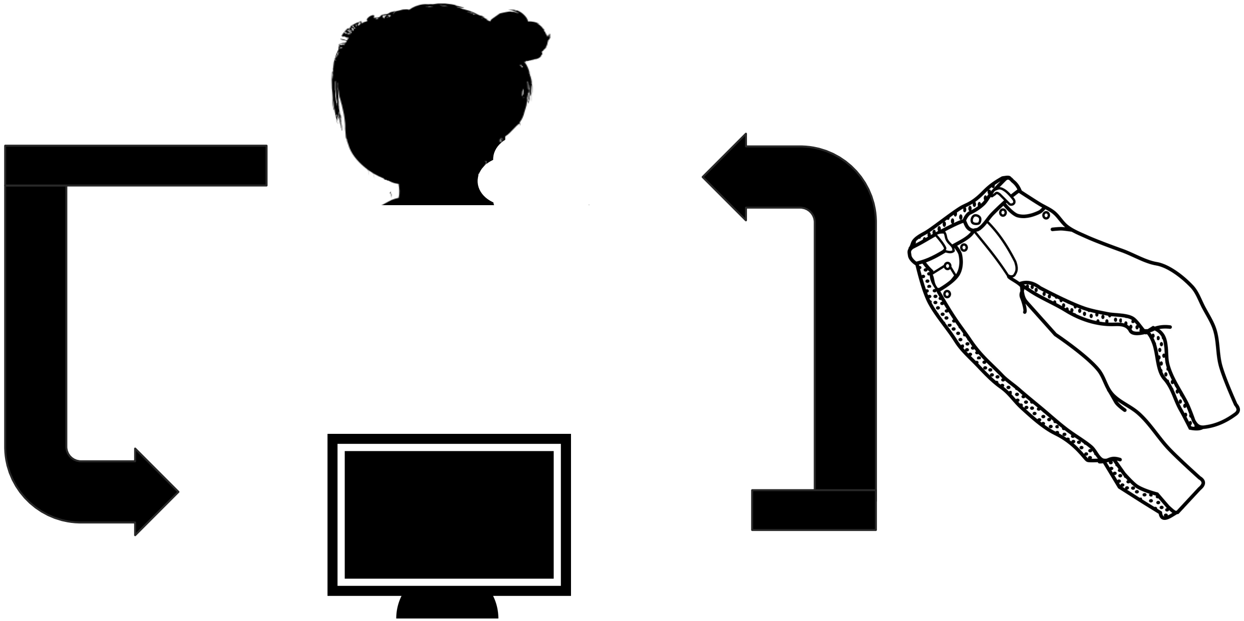
Reinforcement Learning and Reward

Emma Brunskill
CS234
Week 10
Winter 2021

Plan for today

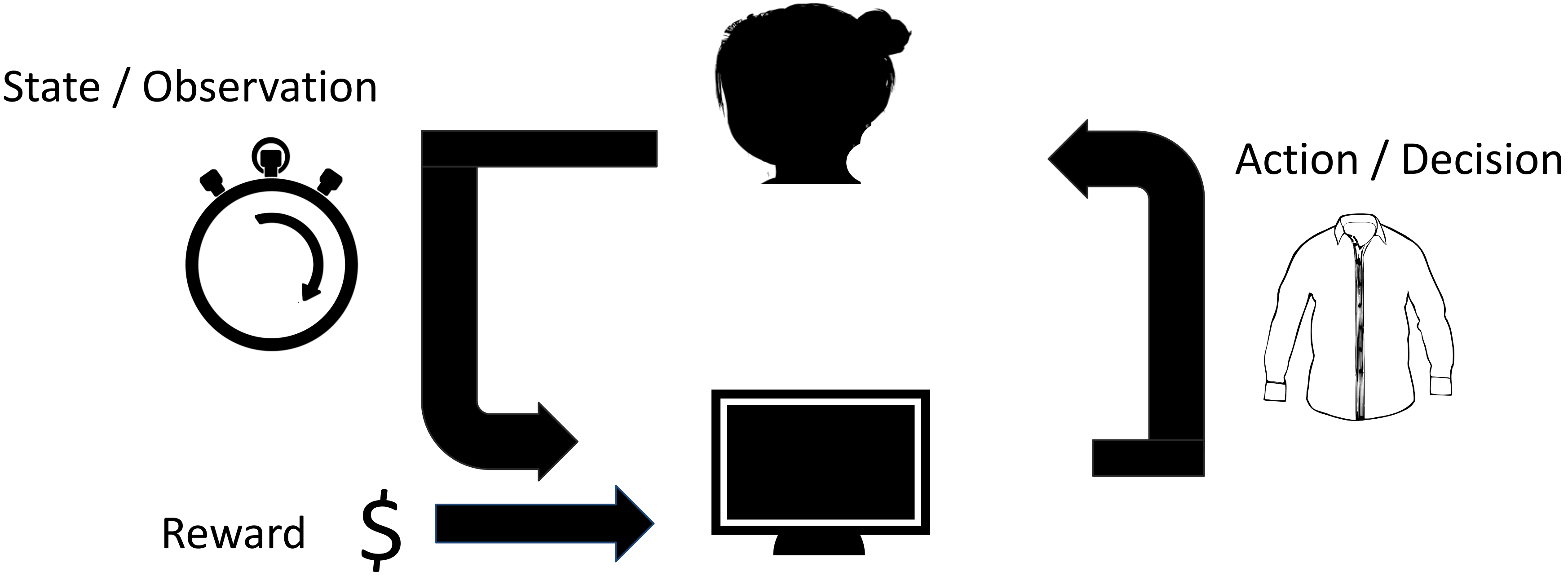
- Reward in RL
- Wrapping up CS234

Reinforcement Learning



Pants image by OpenClipart-Vectors from Pixabay

Decision Policy

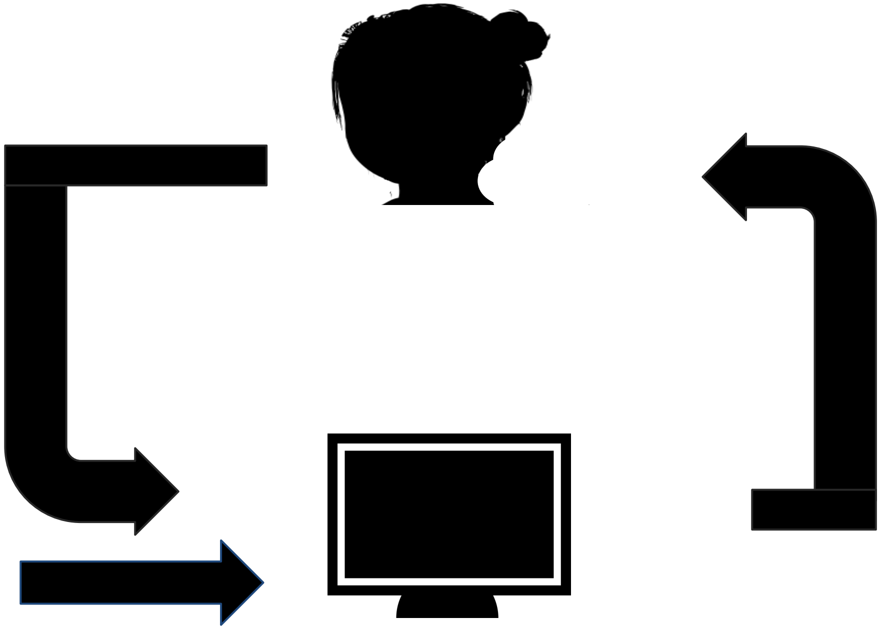


(Decision) Policy: if observe this then do that

Example: If looked at blouse for 10 sec Then show another blouse

stopwatch modified from image by Katarzyna Tyl from Pixabay

Advertising Example



State / Observation:

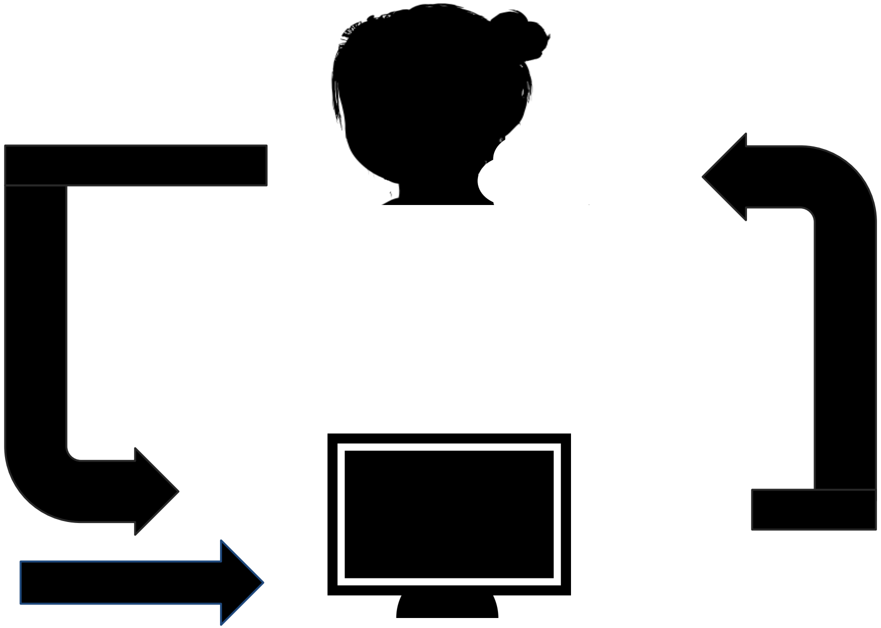
View time

Action / Decision

Choose web ad

Reward: Click on ad

Robot Learning to Unload Dishwasher



State / Observation:

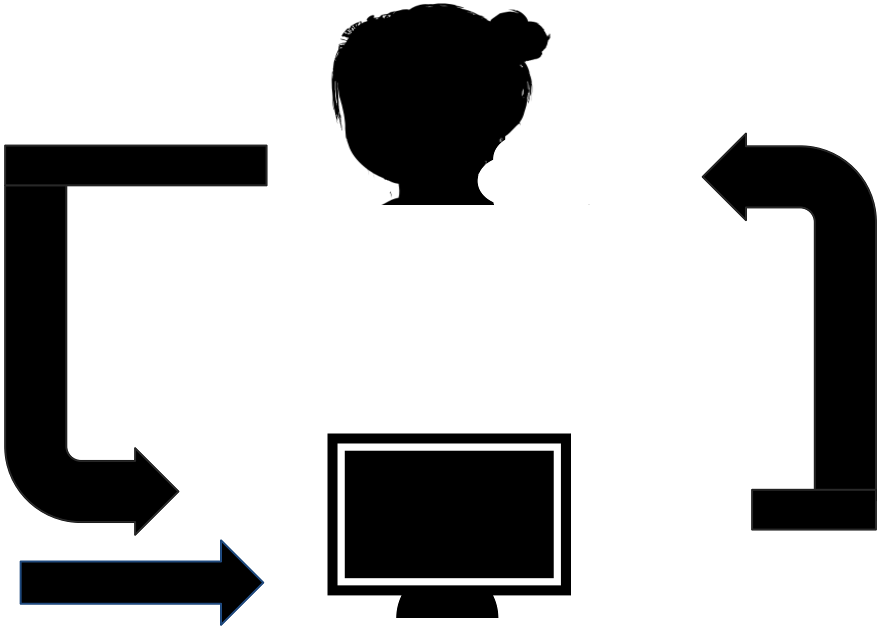
Camera image of kitchen

Action / Decision

Move robot joint

Reward:
If all dishes in dishwasher +1
Else 0

Blood Pressure Management



State / Observation:

Blood pressure
Gender
Location

Action / Decision

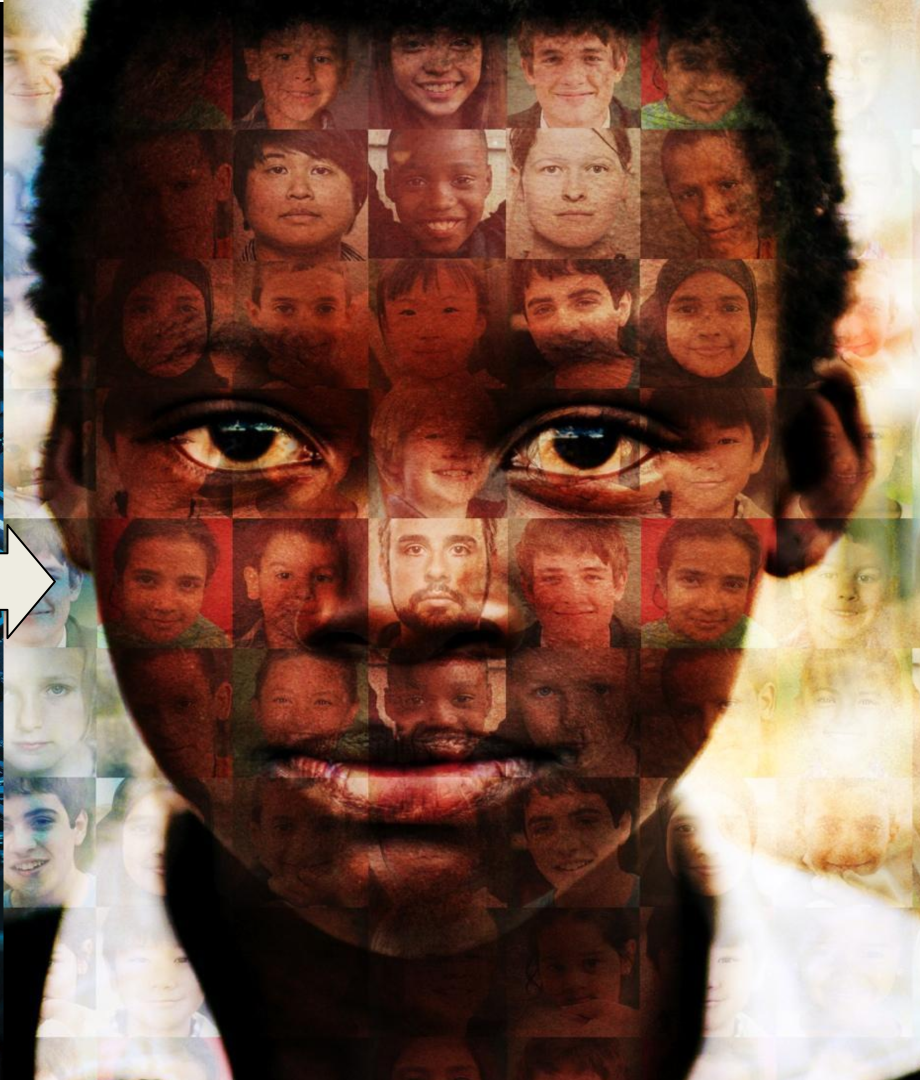
Suggest exercise or meditation

Reward:
If in healthy range: +1
If use medication: -0.05
-

Beyond Expected Reward

- In this class focused on expected scalar reward
- In many real settings
  - Distribution of outcomes (distributional RL, conditional value at risk, …)
  - Multiple-objective (high reward and low cost and …)
  - Constrained maximization (safety, fairness, …)

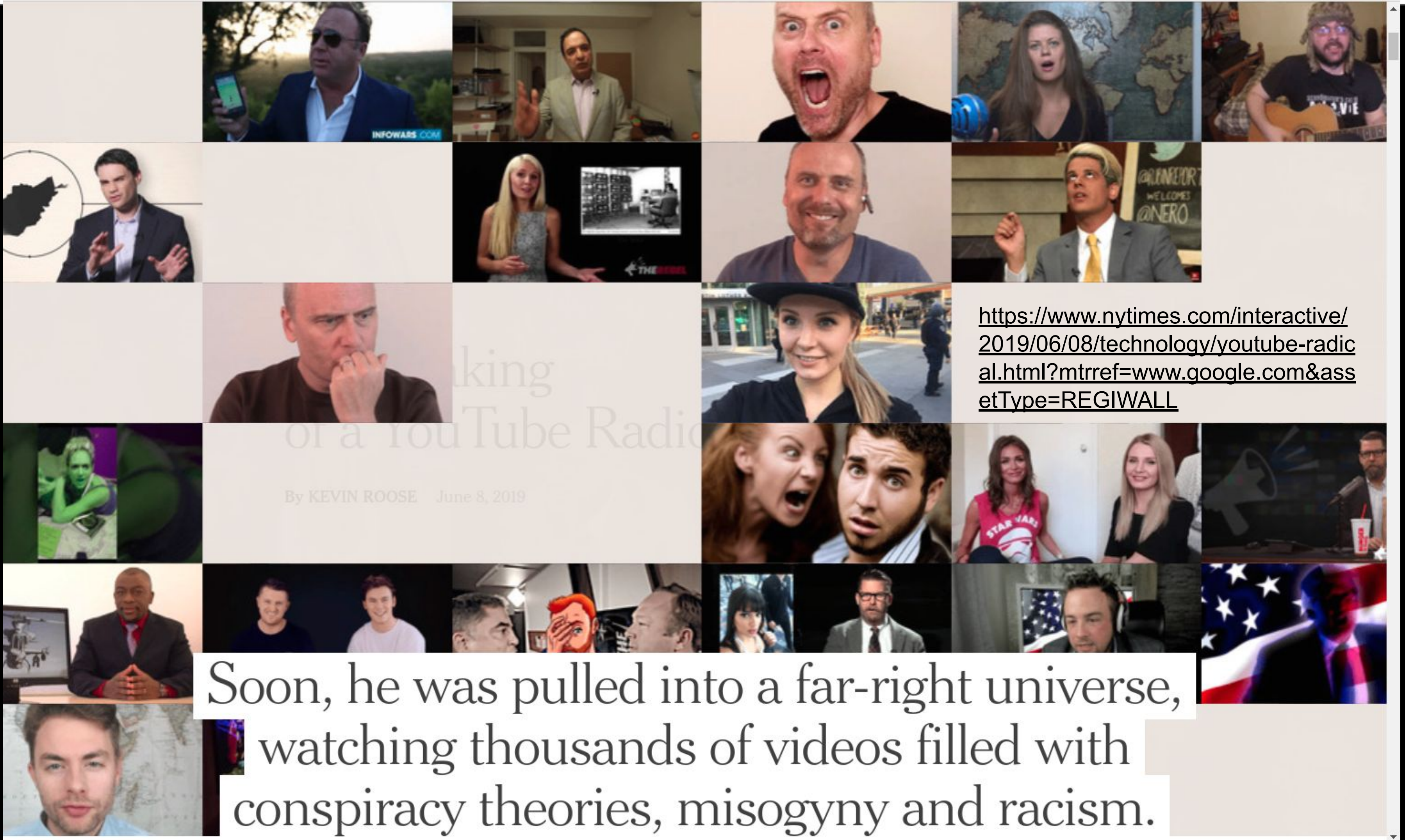
nature
THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE
At last — a computer program that can beat a champion Go player PAGE 484
ALL SYSTEMS GO



Recall Example During My 1st Lecture: AI Teacher

- Student initially does not know addition (easier) nor subtraction (harder)
- Teaching agent can provide activities about addition or subtraction
- Agent gets rewarded for student performance:
  - +1 if student gets problem right,
  - -1 if get problem wrong
- (Think/Discuss) What type of policy would a RL agent learn? Is this what the human designer of this system would likely want?

Caleb Cain was a college dropout looking for direction. He turned to YouTube.



aking
of a YouTube Radic

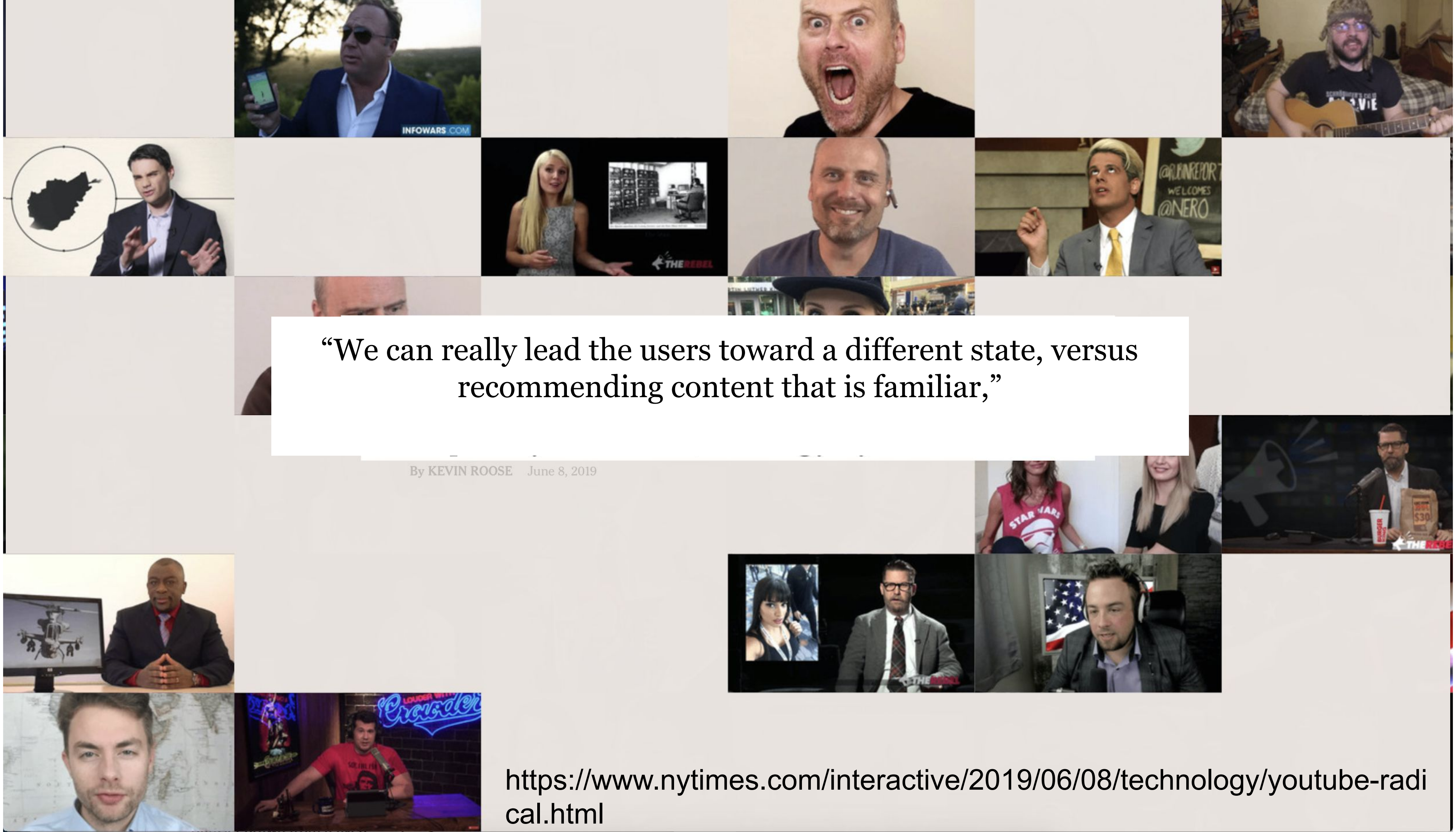
By KEVIN ROOSE June 8, 2019

Soon, he was pulled into a far-right universe, watching thousands of videos filled with conspiracy theories, misogyny and racism.

https://www.nytimes.com/interactive/2019/06/08/technology/youtube-radical.html?mtrref=www.google.com&assetType=REGIWALL

- In last 2 years have been trying out using reinforcement learning
- "… designed to maximize users' engagement over time by predicting which recommendations would expand their tastes and get them to watch not just one more video but many more."

https://www.nytimes.com/interactive/2019/06/08/technology/youtube-radical.html

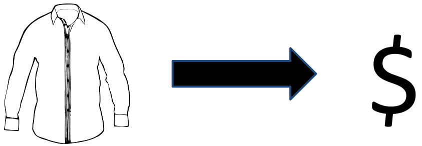


“We can really lead the users toward a different state, versus recommending content that is familiar,”

By KEVIN ROOSE June 8, 2019

https://www.nytimes.com/interactive/2019/06/08/technology/youtube-radical.html

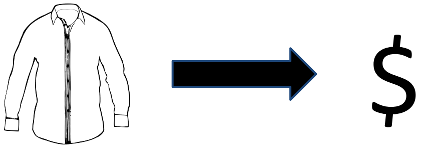
Supervised Learning



$

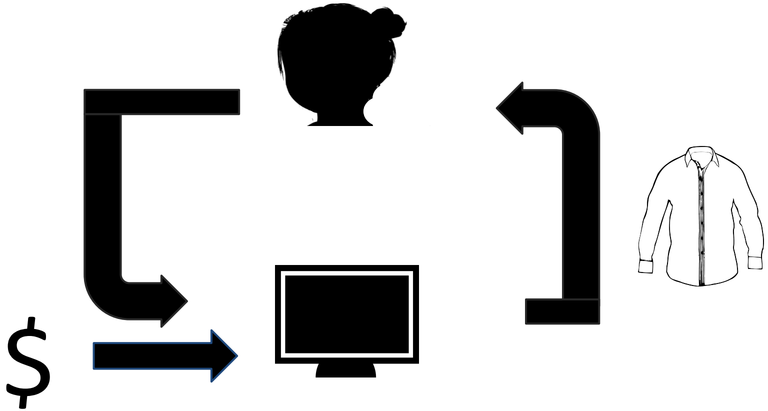
Recommend things people already like*

Supervised Learning



Recommend things people already like*

Reinforcement Learning

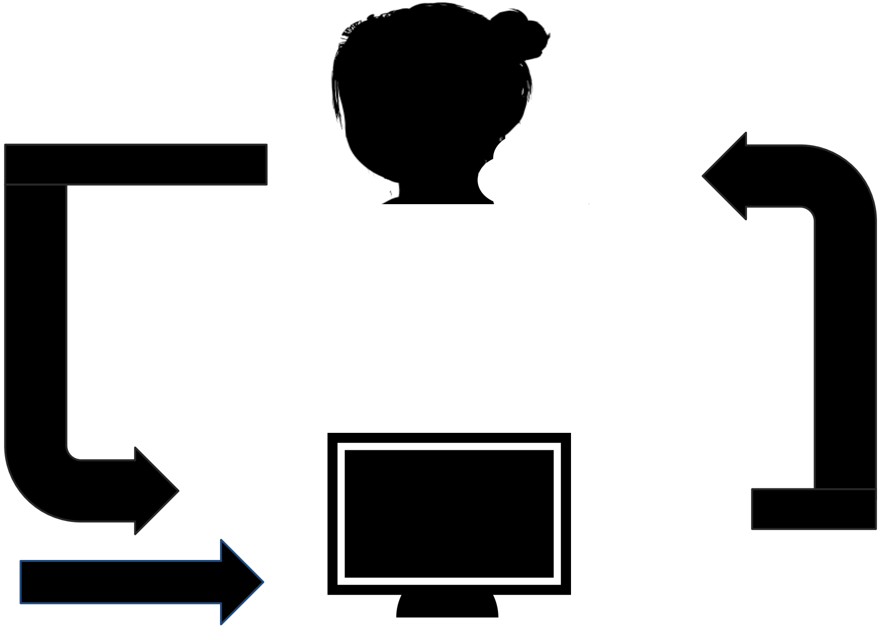


Provide recommendations so people will *(potentially change into people who)* buy more

Reinforcement Learning is Trying to Change (the State of) the World

State / Observation:

Blood pressure
Gender
Location



Action / Decision

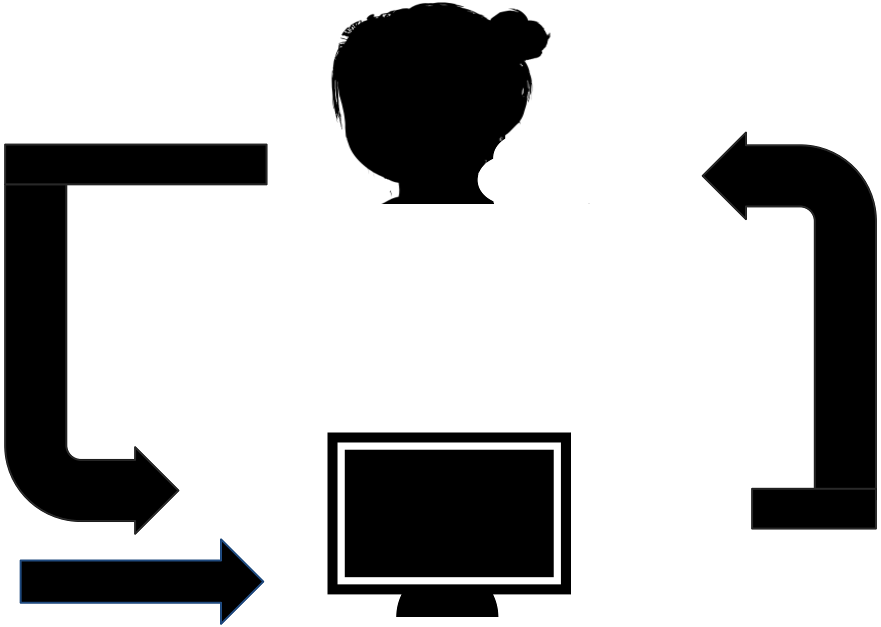
Suggest exercise or meditation

Reward:
If in healthy range: +1
If use medication: -0.05

Reinforcement Learning is Trying to Change (the State of) the World

State / Observation:

Blood pressure
Gender
Location

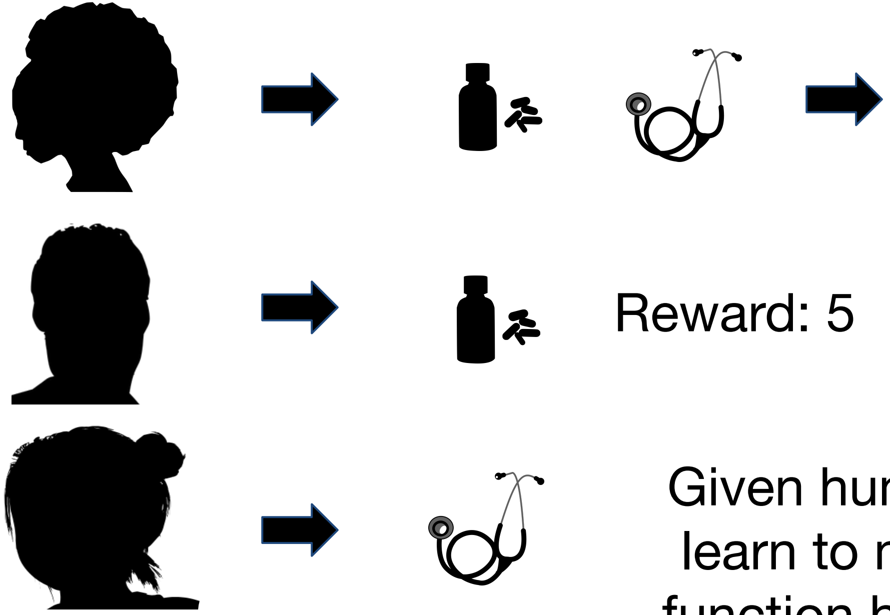


Action / Decision

Suggest exercise or meditation

What is the Reward?

One Idea: Learn the Rewards of People



Reinforcement Learning

Reward: 92

Multi-armed Bandits

Reward: 5

Imitation Learning & Inverse RL

Given human expert decisions, learn to mimic or learn reward function humans are optimizing

Value Alignment

- How can we ensure RL agent is optimizing for our desired rewards?
- Stuart Russell (recent general audience book on this broad topic is Human Compatible: AI and the Problem of Control)
- Anca Dragan, Smitha Milli, Dylan Hadfield-Menell, and others

Wrapping Up CS234

Wrapping Up CS234: Final Parts

- Normal office hours this week
- Homework 4 due 6pm Wed (if you have late days available, deadline is 6pm Friday)

Wrapping Up CS234: Learning Objectives

- Define the key features of reinforcement learning that distinguishes it from AI and non-interactive machine learning (as assessed by the exam).
- Given an application problem (e.g. from computer vision, robotics, etc), decide if it should be formulated as a RL problem; if yes be able to define it formally (in terms of the state space, action space, dynamics and reward model), state what algorithm (from class) is best suited for addressing it and justify your answer (as assessed by the exam).
- Implement in code common RL algorithms (as assessed by the assignments).
- Describe (list and define) multiple criteria for analyzing RL algorithms and evaluate algorithms on these metrics: e.g. regret, sample complexity, computational complexity, empirical performance, convergence, etc (as assessed by assignments and the exam).
- Describe the exploration vs exploitation challenge and compare and contrast at least two approaches for addressing this challenge (in terms of performance, scalability, complexity of implementation, and theoretical guarantees) (as assessed by an assignment and the exam).

Wrapping Up CS234: Classes to Learn More

- CS332 Advanced Survey of Reinforcement Learning (me)
- AA 203: Optimal and Learning-based Control (Marco Pavone)
- MS&E / EE: Ben Van Roy sometimes offers a theory-oriented RL class
- To learn more about theory of RL, Sham Kakade (UW) & Wen Sun (Cornell) taught a class in Fall 2020 based on a joint working book https://rltheorybook.github.io https://wensun.github.io/CS6789.html

Wrapping Up CS234: The End!

- Thank you so for your questions and participation in this class!
- Particularly under so many challenges (remote work, pandemic, power cuts…) all of you should be proud of how much you accomplished
- Please fill in the course evaluations-- I greatly value your feedback and knowing what helped you learn and what could use improvement helps me improve the class for later students
- Keep in touch! I love hearing about what students do next, and if there are ideas in RL they keep using or advancing