Reinforcement Learning & Learning to Promote Learning

Emma Brunskill
Stanford University

AI to Automate Humans

<u>Will Robots Replace Human Drivers, Doctors and Other Workers?</u>

<u>Robots could take over 38% of U.S. jobs within about 15 years</u>

<u>Will robots replace workers by 2030?</u>

https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=newssearch&cd=6&ved=0ahUKEwjI98y9hO7VAhU
SyWMKHRnOAwUQqQIINygAMAU&url=https%3A%2F%2Fwww.wired.com%2Fstory%2Fdriverless-cars-need-e
ars-as-well-as-eyes%2F&usg=AFQjCNHYIWQaBgaSNcJlLzHIY_i8kIygdw

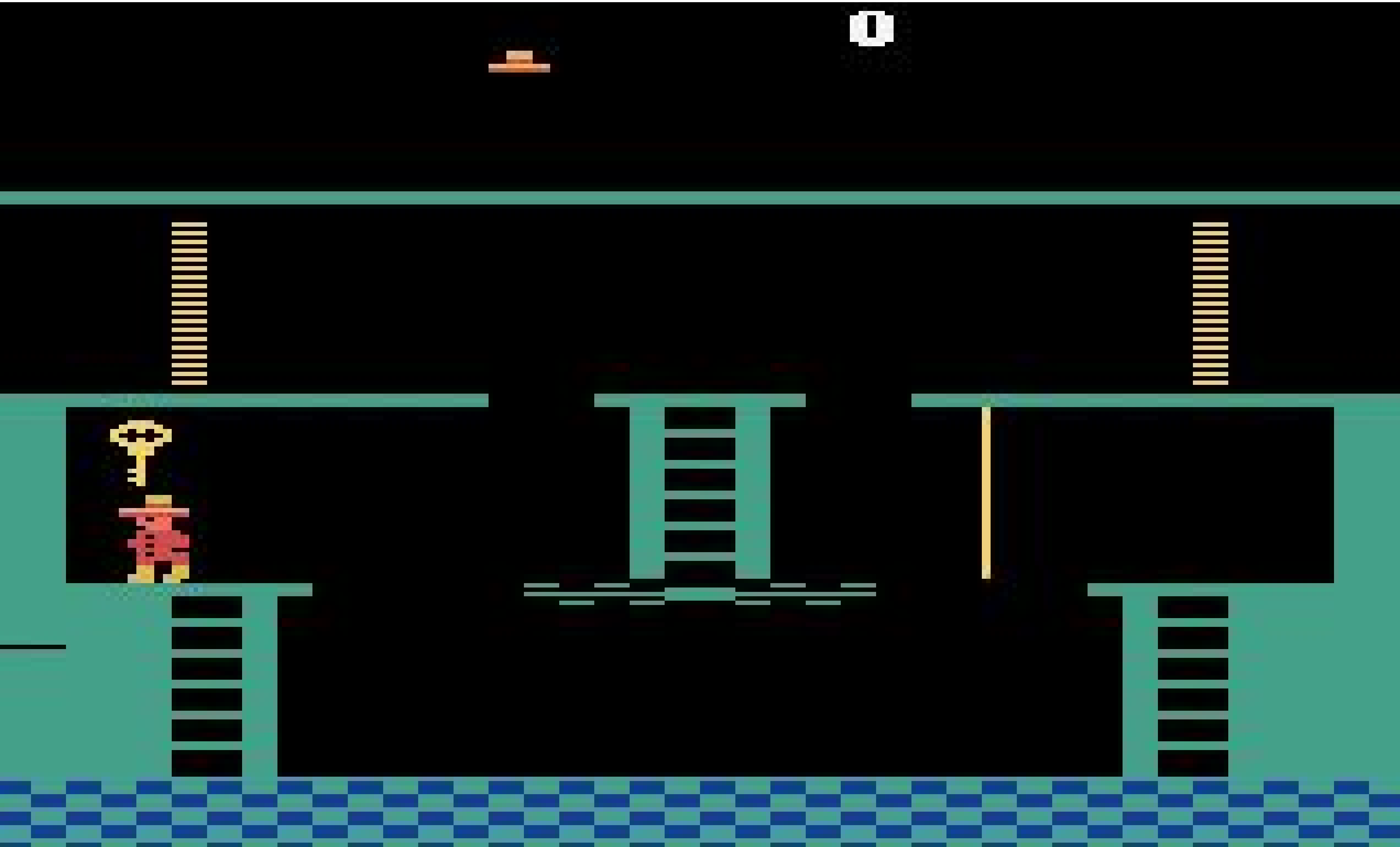
http://money.cnn.com/2017/08/18/news/economy/us-farmers-immigration-automation/index.html

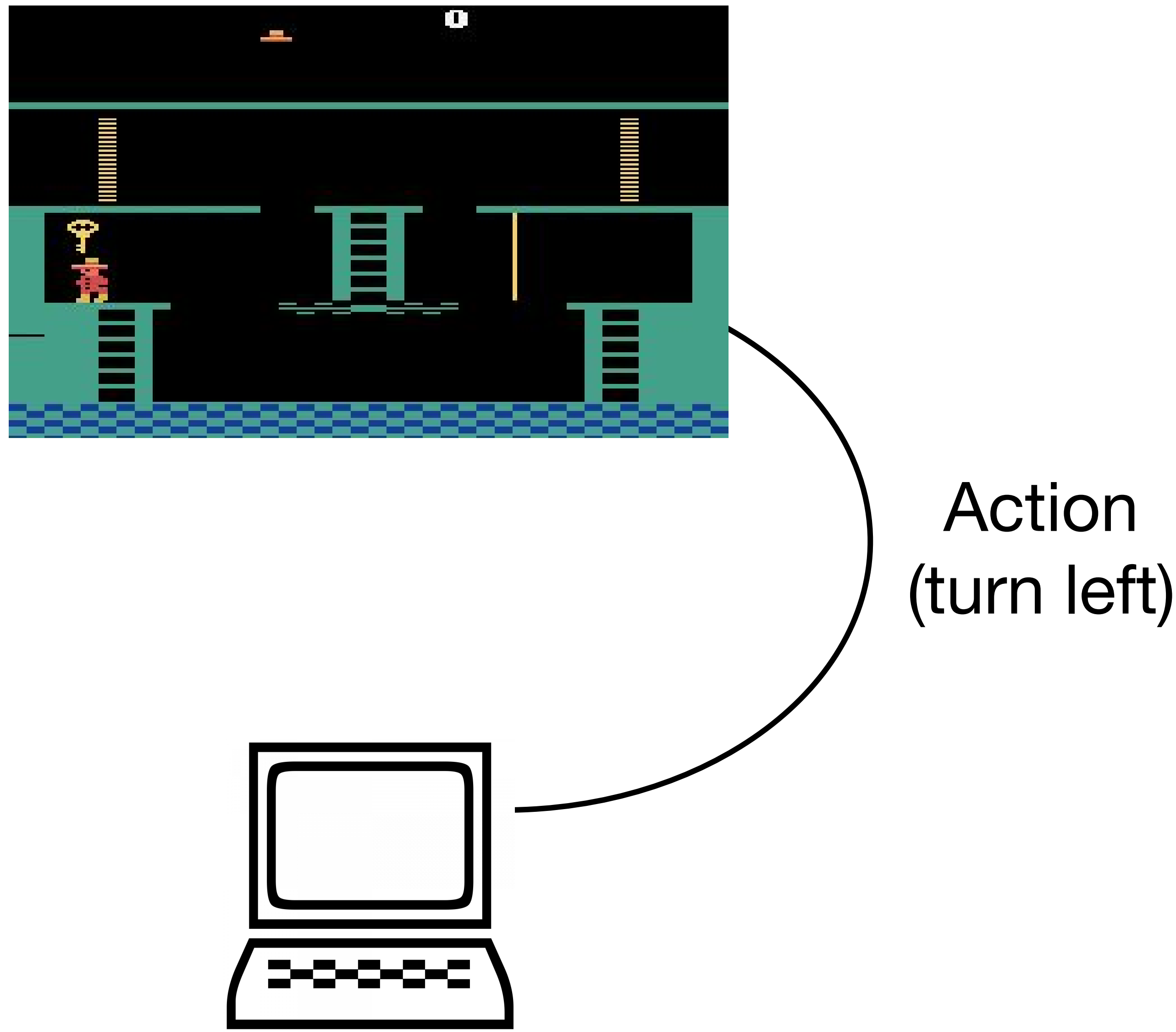
DAMIANGRETKA VIA GETTY IMAGES

http://www.huffingtonpost.co.uk/entry/artificial-intelligence-helping-doct
ors-diagnose-skin-cancer-faster_uk_599d428be4b0a296083b0778

Artificial Intelligence to Amplify People

Reinforcement Learning




Action
(turn left)

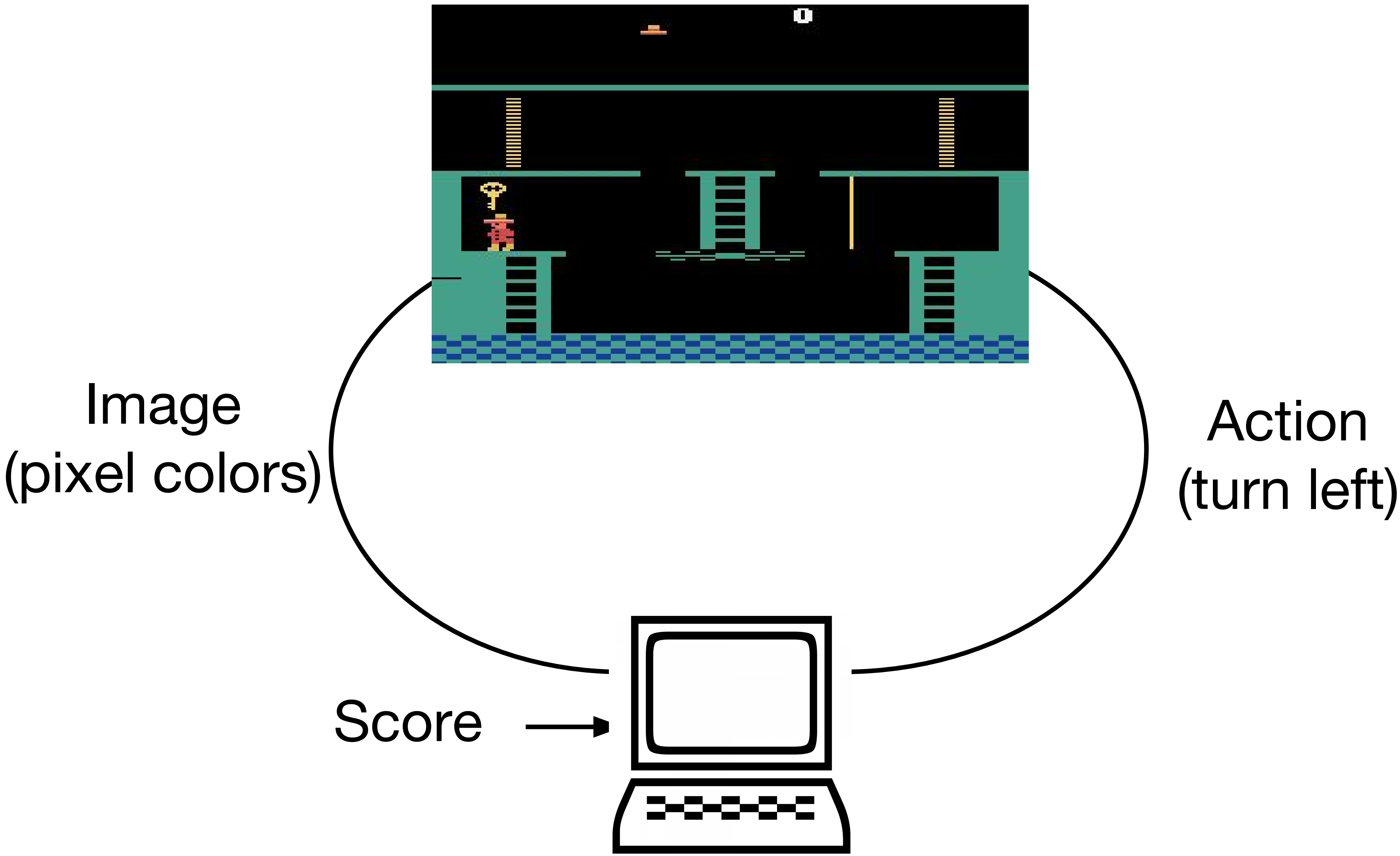
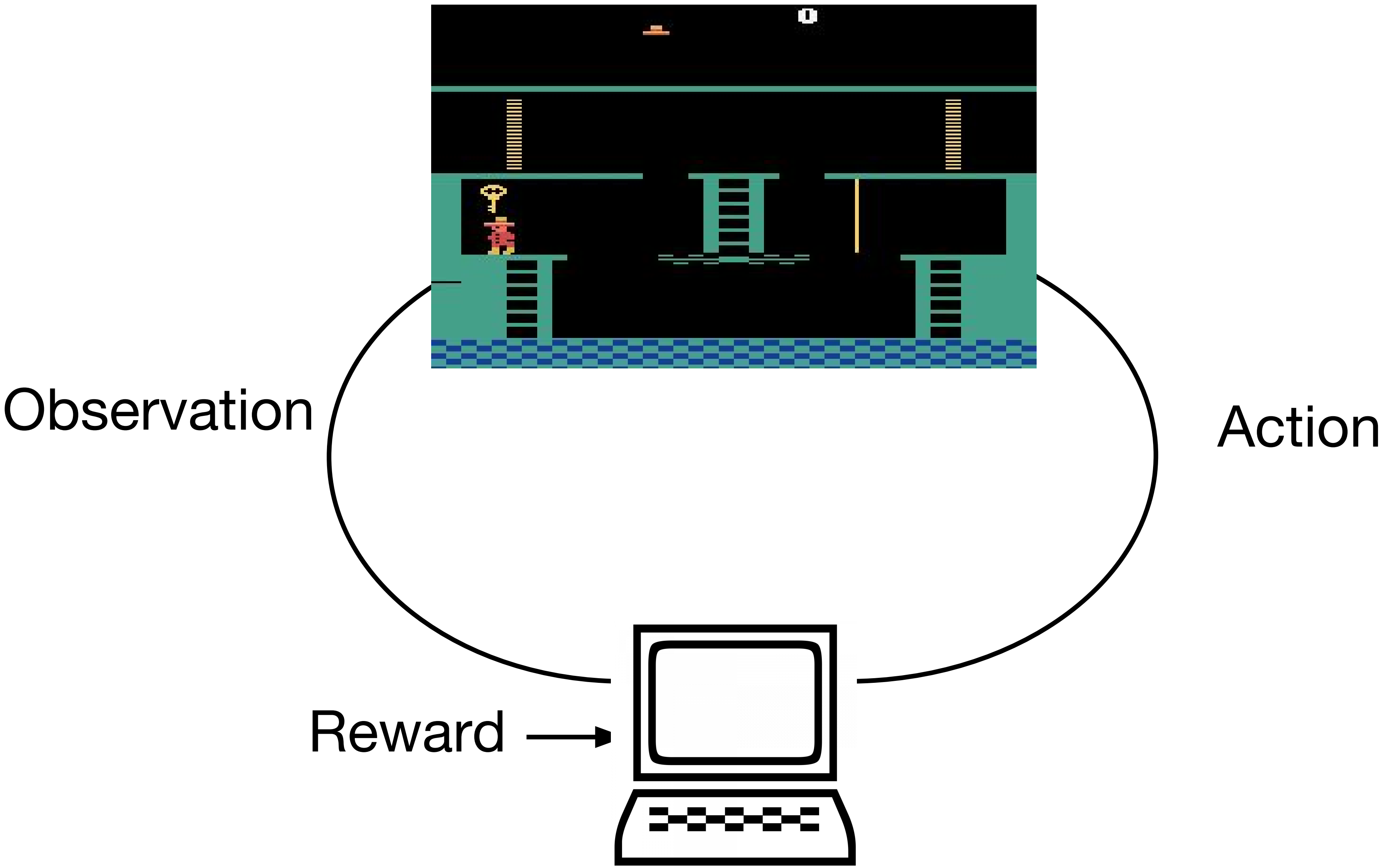
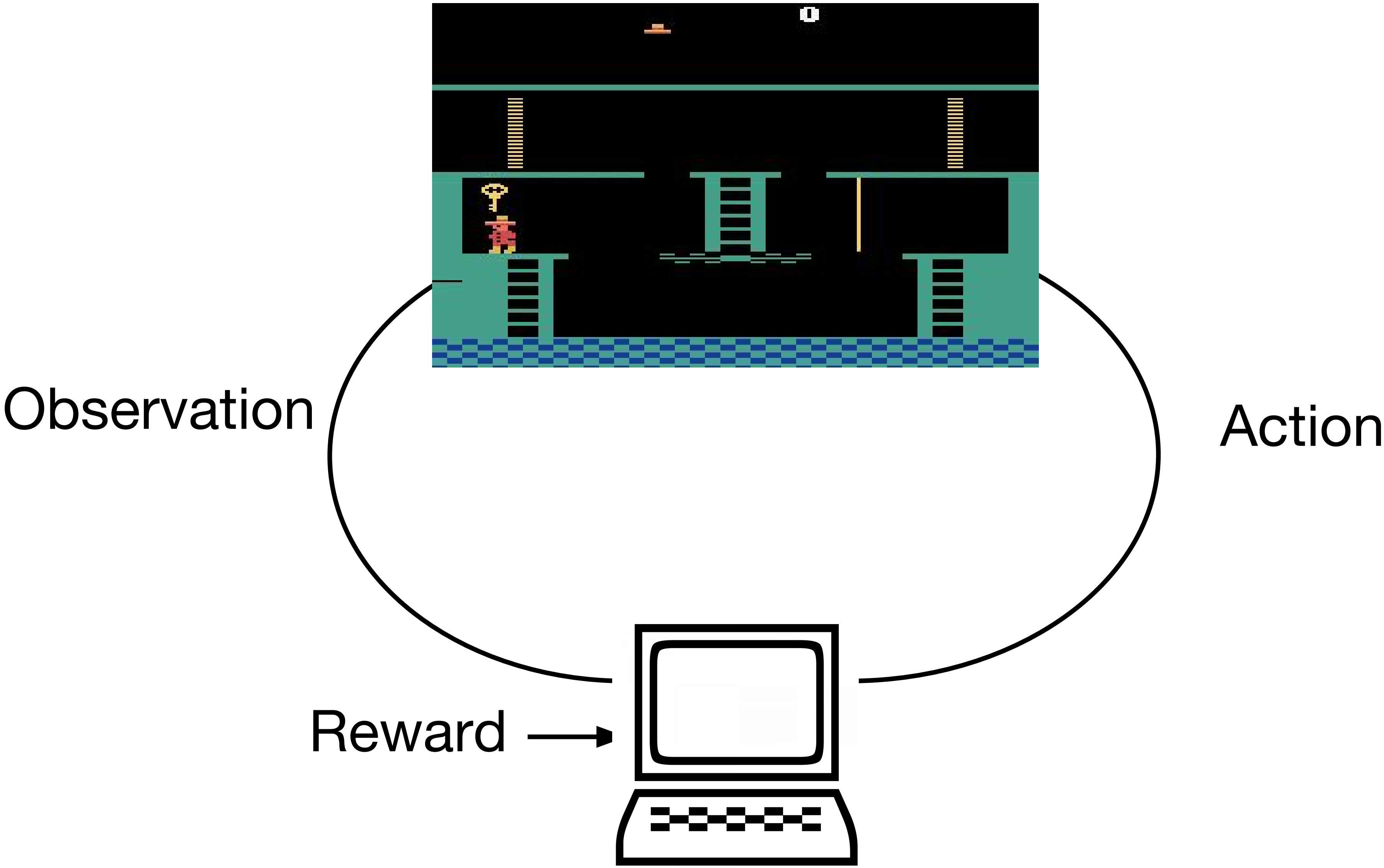


Image
(pixel colors)
Action
(turn left)
Score

Reinforcement Learning

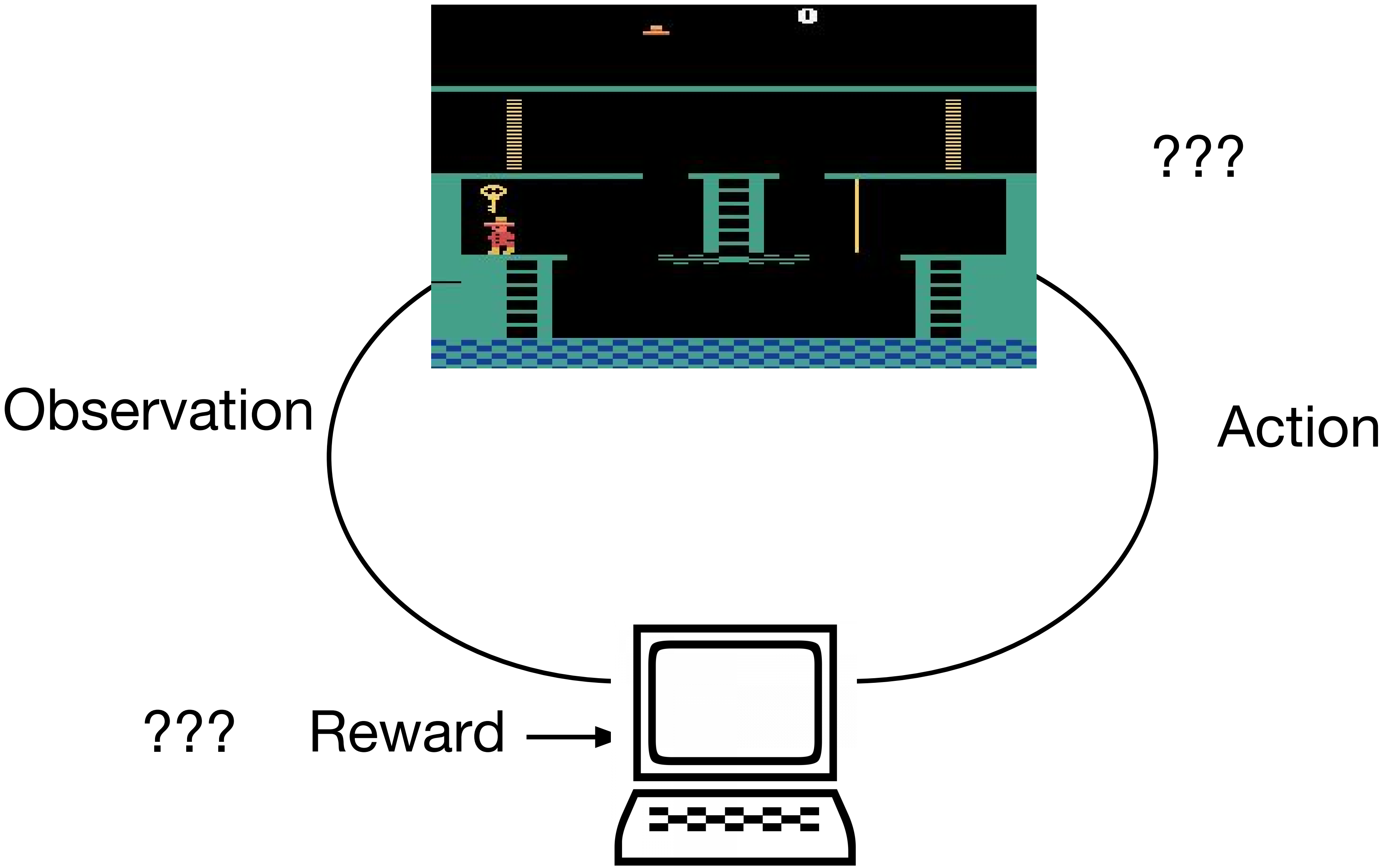


Reinforcement Learning



Policy: Map Observations → Actions
Goal: Choose actions to maximize expected rewards

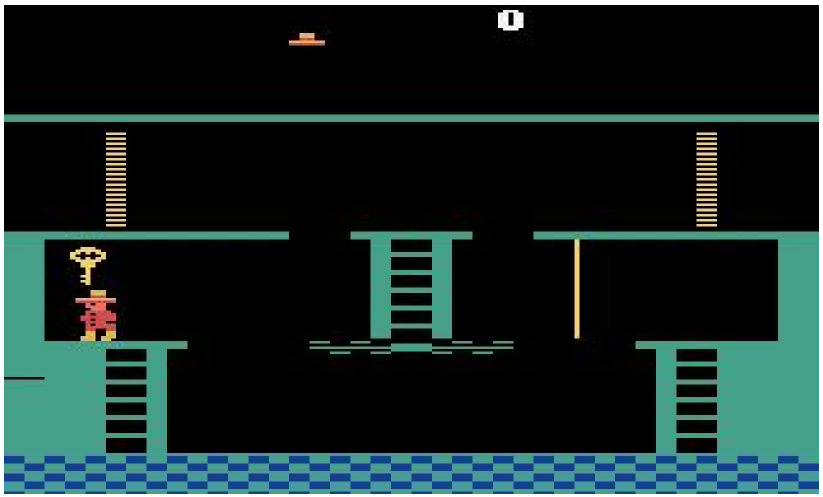
But Don’t Know How World Works!



Policy: Map Observations → Actions

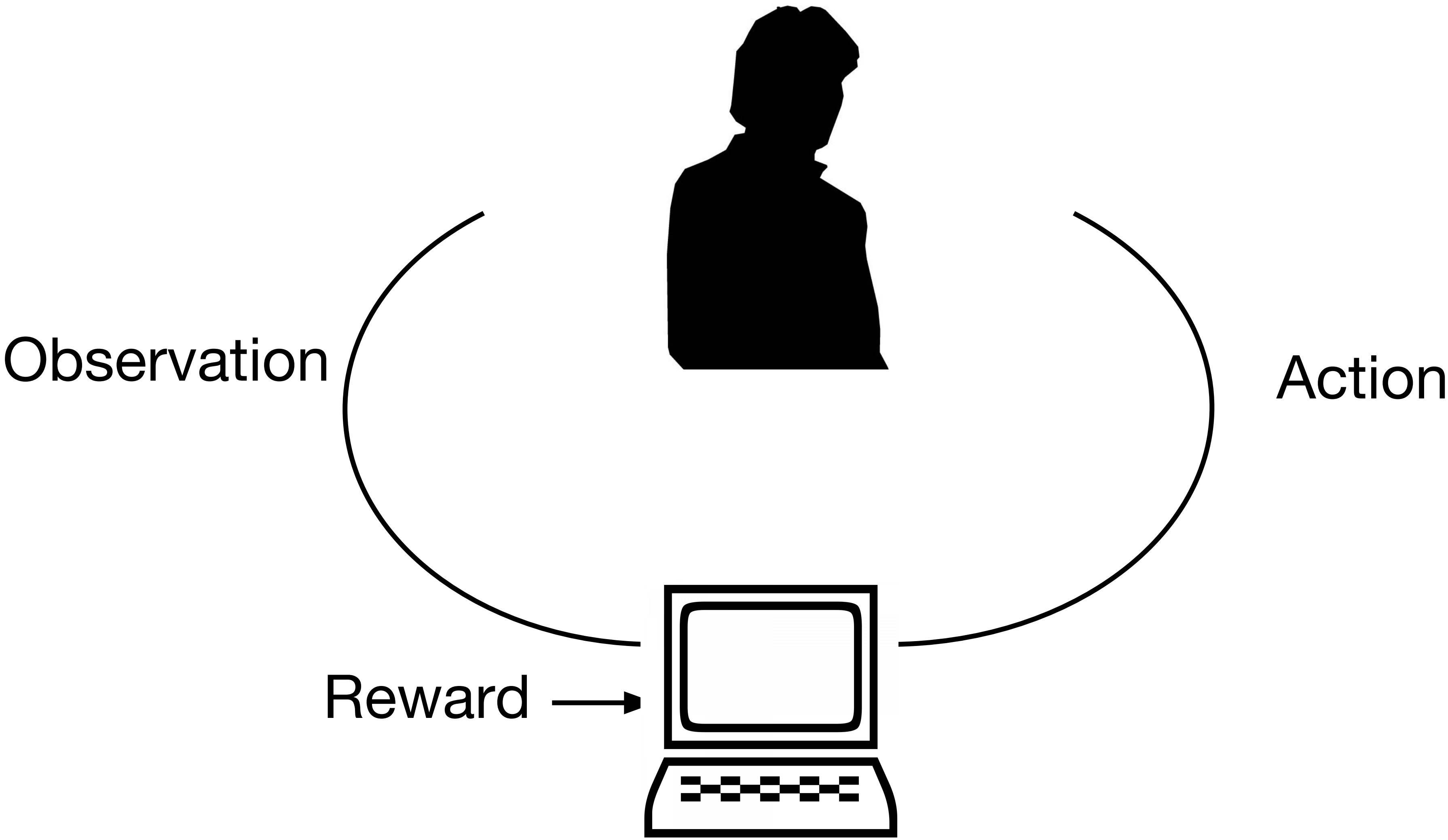
Goal: Choose actions to maximize expected rewards

Reinforcement Learning Progress



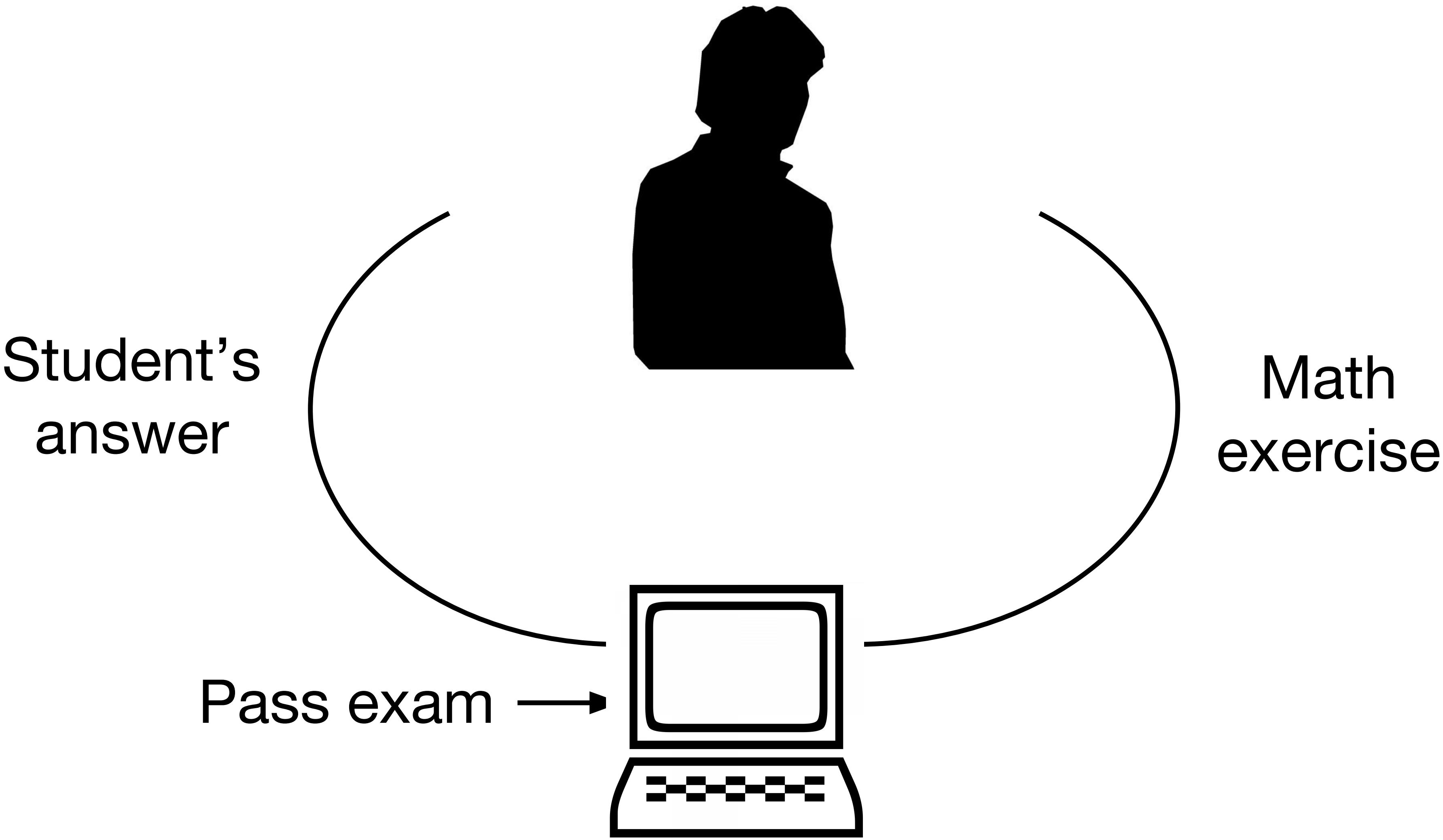
Real Potential: Humans & AI

Reinforcement Learning with and for People



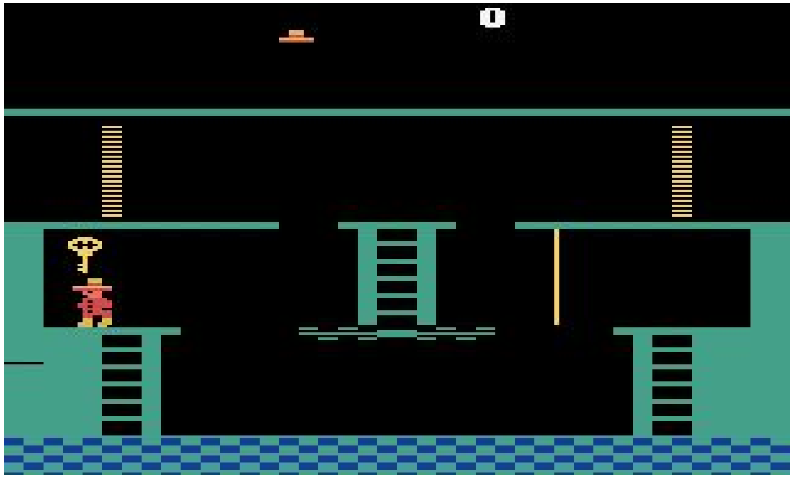
Policy: Map Observations → Actions
Goal: Choose actions to maximize expected rewards

Reinforcement Learning with and for People



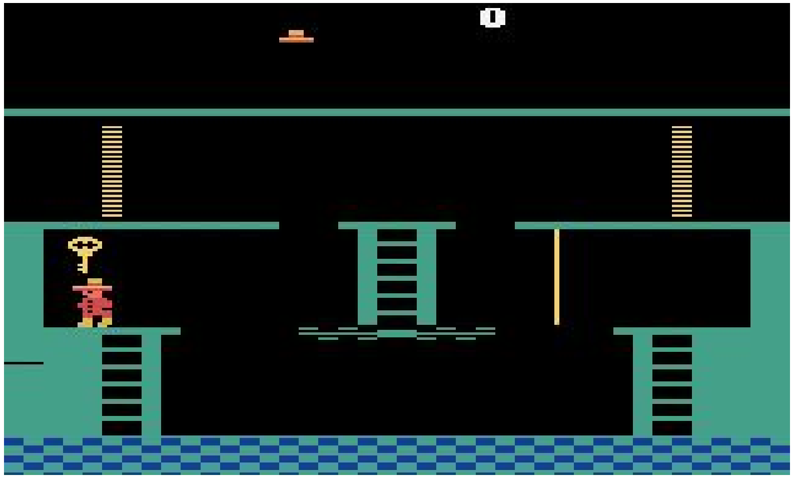
Policy: Map Observations → Actions

Goal: Choose actions to maximize student outcomes




r3d24
r3d32

≠



≠

Cheap to try things, or
Simulate

High stakes
Hard to model

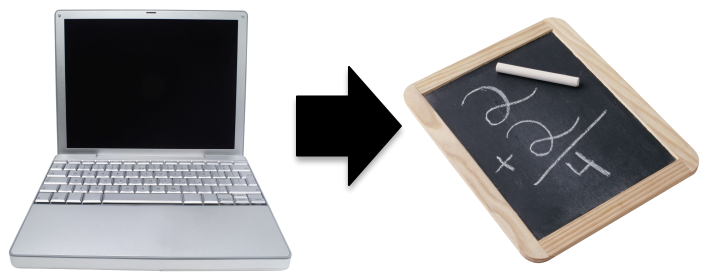
Reinforcement Learning & Learning to Promote Learning

Making better decisions by

1) Learning from past experience
2) Having humans help machines

A Classrooms

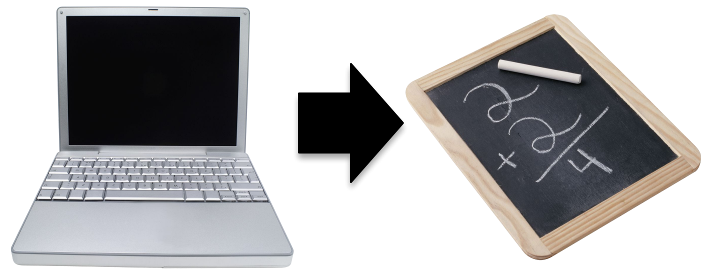
Avg Score: 95

A Classrooms  Avg Score: 95

B Classrooms  Avg Score: 92

What should we do for a new student?

A Classrooms

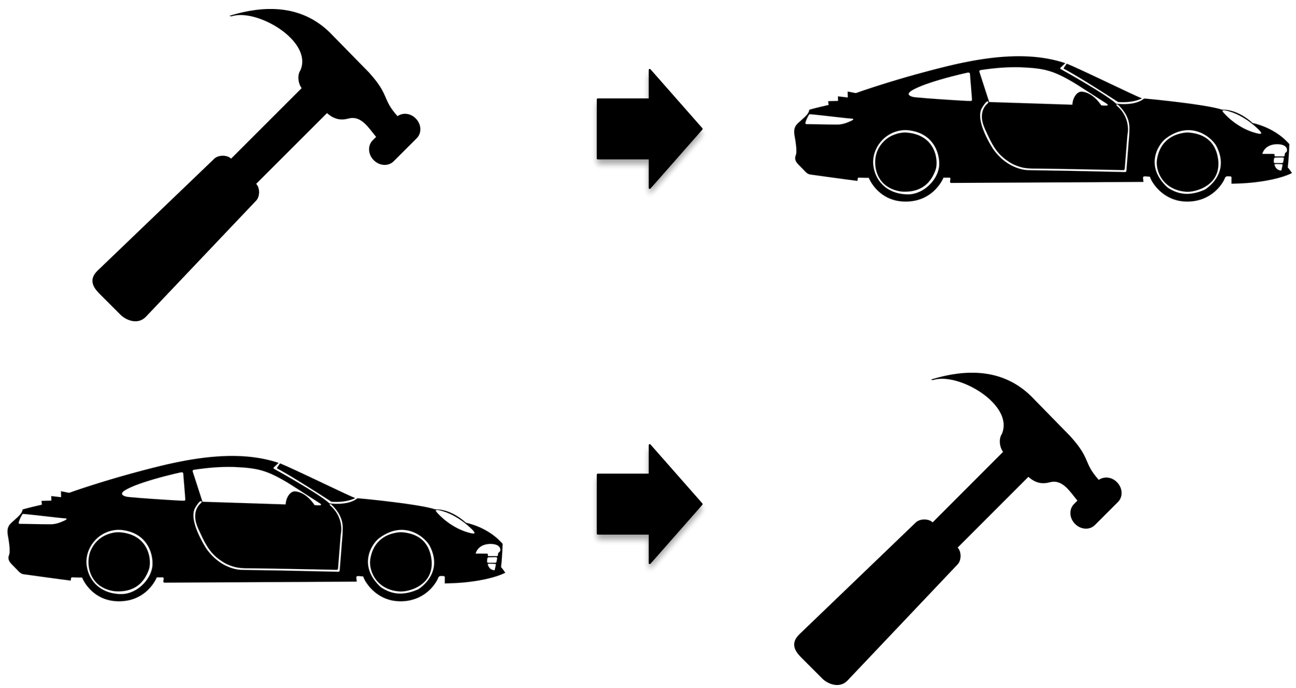


Avg Score: 95

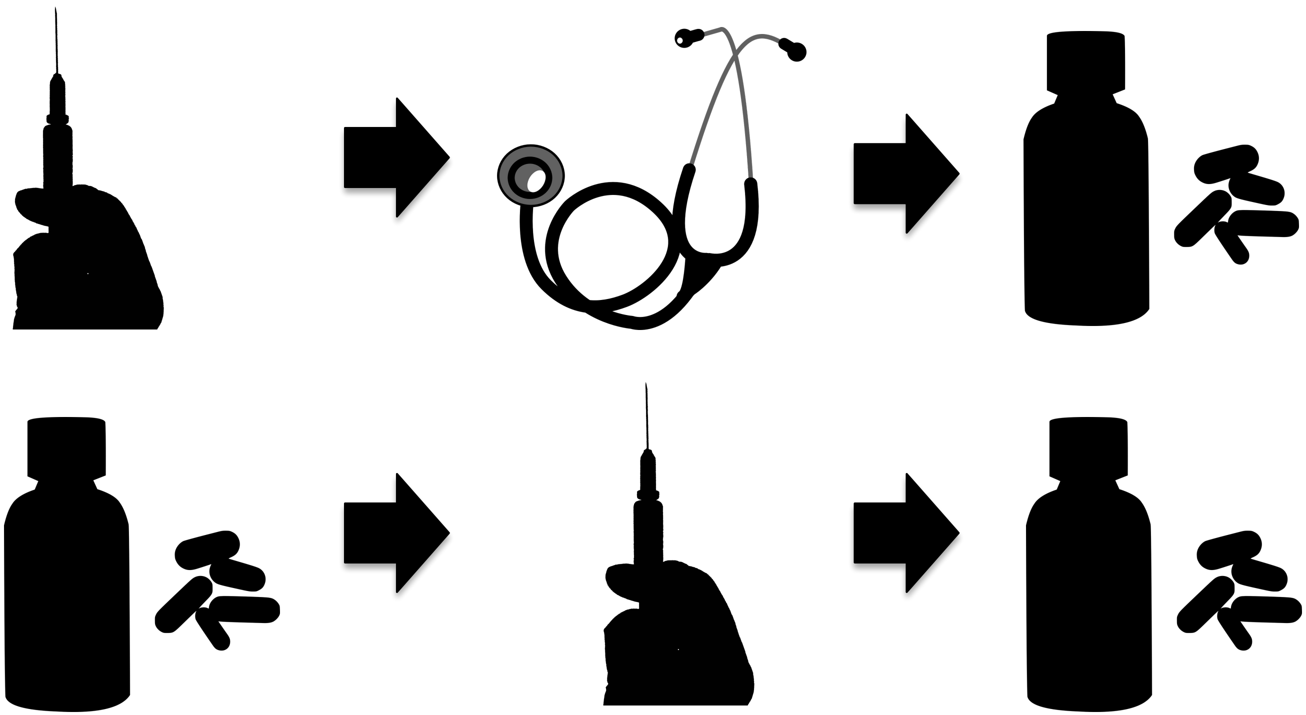
B Classrooms

Avg Score: 92

Comes Up in Many Domains: e.g. Equipment Maintenance Scheduling



Comes Up in Many Domains: e.g. Patient Treatment Ordering



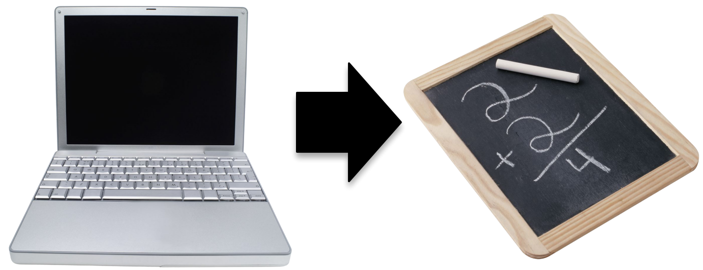
Core Aspect of Intelligent Behavior

Data about past decisions & outcomes

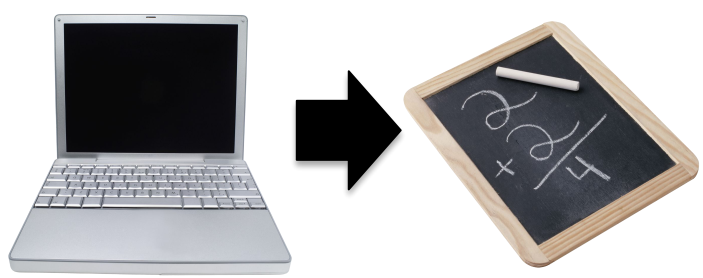
How best to act in the future?

Image: https://upload.wikimedia.org/wikipedia/commons/f/f0/DARPA_Big_Data.jpg

Challenge: Counterfactual Reasoning

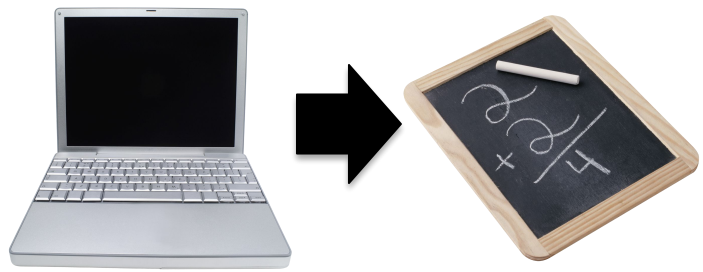
A Classrooms  Avg Score: 95

B Classrooms  Avg Score: 92

B Classrooms  Avg Score: ????

Challenge: Generalization to Untried Policies

A Classrooms

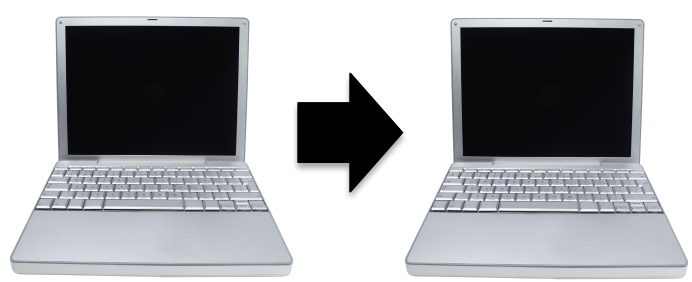


Avg Score: 95

B Classrooms

Avg Score: 92

B Classrooms



Avg Score: ????


Level 1:8
Fork
1
1/2
1/2
1/2
1/2
1/2
1/2
1/2
1/2
MENU
OPTIONS

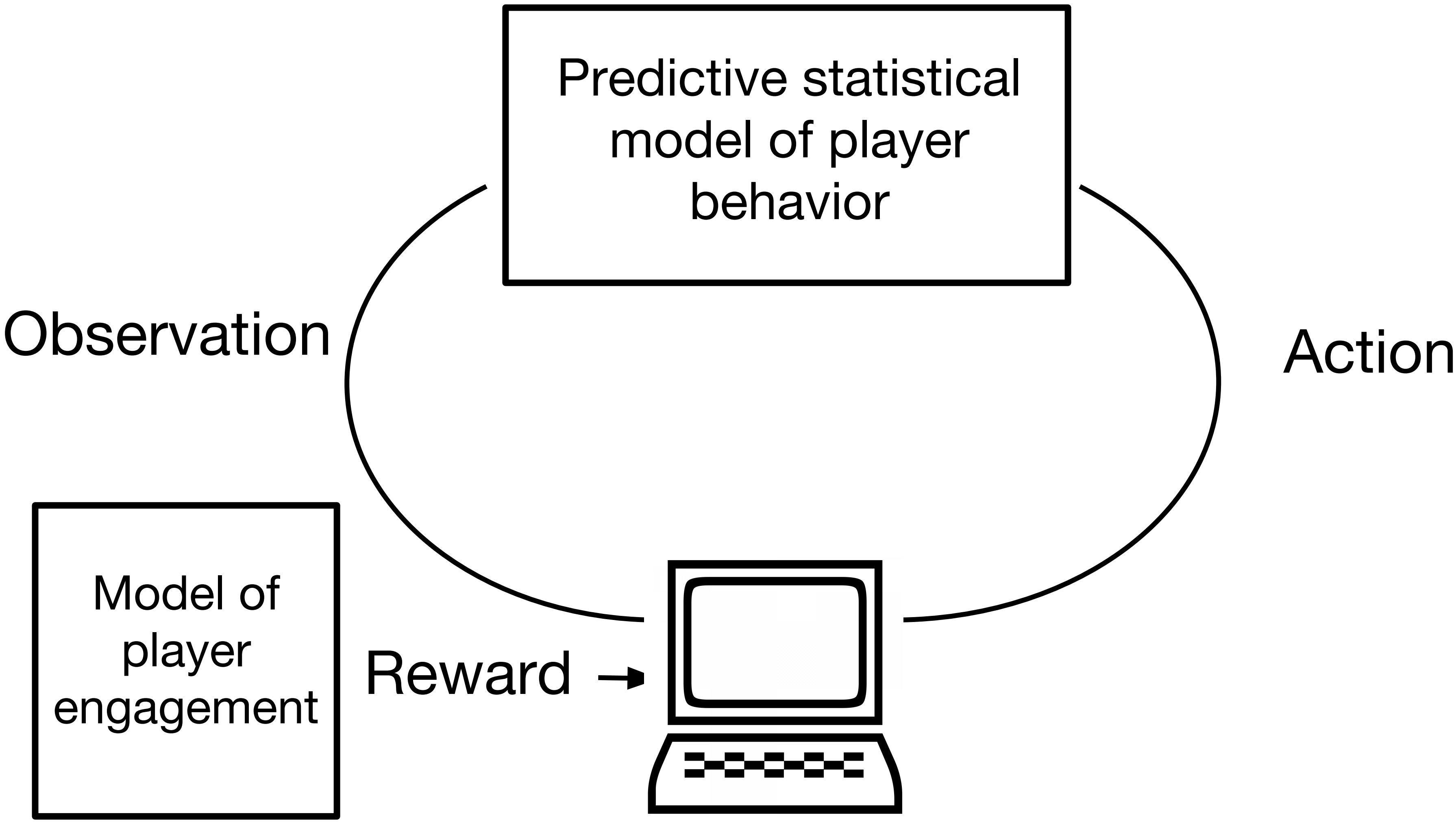
Policy: Player state → level
Goal: Maximize engagement
Old data: ~11,000 students

Statistical Predictive model

(e.g. Predict if student will get next level correct)

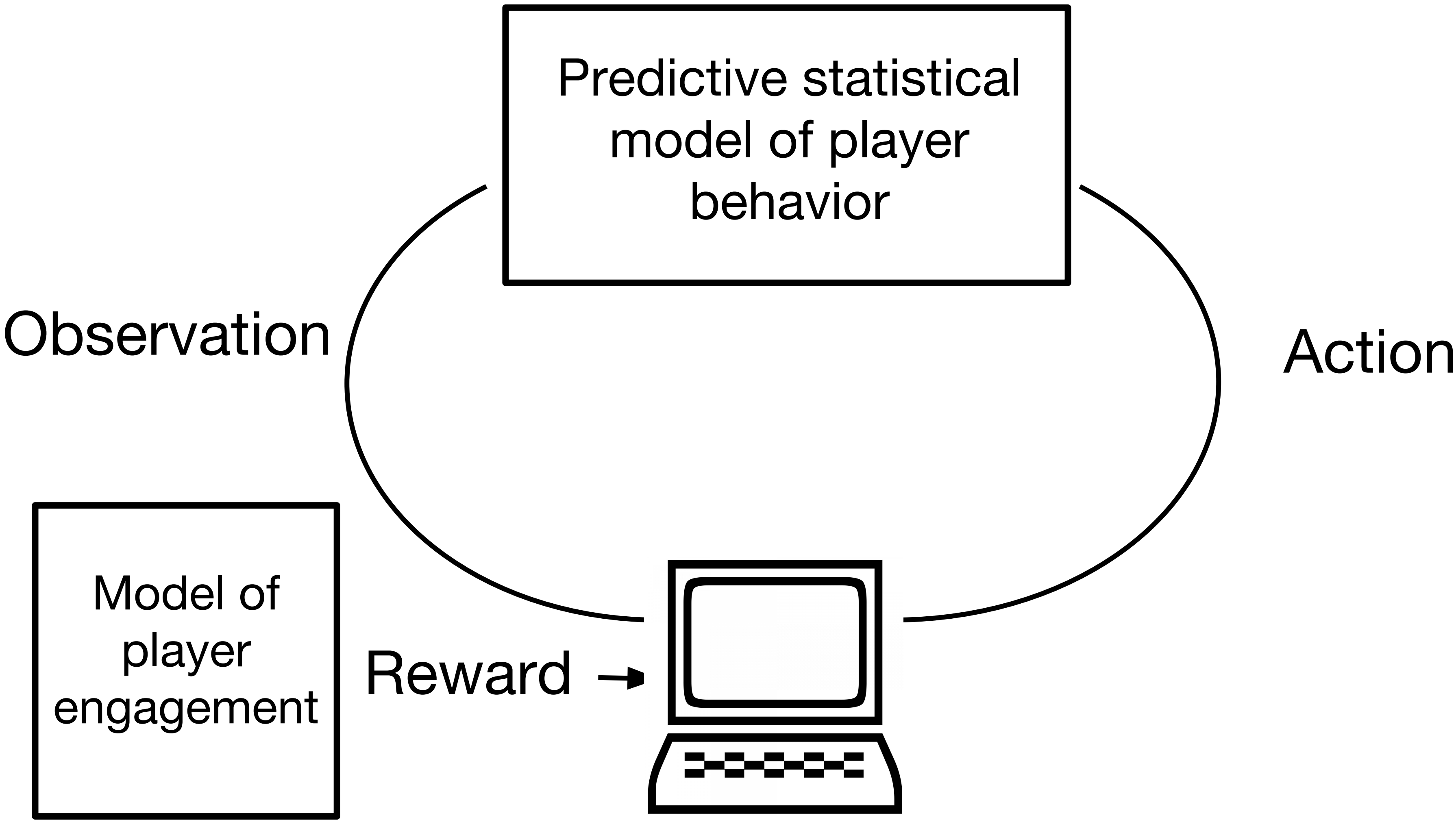
Image: https://upload.wikimedia.org/wikipedia/commons/f/f0/DARPA_Big_Data.jpg

Use Models as a Simulator



Goal: Choose actions to maximize expected rewards

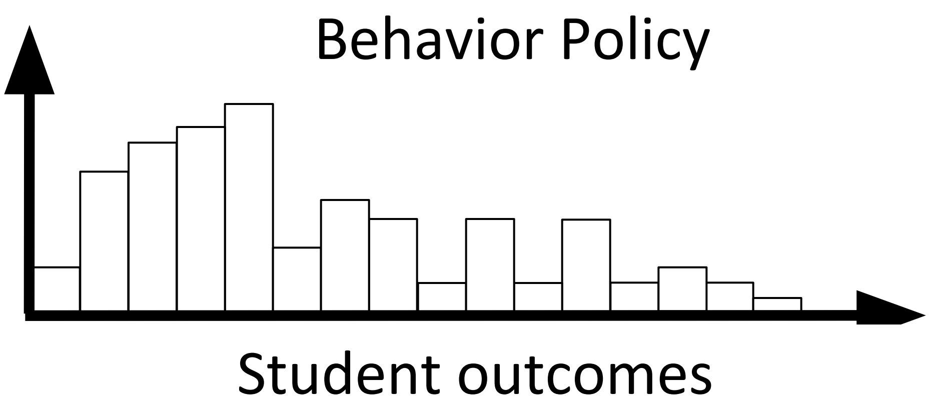
Problem: When is a Model Good Enough?

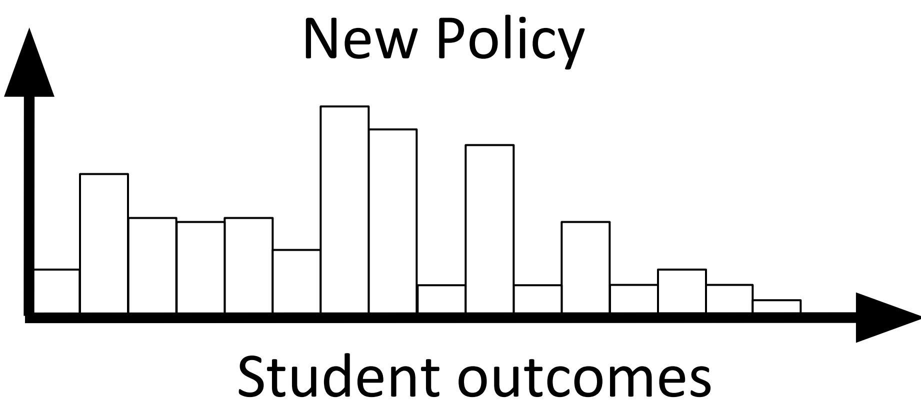


Goal: Choose actions to maximize expected rewards

Alternative: Reweigh Old Experience to Look Like New Policy

- No statistical predictive model assumptions

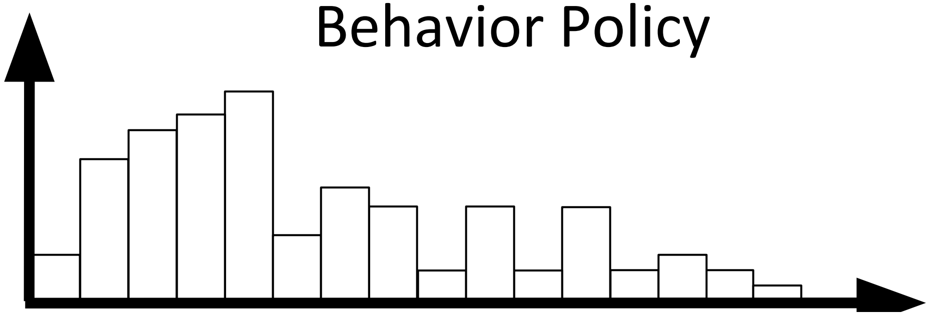




Alternative: Reweigh Old Experience to Look Like New Policy

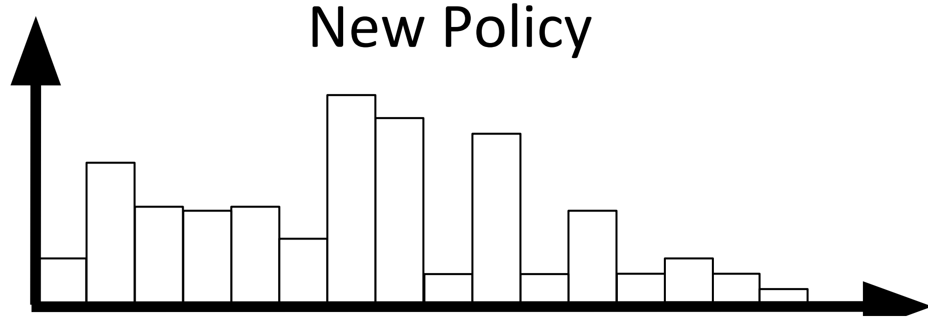
- No statistical predictive model assumptions

Behavior Policy



Student outcomes

New Policy



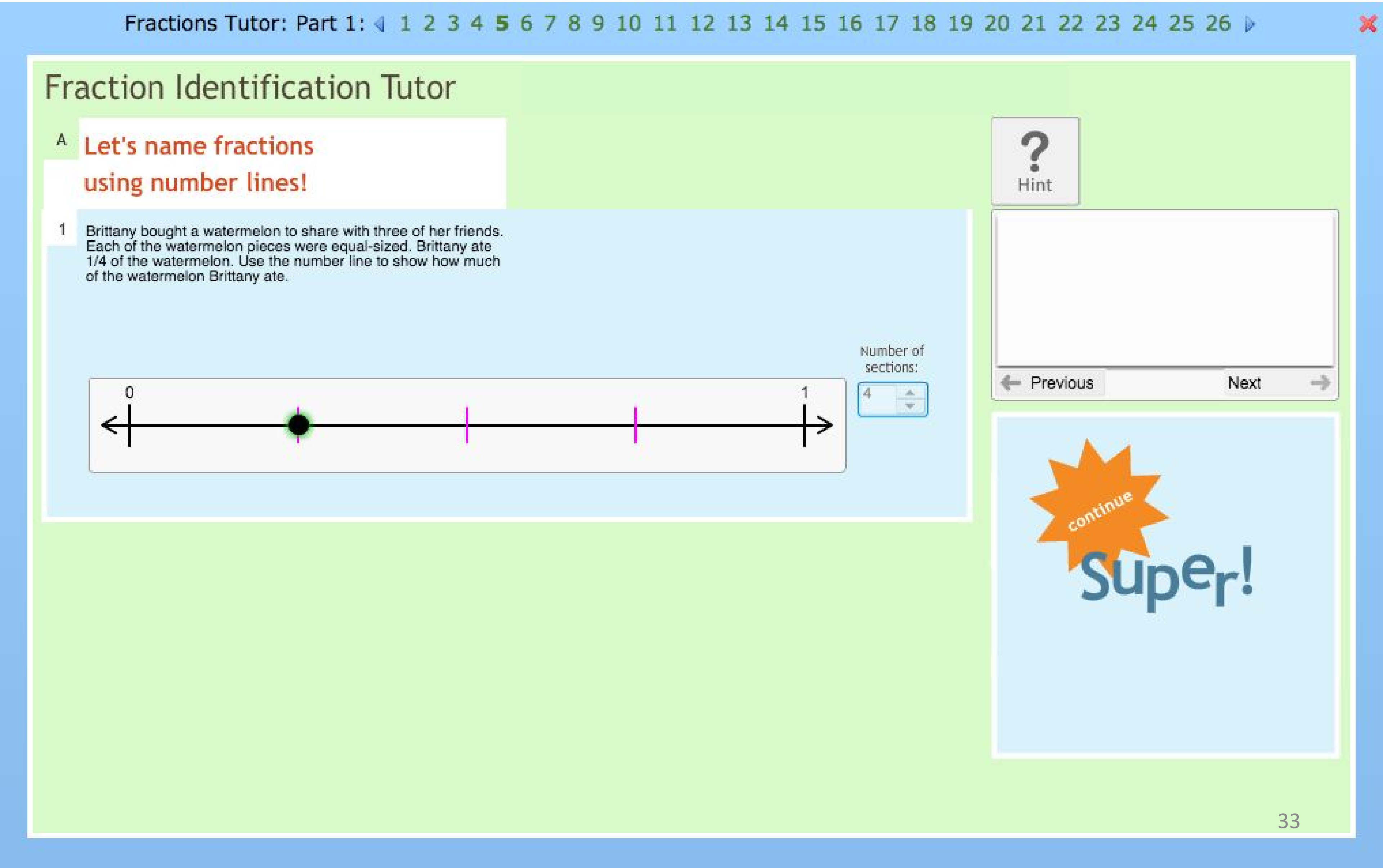
Student outcomes

- Unbiased* estimate of new policy's performance

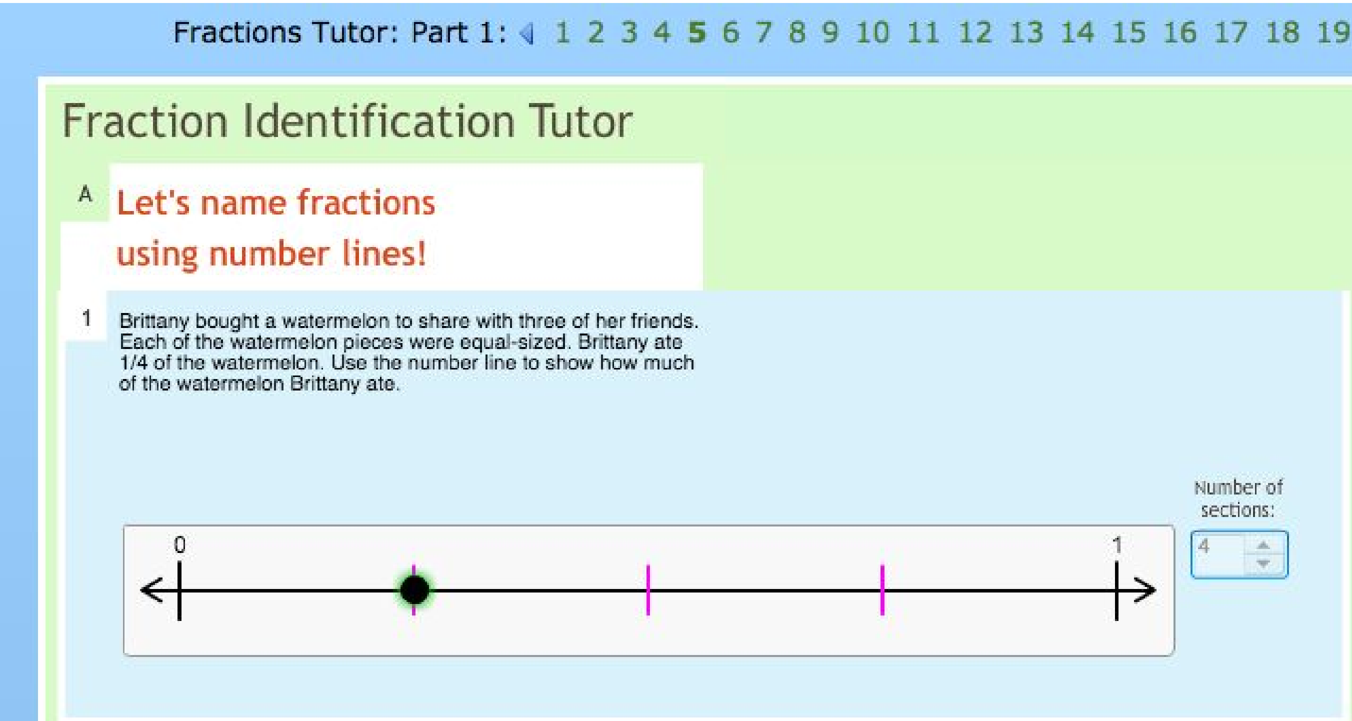
*Under mild assumptions

We used to find a policy with 30% higher engagement (Mandel et al. 2014)

When Making Many Decisions...



Towards Better Estimates of New Policies



- Trade bias and variance
- New methods to combine models & direct evaluation (Guo, Thomas, B 2017; Thomas and B 2016)

Towards Using Old Data to Confidently Identify Better Policies for Future Use

Data about past decisions & outcomes

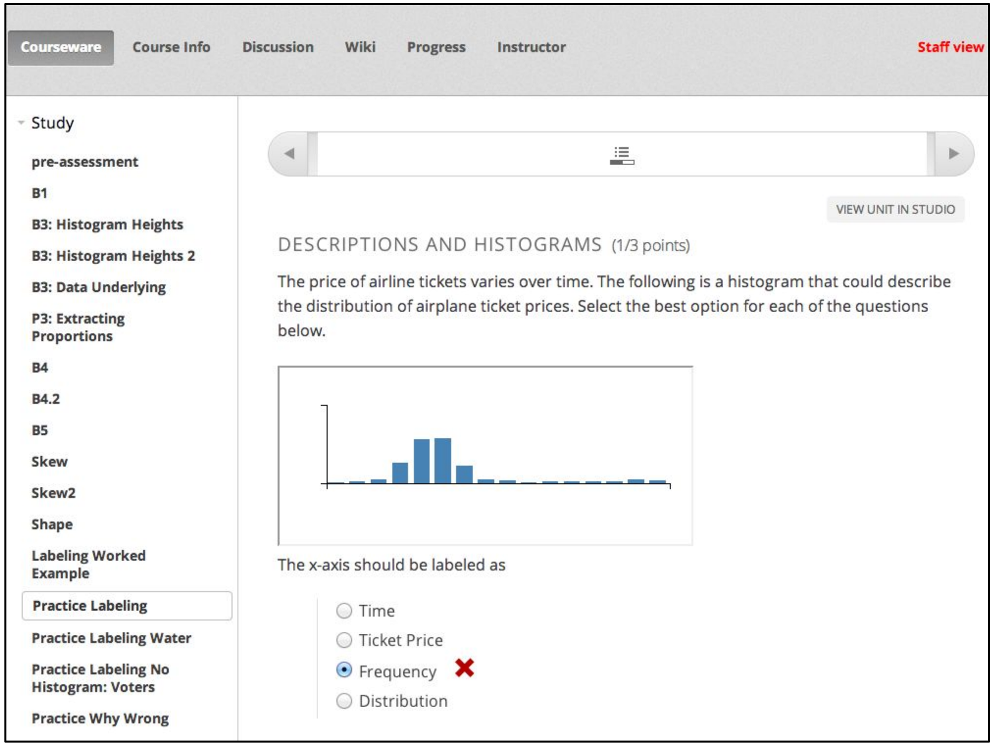
➡ How best to act in the future?

Image: https://upload.wikimedia.org/wikipedia/commons/f/f0/DARPA_Big_Data.jpg

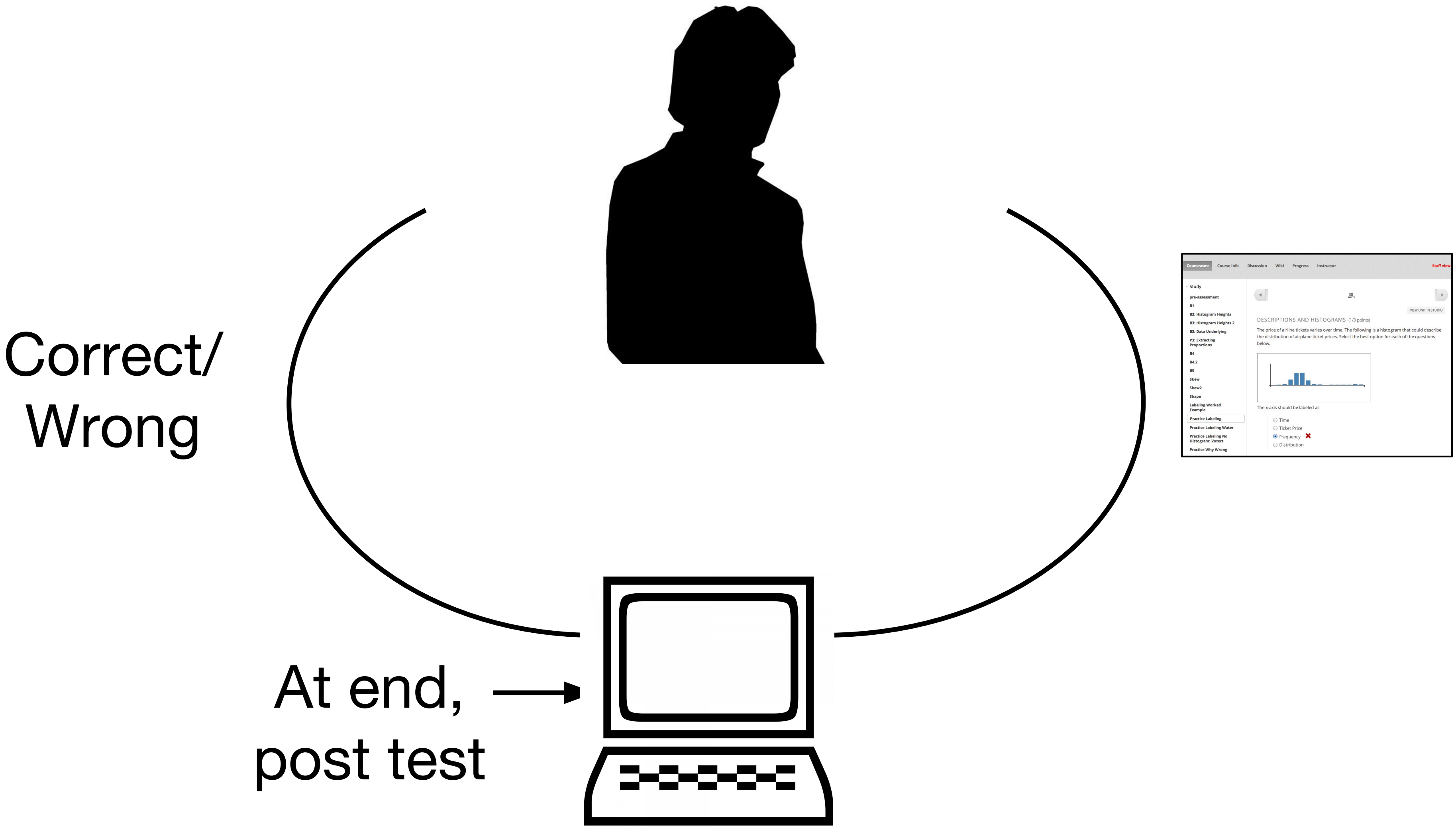
Reinforcement Learning & Learning to Promote Learning

- Making better decisions by
  - Learning from past experience
  - Having humans help machines

Histogram Tutor

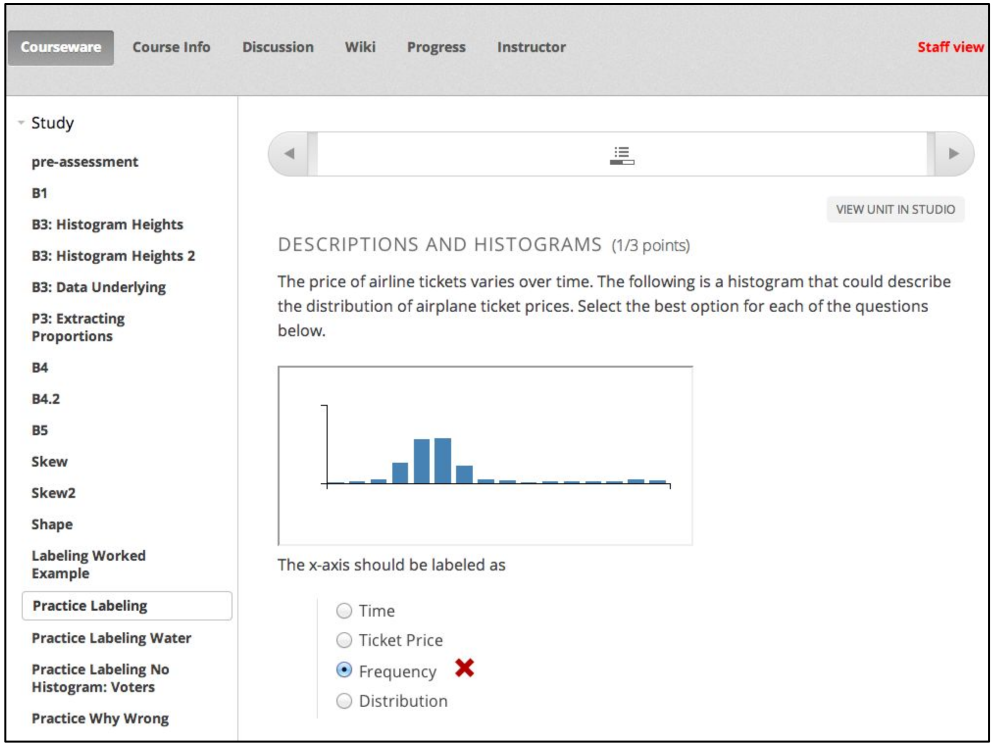


Continually Improving Tutoring System

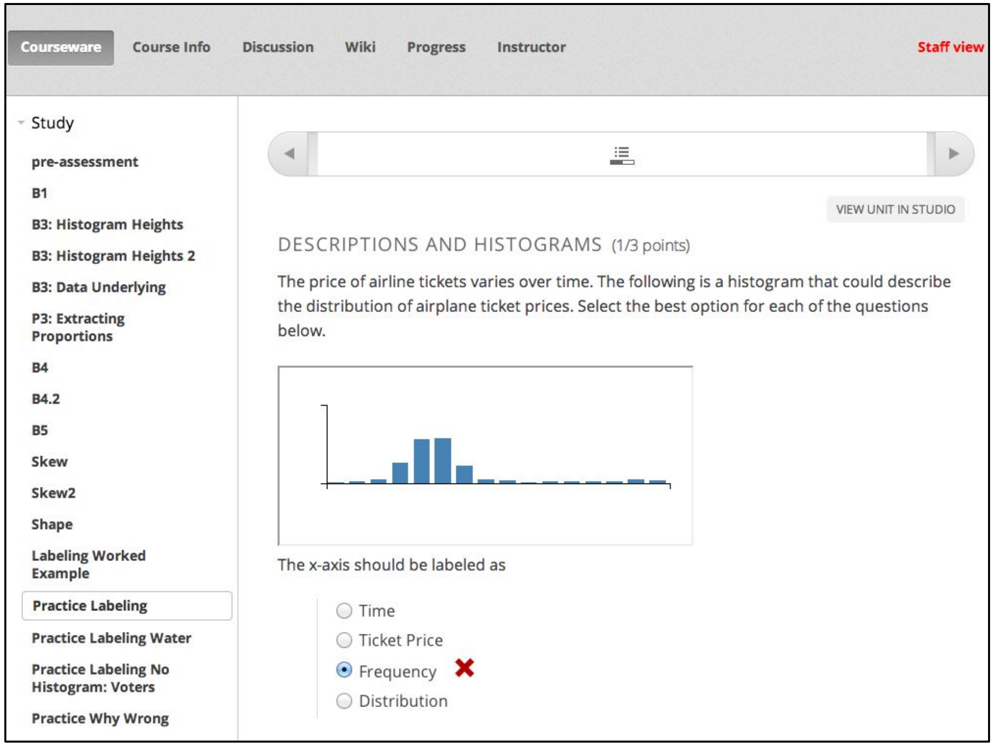


Improving Across Many Students

Over Time Tutoring System Stopped Giving Some Problems to Students



System Self-Diagnosed that Problems Weren't Helping Student Learning

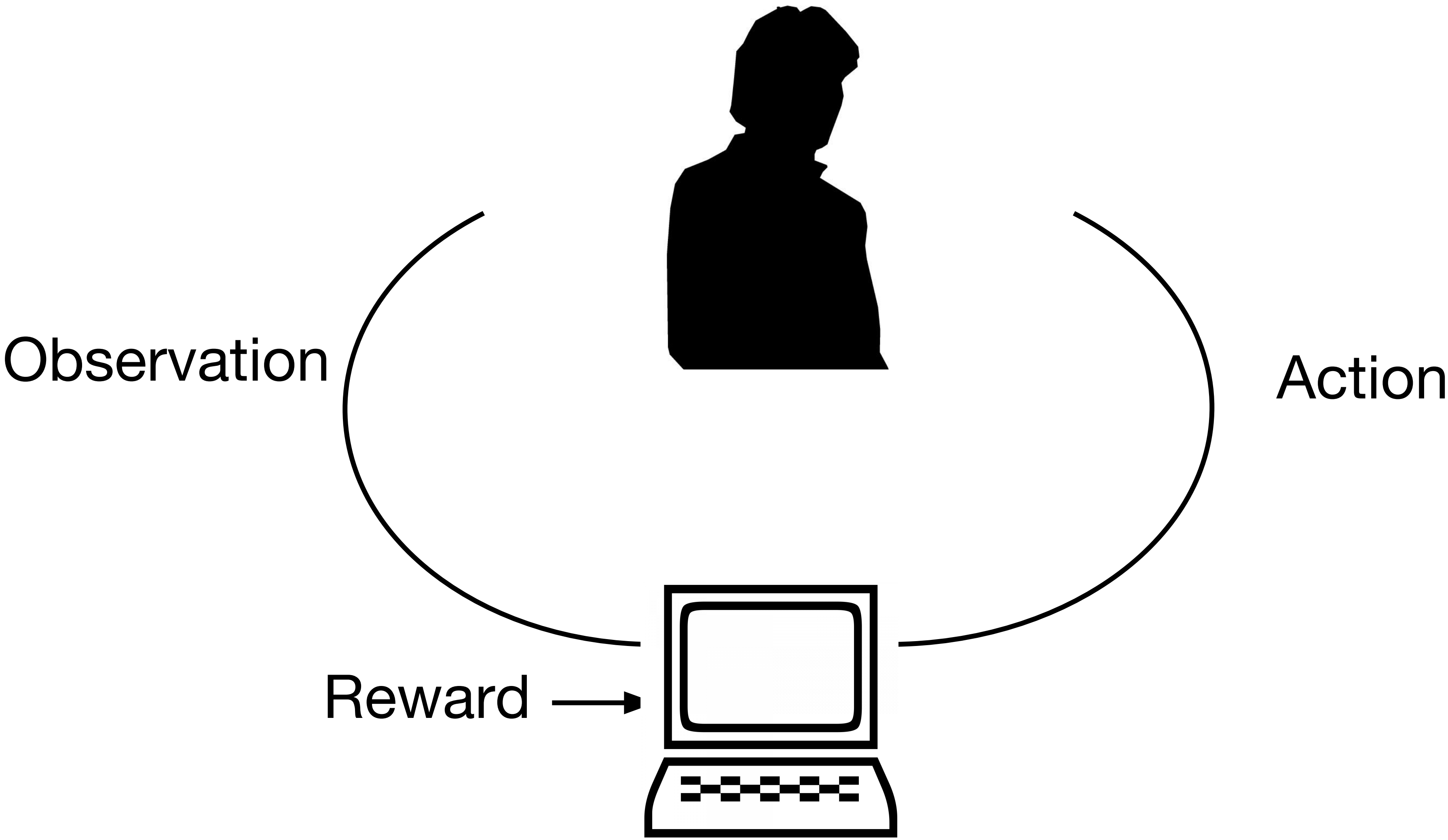


Humans are Invention Machines

New actions

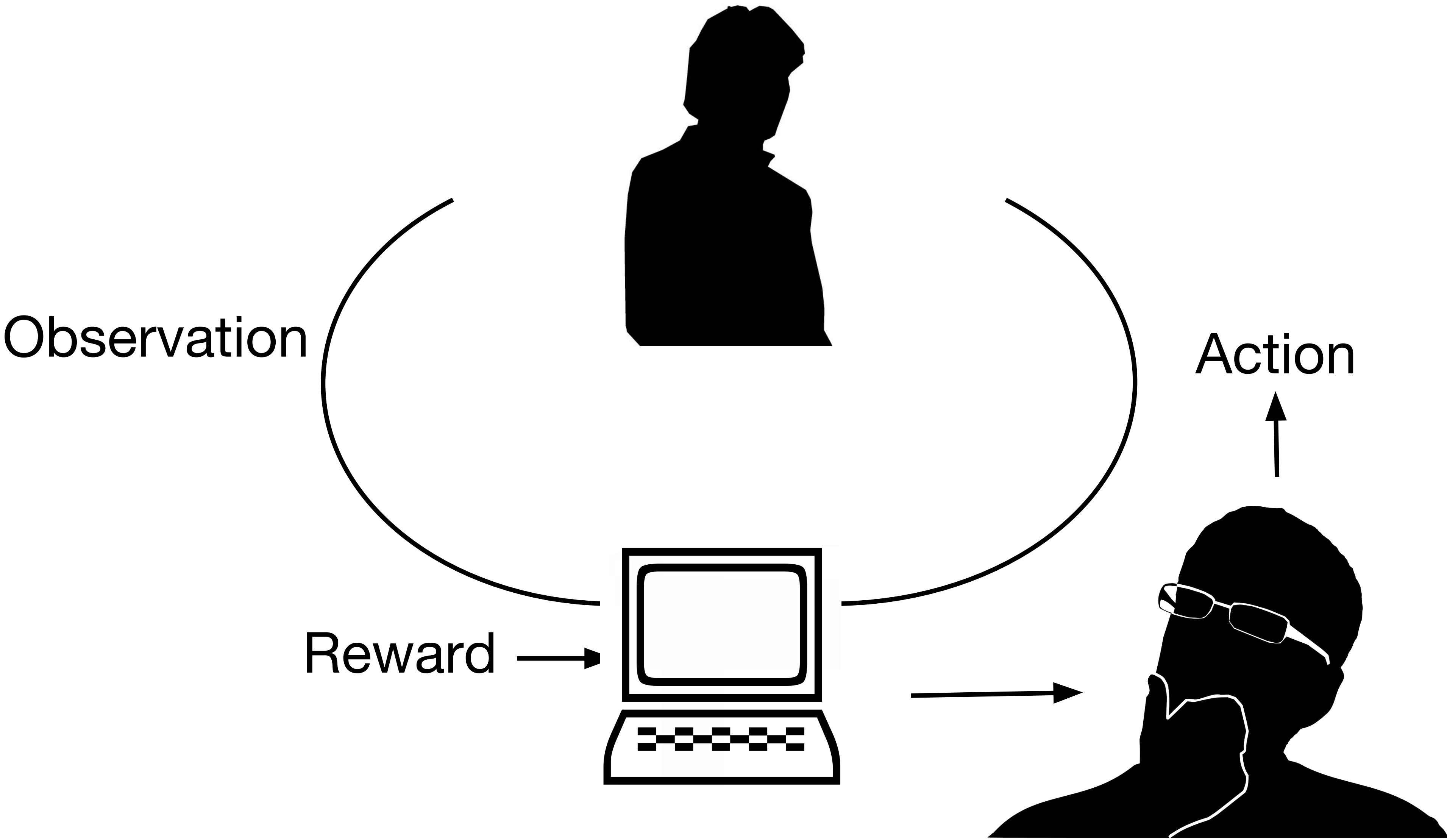
New sensors

Reinforcement Learning



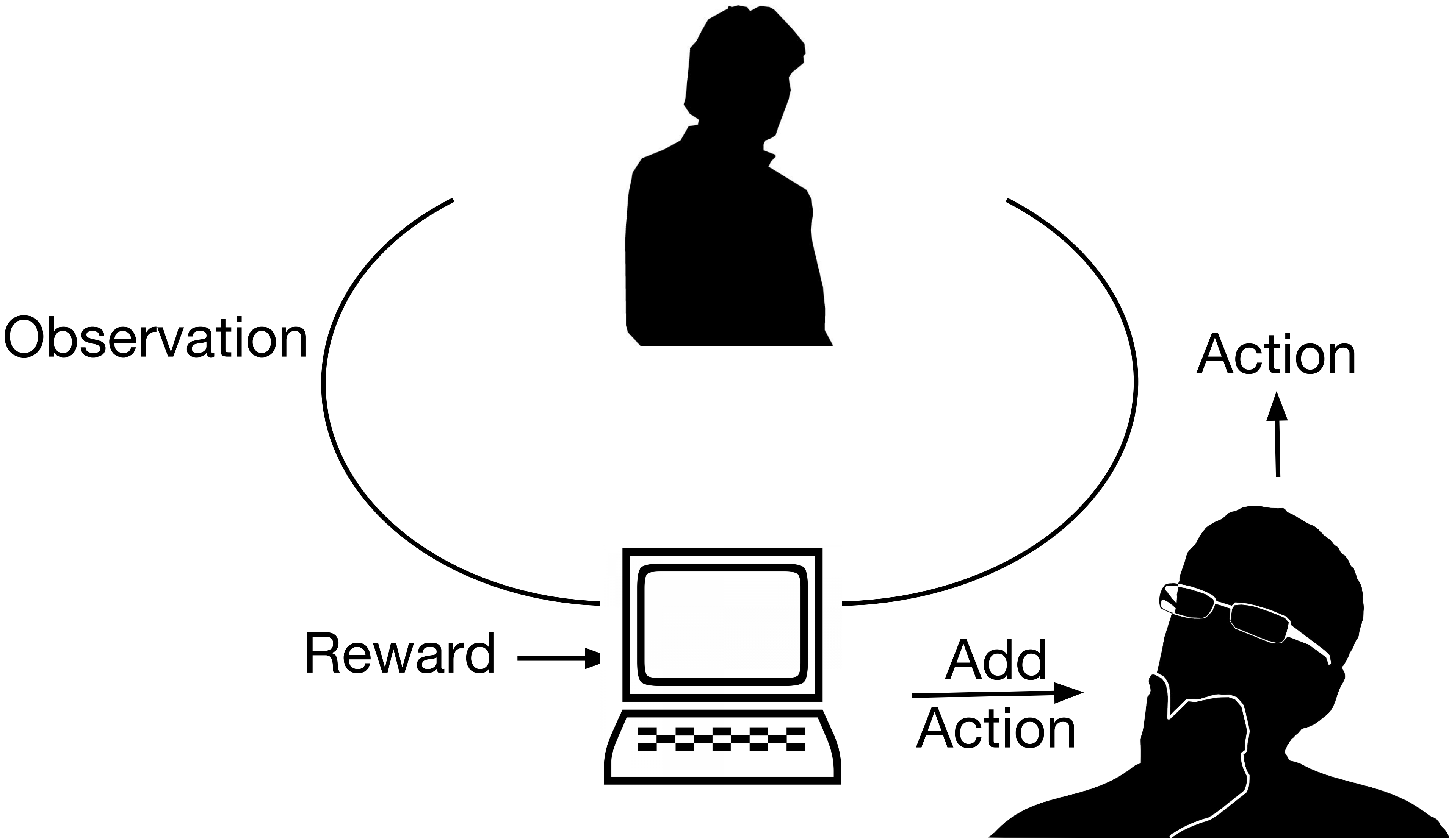
Goal: Choose actions to maximize expected rewards

Human in the Loop Reinforcement Learning



Goal: Choose actions to maximize expected rewards

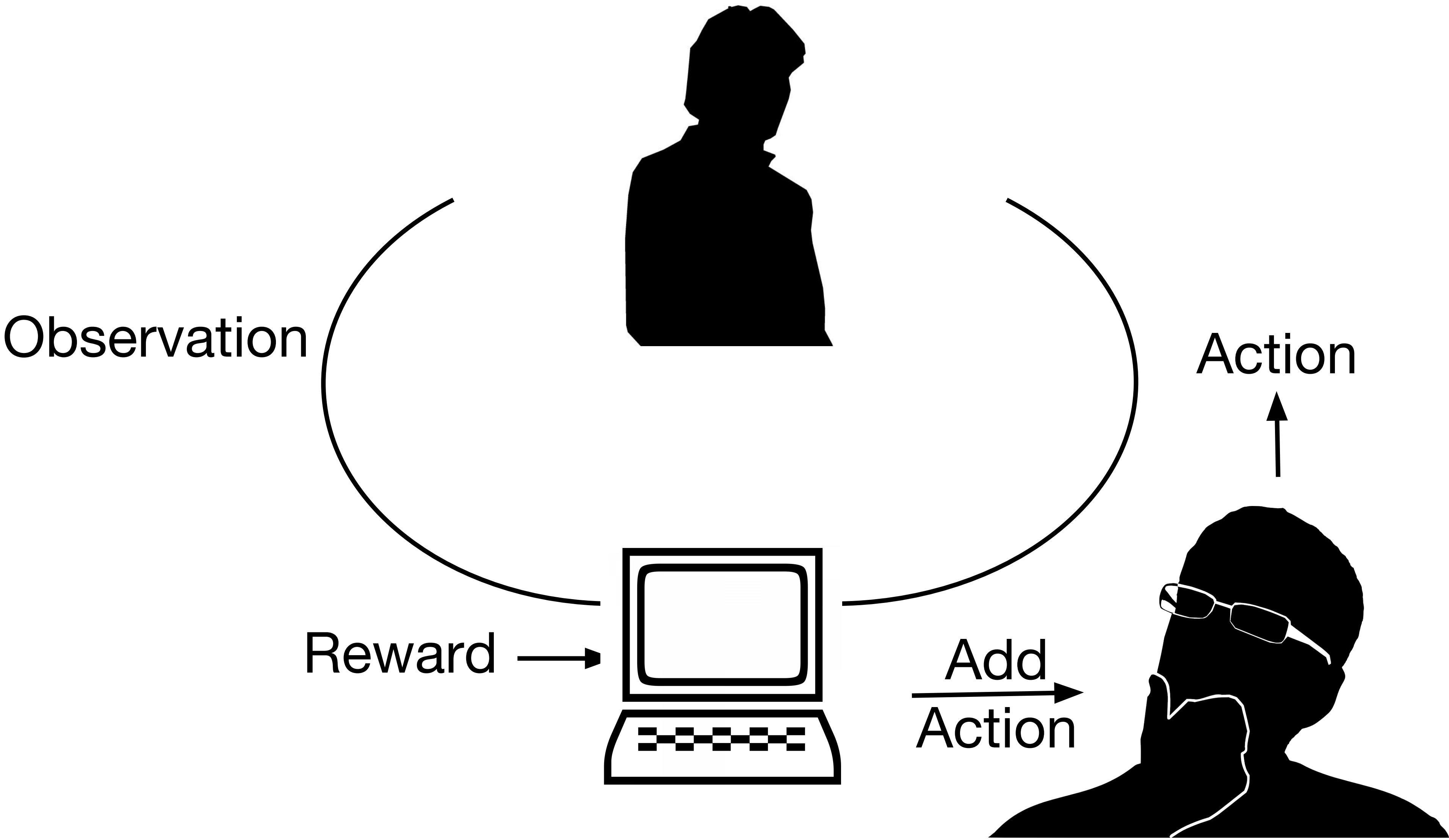
Human in the Loop Reinforcement Learning



Goal: Choose actions to maximize expected rewards

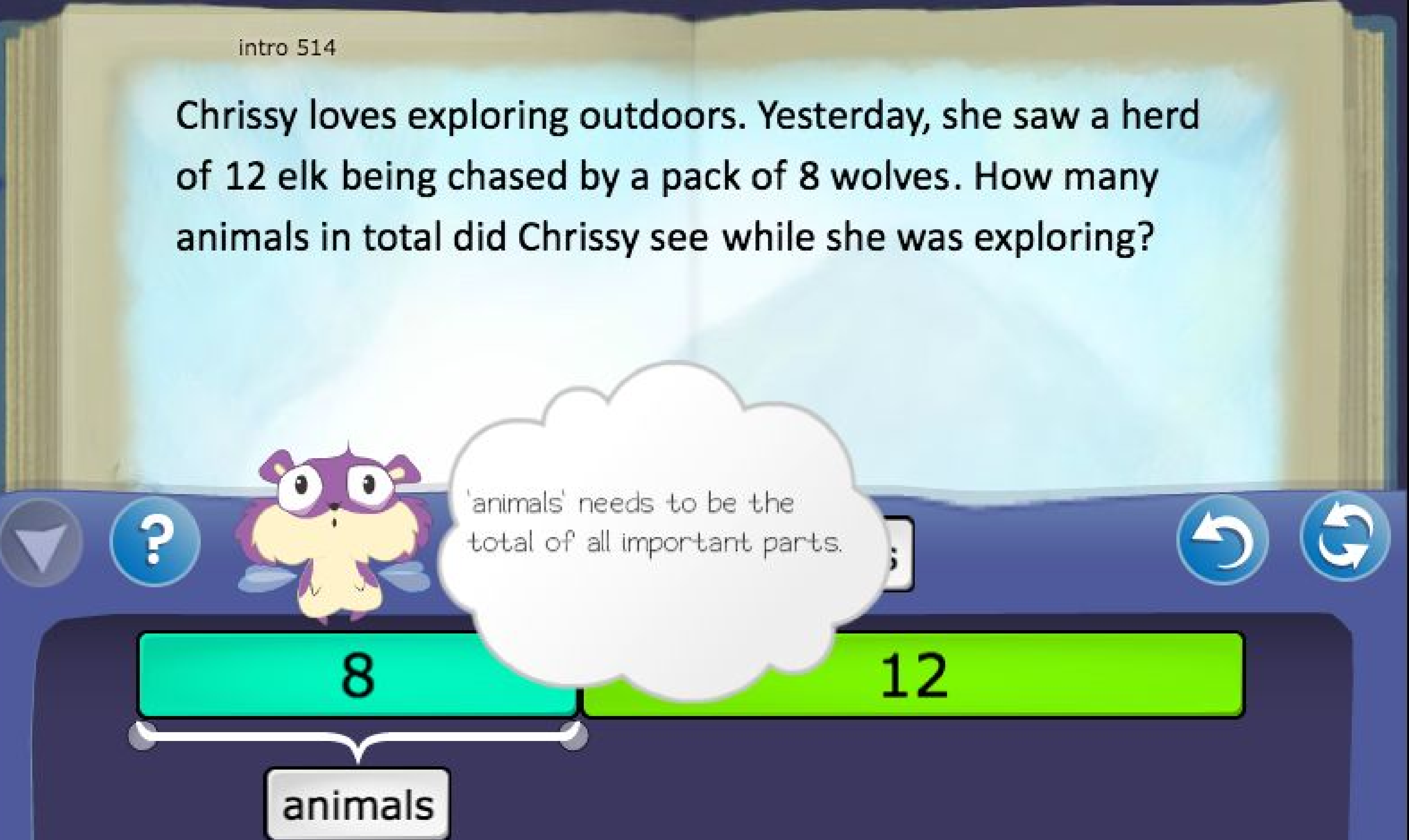
Where to Ask for New Actions?

Mandel, Liu, Brunskil & Popovic, AAAI 2017



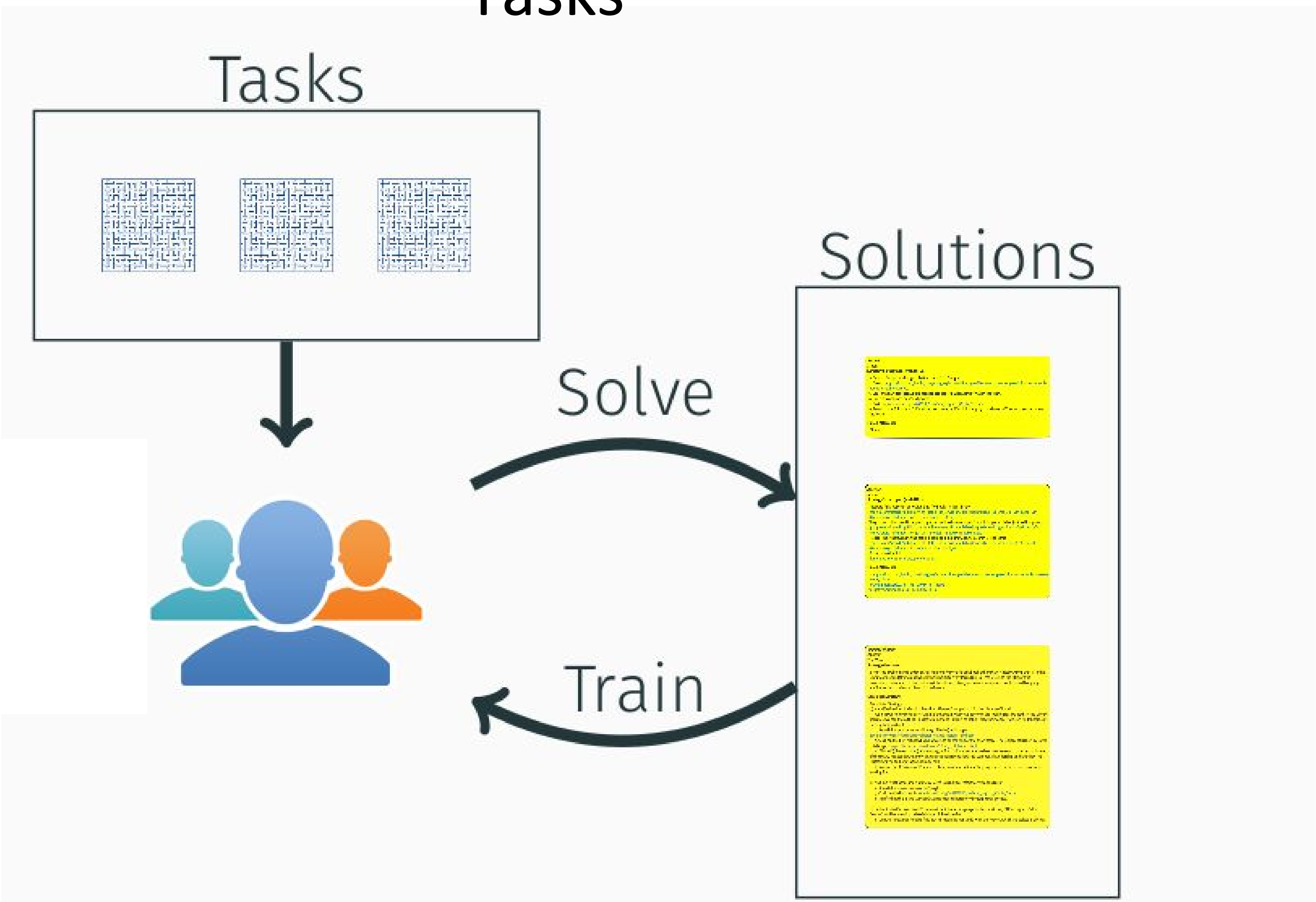
Goal: Choose actions to maximize expected rewards

Chrissy loves exploring outdoors. Yesterday, she saw a herd of 12 elk being chased by a pack of 8 wolves. How many animals in total did Chrissy see while she was exploring?

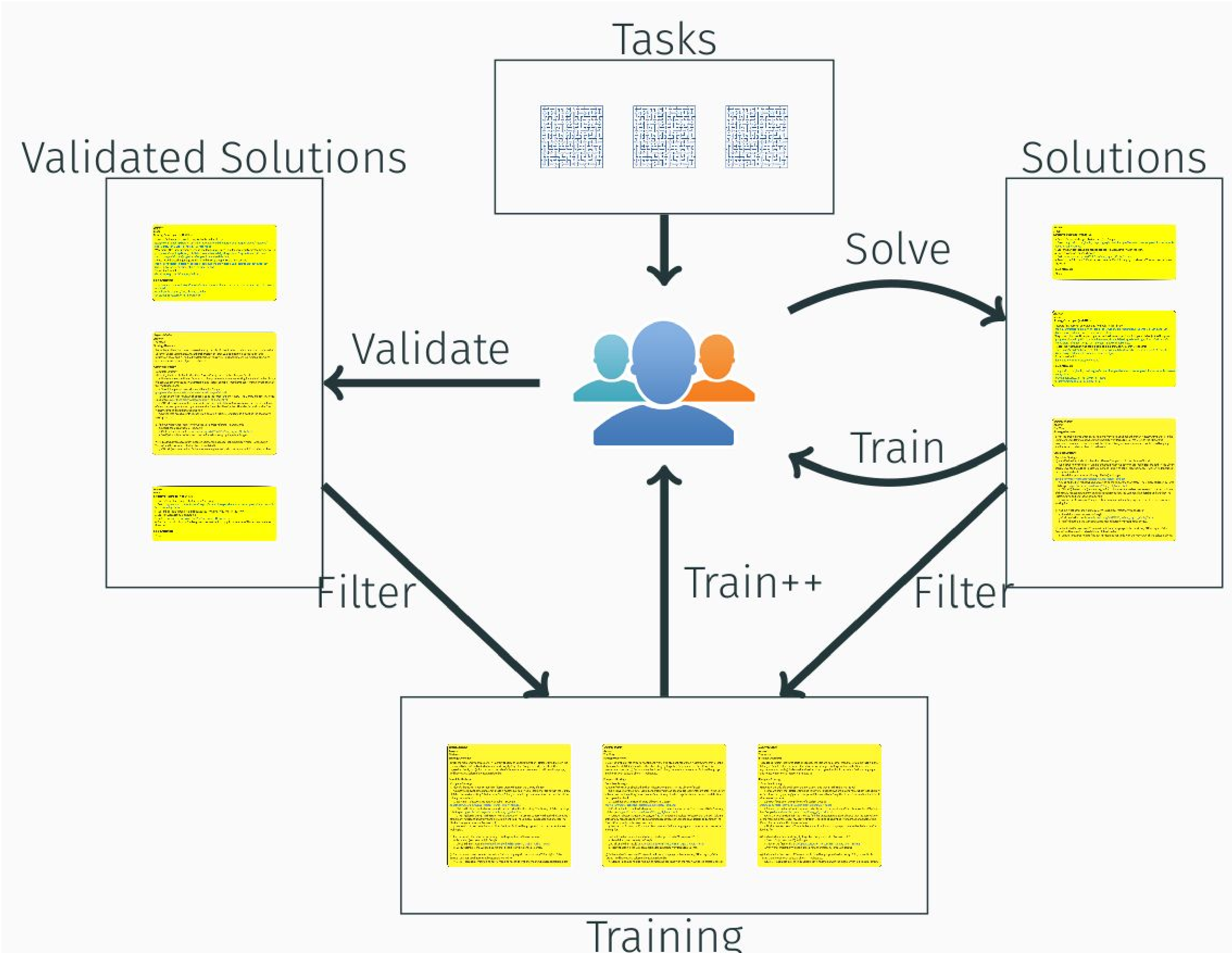


- New actions = new hints
- Learning where to ask for new hints
- **People helping computers to teach people**

People Helping Computers to Teach People Tasks



Doroudi, Kamar, Brunskill, Horvitz, CHI 2016



slide from Doroudi

Reinforcement Learning & Learning to Promote Learning

- Making better decisions by
  - Learning from past experience
  - Having humans help machines

Thanks to

and Karan Goel, Travis Mandel, Yun-En Liu, NSF, ONR, Microsoft, Google, Yahoo & IES

Reinforcement Learning & Learning to Promote Learning

- Making better decisions by
  - Learning from past experience
  - Having humans help machines