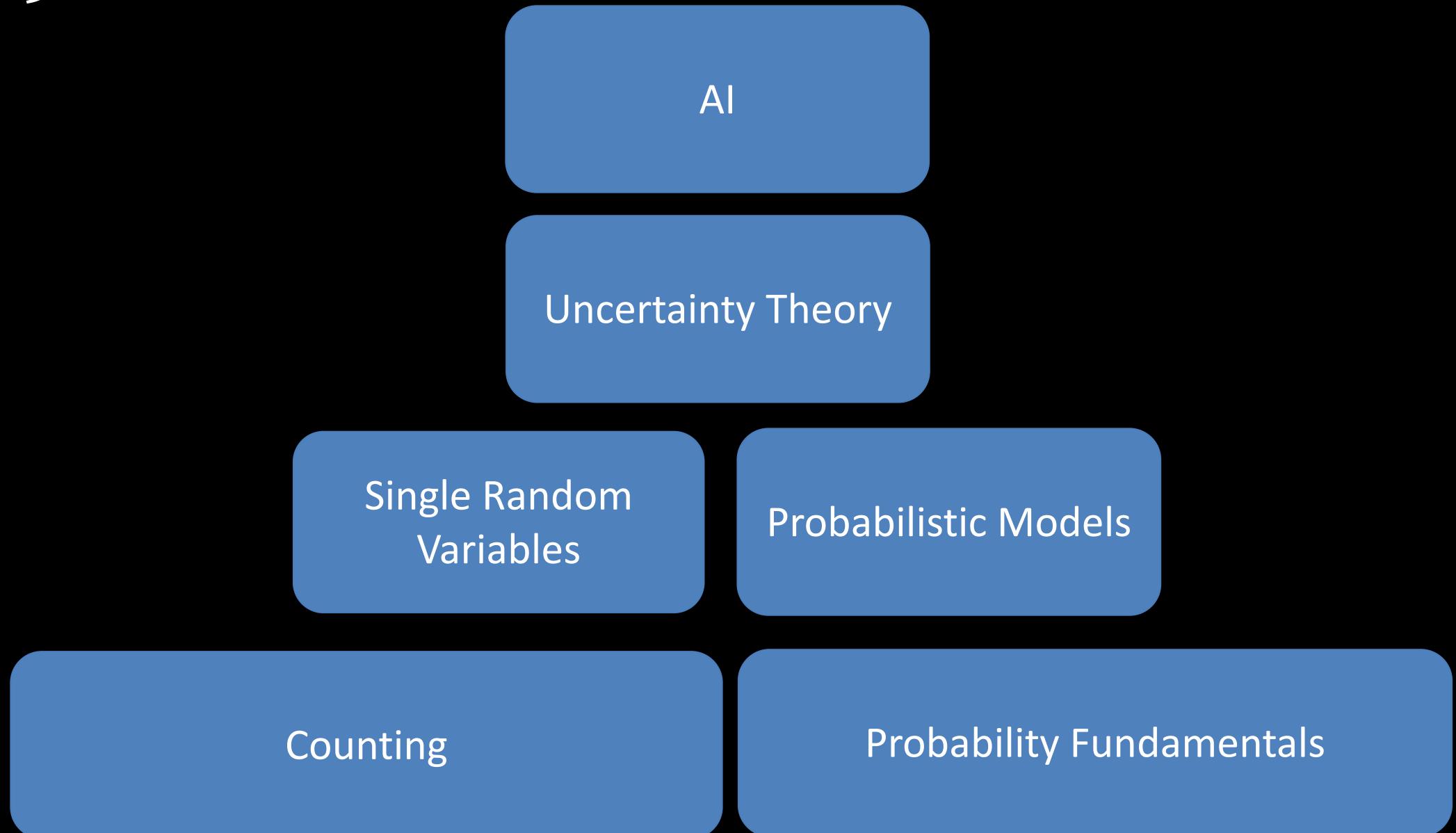


Future of Probability

Chris Piech

CS109



Learning Goal: Abundance of important problems



Chose k examples from a dataset of lots of student work

The screenshot shows a web-based programming assignment interface. On the left, a sidebar lists numbered items from 1 to 11, with items 1, 3, 5, 6, 9, and 11 having green checkmarks. The main area displays a question about the `numpy.random.choice` function. It includes a code editor with a Python script, documentation links, and a note about set equality. On the right, there are two answer submission sections: one for a numeric answer (set to 4) and another for an explanation (containing "asdfasdf!!!???" followed by $Y = 5$). Below these are two code snippets with run buttons.

PS1

numpy.random.choice

How many unique sets could be returned by the following `main()` function:

```
import numpy as np

def main():
    options = [0,1,2,3,4,5,6,7,8,9]
    ordered = np.random.choice(options,
                               size = 4, replace=False)
    unordered = set(ordered)
    return unordered
```

Here is documentation for numpy's random choice function:
[numpy.random.choice](#)

Note that sets are considered unordered. If two sets have the same elements, they are not distinct.

Answer Editor Solution

Numeric Answer: 4

Explanation:
asdfasdf!!!???

$Y = 5$

1 print('test')

Run Show

abcdef

$Y =$

$Y = 25$

1 for i in range(100):

2 print(i)

Run Show

Previous Question Next Question

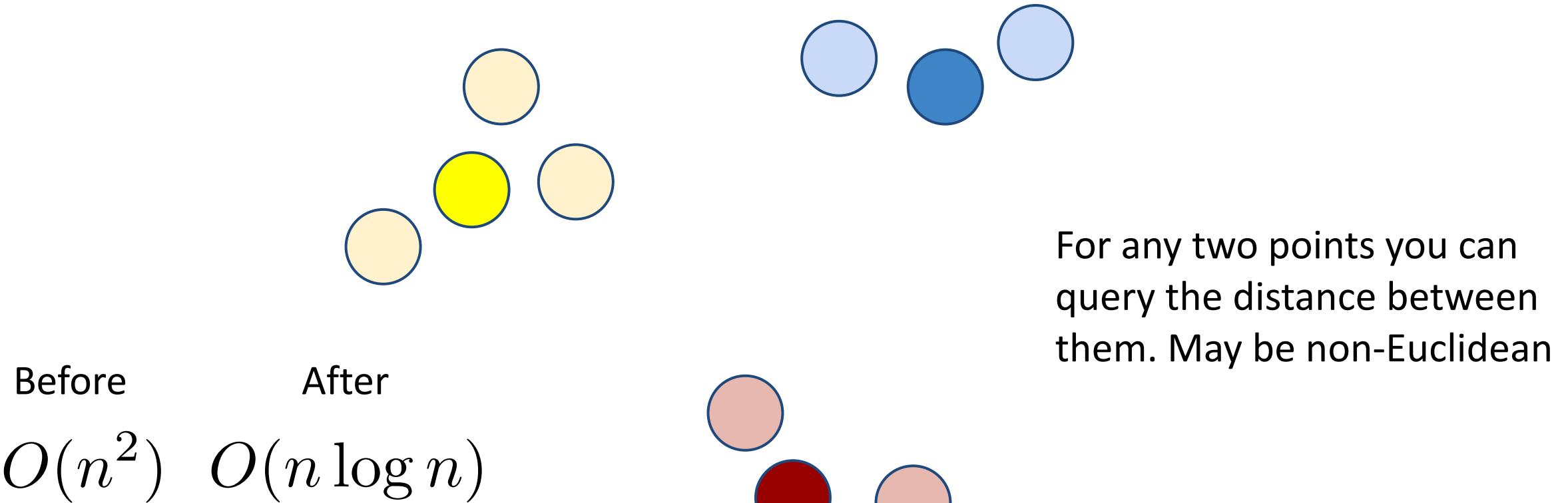
300 solutions

Find the 10 solutions which are most representative

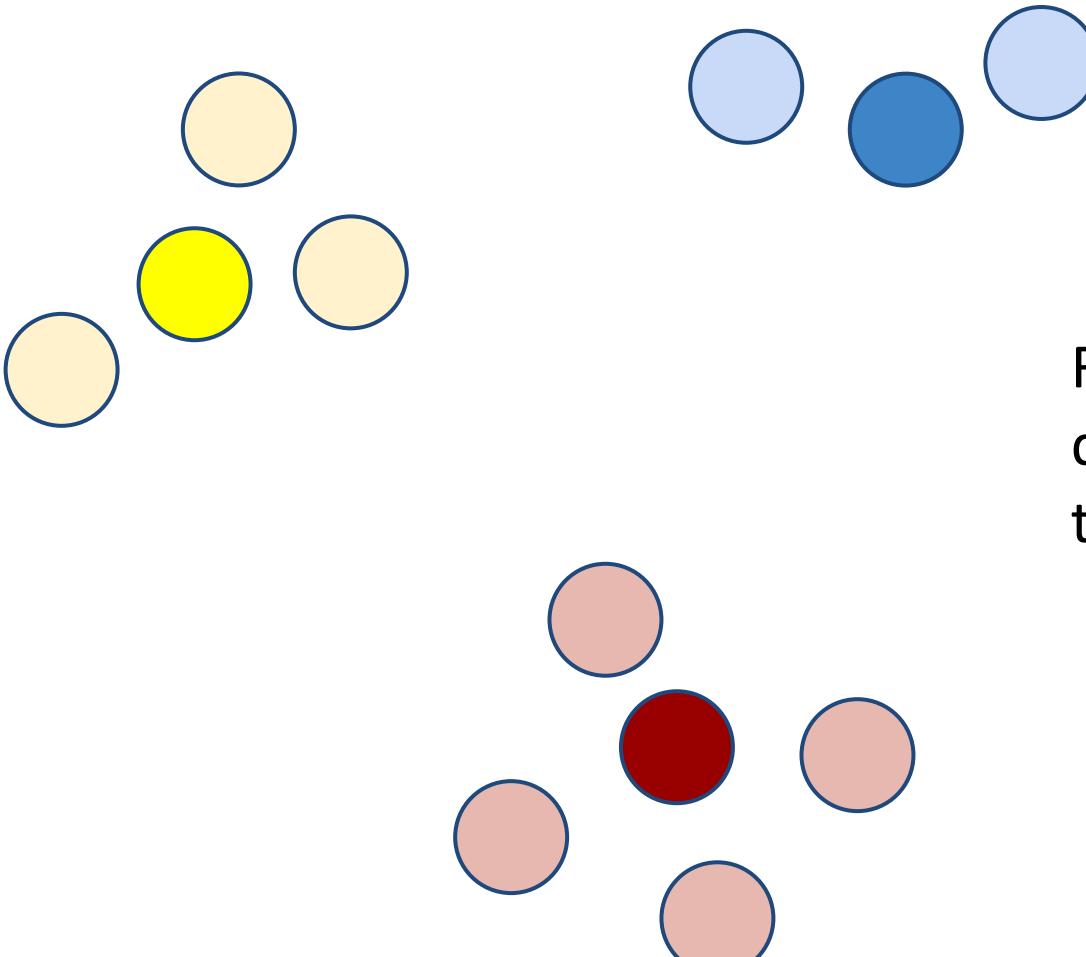


K Medoids: A Classic Algorithm

Choose the k nodes such that the sum of minimized distances is as small as possible



Step 1: Choose the node closest to the rest (on average)



For any two points you can query the distance between them. May be non-Euclidean



The inner loop can be thought of as Thompson Sampling



3, 4

10, 11, 9, 8

12, 11, 10, 14

2, 0



Allowed us to revisit several core algorithms

- k -Medoids
 - **Mo Tiwari, ... Chris Piech.** "*BanditPAM: Almost Linear Time k -medoids Clustering via Multi-Armed Bandits*", NeurIPS 2020
- Random Forests
 - **Mo Tiwari, ... Chris Piech.** "*MABSplit: Faster Forest Training Using Multi-Armed Bandits*", NeurIPS 2022
- Maximum Inner Product Search
 - **Mo Tiwari,... Chris Piech.** "*Faster Inner Product Search in High Dimensions*", Submitted to AAAI 2023 (favorable reviews)



Today
Digital Future of Probability

Themes to look out for:

Probabilistic Modelling vs Deep Learning
Distributions vs Point Estimates

Let me tell you a story about one particular problem.

It will give you a sense of what it takes to do research in **applied probability**, and our desire to solve the problems, will lead us to deep **theoretical** challenges in modern AI.

Application -> Theory

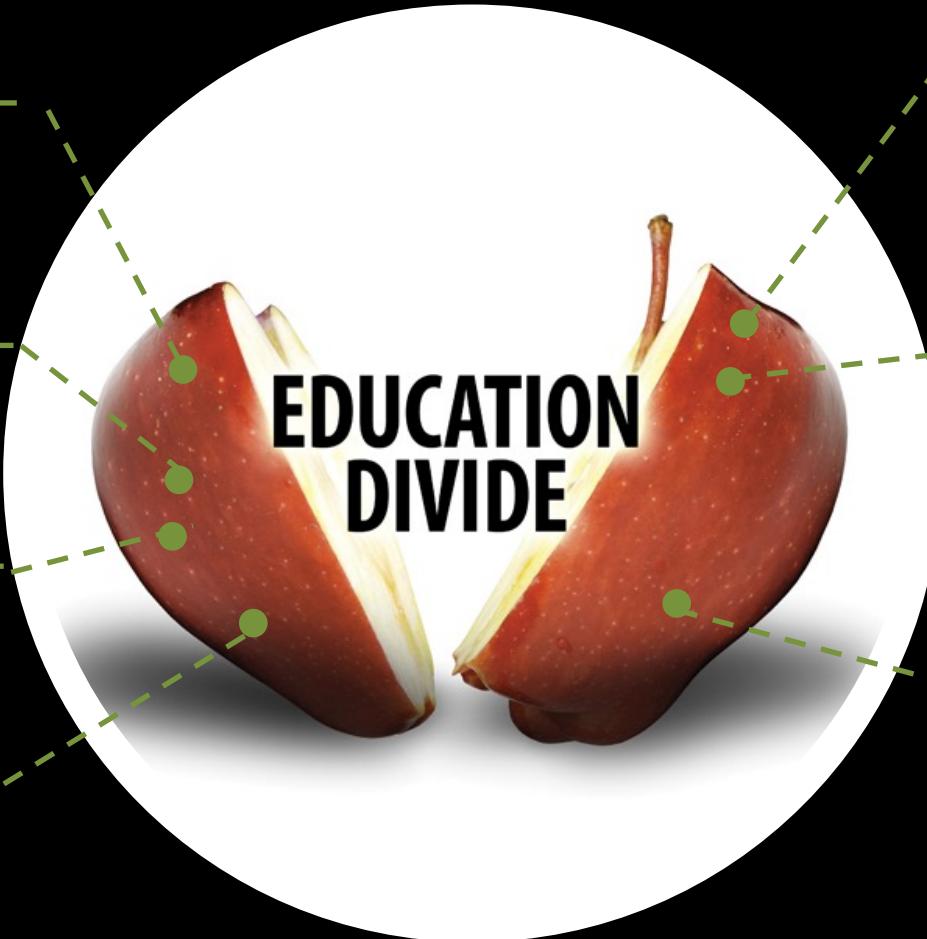
Quality Education Gap

90% of children
enroll in primary
education [1]

40% in secondary
education [1]

20% in tertiary
education [1]

Dramatic quality
differences



375 million workers need to
be retrained by 2030 [3]

Half a million unfilled
computer science jobs
(60% of STEM jobs) [2]

For all learners we
want *quality*.

[1] *World development indicators 2015*. World Bank Publications, 2015.

[2] USA Bureau of Labor Statistics Employment Projections , 2016.

[3] Jobs lost, jobs gained. McKinsey Global Institute, 2017.



Smart Phone Access

Advanced Economies



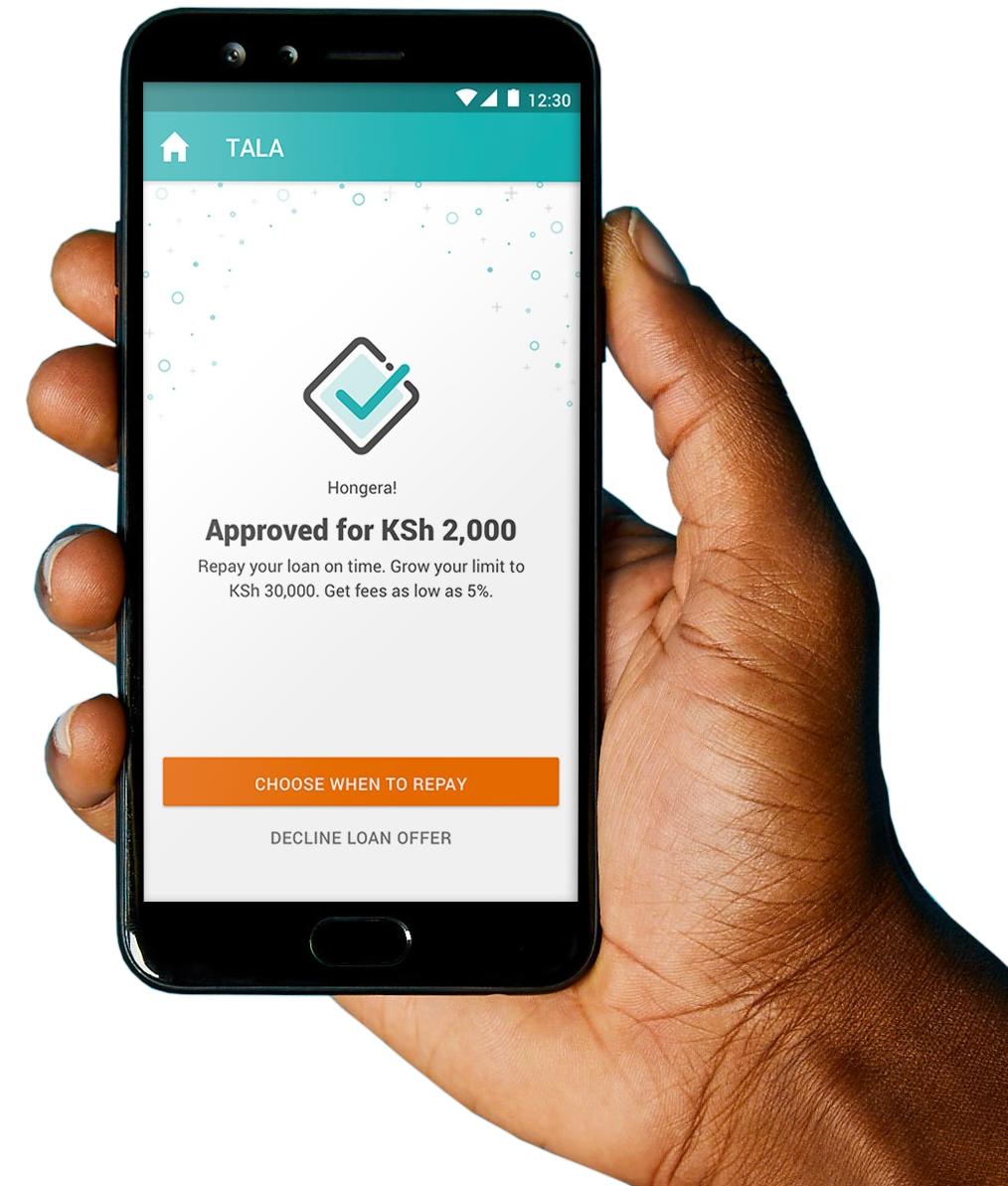
Emerging Economies



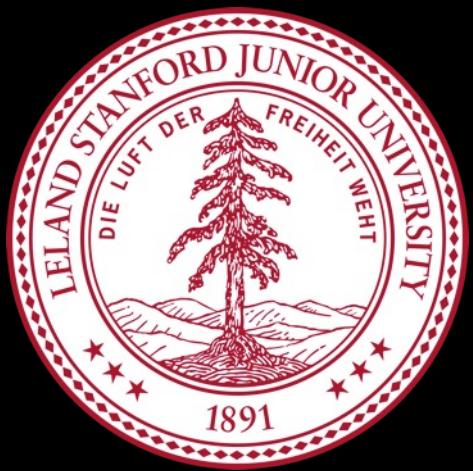
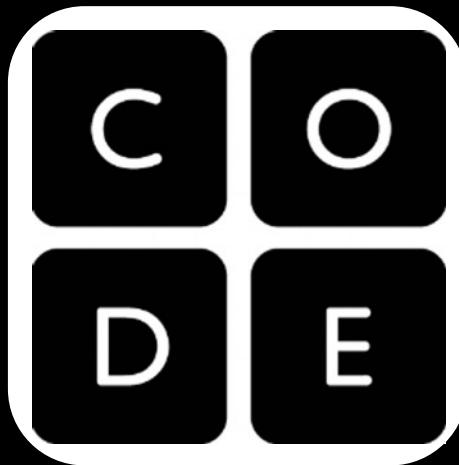
Smartphone

Mobile

No phone



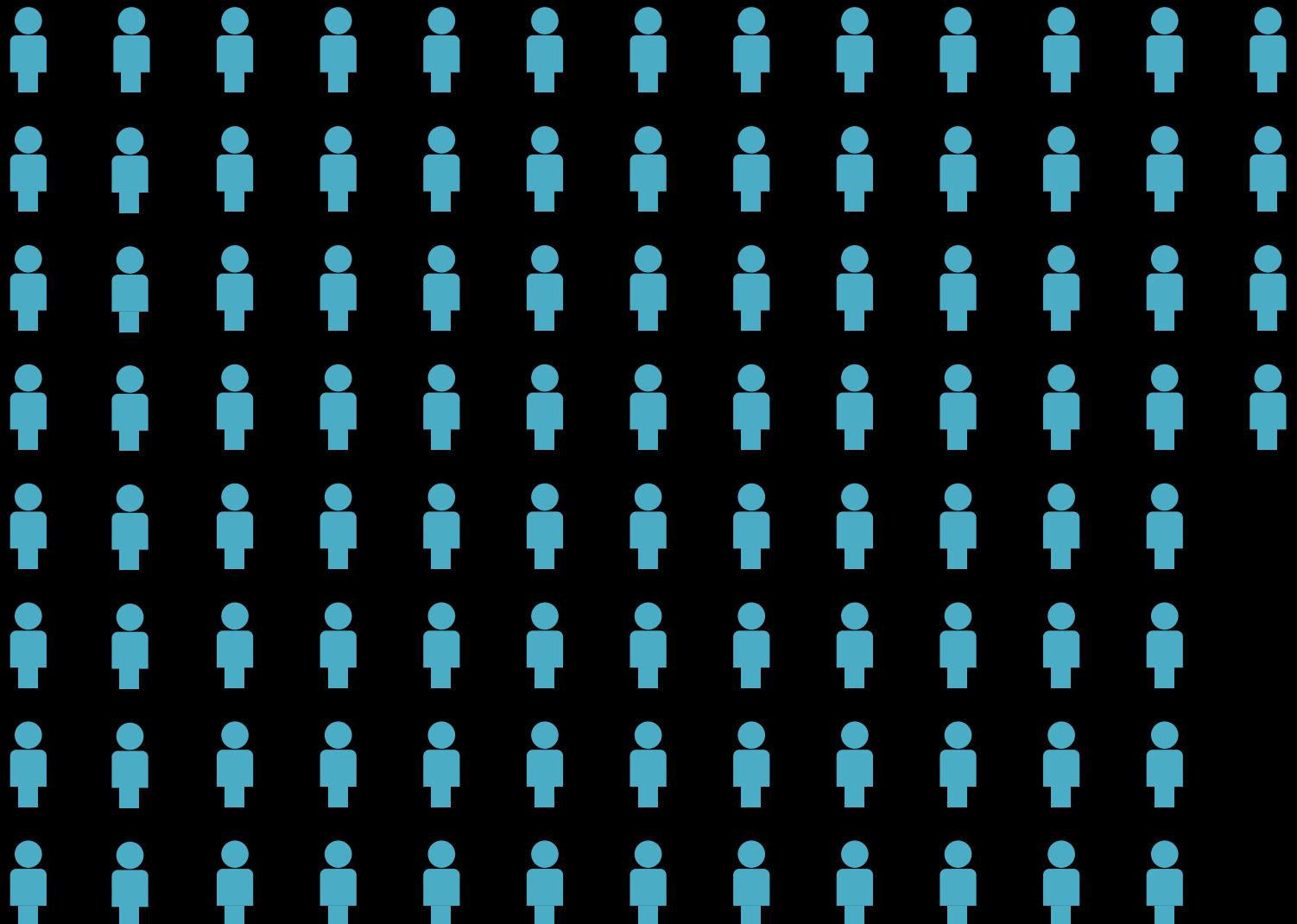
Unprecedented Data



Over 50 million learners

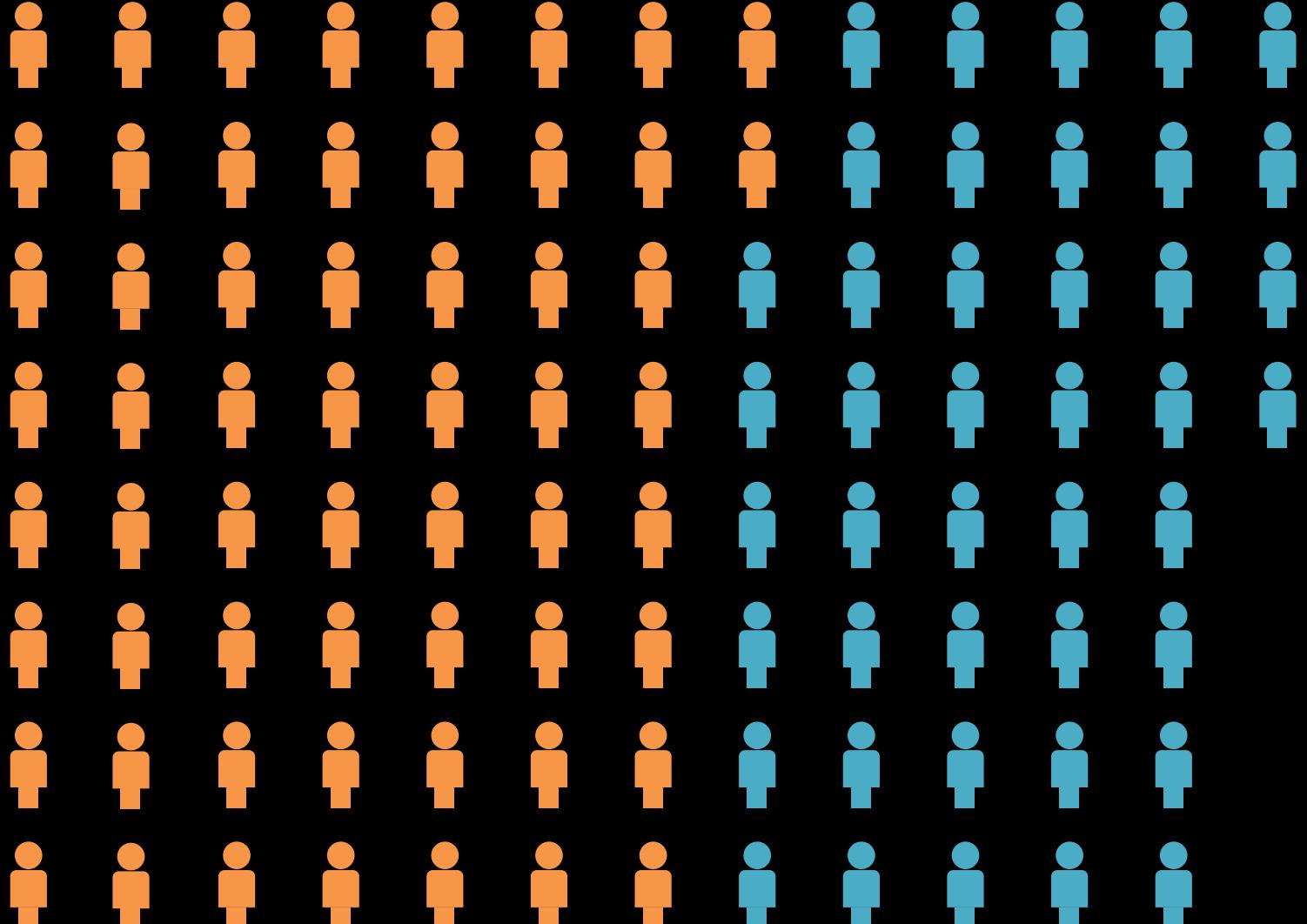


US K-12 Students



= 500,000 learners

Code.org Students



= 500,000 learners



Code.org by the numbers

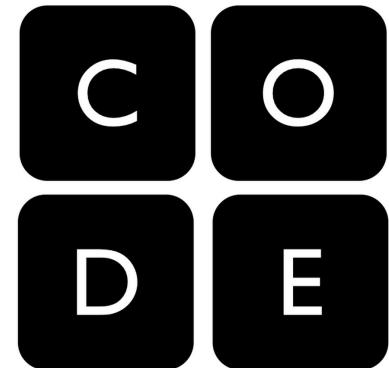
1,234,127 teachers

42M unique enrolled students

Used in **180+** countries

832M hour of code sessions

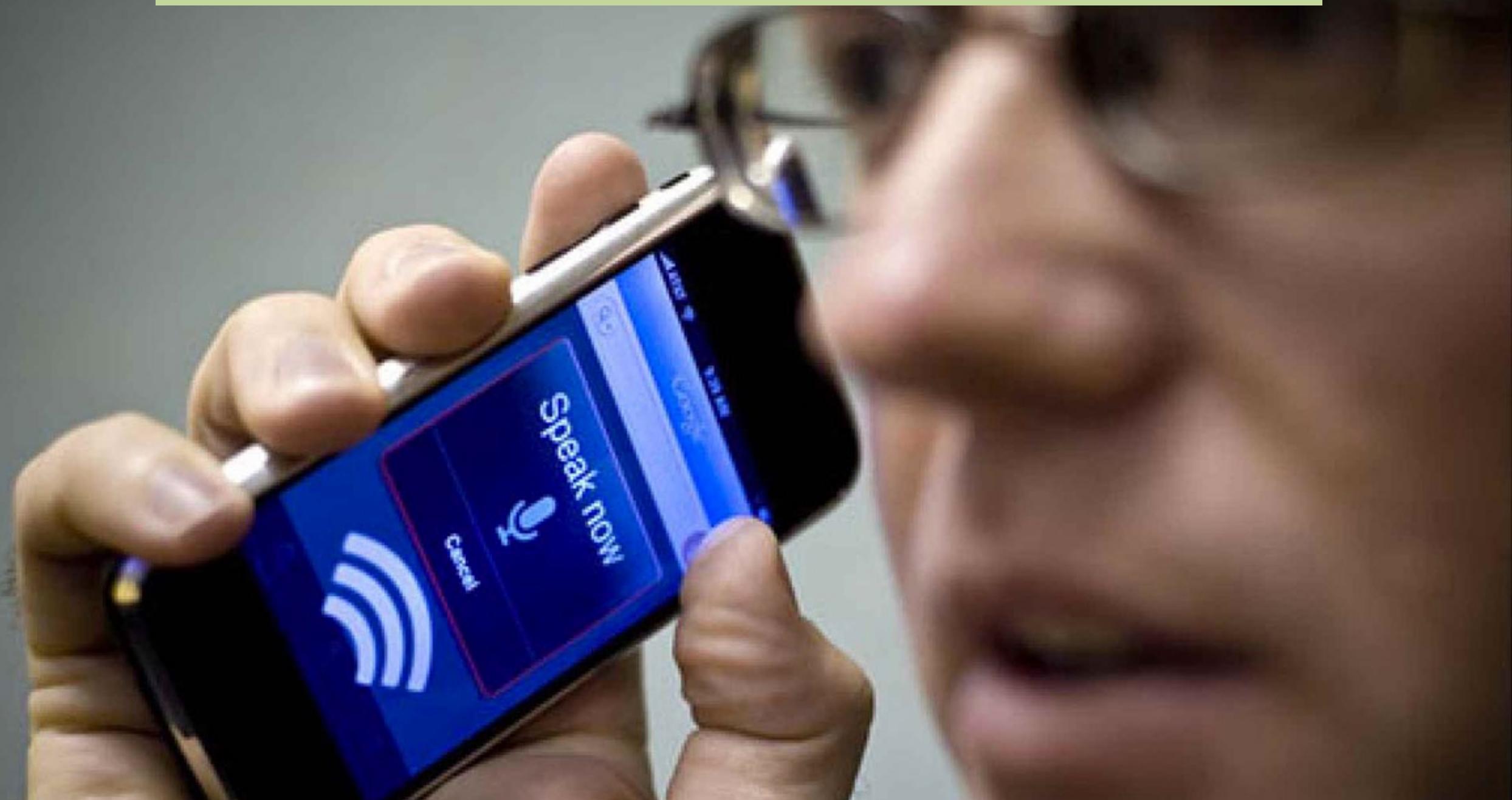
4 papers publish with our lab



50M K12 students in the US



Speech Recognition



Grand Challenge in Education

**Autonomously support
education by better
understanding students.**

Clear
Societal Need

Scale
education

New
Datasets of
Learning

Online
assignments

Deep
Learning

AI
Renaissance



Feedback is Labor Intensive



Online classes have not solved the feedback problem [1].



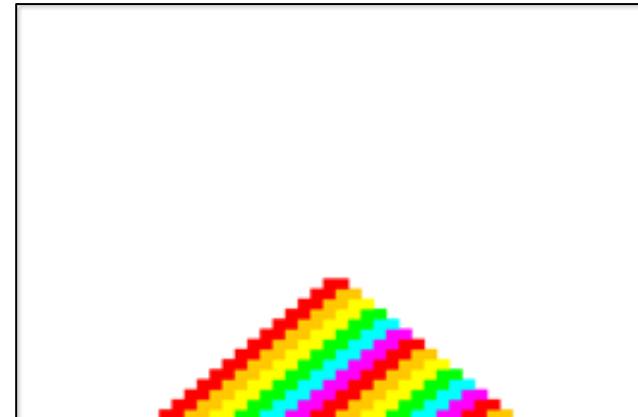
[1] Deconstructing Disengagement. Analysing learner subpopulations in MOOCs. Kizilcec, Piech, Schneider. Over 600 citations since 2013



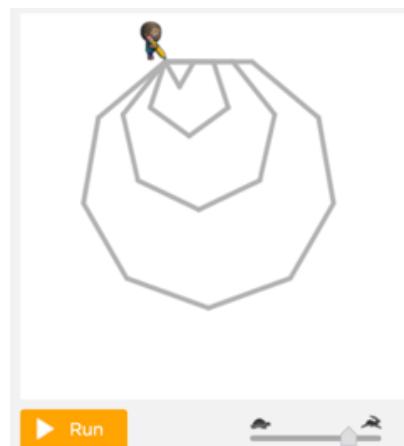
Many domains of student work

$$\begin{aligned}9 \times 6 &= (10 - \boxed{}) \times 6 \\&= 10 \times 6 - \boxed{} \times 6 \\&= 60 - \boxed{} \\&= \boxed{}\end{aligned}$$

```
when run
repeat until [acorn icon]
  do [if path ahead
    do [turn left
    else move forward]]
```



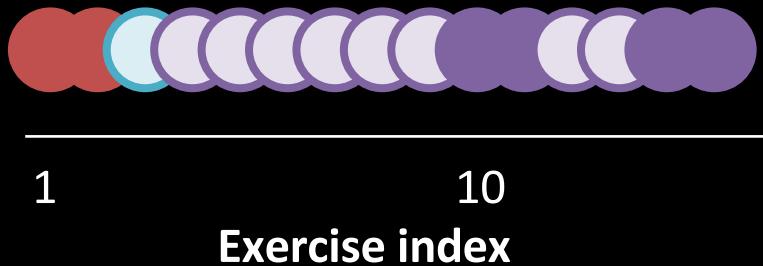
Why did the original pilgrims come to America?



Chapter 0: Always start simple

First deep learning for education

KHAN
Student

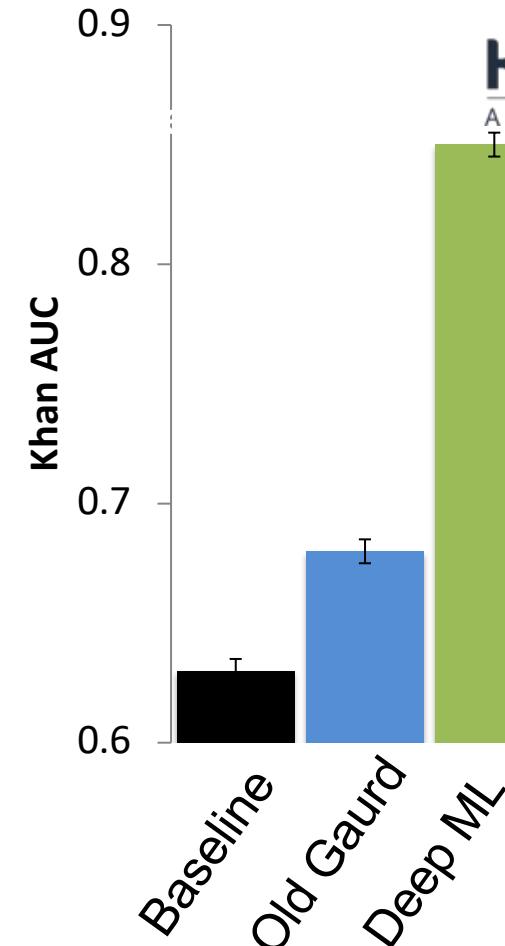


Exercise Type:

- Solving for x-intercept
- Solving for y-intercept
- Graphing linear equations
- Square roots
- Slope of a line

Answer:

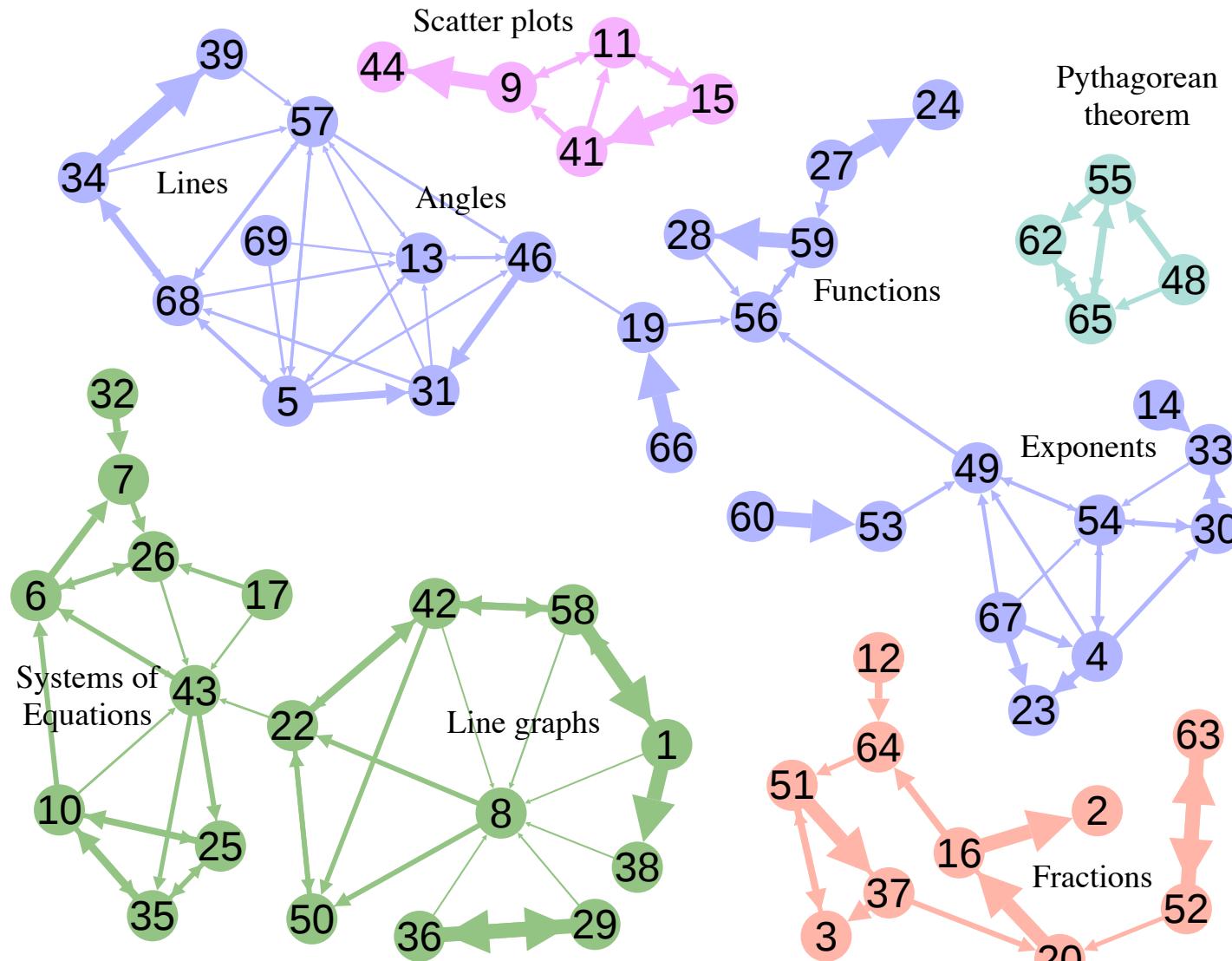
- Correct
- Incorrect



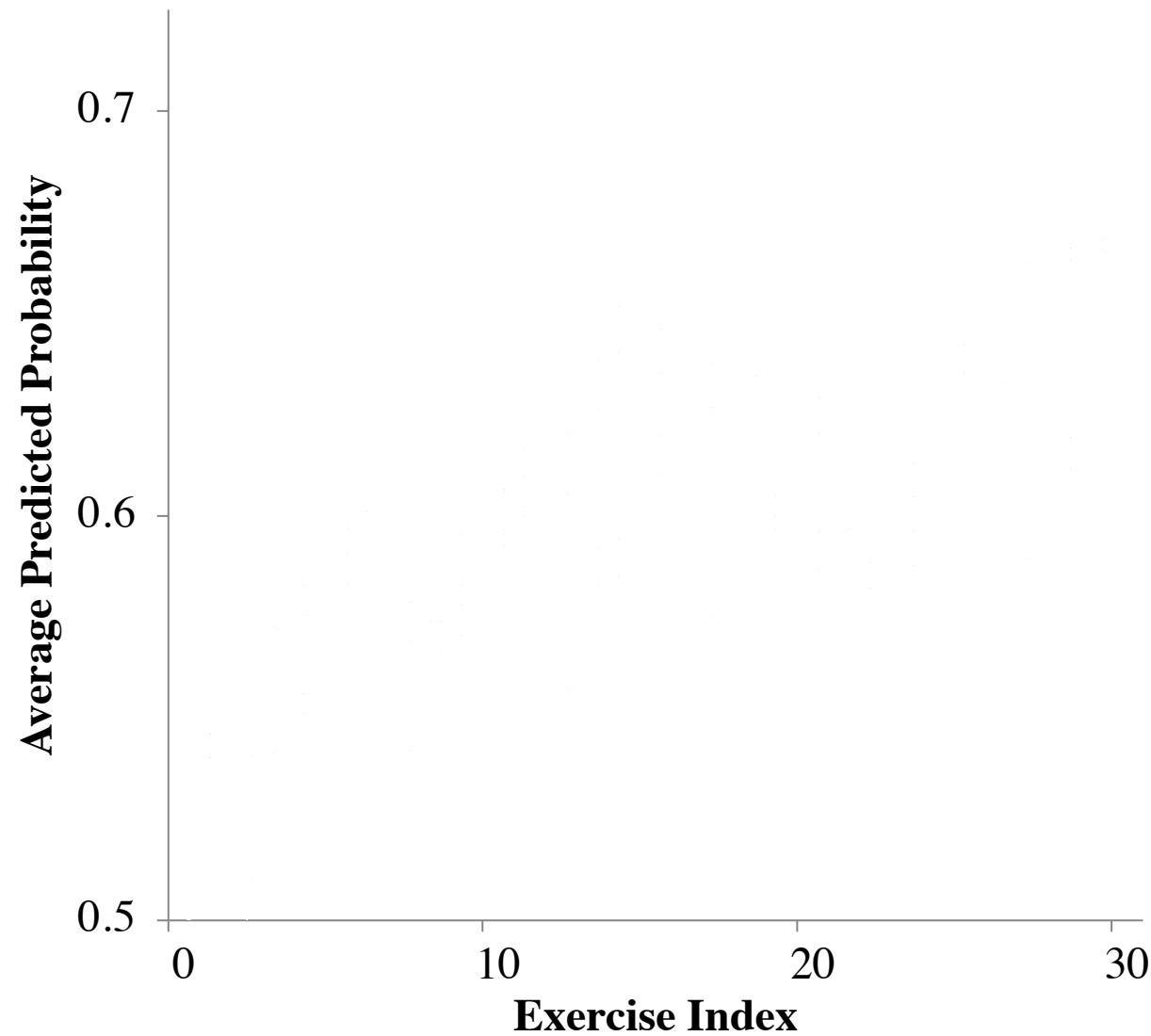
Old Problem



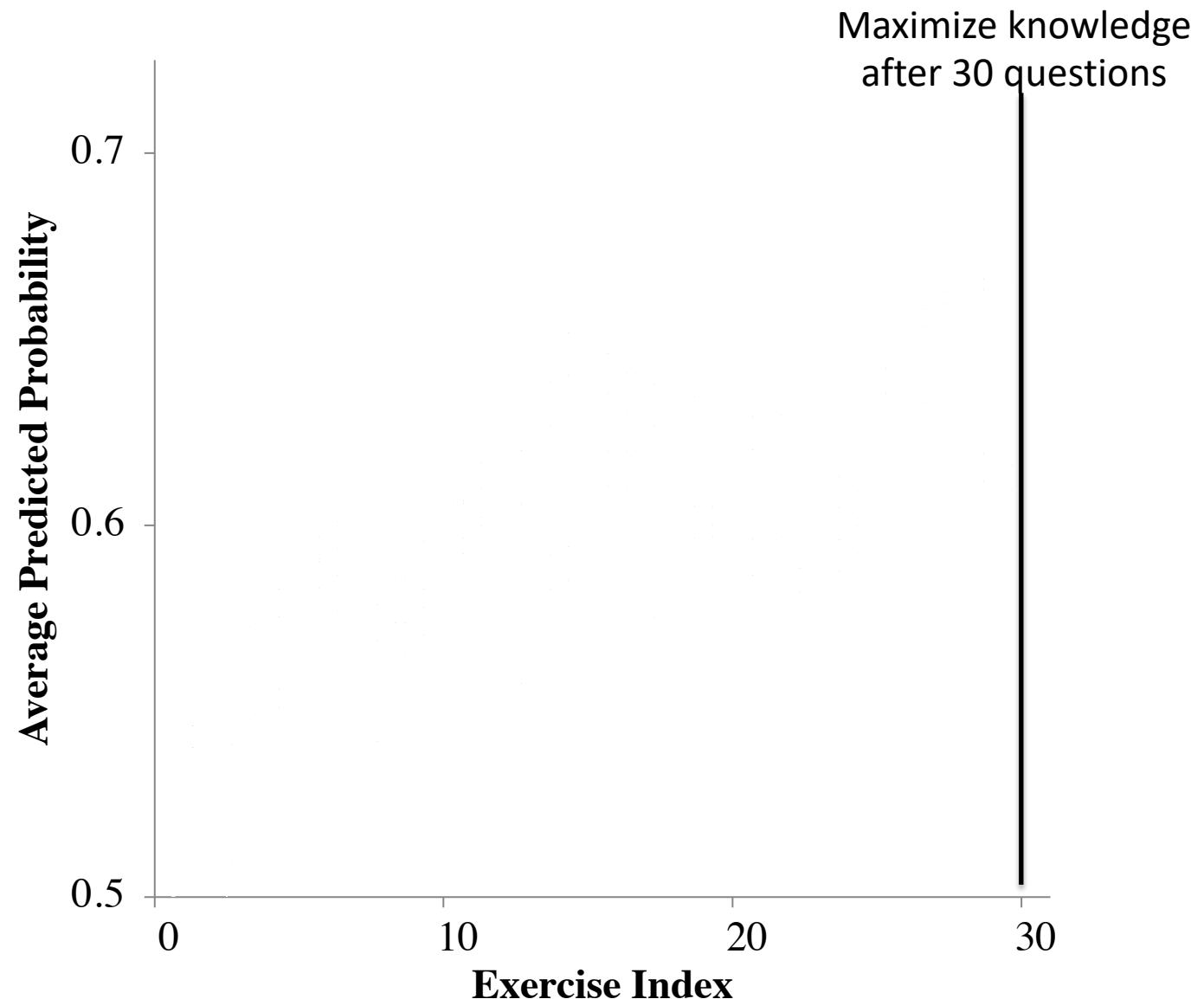
Learns Concept Relationships



Optimal Teaching

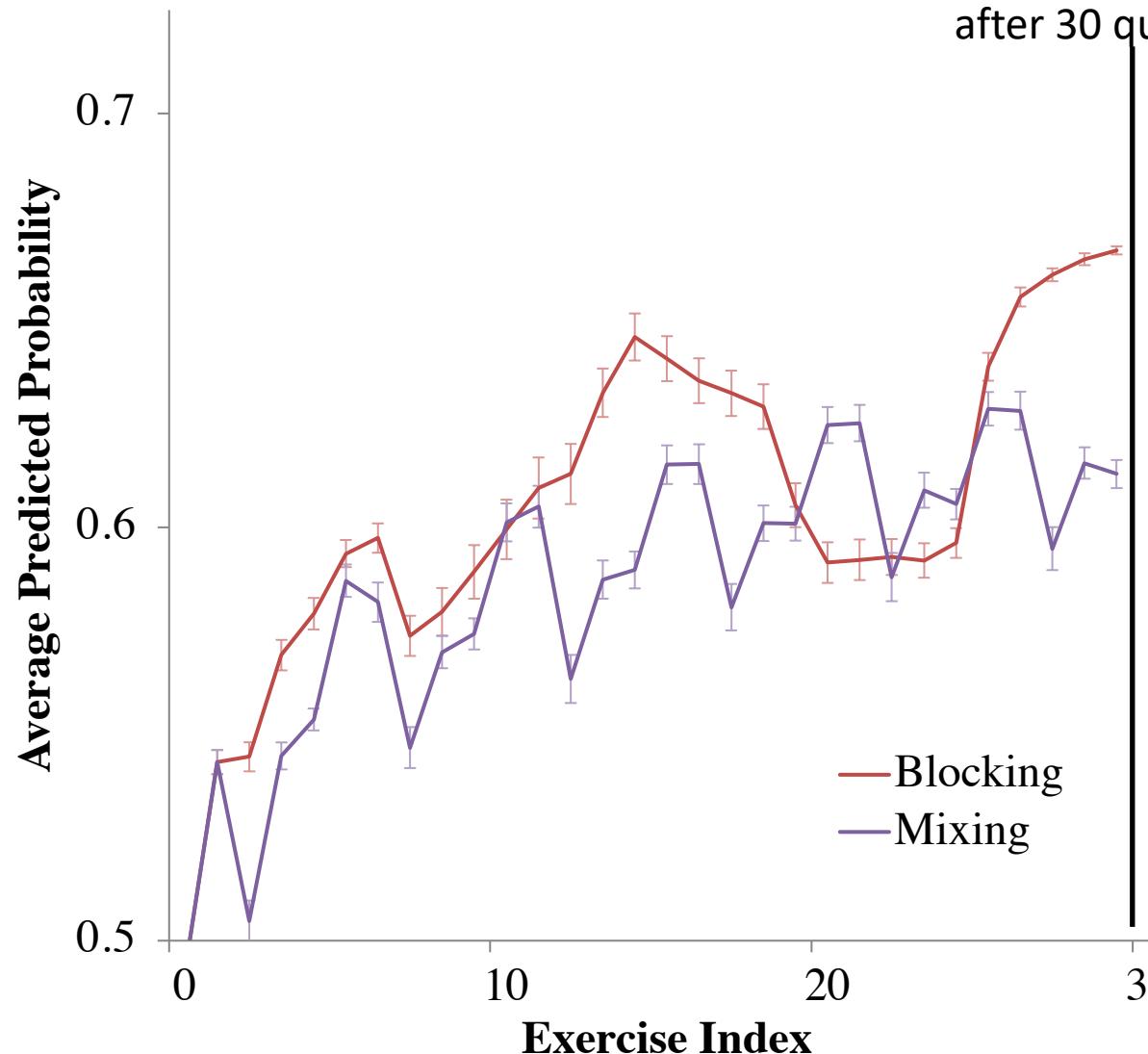


Optimal Teaching



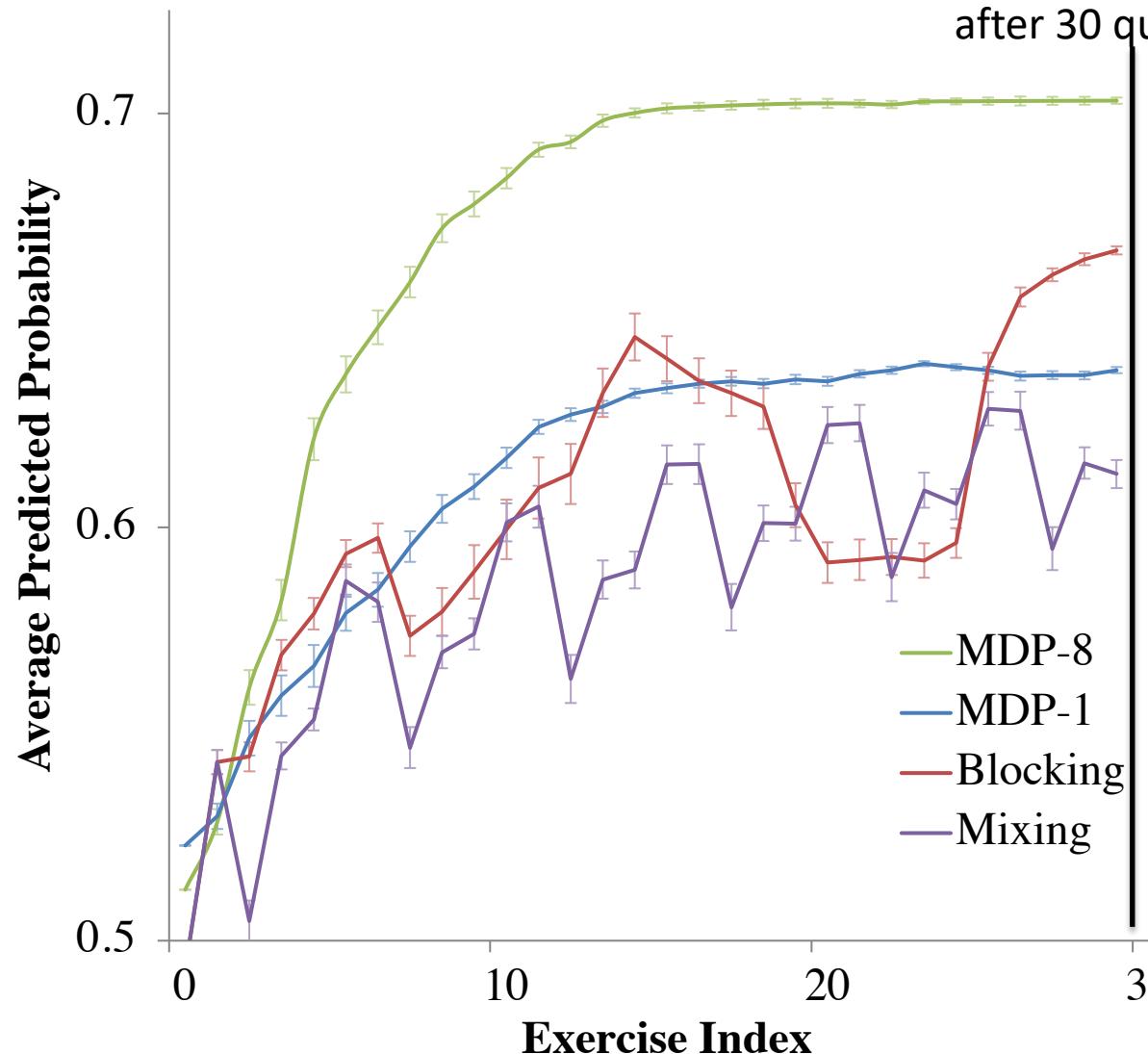
Optimal Teaching

Maximize knowledge
after 30 questions



Optimal Teaching

Maximize knowledge
after 30 questions



We truly would rather move beyond
correct / incorrect



Some domains are very hard

Introduction to Python Upgrade to Pro

script.py

```
1 my_name = "Codecademy"
2 print("Hello and welcome " + my_name + "!")
3
```

File "script.py", line 3

SyntaxError: unexpected EOF while parsing

Hi! I'm Elsa of Arendelle. Help me create a single line.

Not quite. You have to use a block you aren't using yet.

Blocks Workspace: 3 / 2 blocks Start Over Show Code

move [forward ▾ by 100 ▾ pixels]

turn [right ▾ by 90 ▾ degrees]

turn [left ▾ by 90 ▾ degrees]

when run

turn [left ▾ by 90 ▾ degrees]

turn [left ▾ by 90 ▾ degrees]



Can you understand this code?

Top Secret Not Secure

Question Solution

Instructions

- If there are many moves, focus on the first one
- Random code strategy is for when the student seems to be trying things randomly
- Lookout for students who don't get nesting or pre/post conditions. Often extra blocks in a body is an indication that they don't get that the post of the loop has to match the precondition

Question



```
import code.org.*;  
  
public class MySoln {  
    public void run() {  
        move(50);  
        for(int i=0; i<4; i++){  
            if(frontIsClear()) {  
                turnLeft(90);  
            }  
            for(int j=0; j<i; i++){  
                move(i * 20);  
                turnRight(120);  
                move(10);  
            }  
        }  
    }  
}
```

Student 0

Label Console

✓ Num Done: 8273

Strategy

Beeper Boundary (most people do this)
 Triangle Strategy
 Recursive Strategy

Looping

Correct use of looping
 Doesn't use a while
 Doesn't have correct stop condition

Stanford TAs label
800 submissions

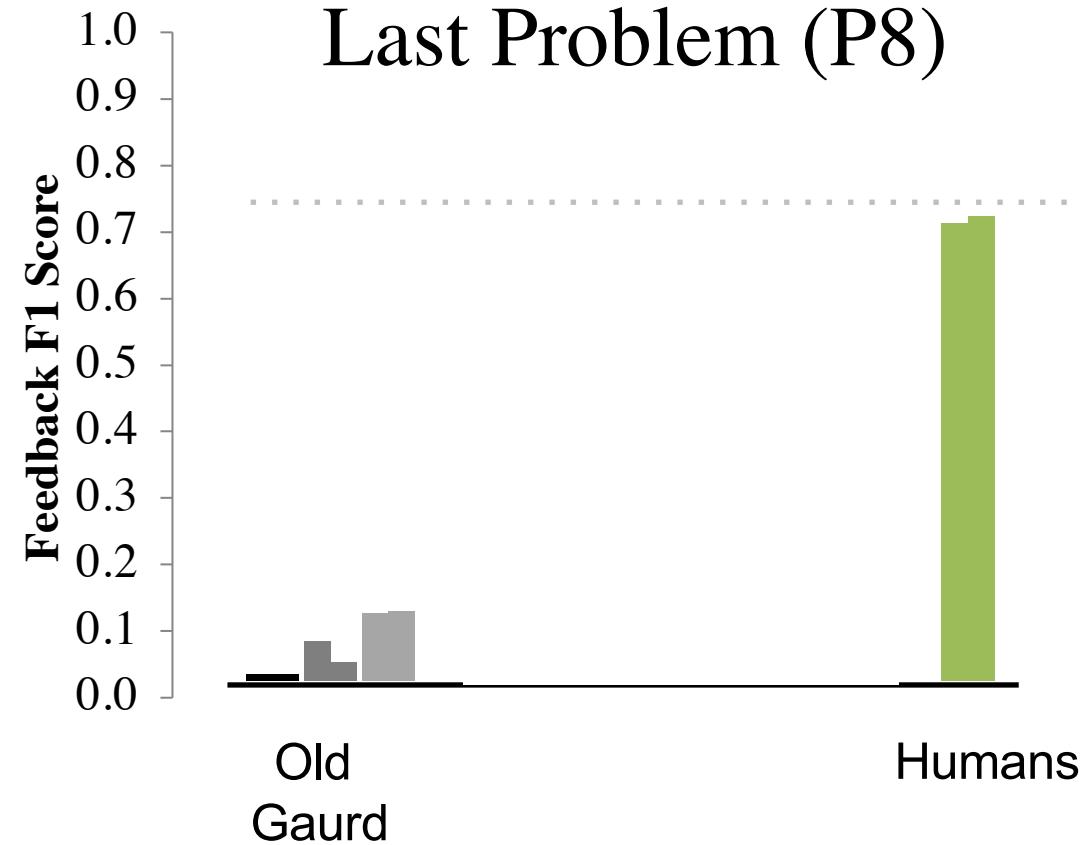
Loop post condition doesn't match precondition
 Repetition of bodies

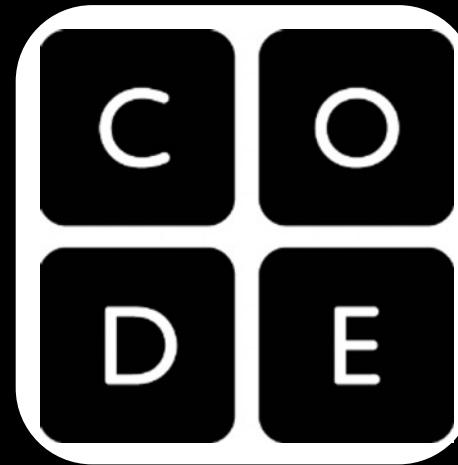
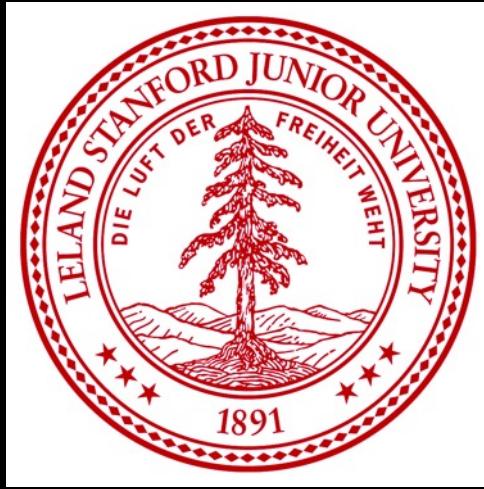
Cleanup

Record label

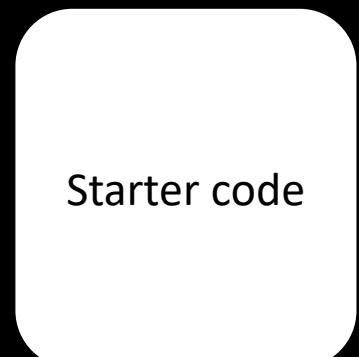
Traditional Deep Learning Doesn't Work

Label student code





Can we provide feedback
by dynamic analysis?



...



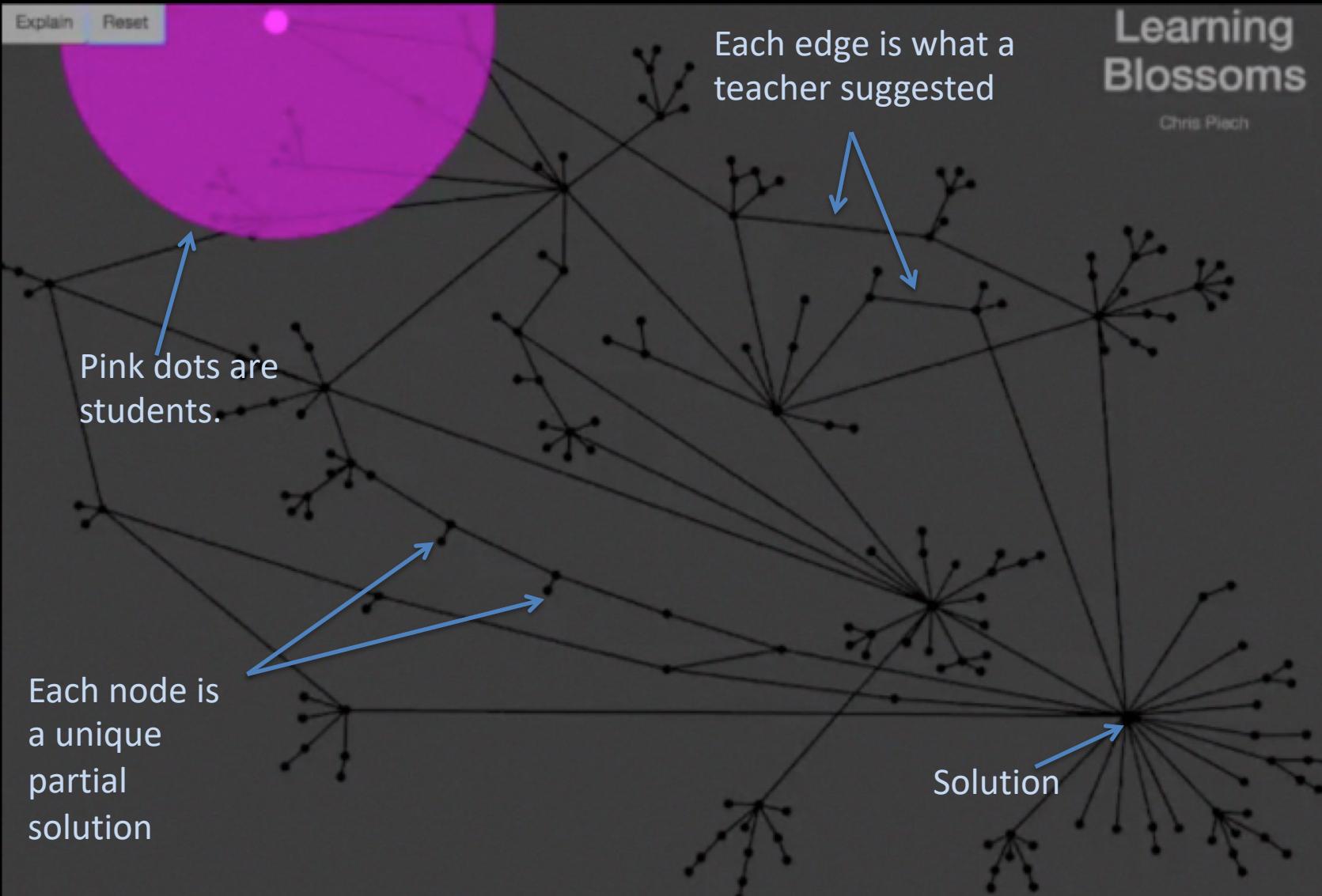
Chapter 1: Better data source?

Explain Reset

Learning Blossoms

Chris Piatek

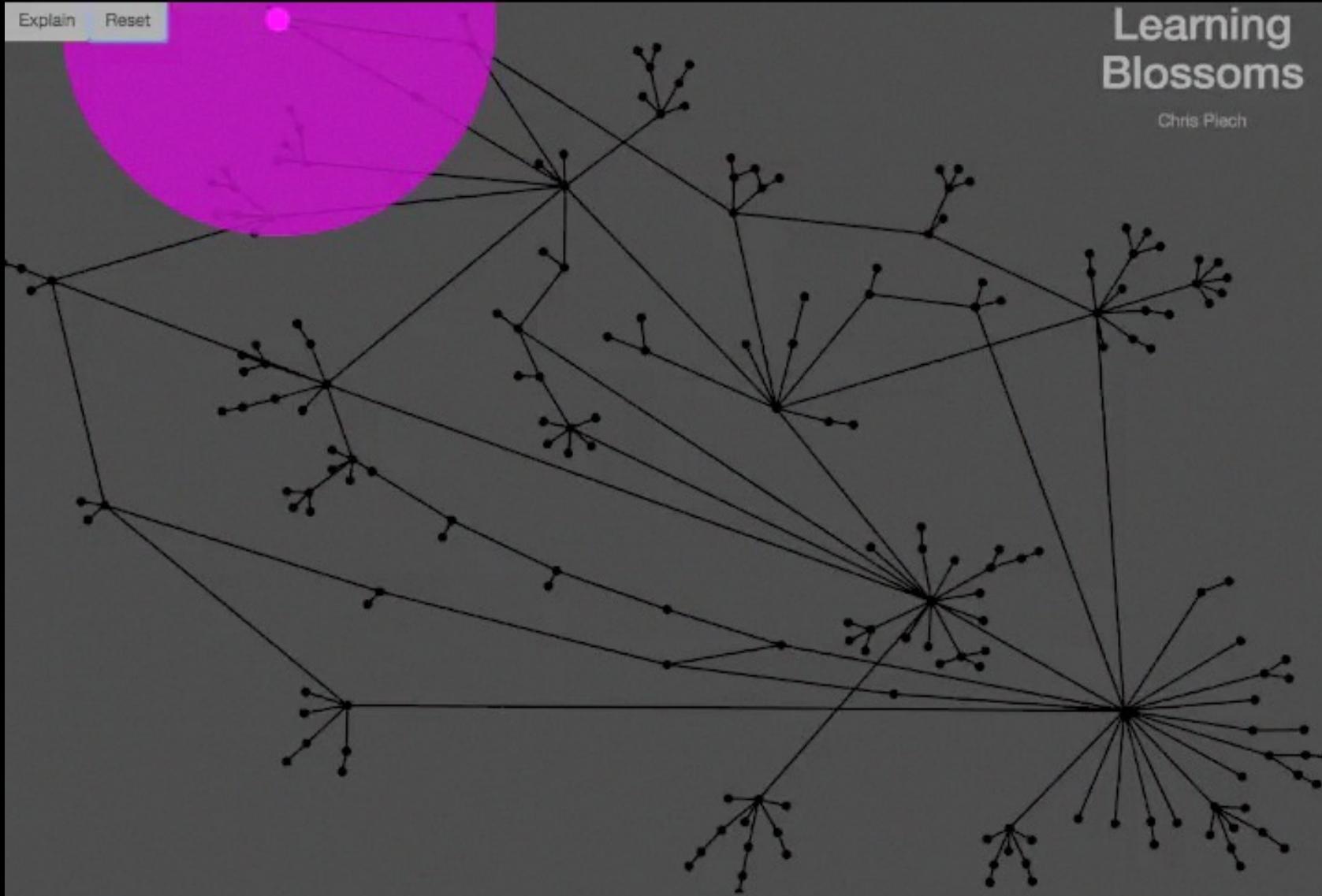




Explain Reset

Learning Blossoms

Chris Piech



The Crowd is Un-wise

Temporal methods tried:

- Shortest path
- Min Time
- Expected Success
- Reinforcement learning
- Most Common Next
- Most Popular Path



when run
move forward
move forward
turn left ⌂ ▽

when run
move forward
turn left ⌂ ▽

when run
move forward
move forward
turn right ⌂ ▽

when run
move forward
move forward
turn left ⌂ ▽
move forward

18%

45%

12%



Desirable Path Algorithms

Poisson Common Path

$$\gamma(s) =$$



Predicted next
partial solution

First step in the *most frequent* path to
the solution from s , taken by *average*
students. Assume poison process.



Desirable Path Algorithms

Poisson Common Path

$$\gamma(s) = \arg \min_{p \in Z(s)} \sum_{x \in p} \frac{1}{\lambda_x}$$

Path Cost

Submission count of partial solution

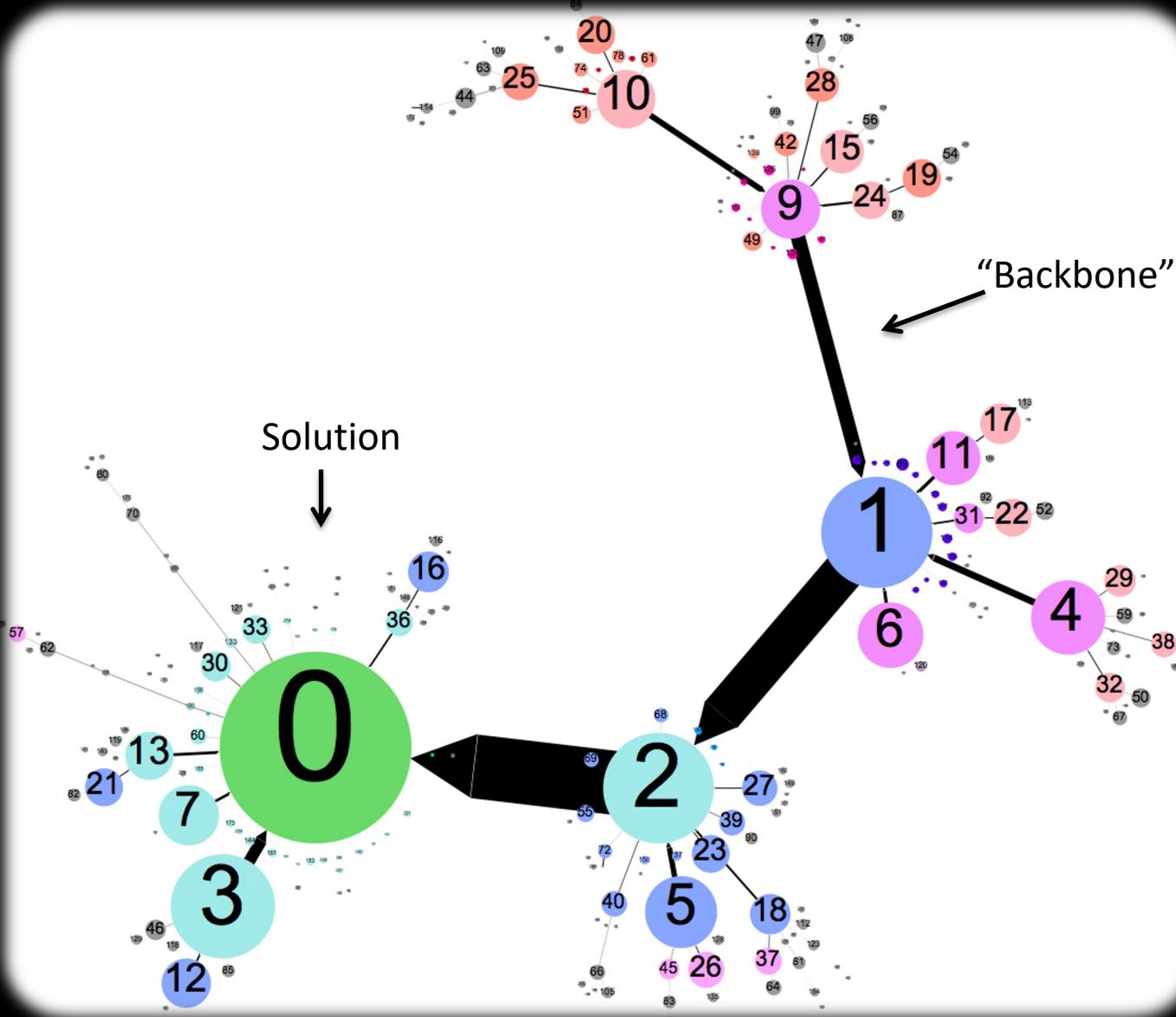
Predicted next partial solution

Paths to solution

Partial solutions in the path



Learned Problem Solving Policy



Only worked well for 6 line programs...

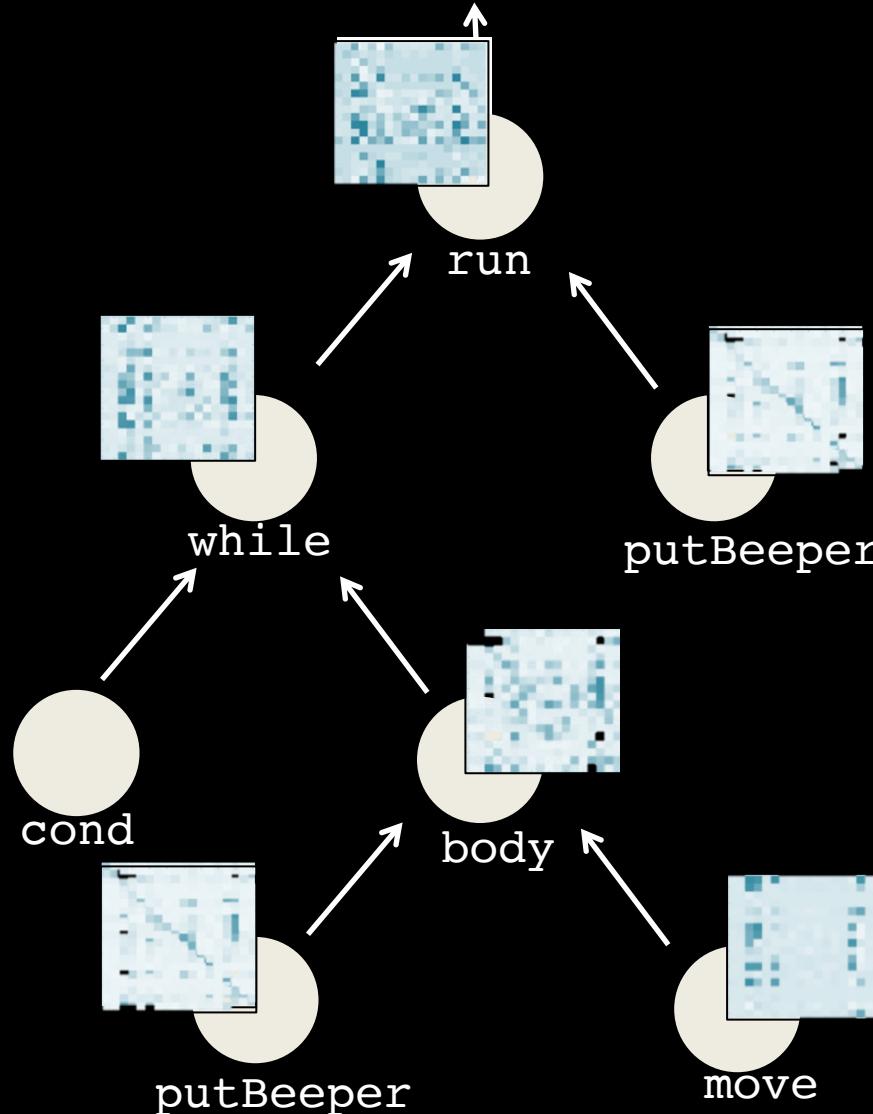
Chapter 2: Start to invent new algorithms...



Neural Network to Encode a Program

```
// User defined method  
private void run() {  
    while(isClear()){  
        putBeeper();  
        move();  
    }  
    putBeeper();  
}
```

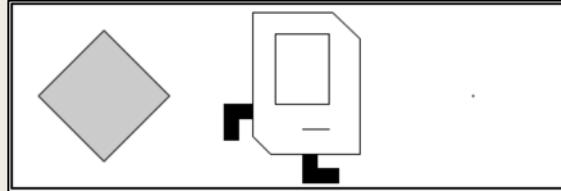
It looks like you have a fencepost error!



*Note: this was coded pre-tensor flow



Collect Triples



Precondition

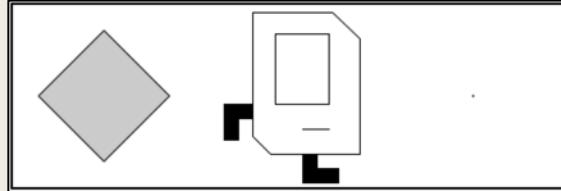
```
putBeeper();  
move();
```

Code

About 5 million triples per assignment



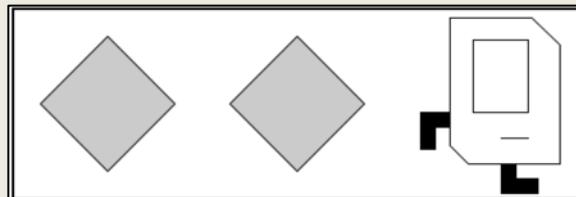
Collect Triples



Precondition

```
putBeeper();  
move();
```

Code



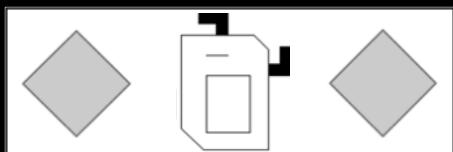
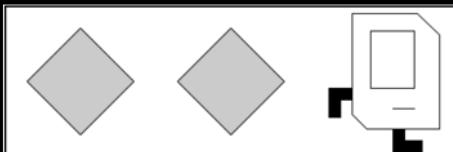
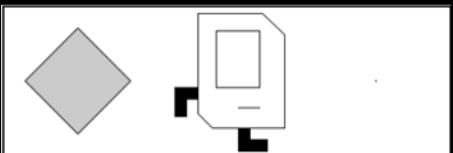
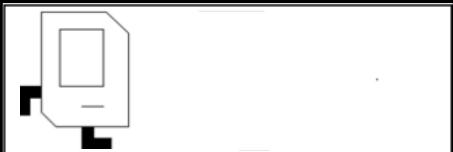
Postcondition

About 5 million triples per assignment



A Code Phrase is a Mapping

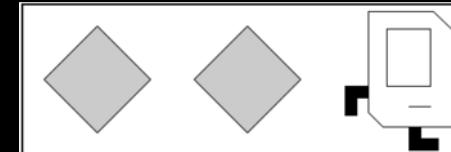
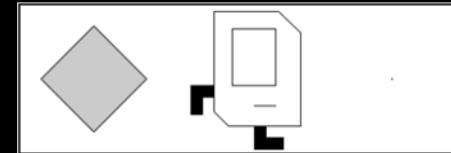
All possible preconditions



⋮



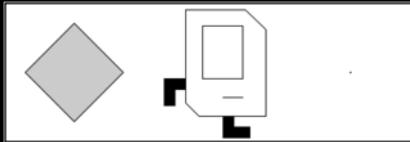
All possible postconditions



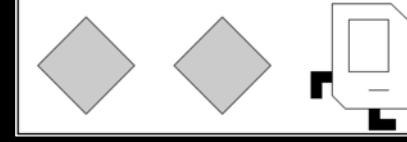
⋮



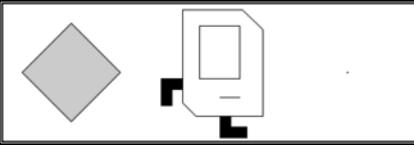
Neural Network for Programs



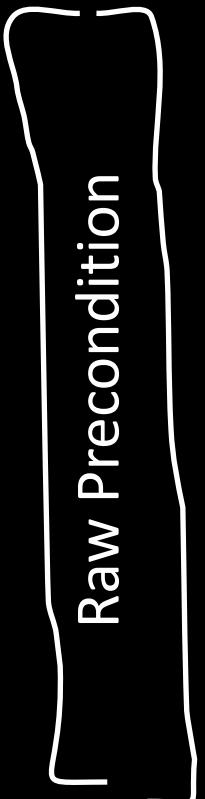
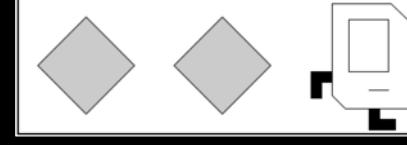
```
method step() {  
    putBeeper();  
    move();  
}
```



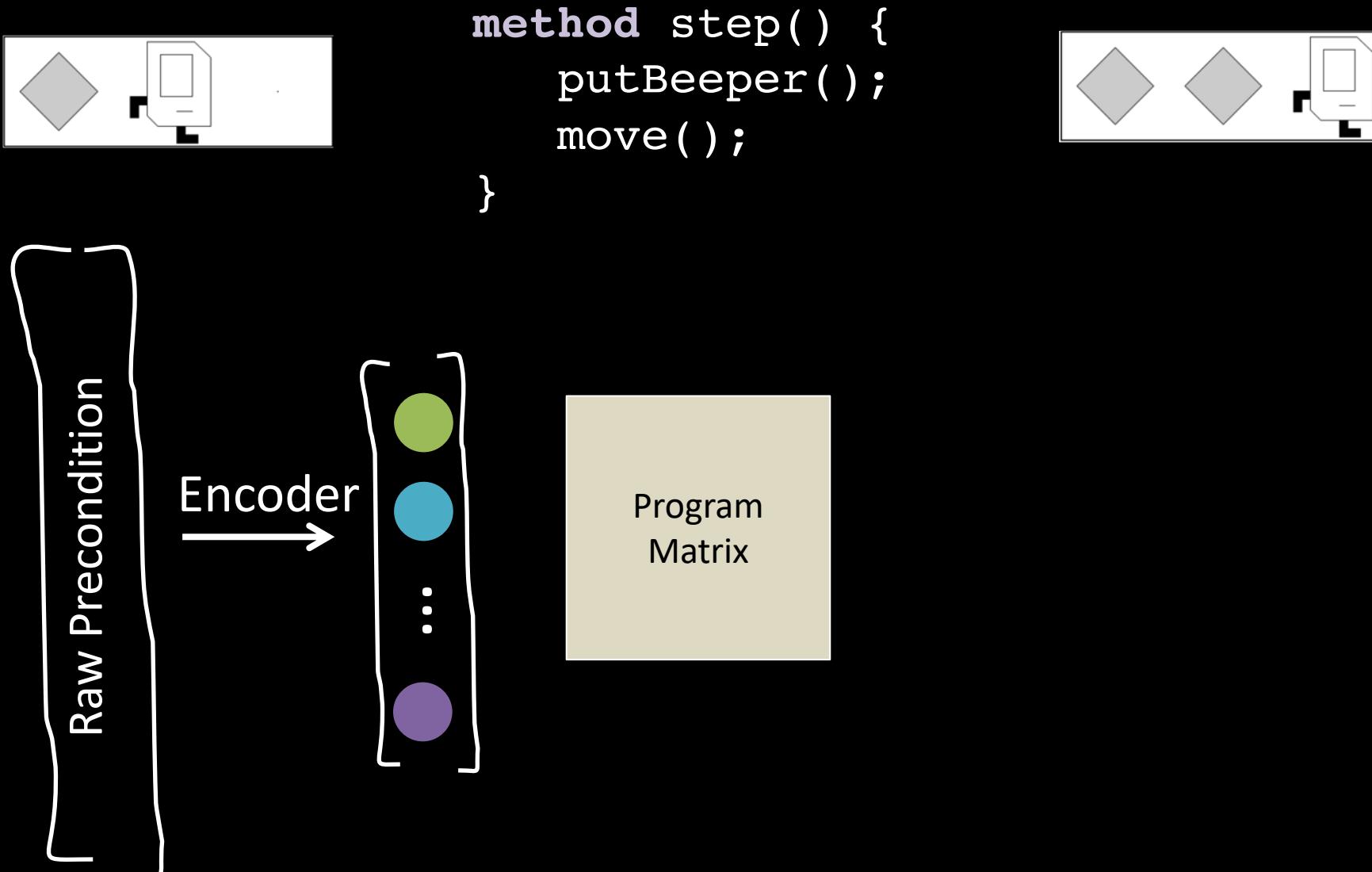
Neural Network for Programs



```
method step() {  
    putBeeper();  
    move();  
}
```



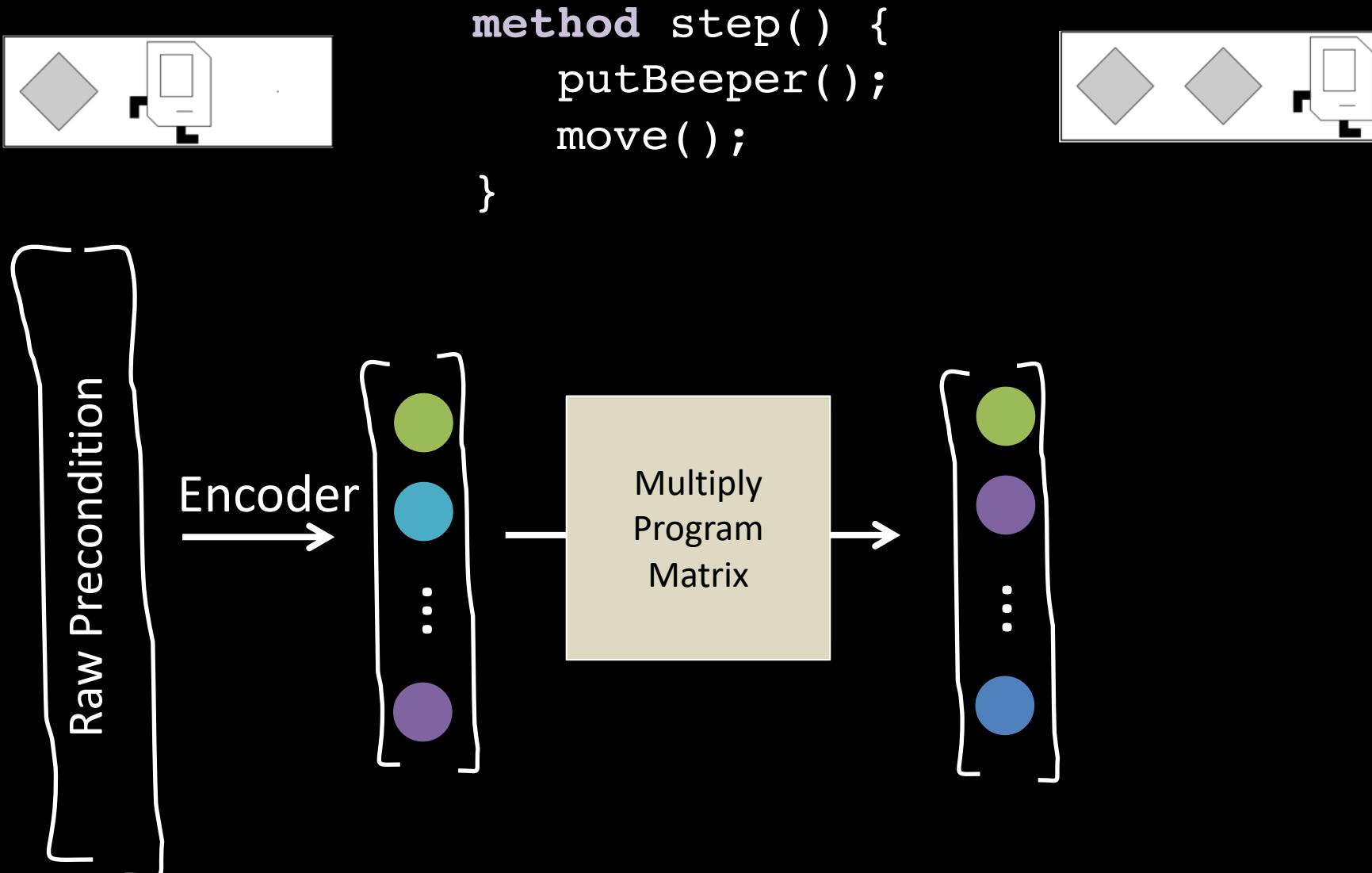
Neural Network for Programs



*coded pre-tensor flow



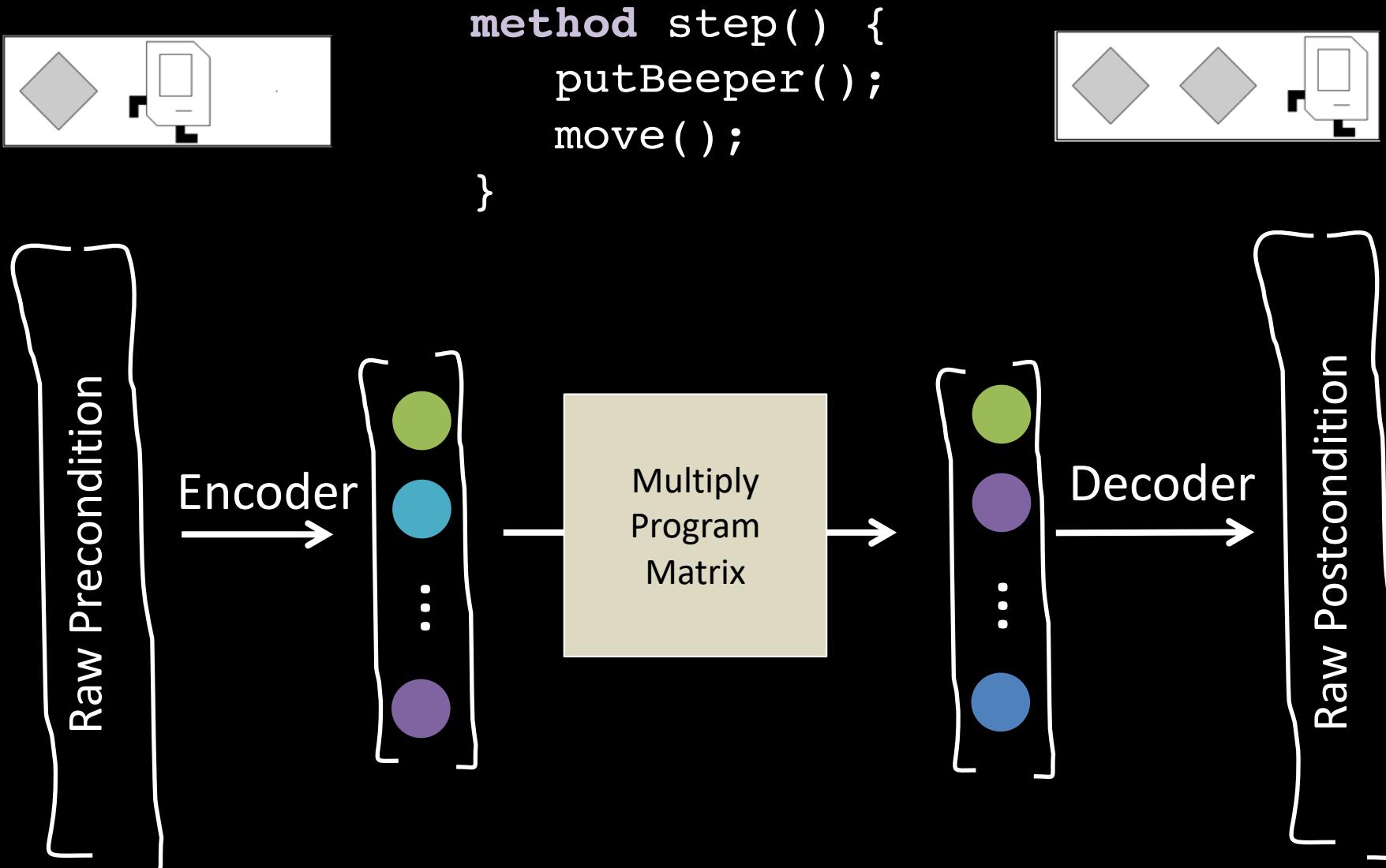
Neural Network for Programs



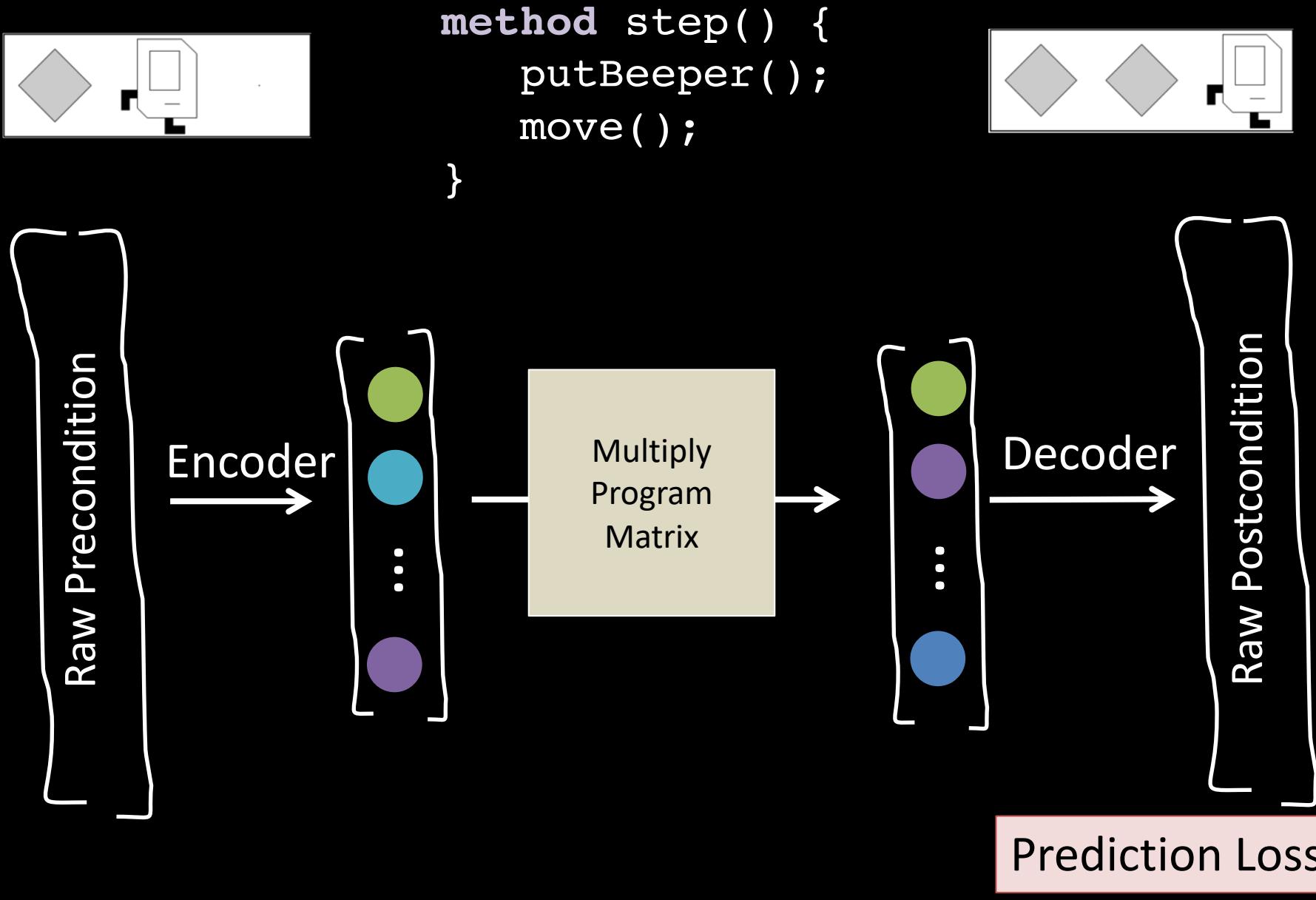
*coded pre-tensor flow



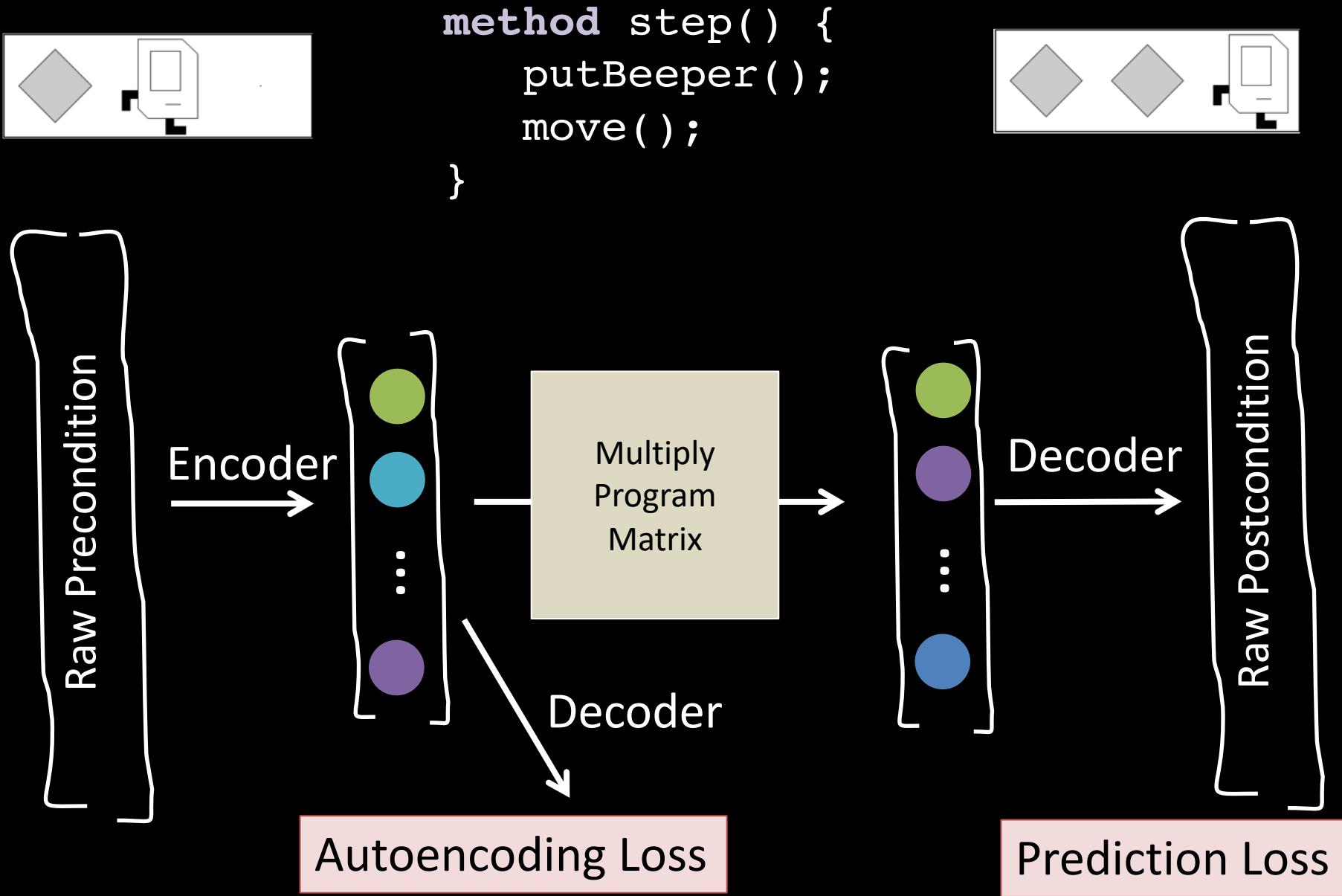
Neural Network for Programs



Neural Network for Programs



Neural Network for Programs

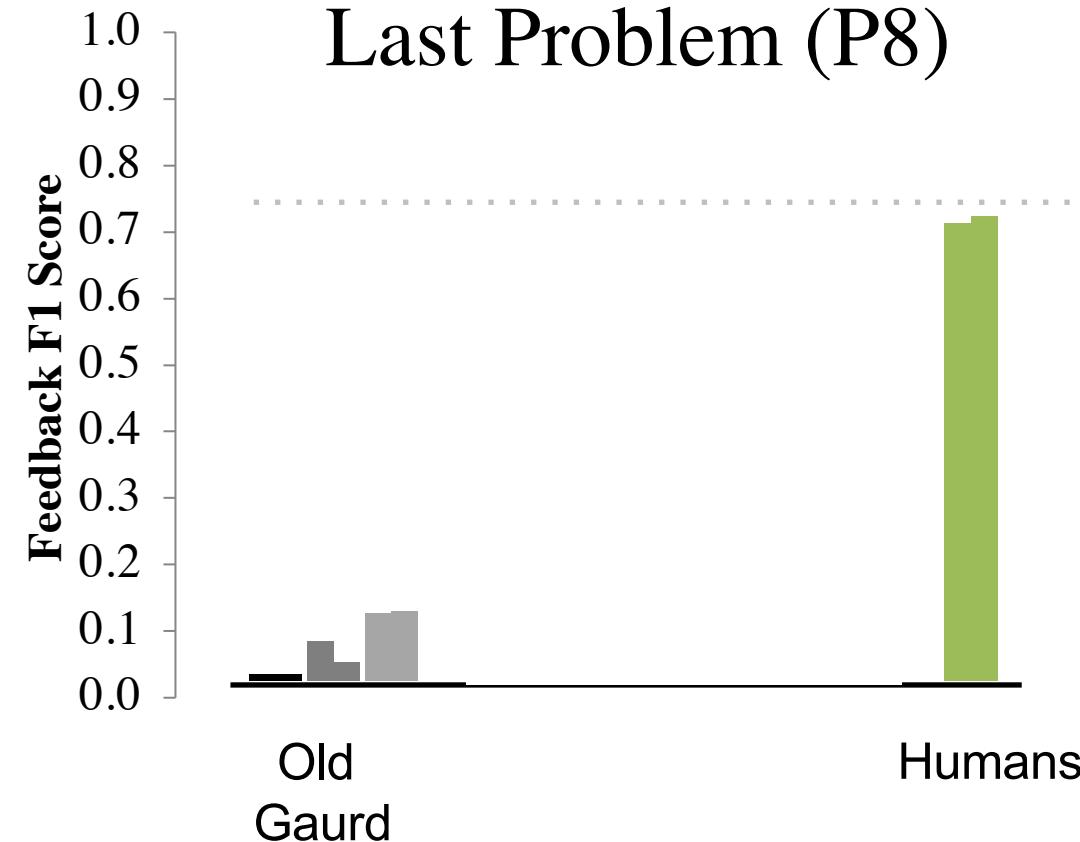


Does it work?



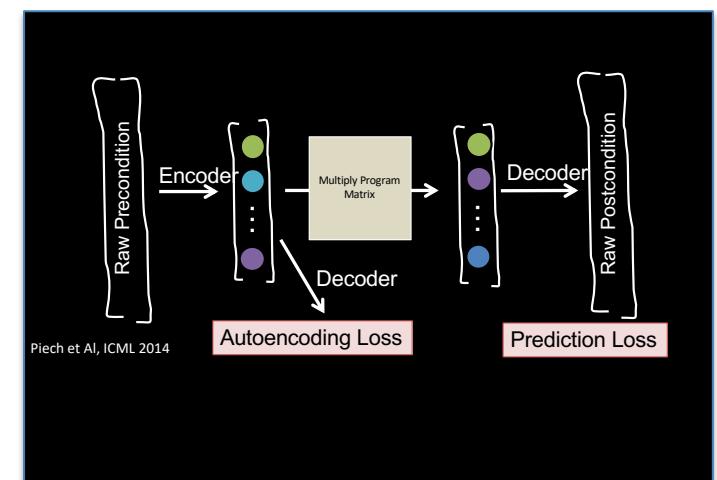
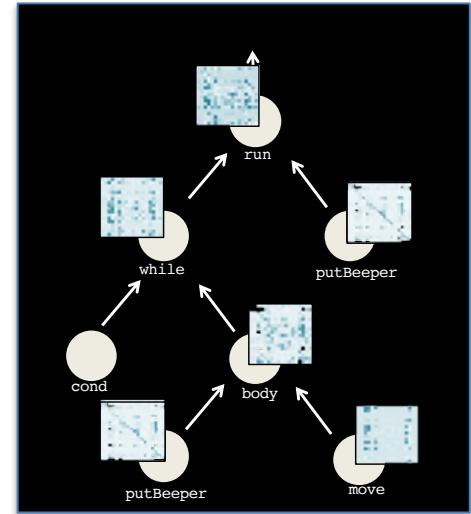
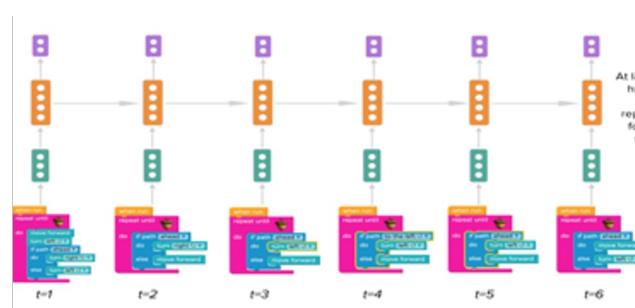
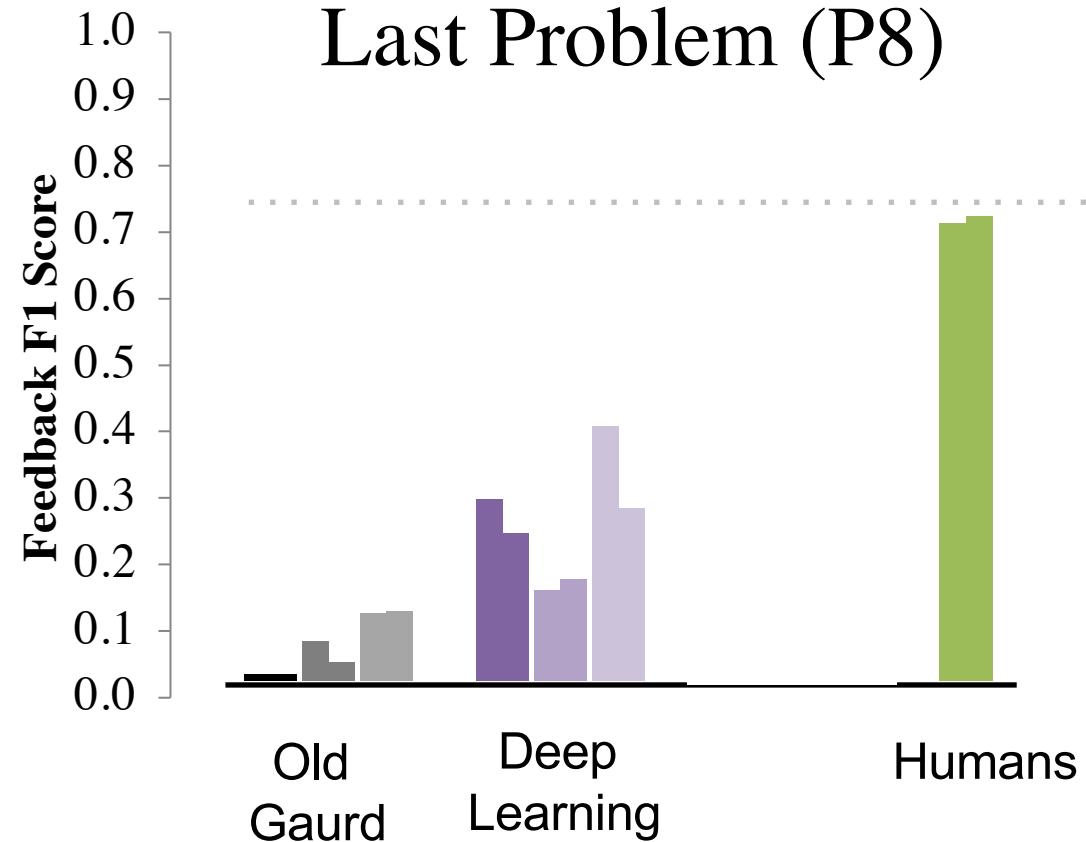
Traditional Deep Learning Doesn't Work

Label student code



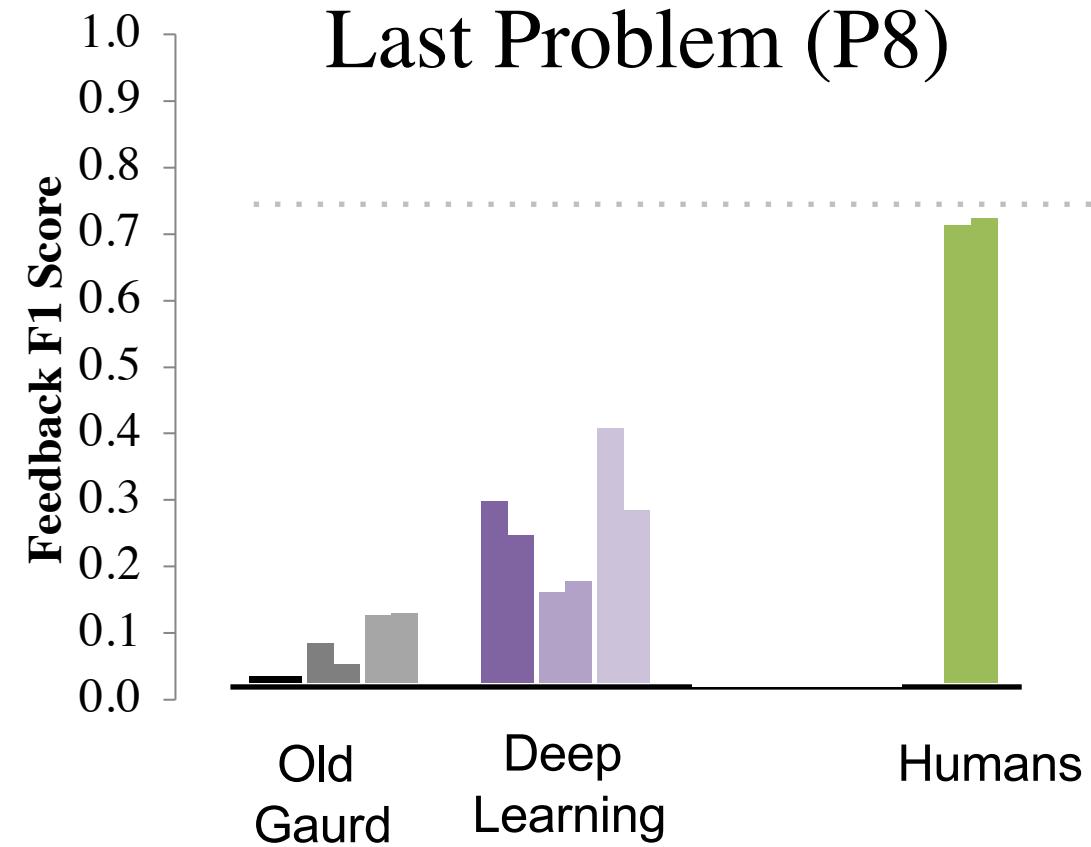
Inaccurate, Uninterpretable, and Data Hungry

Label student code



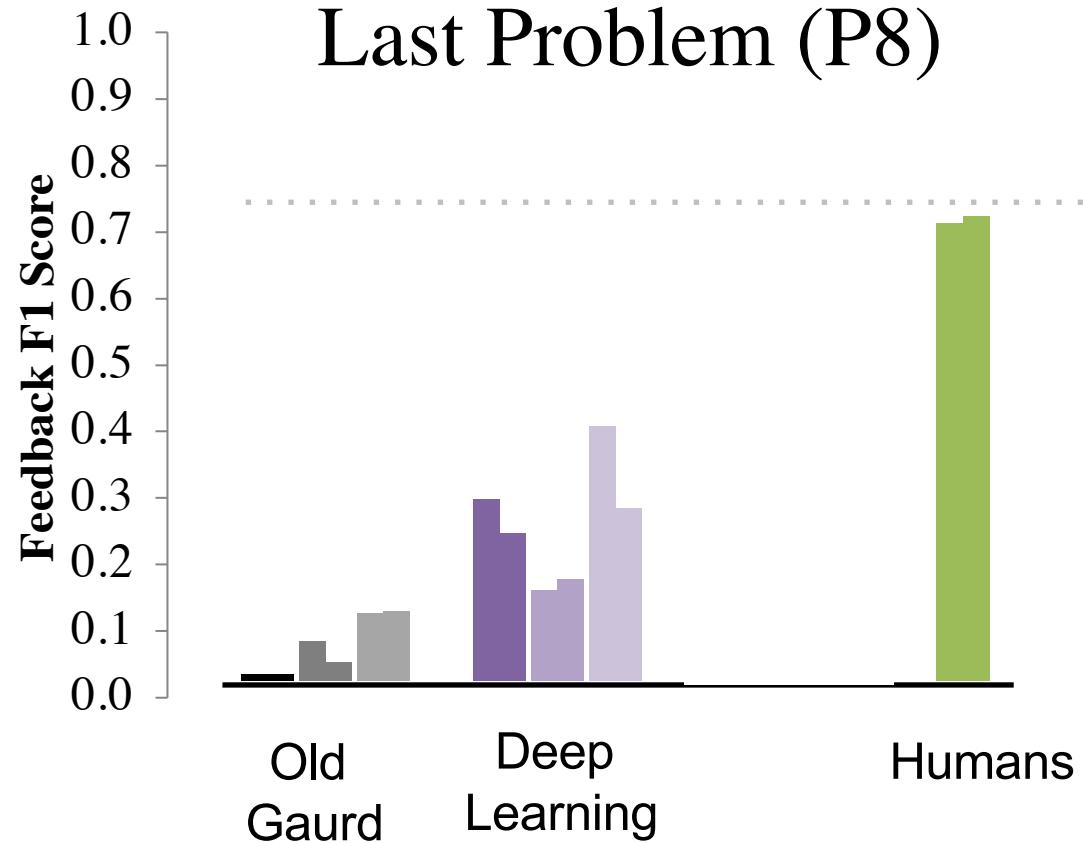
Inaccurate, Uninterpretable, and Data Hungry

Label student code



Inaccurate, Uninterpretable, and Data Hungry

Label student code



We need one shot learning

We need verifiability



Why is it so hard?

Hard Problem

Brute force solution?

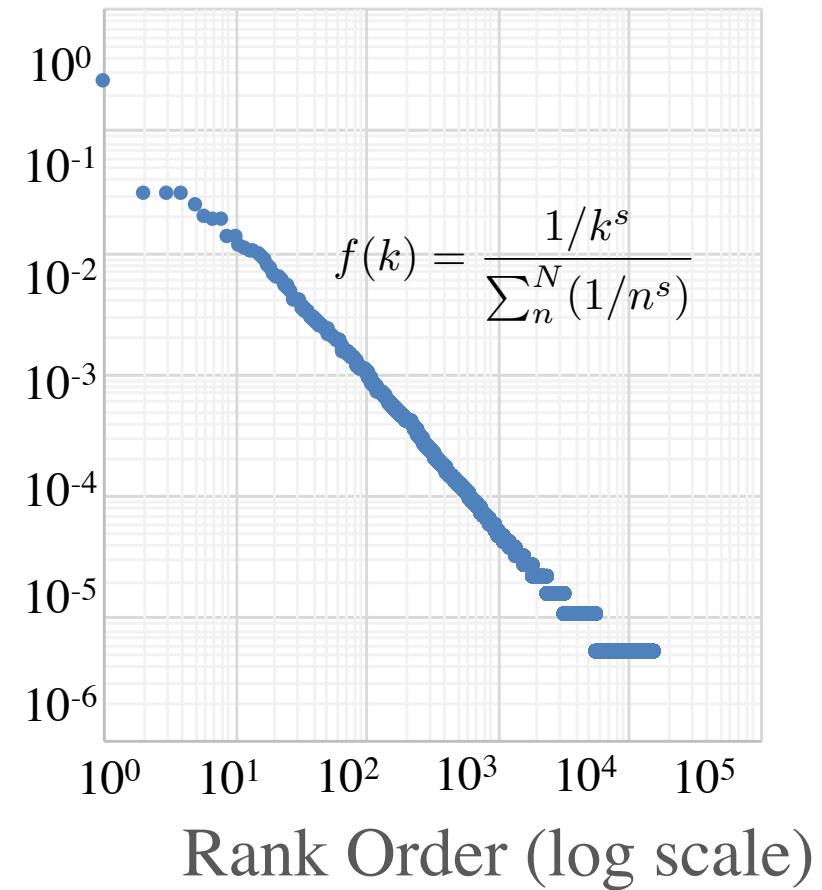


1 million unique solutions to
programming Linear Regression

WWW 2014

Probability Mass
(log scale)

Code Zipf Plot



They are all Zipf!

(a) Datasets in Computational Education

Code.org Problem 8

Draw me!

```
for [i v] from [3] to [10] by [2]
  do
    for [j v] from [0] to [i]
      do
        Move forward [10 x:i]
        Turn right (360 / i)
      end
    end
  end
end
```

for [i v] from [3] to [9] by [2]
do
repeat for [i]
do
 move forward [10 x:j]
 turn right (360 / i)
end
end
end

Powergrading P13

What is one reason the original colonists came to America?

- declared our independence from england
- religeous freedom
- as a criminal punishment

- Religious freedom
- For religious freedom
- Freedom

- to create a new colony
- to find better economic prospects
- to break away from the church in great britain

CS1: Liftoff

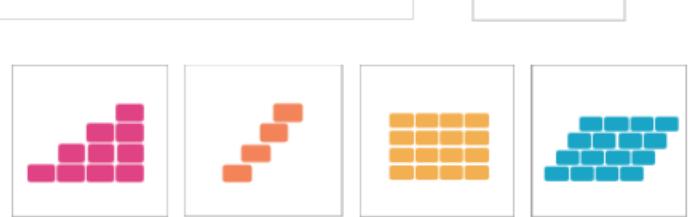
Write a Java Program to print the numbers 10 down to 1 and then write liftoff. You must use a loop.

```
public void run() {
  for (int i=START; i>0; i--) {
    println(i);
  }
  println("Liftoff!");
}
```

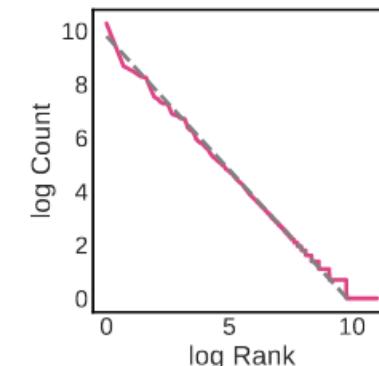
```
public void run() {
  for(int i = START; i>0; i--)
  {
    println(i);
    pause(1000);
  }
  println("Liftoff!");
}
```

PyramidSnapshot

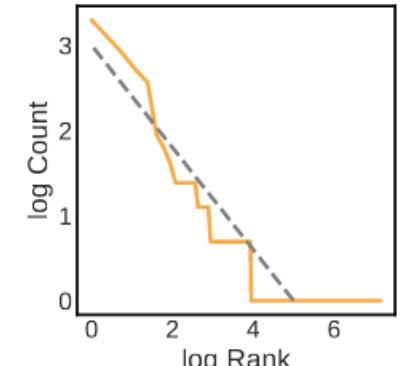
Use the graphics library to construct a symmetric and centered pyramid with a base width of 14 bricks.



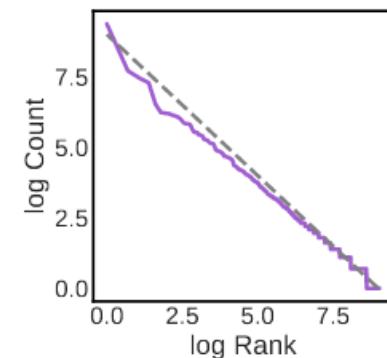
(b) Code.org P8



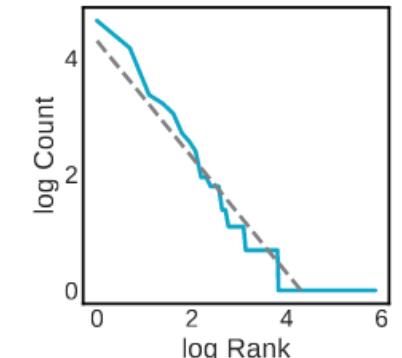
(c) CS1: Liftoff



(d) Pyramid



(e) Powergrading



[Suspense]

Chapter 3: Back to the drawing board

Humans Don't Need Much Data

Single training example:

ବୁ

Test set:

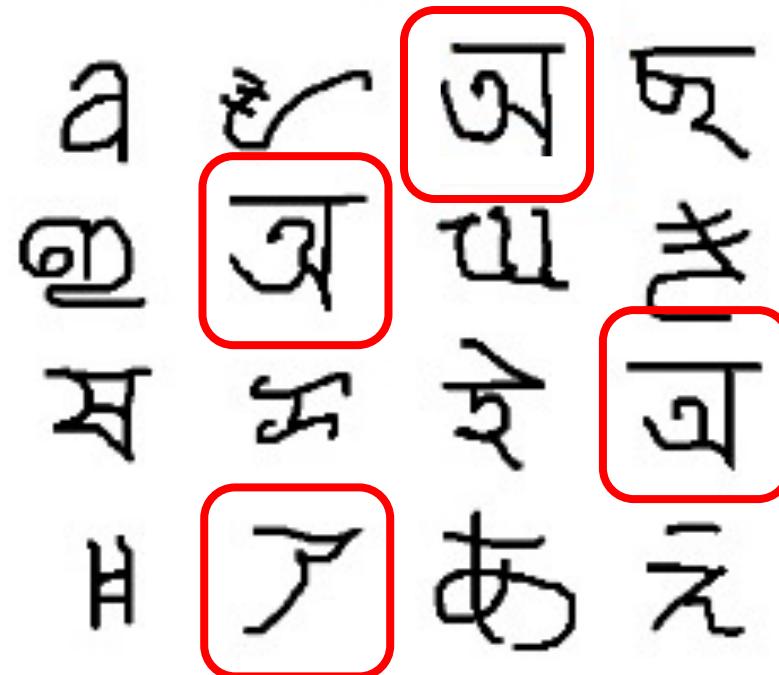
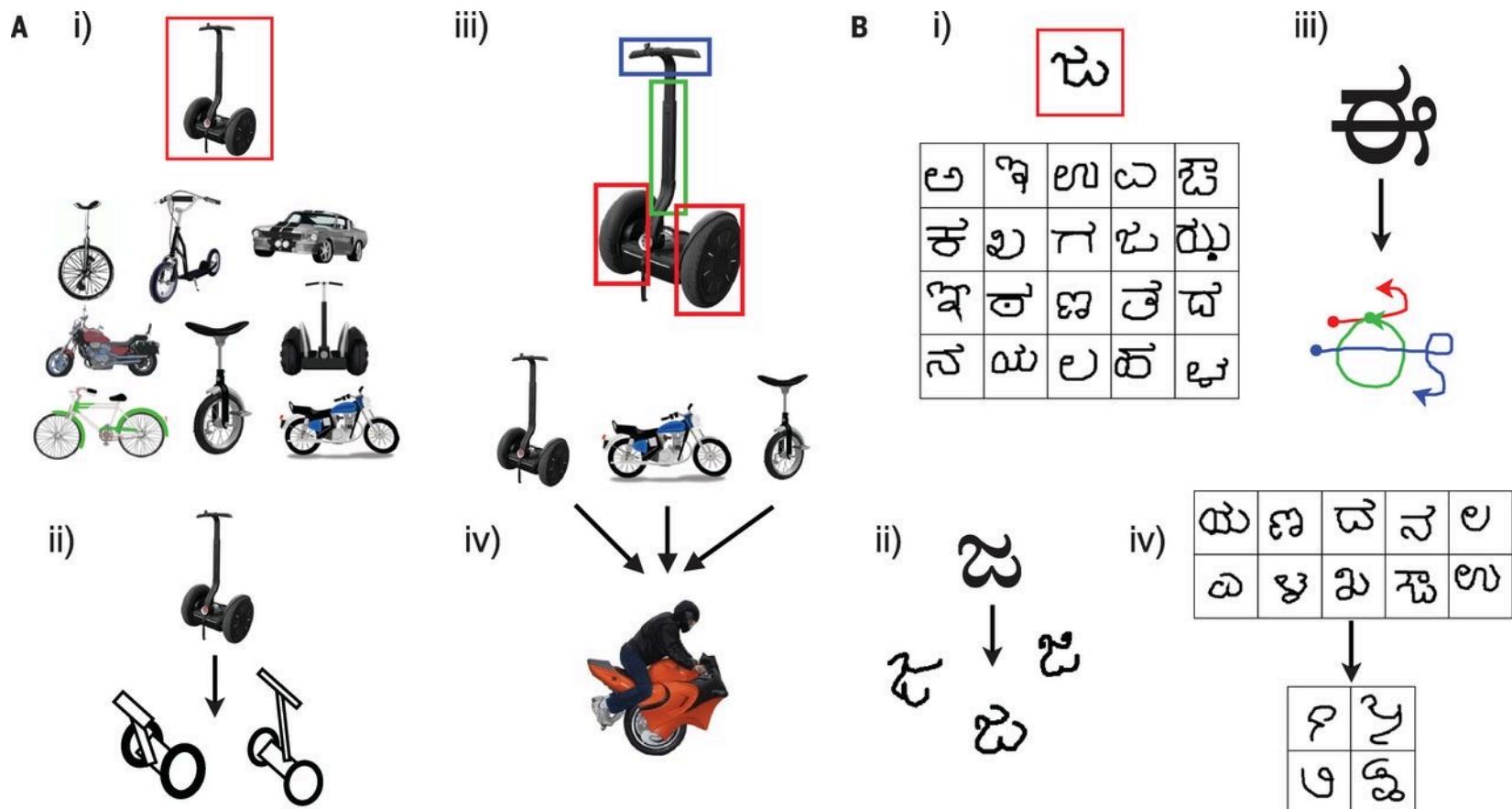


Fig. 1 People can learn rich concepts from limited data.



Brenden M. Lake et al. Science 2015;350:1332-1338

Copyright © 2015, American Association for the Advancement of Science

Science
AAAS



Fig. 2 Simple visual concepts for comparing human and machine learning.

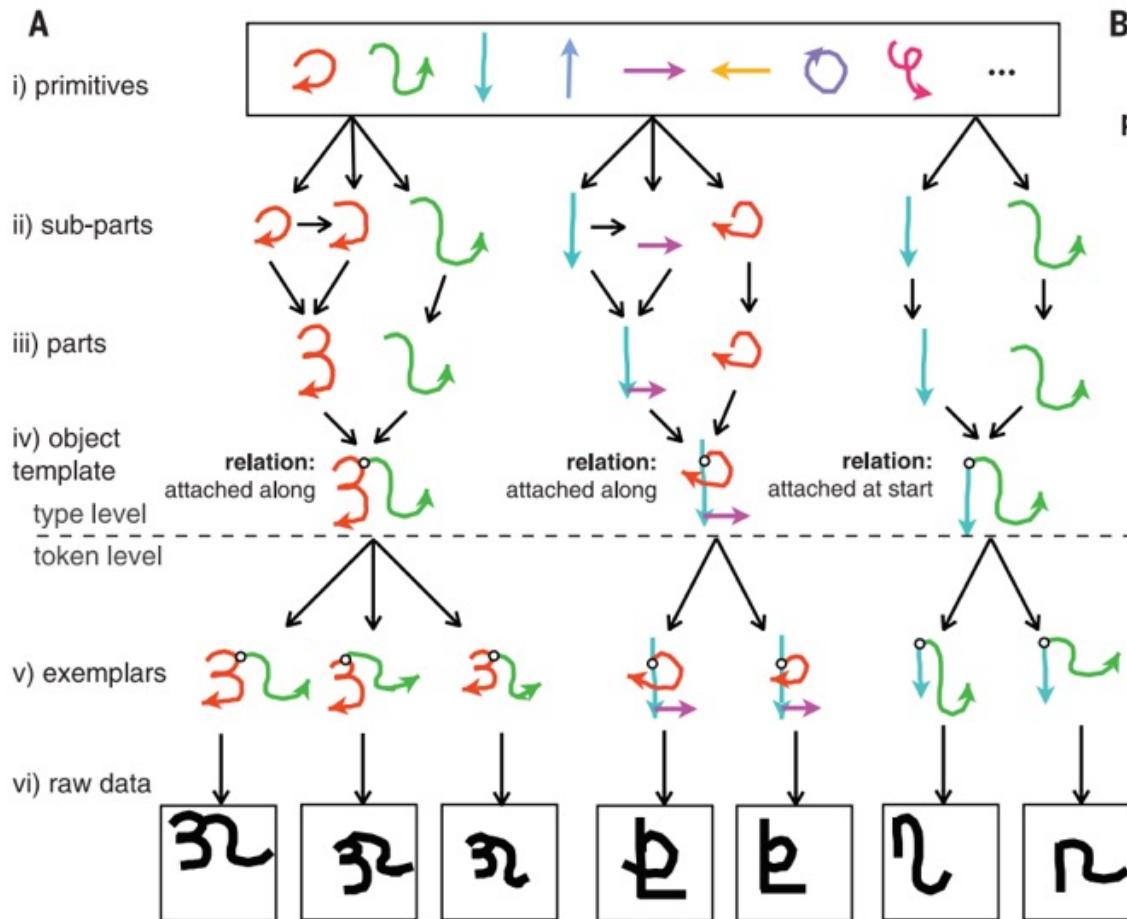


Brenden M. Lake et al. Science 2015;350:1332-1338

Copyright © 2015, American Association for the Advancement of Science

Science
AAAS

Bayesian Program Learning



B

```

procedure GENERATETYPE
   $\kappa \leftarrow P(\kappa)$                                 ▷ Sample number of parts
  for  $i = 1 \dots \kappa$  do
     $n_i \leftarrow P(n_i|\kappa)$                           ▷ Sample number of sub-parts
    for  $j = 1 \dots n_i$  do
       $s_{ij} \leftarrow P(s_{ij}|s_{i(j-1)})$  ▷ Sample sub-part sequence
    end for
     $R_i \leftarrow P(R_i|S_1, \dots, S_{i-1})$            ▷ Sample relation
  end for
   $\psi \leftarrow \{\kappa, R, S\}$ 
  return @GENERATETOKEN( $\psi$ )                         ▷ Return program

```

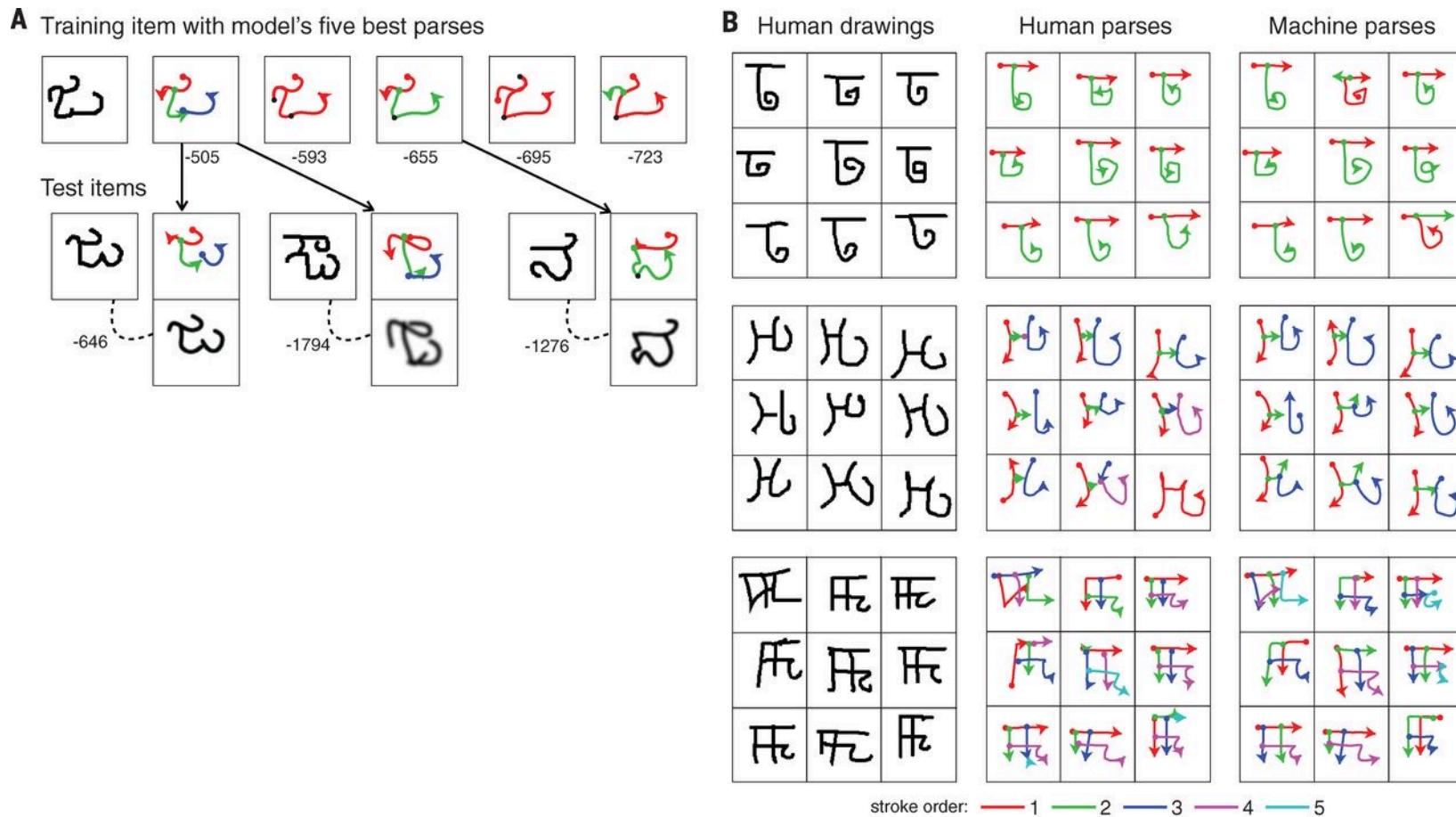
```

procedure GENERATETOKEN( $\psi$ )
  for  $i = 1 \dots \kappa$  do
     $S_i^{(m)} \leftarrow P(S_i^{(m)}|S_i)$                 ▷ Add motor variance
     $L_i^{(m)} \leftarrow P(L_i^{(m)}|R_i, T_1^{(m)}, \dots, T_{i-1}^{(m)})$  ▷ Sample part's start location
     $T_i^{(m)} \leftarrow f(L_i^{(m)}, S_i^{(m)})$  ▷ Compose a part's trajectory
  end for
   $A^{(m)} \leftarrow P(A^{(m)})$                         ▷ Sample affine transform
   $I^{(m)} \leftarrow P(I^{(m)}|T^{(m)}, A^{(m)})$           ▷ Sample image
  return  $I^{(m)}$ 

```



Fig. 4 Inferring motor programs from images.



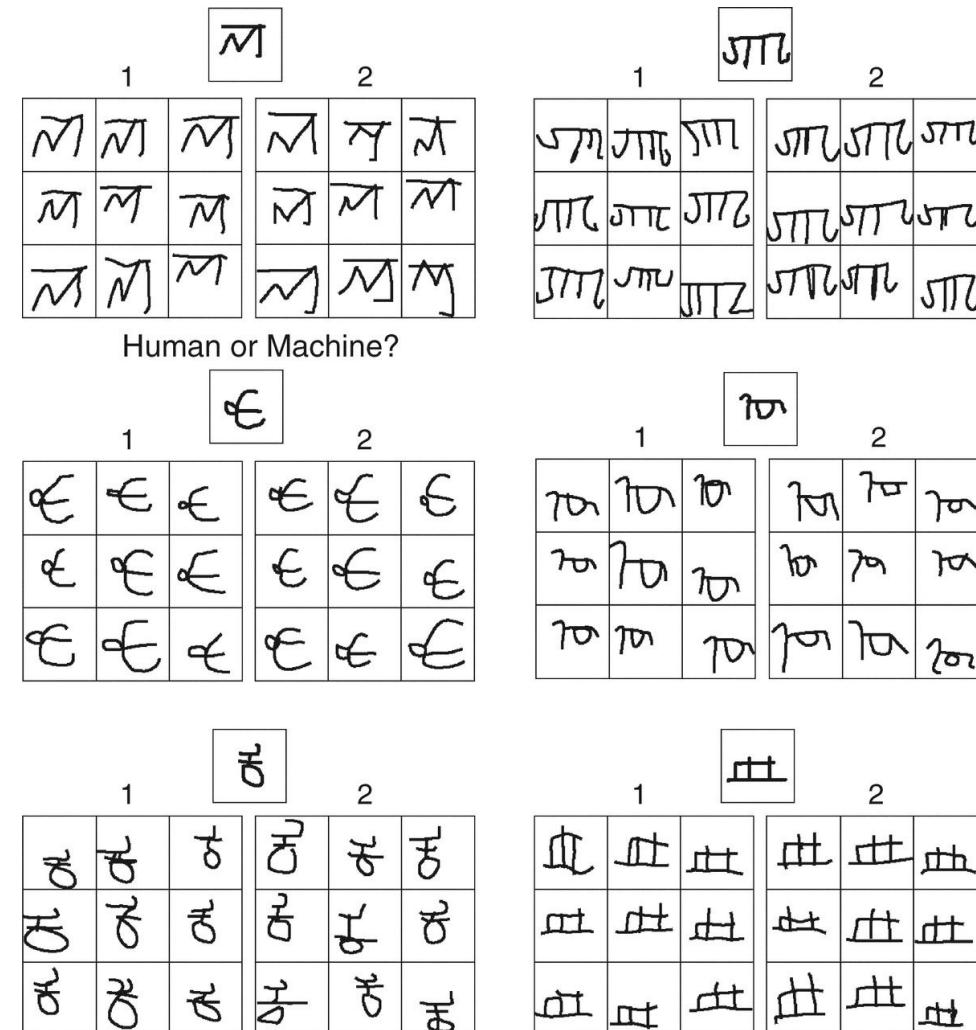
Brenden M. Lake et al. Science 2015;350:1332-1338

Copyright © 2015, American Association for the Advancement of Science

Science
AAAS



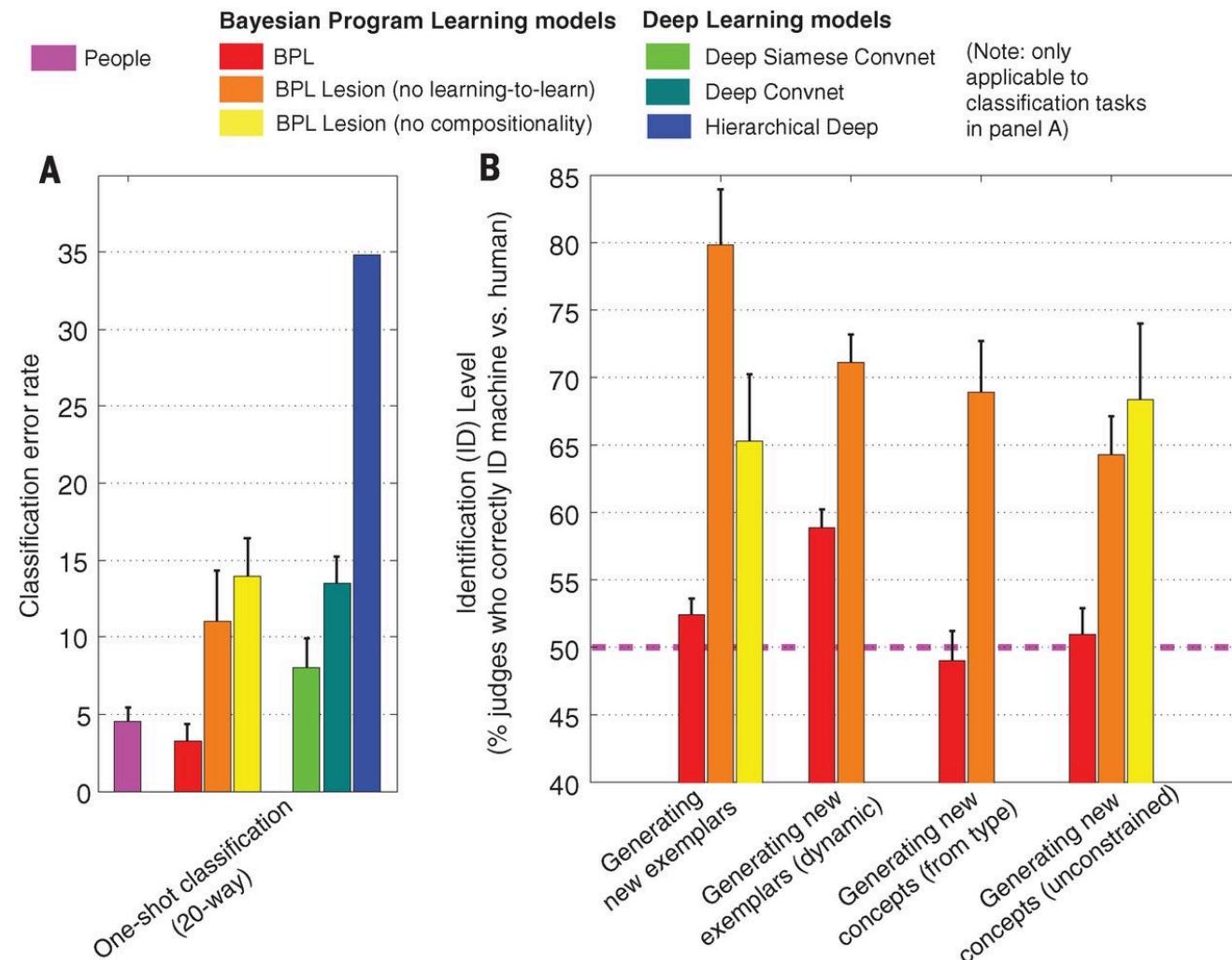
Fig. 5 Generating new exemplars.



Brenden M. Lake et al. Science 2015;350:1332-1338



Fig. 6 Human and machine performance was compared on (A) one-shot classification and (B) four generative tasks.



Brenden M. Lake et al. Science 2015;350:1332-1338

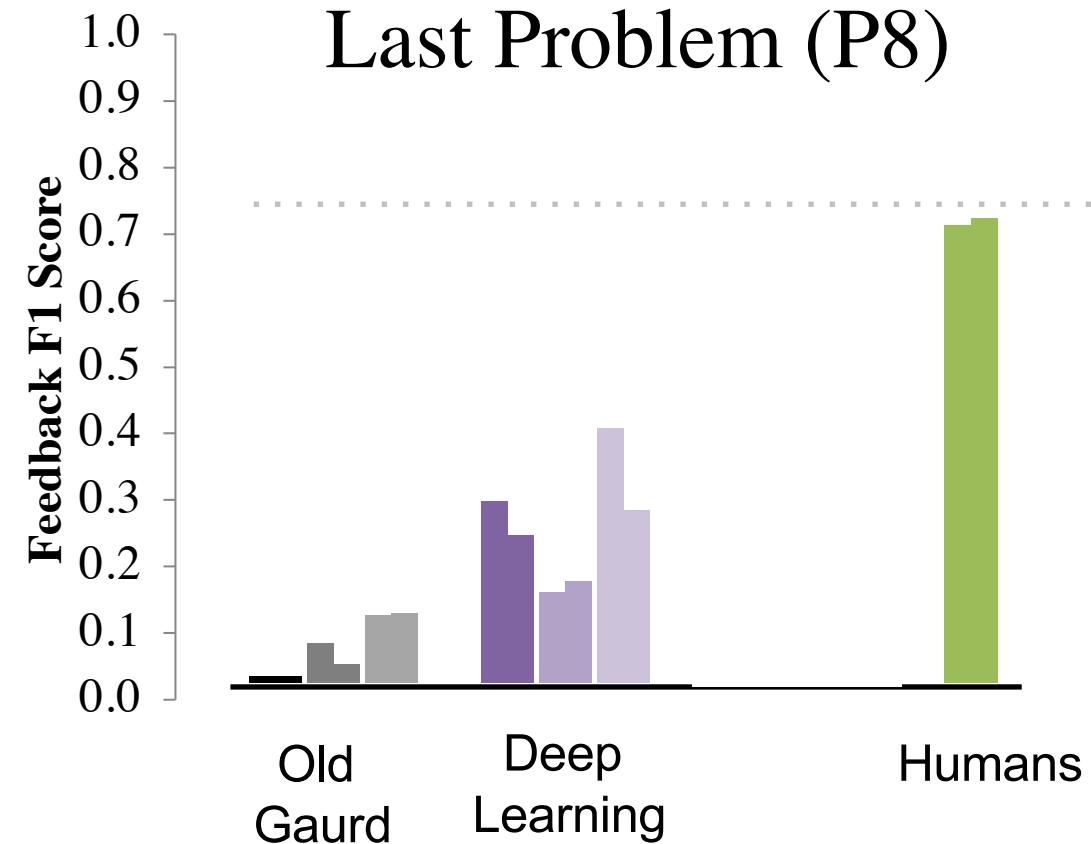
Copyright © 2015, American Association for the Advancement of Science

Science
AAAS



Generative Understanding

Label student code

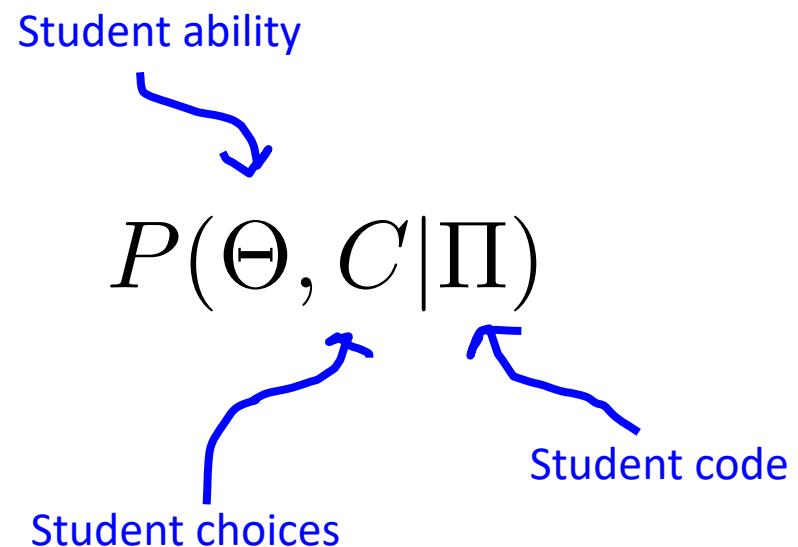


Grading is Hard

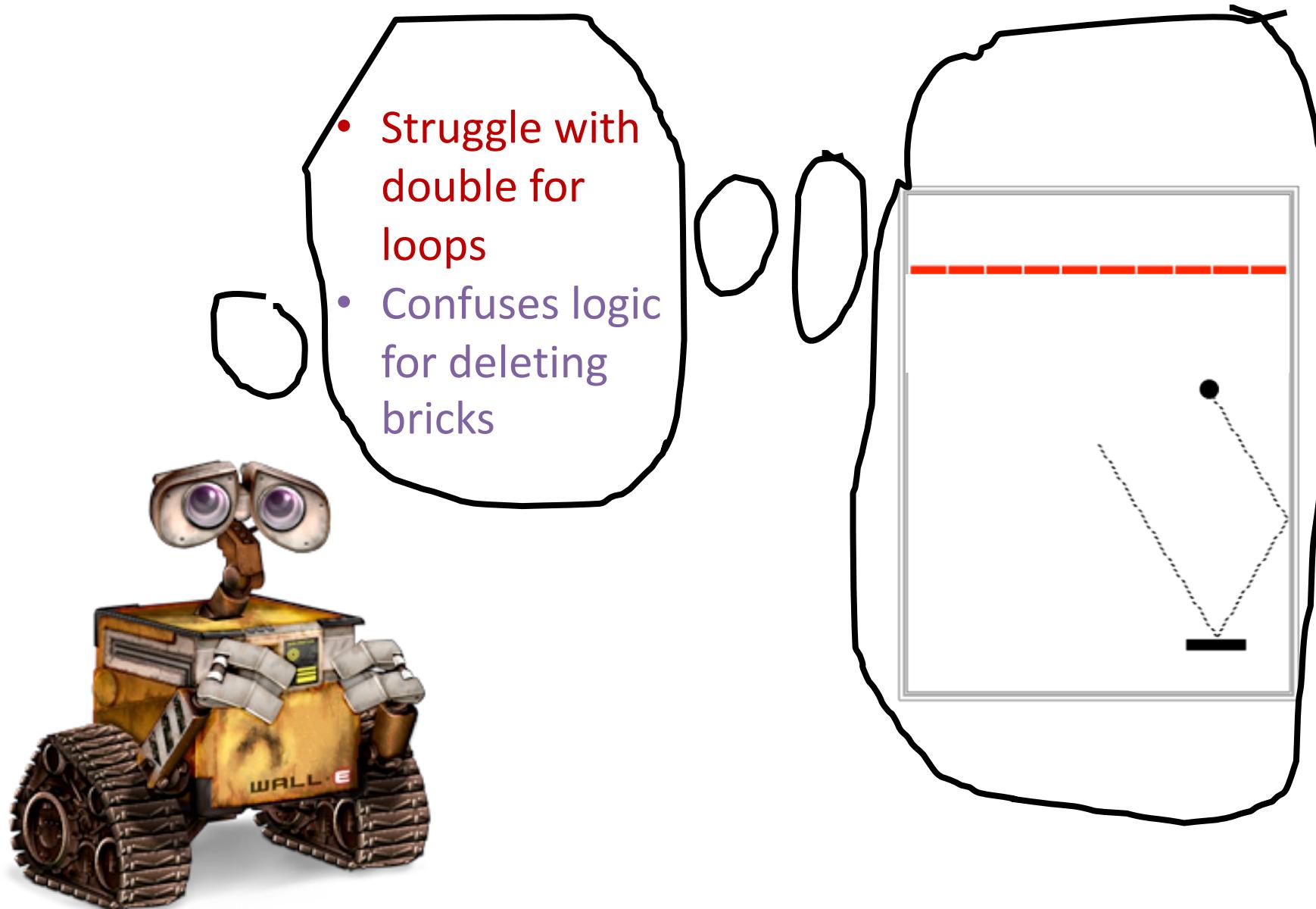


Grading:

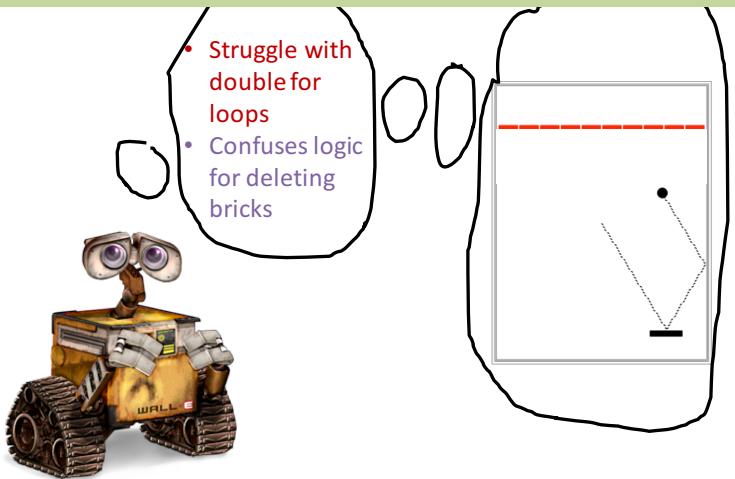
Infer ability and
choices from code



Generative Probabilistic Model?



Imagine Students



This is easy and exponential
human impact

A students
“ability”

$$\Theta \sim \text{pythonSample}$$

A students
“choices”

$$C \sim \text{pythonSample} | \Theta$$

The resulting
code

$$\Pi \sim \text{pythonSample} | C$$



This is hard and linear
human impact

$$P(\Theta, C | \Pi)$$

Infer ability and
choices from code



Bayesian Programming Language



ideaToText



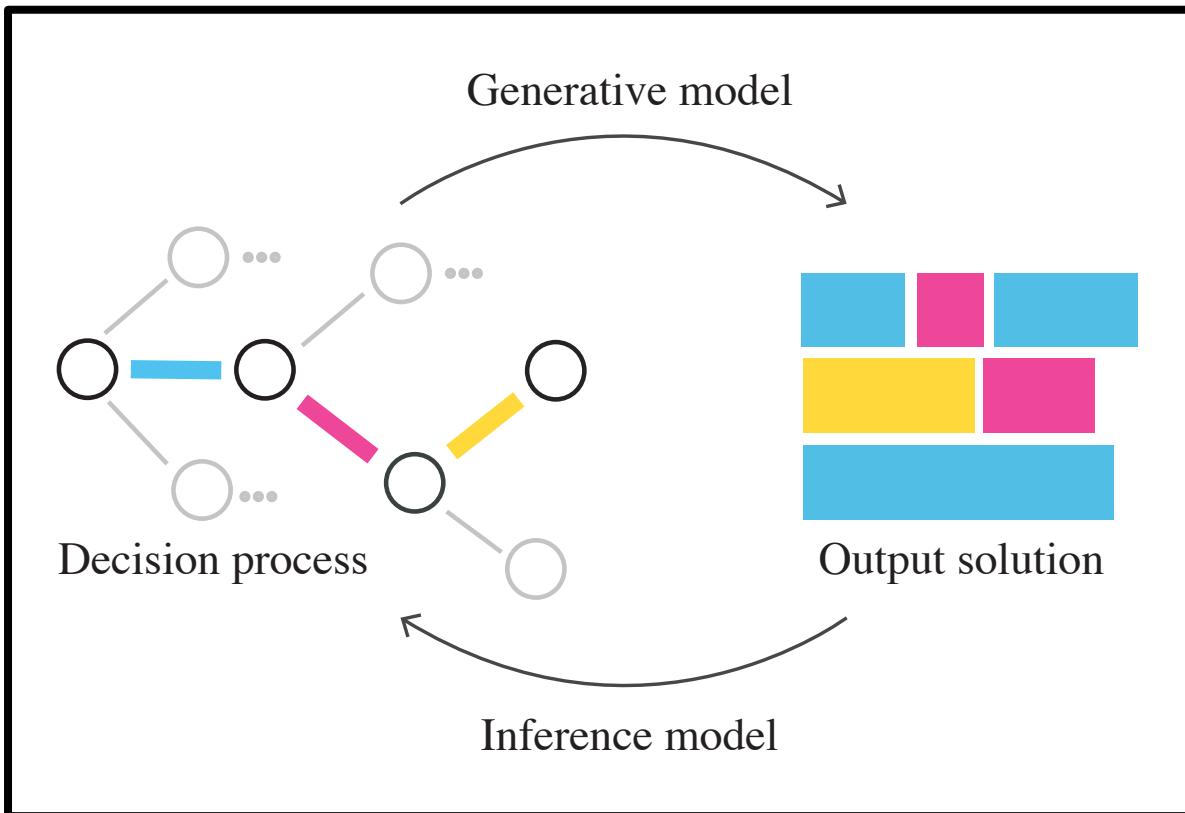
Teachers Articulate N misconceptions

```
7 # This python class is a RubricSampling Decision
8 # it generates programs that print the numbers 10 -> 1
9
10 class Countdown(Decision):
11
12     def registerChoices(self):
13         # these are the main strategies for printing out a
14         # countdown
15         self.addRubricChoices('loop-style', {
16             'for' : θ1,
17             'while' : θ2,
18             'none' : θ3,
19             'empty' : θ4
20         })
21
22     # we can make some grading choices based on which
23     # strategy they chose (did they actually use a loop?)
24     def processChoices(self):
25         style = self.getChoice('loop-style')
26         hasLoop = style != 'none' and style != 'empty'
27         self.addLabel('rubric-hasLoop', hasLoop)
28
29
30     # Based on their strategy render a different decision
31     def renderCode(self):
32         style = self.getChoice('loop-style')
33         if style == 'for': return '{ForSoln}'
34         if style == 'while': return '{WhileSoln}'
35         if style == 'none': return '{NoLoopSoln}'
36         if style == 'empty': return ''
37
```

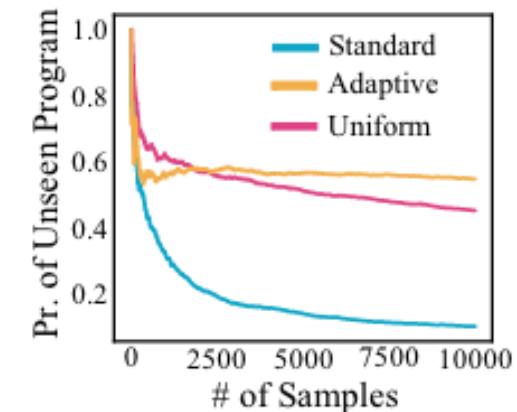
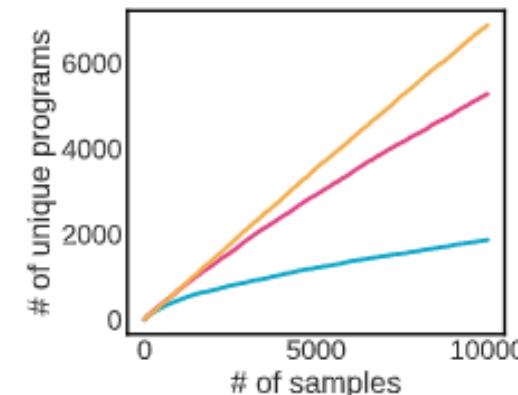
1. This is code for a single decision point
2. Give a name to the choice that the student is making (it is a random variable)
3. How do those choices translate into feedback?
4. What does the code look like? Often evokes other decision points



Generative Understanding



Idea: (1) sample a ton, then(2) build a neural network to learn to predict decisions



Decision process

Next choice

$$p_{\mathcal{G}}(x_{a_1}, \dots, x_{a_T} | y) = \prod_{t=1}^T p_{\mathcal{G}}(x_{a_t} | y, \mathbf{x}_{<a_t})$$

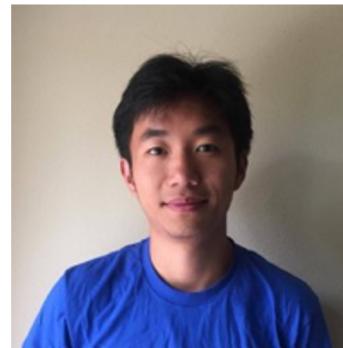
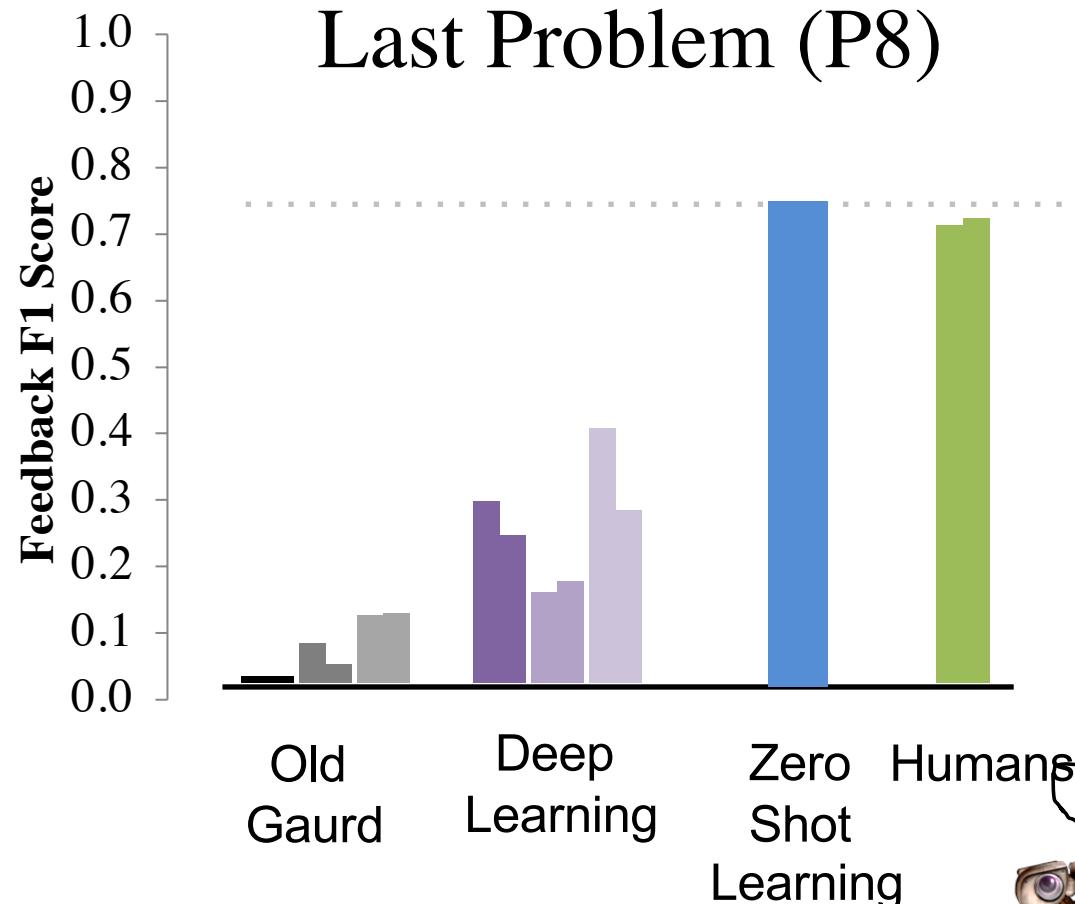
solution

Previous choices

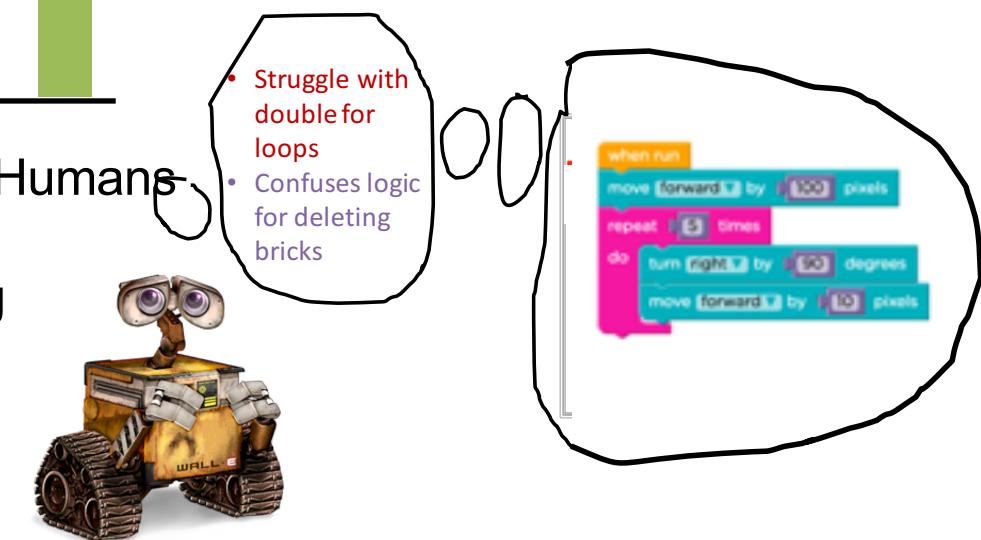


Generative Understanding

Label student code



*Outstanding Student
paper award, AAAI 2019*



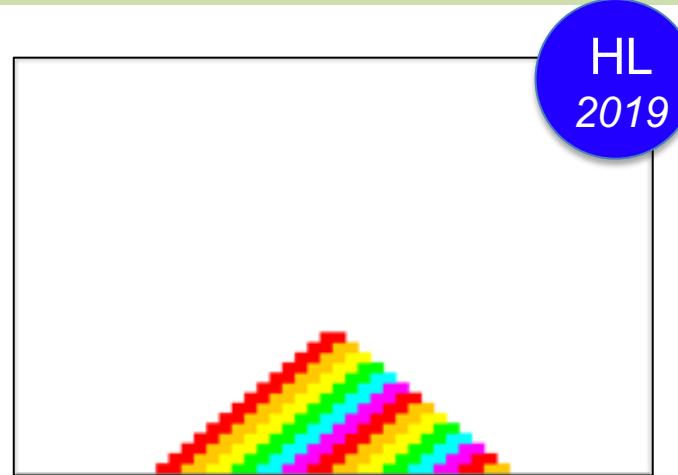
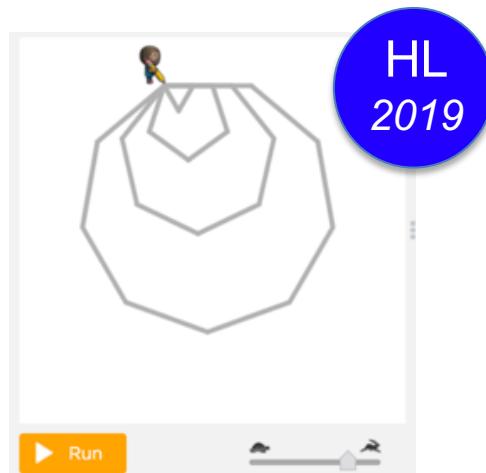
Not just for code

Results from early 2019

Many domains of student work

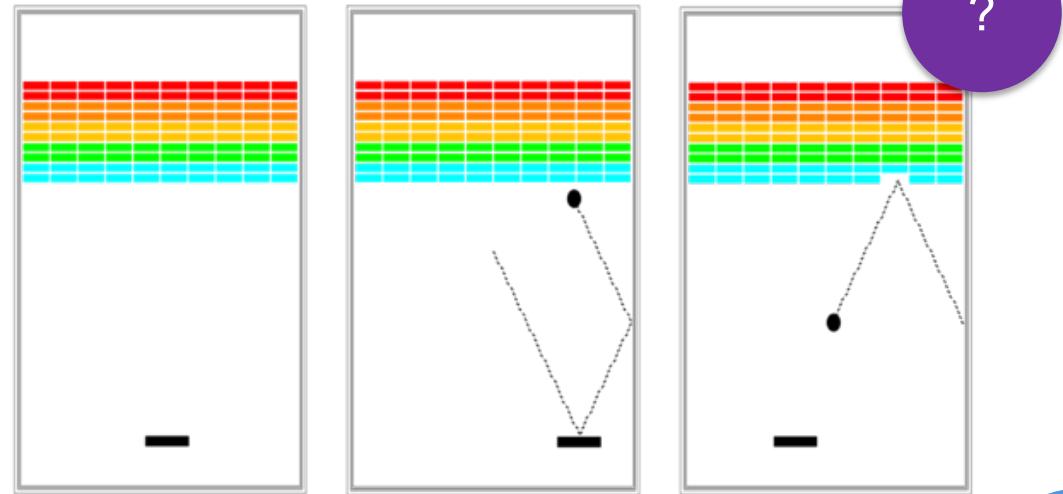
HL
2017

$$\begin{aligned} 9 \times 6 &= (10 - \square) \times 6 \\ &= 10 \times 6 - \square \times 6 \\ &= 60 - \square \\ &= \square \end{aligned}$$



NH
2019

Why did the original colonists come to America?



NH

Near Human

HL

Human Level

SH

Super Human Level



So what?

What does this mean for me?!?

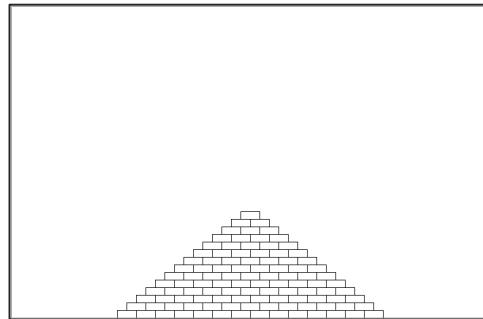
Understanding Process

2,600 students

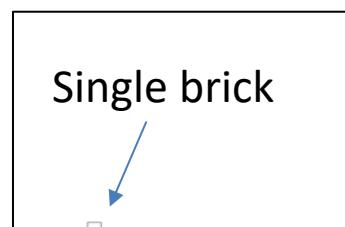
130,000 partial solutions

μ snapshots per student = 50

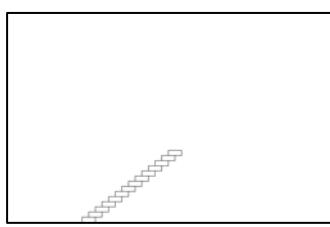
μ time per student = 2 hours



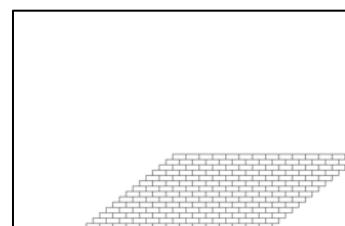
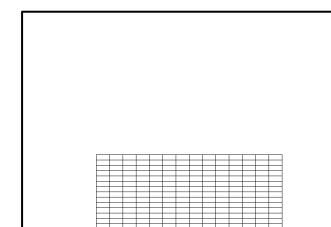
Step 1



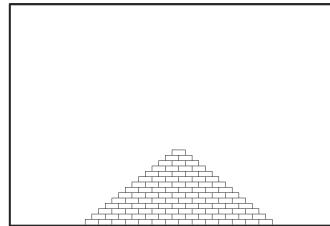
Step 2



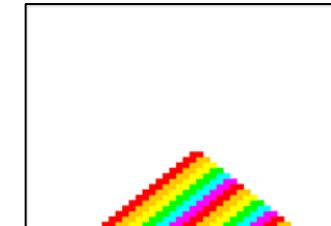
Step 3



Step 4



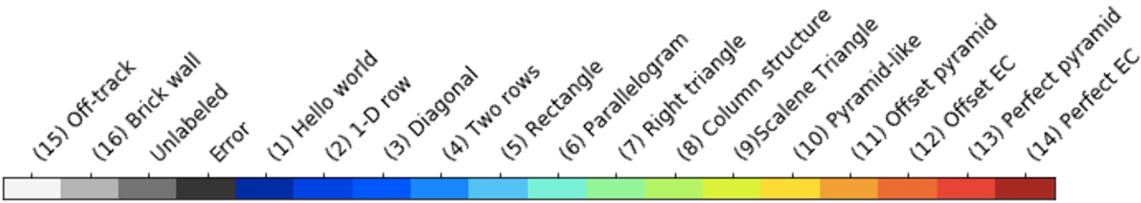
Step 5



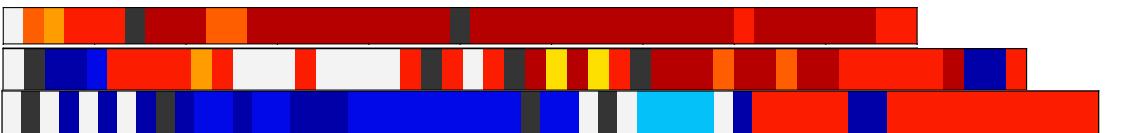
Step 6



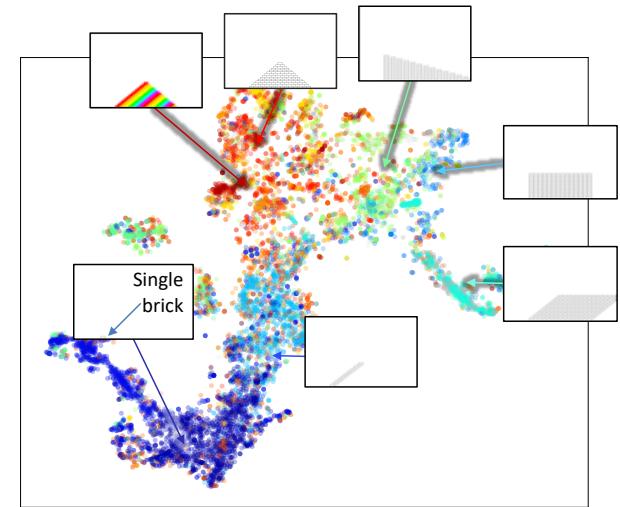
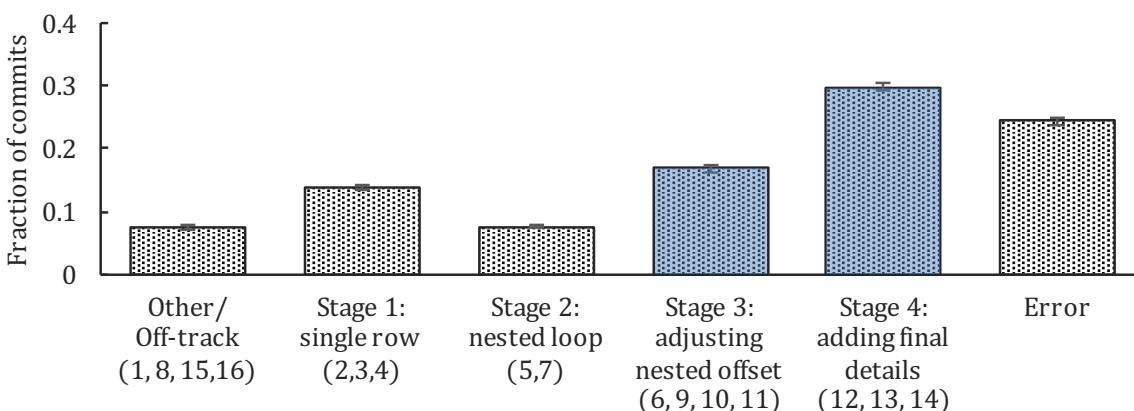
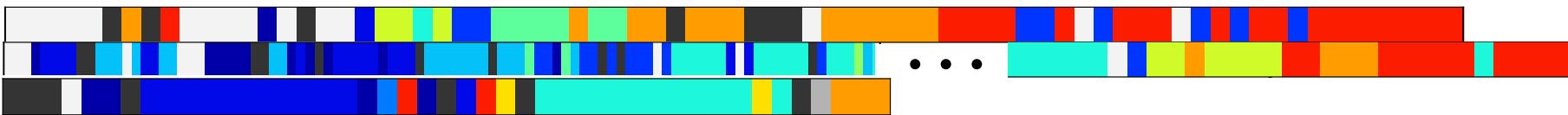
Understanding Process



Students scoring in 99th percentile on midterm exam



Students scoring in $\leq 3^{\text{rd}}$ percentile on midterm exam



Understanding Process

The screenshot shows the CS106A Pensieve interface. On the left, a list of code commits is shown, each with a timestamp and a colored square icon. A vertical bar indicates the current commit being viewed. The central area displays the `Pyramid.java` file content. The right side features a preview of a pyramid graphic made of small squares and a line chart titled "SourceLength" showing the evolution of the code's character count over time.

Commit	Time	Icon
14	0h 13m	Yellow
15	0h 16m	Purple
16	0h 16m	Cyan
17	0h 16m (0h 11m)	Blue
18	0h 16m (66h 38m)	Magenta
19	0h 16m (18h 7m)	Yellow
20	0h 16m (8h 9m)	Magenta
21	0h 18m	Magenta
22	0h 20m	Magenta
23	0h 21m	Magenta
24	0h 30m	Magenta
25	0h 36m	Magenta
26	0h 36m	Magenta
27	0h 36m (1h 25m)	Red
	Milestone 13	
28	0h 36m (8h 5m)	Red
29	0h 38m	Red
30	0h 44m	Red
31	0h 44m (0h 45m)	Red

```
/*
 * File: Pyramid.java
 * Name: A.J. Aldana
 * Section Leader: Kaitlyn Lagattuta
 *
 * -----
 * This file is the starter file for the Pyramid problem.
 * It includes definitions of the constants that match the
 * sample run in the assignment, but you should make sure
 * that changing these values causes the generated display
 * to change accordingly.
 */

import acm.graphics.*;
import acm.program.*;
import java.awt.*;

public class Pyramid extends GraphicsProgram {
    /** Width of each brick in pixels */
    private static final int BRICK_WIDTH = 30;

    /** Height of each brick in pixels */
    private static final int BRICK_HEIGHT = 12;

    /** Number of bricks in the base of the pyramid */
    private static final int BRICKS_IN_BASE = 18;

    public void run() {
        int x = (getWidth() / 2) - (BRICKS_IN_BASE / 2) * BRICK_WIDTH;
        int y = getHeight() - BRICK_HEIGHT;
        for(int row = BRICKS_IN_BASE; row > 0; row--) {
            layBricks(row, x, y);
            y -= BRICK_HEIGHT;
            x += BRICK_WIDTH / 2;
        }
    }

    private void layBricks(int row, int x, int y) {
        for(int col = 0; col < row; col++) {
            add(new Rectangle(x + col * BRICK_WIDTH, y, BRICK_WIDTH, BRICK_HEIGHT));
        }
    }
}
```

SourceLength

Characters

Time into Problem (hours)

Selection Comments Code

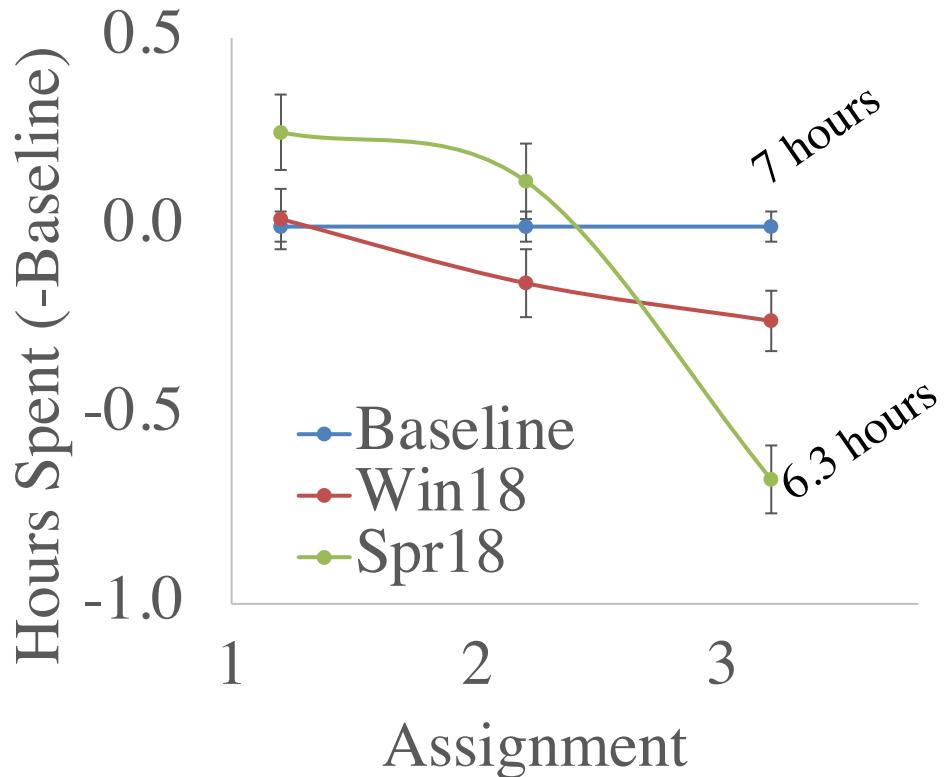
One quarter, all students were *shown* their progress:

- Early correction of bad habits.
- Chance to teach the art of programming.
- Academic dishonesty becomes much harder.



Understanding Process

Using assignment *timing* as pre-post

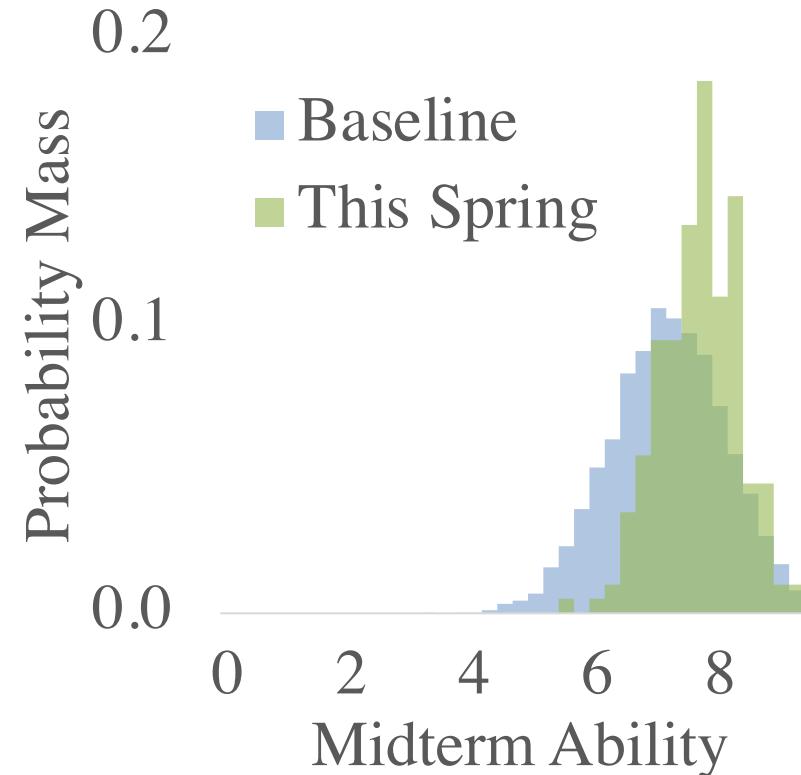


$$E[\hat{X}_3|X_1] - E[X_3] = 42 \text{ mins}$$

Predicted time Actual time $p < 0.00001$

Assignments are taking less time

Item Response Theory based ability assessment



$$S_{i,j} = n \cdot \sigma(a_i - d_j)$$

Score points ability difficulty

Students perform better than expected



Feedback on *process* is an open problem

Our understanding of student work is not perfect

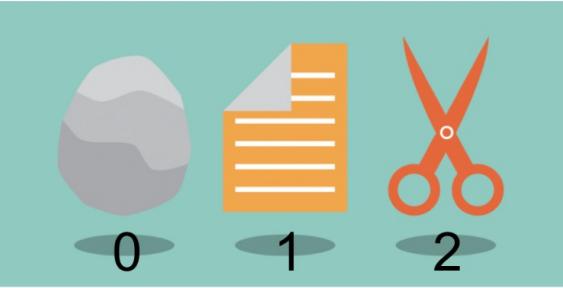
Hit a ceiling with cs106a midterm

Midterm Grading Challenge

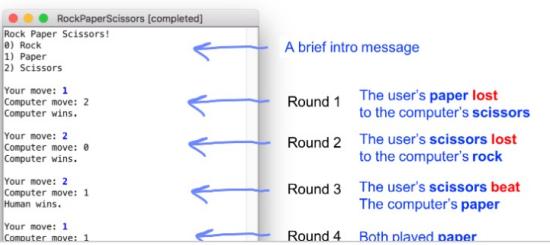
Browse Q3 DynamoDB · AWS Console localhost:3000/#/view/cs106aMidWin19/3/abigail8/cpiech/8daa714e79a10cb

Question Solution Starter Student Answer

Rock Paper Scissors (30 points)



Write a `ConsoleProgram` that has a user play rock paper scissors against a computer until either the user or the computer has **three** "wins". To make the code simpler, use integers to represent the different plays (0 is rock, 1 is paper, 2 is scissors). Example run:



```
public class RockPaperScissors extends ConsoleProgram {
    /* constants */
    private static final int ROCK = 0;
    private static final int PAPER = 1;
    private static final int SCISSORS = 2;
    private static final int N_WINS = 3;

    private RandomGenerator rg = new RandomGenerator();

    public void run() {
        introMessage();
        for (int i = 0; i < N_WINS; i++) {
            inputNumber();
            roundWinner();
        }
        gameWinner();
    }

    private void introMessage() {
        println("Rock Paper Scissors!");
        println("0) Rock");
        println("1) Paper");
        println("2) Scissors");
        println(" ");
    }

    private void inputNumber() {
        int a = readInt("Your move: ");
        int computerMove = rg.nextInt(3);
        int b = readInt("Computer move: " + computerMove);
    }

    private int computer = 0;
    private int human = 0;

    private void roundWinner() {
        if (a == 1 && b == 0) {
            println("Human wins ");
        }
    }
}
```

Browse Answers

Grade: 26/30 points ✓ grade submitted!

Rubric: Standard

- For loop instead of a while loop (should go up to 3 wins by computer or user not 3 games)

User Inputting Move (3 points)

Perfect (0 points)
 Minor Error (1 points)
 Major Error (2 points)
 Totally Wrong (3 points)

Computer Generating Move (4 points)

Perfect (0 points)
 Minor Error (1 points)
 Major Error (2 points)
 Major Errors/No Attempt (4 points)

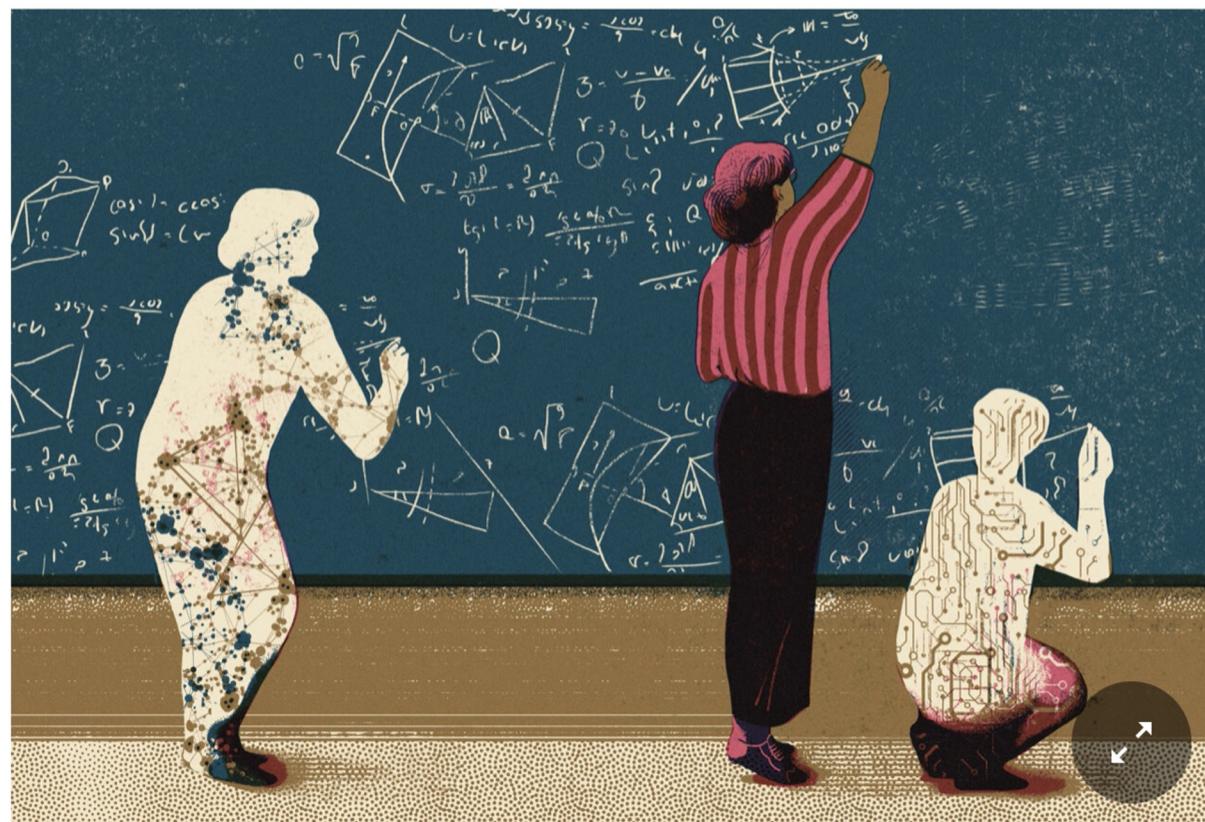
Determining the winner of a round (6 points)

Perfect (0 points)
 Minor Error (1 points)
 Major Error (2 points)

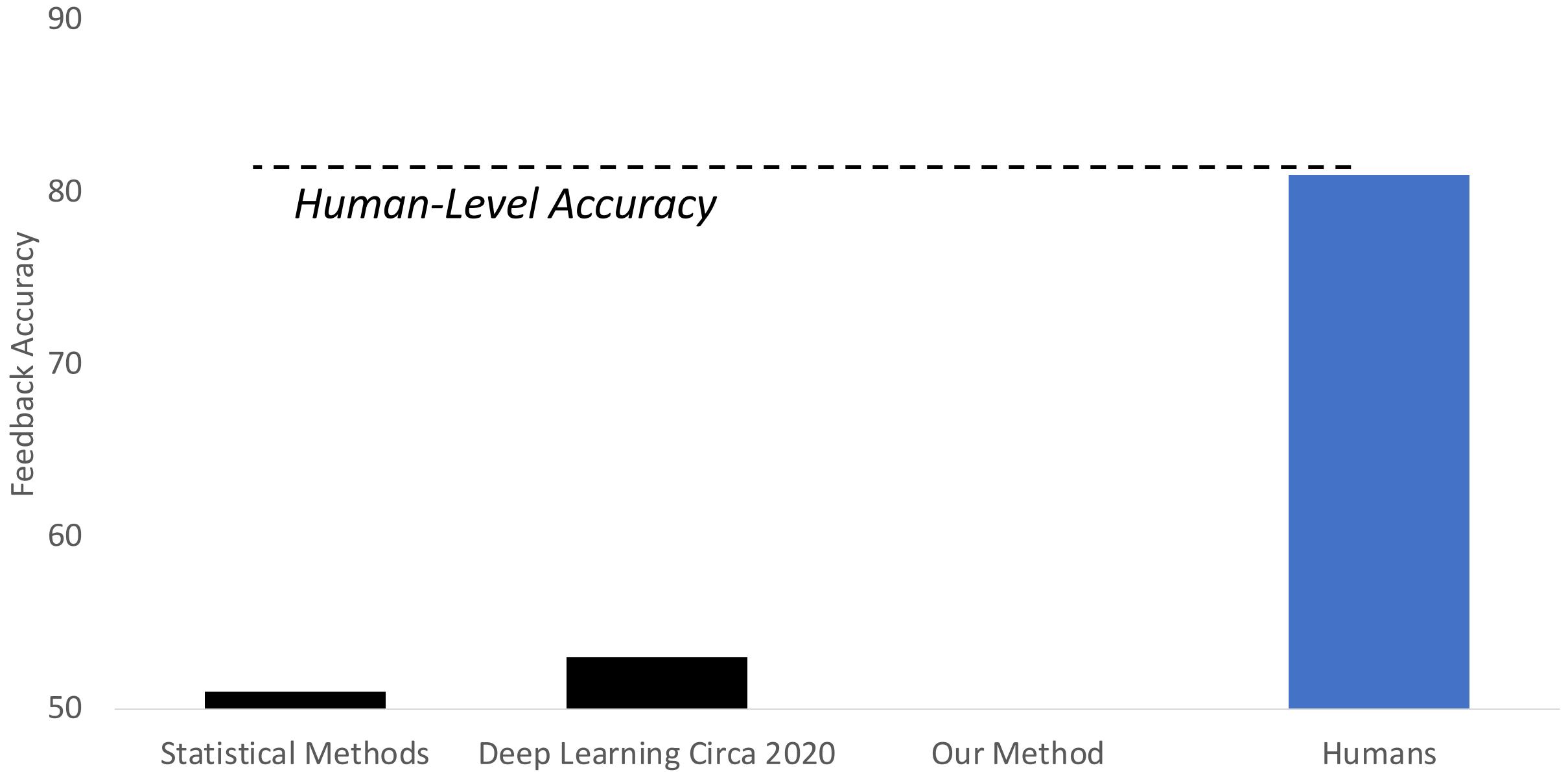
Exam / Question 3 / Student < Previous Submit Grade Next >

Can A.I. Grade Your Next Test?

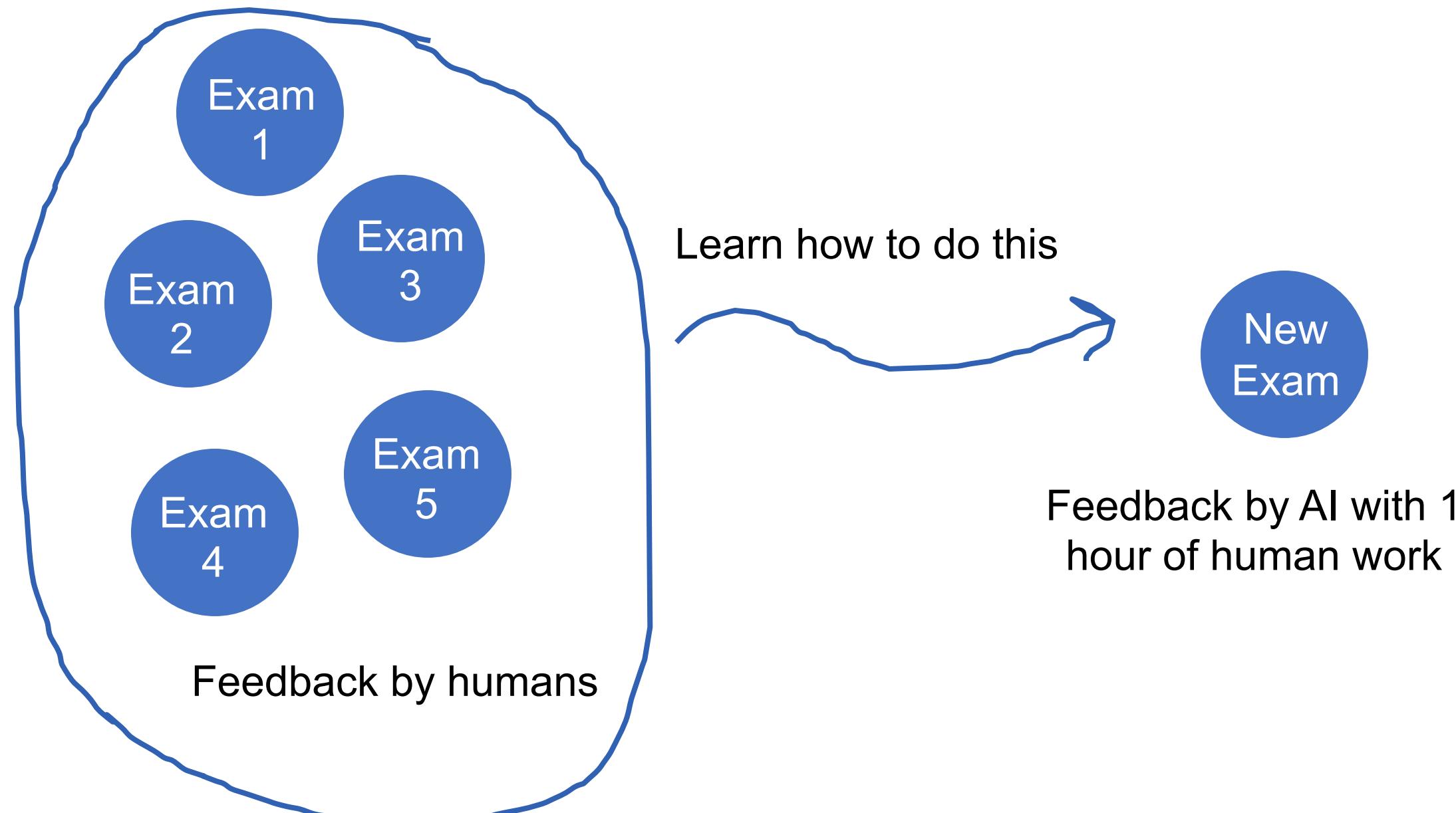
Neural networks could give online education a boost by providing automated feedback to students.



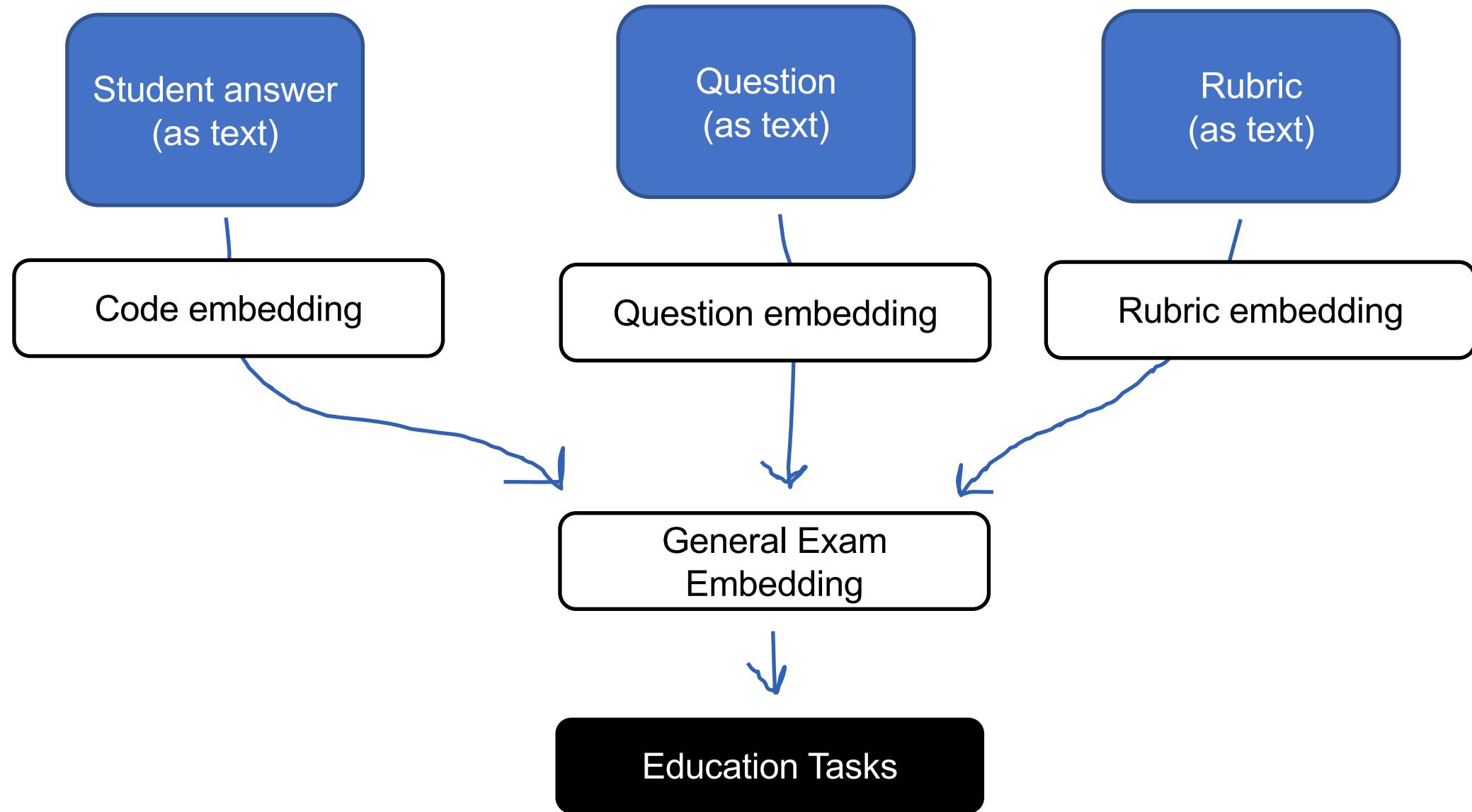
Rubric Level Accuracy on Few-Shot Grading a Novel Question



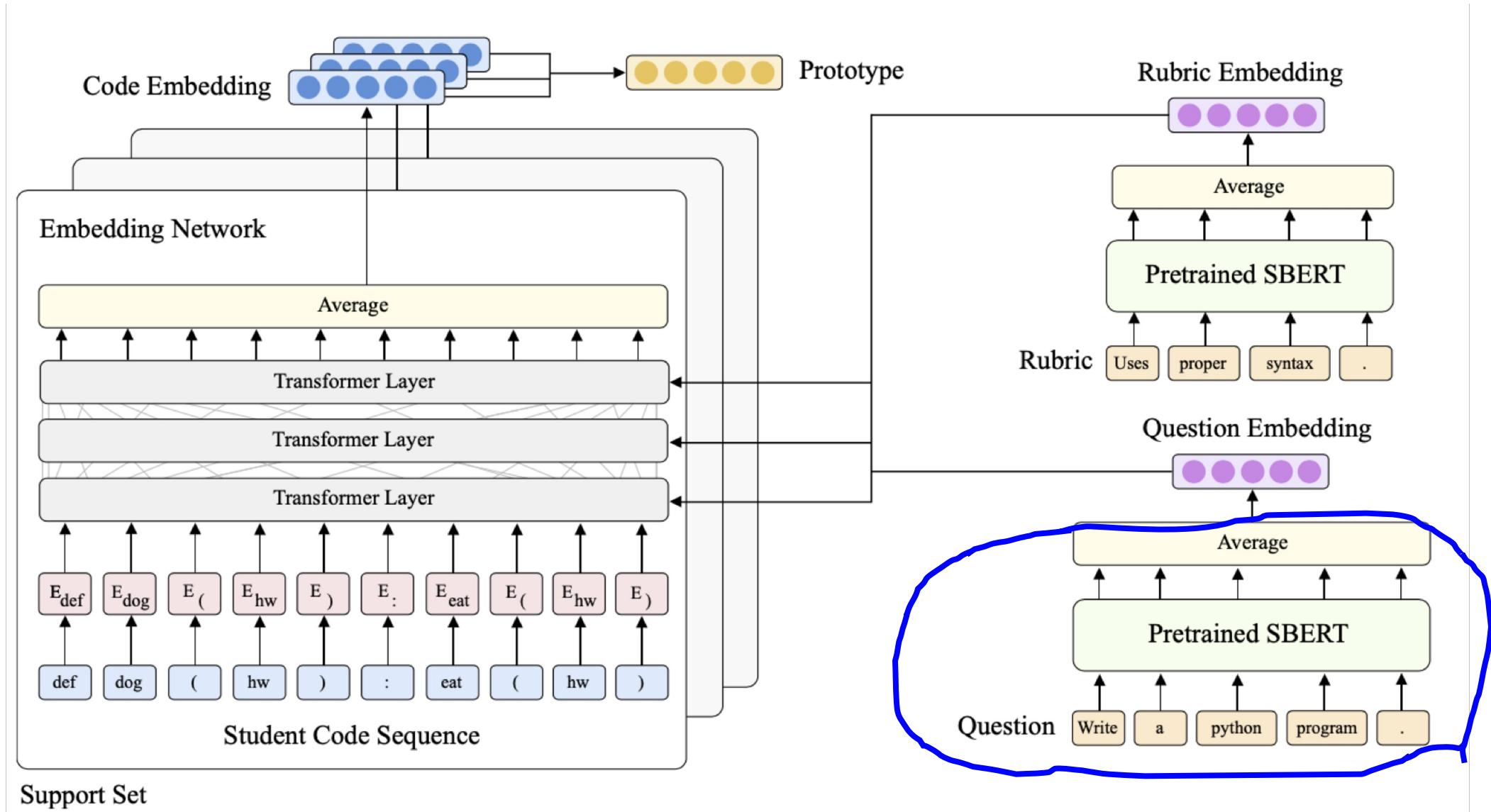
Give Feedback on Fresh Stanford Midterm



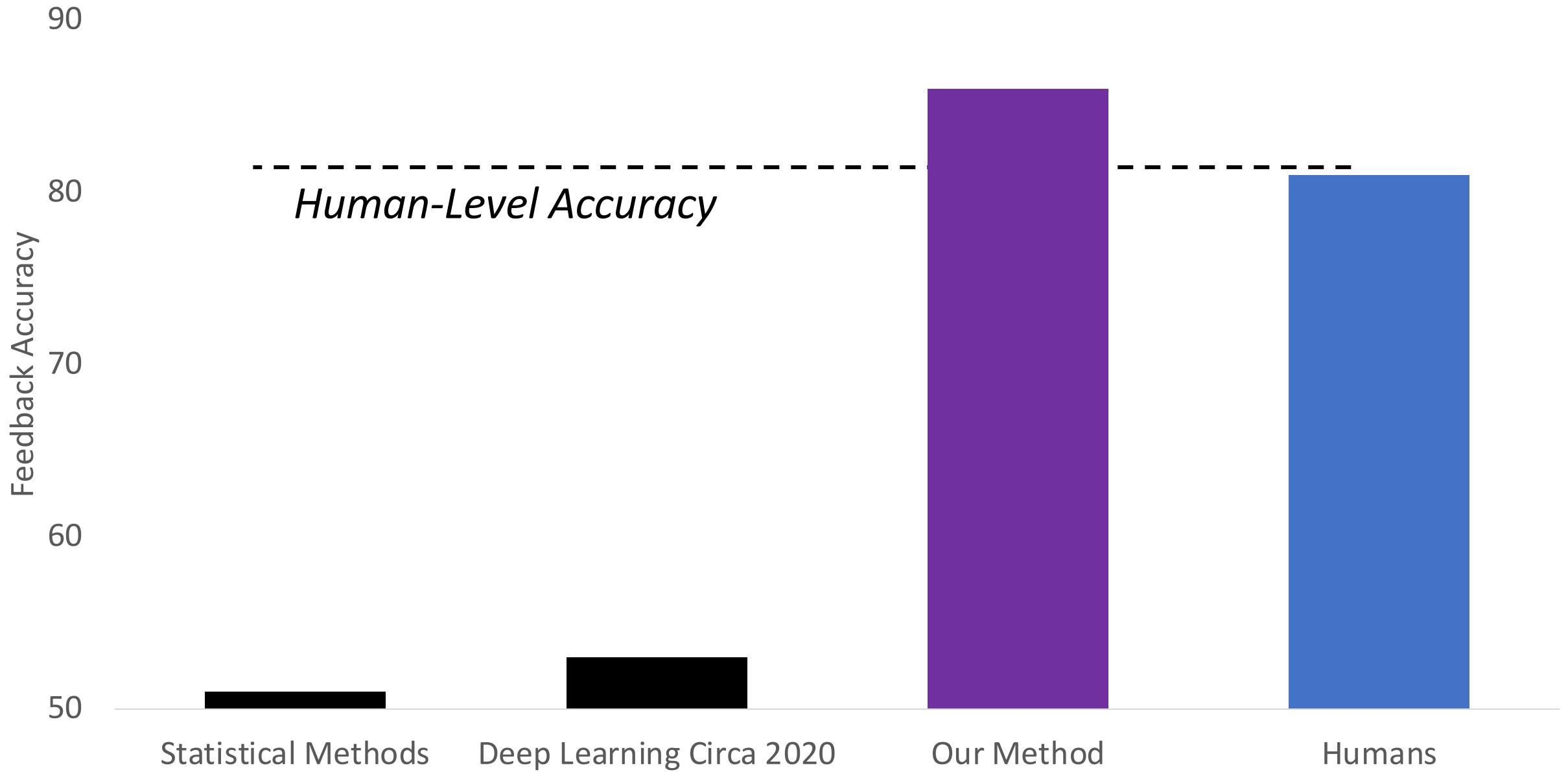
General Exam Grading Model



Invented the Proto-Transformer



Rubric Level Accuracy on Few-Shot Grading a Novel Question



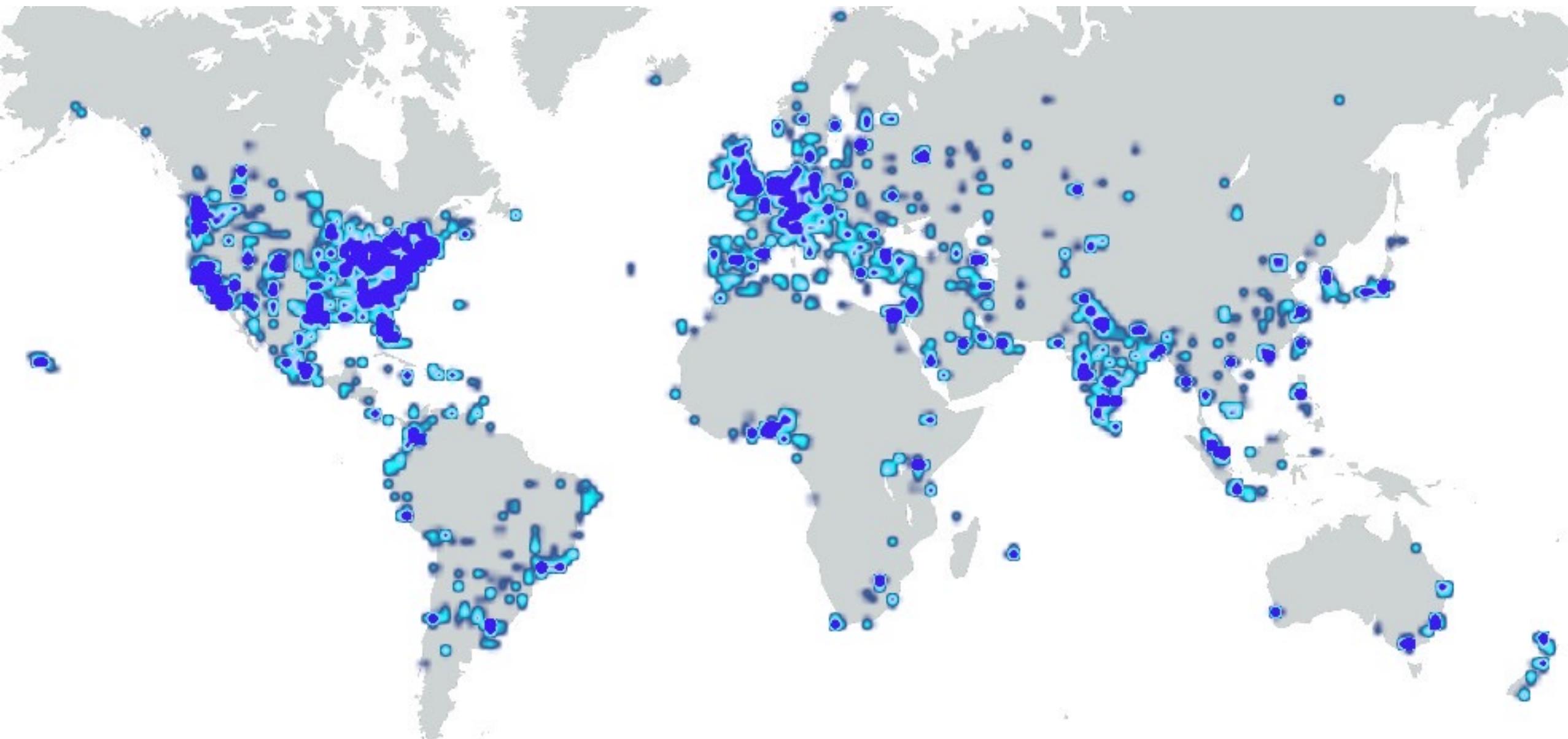
Stanford Code in Place:



1100+ section leaders teach
12,000+ students
1/2 of CS106A
As Community Service

Featured in





Gave Feedback to 3,500 Real Students

Do you agree? AI feedback 97.9%. Human feedback 96.7%

The screenshot shows a web browser window titled "Code in Place Feedback". The URL is "codeinplace.stanford.edu/diagnostic/feedback". The tab bar shows "Question 1" is active, followed by Question 2, Question 3, Question 4, Question 5, and Wrap-Up.

Feedback Section:

- GETTING INPUT FROM USER:** A question asking the student to get input from the user, convert it to a number, and save it as a variable.
- Close.** There is a minor error with your logic to get input from user. This could be something like forgetting to convert user input to a float
- Do you agree with the feedback in the purple box?**
- Buttons:** Thumbs up and thumbs down.
- Please explain (optional):** A text input field.

Your Solution:

```
def main():
    # TODO write your solution here
    height=input("Enter your height in meters: ")
    if height < 1.6:
        print("Below minimum astronaut height")
    if height > 1.9:
        print("Above maximum astronaut height")
    if height >= 1.6 and height <= 1.9:
        print("Correct height to be an astronaut")
```

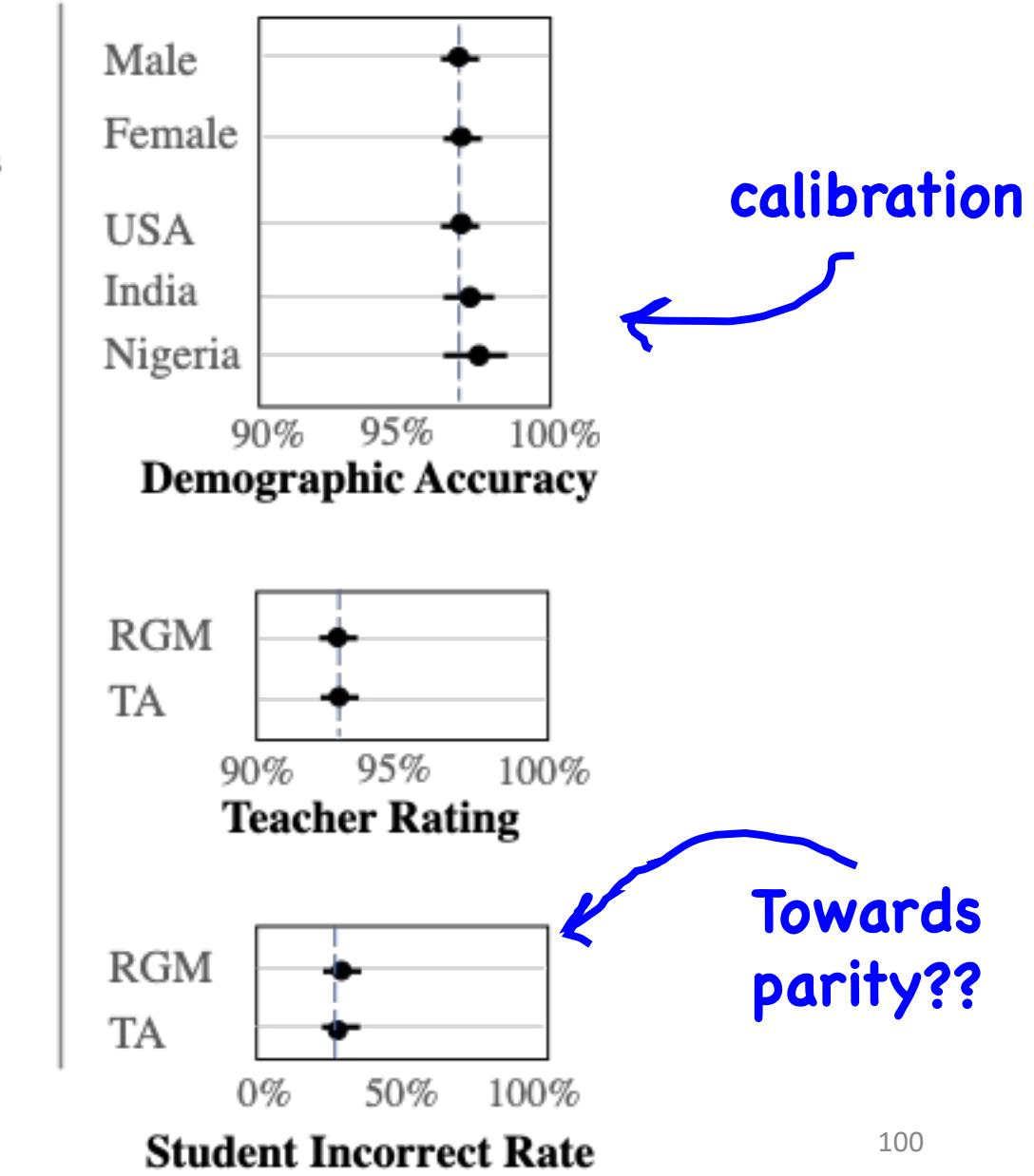
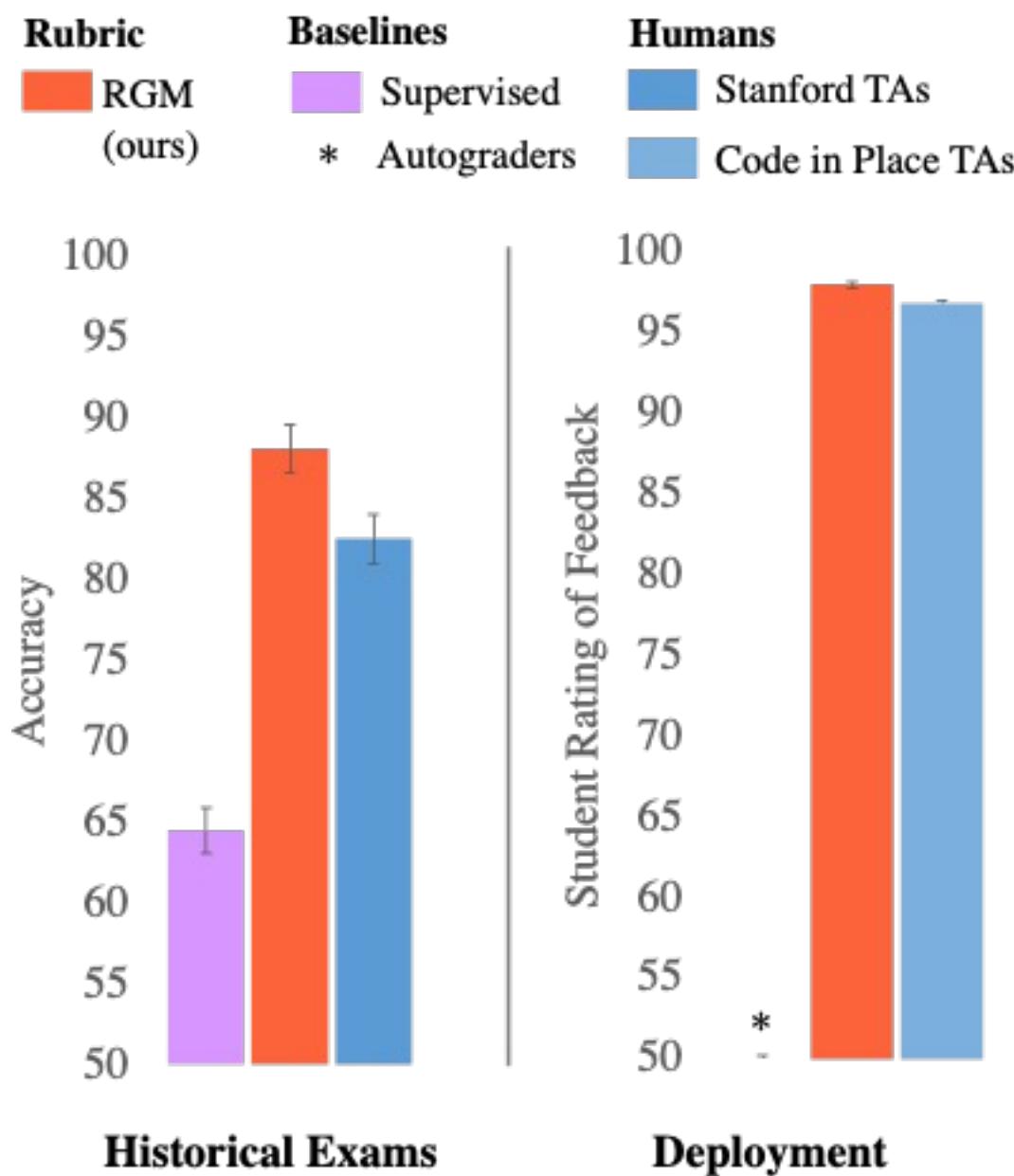
Annotations:

- A blue arrow points from the text "AI generated feedback" to the purple feedback box.
- A blue arrow points from the text "Students evaluate the feedback" to the thumbs up/down buttons.
- A blue arrow points from the text "Algorithm uses attention to highlight where in the code the error comes from" to the highlighted line "height=input("Enter your height in meters: ")".
- A blue arrow points from the text "Syntax error (missing ")" here would prevent auto graders from being useful." to the line "if height < 1.6:".

Algorithm uses
attention to highlight
where in the code the
error comes from

Syntax error (missing ")"
here would prevent auto
graders from being
useful.

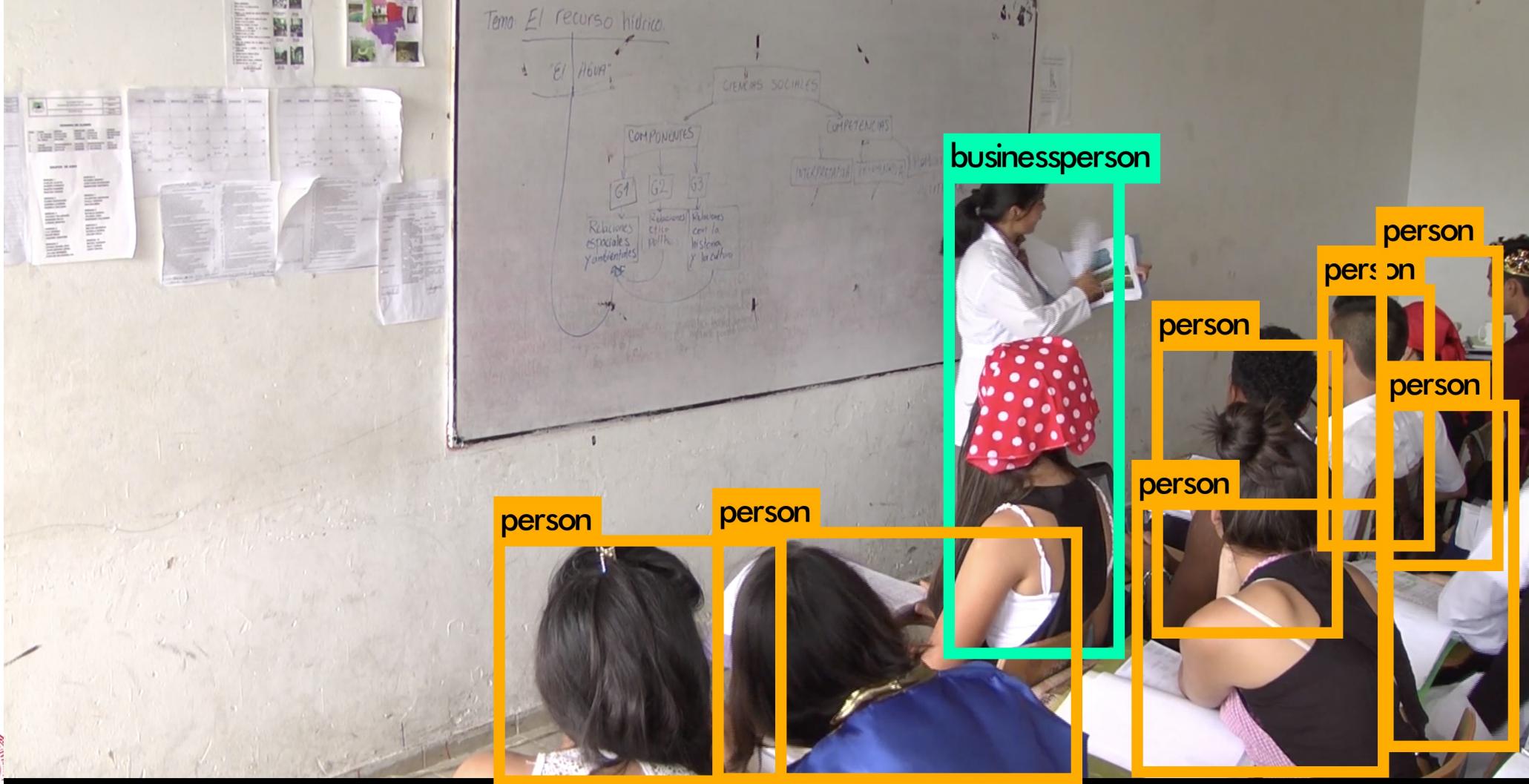
Impact and Fairness Analysis (work in progress)



That is just one example:
New problems

Feedback for Teachers

200,000 videos of teachers in Colombia, Chile and USA teaching



AI Teacher Training

1

Record Class



2

Transcribe & Anonymize
Recording



3

Analyze
Transcripts



4

Generate
Insights



Causal impact of AI teacher Training

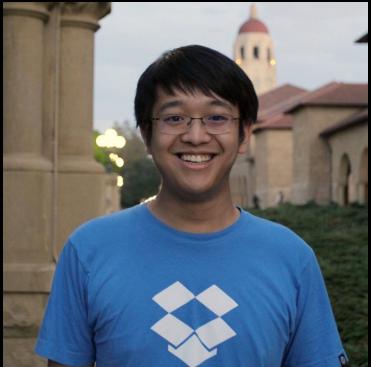
Within 4 lessons, as a result of training

1. Teachers asked 10% **more questions**
2. Teachers “**took up**” student ideas 10% more
3. Students were significantly more likely to:
recommend the class ($p < 0.05$),
find sections helpful ($p < 0.05$)



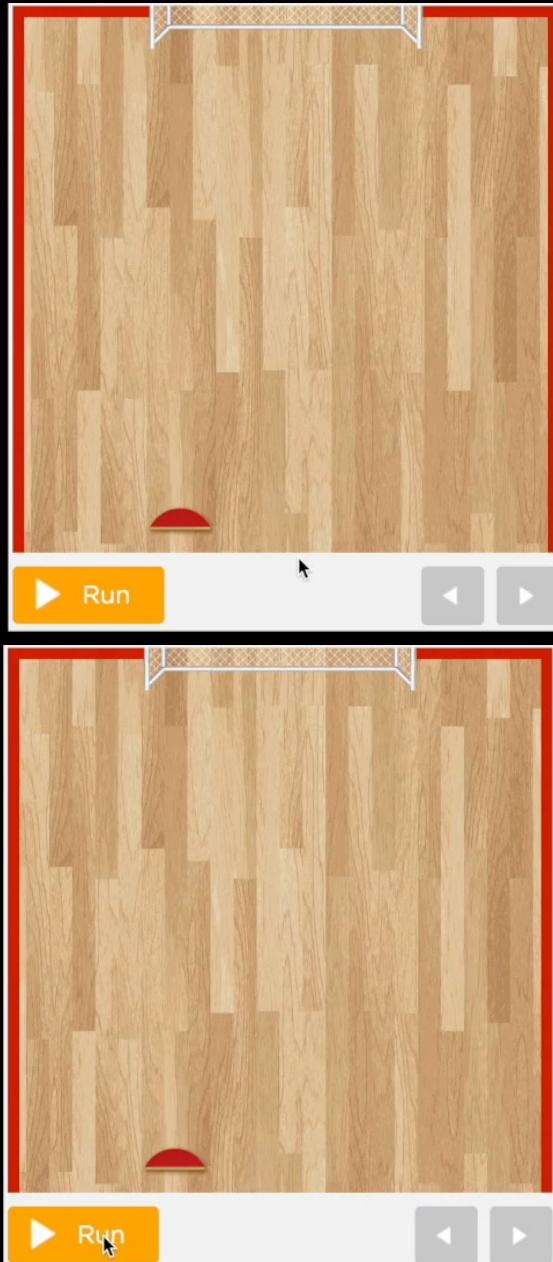
Play 2 Grade

Allen Nie



Impact:

Immediately change
what sort of assignments
are auto-gradable



Problem:

Grade the ~1M unique student implementations of this problem on code.org

Input:

Teacher gives you one example of each mistake on their rubric and one example of invariances

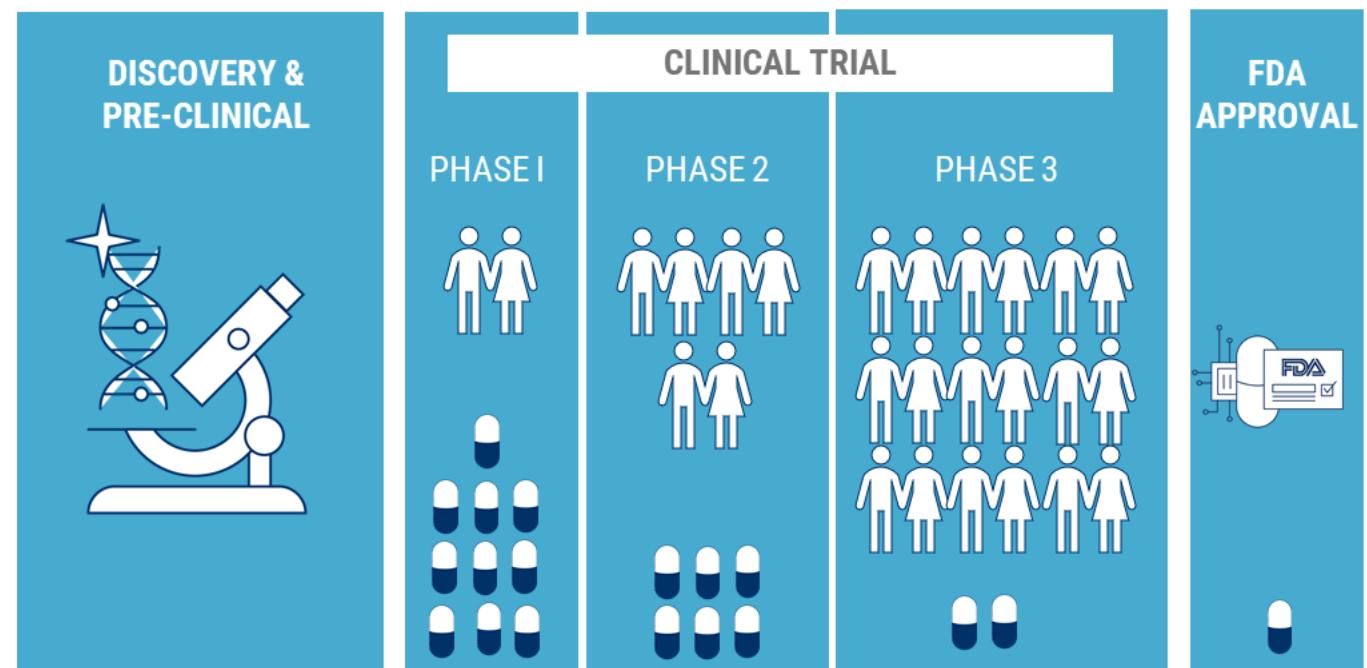
Theory contribution:

First model to build deep RL for a classification task . *Instead of learning an environment you are learning to test an environment*

More than education



 Bringing a drug to market is a drawn-out process



Source: cbinsights.com

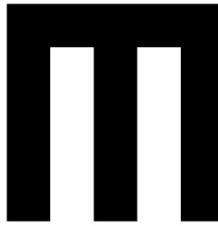
 CBINSIGHTS



More than education

Vision Test myeyes.ai/measure

Left Eye



Featured in
THE LANCET

Progress: 10%

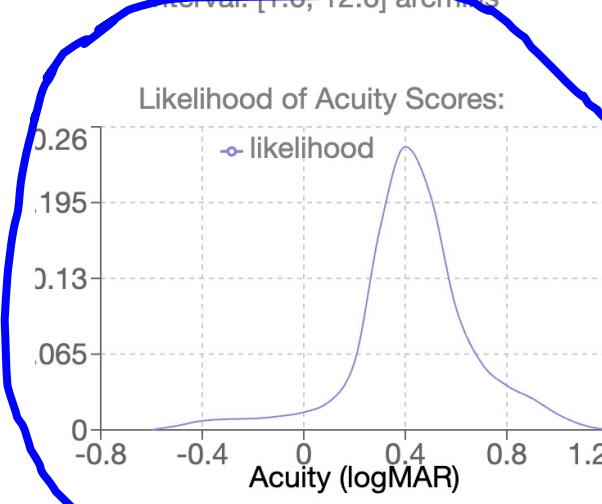
StAT Algorithm

N done: 2
Curr size: 3.3 arcmin
Curr size: 0.5 logMAR

MAP acuity: 2.5 arcmin
MAP acuity: 0.4 logMAR

Interval: [1.0, 12.0] arcmins

Likelihood of Acuity Scores:



Acuity (logMAR)



What else should be a **random variable**?

Grades??

Application -> Theory

Understand social science,
especially with small data

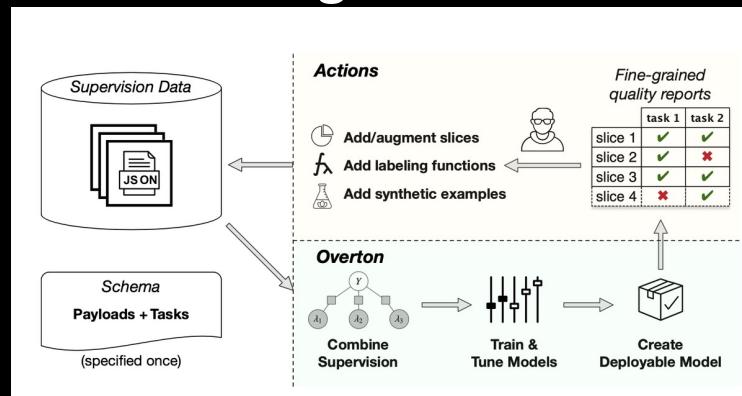
Understand language

What are things that AI
currently can't do?

Design itself

Explain why it made the
choices it did

Teach humans based on
what it has learned



What should you do
next?

Go solve amongst the abundance of
important problems



Final Project | AA228/CS238

Stanford University

AA228/CS238

Decision Making under Uncertainty

 MENU

Final Project

The objective of the final project is to explore topics in decision making under uncertainty in greater depth than is permitted in class. The choice of topic is up to you, but it should be related to the general themes of the course. As part of the project you should:

- * describe an approach (existing or newly developed),
- * apply the approach to a problem of interest (which may or may not be related to aerospace), and
- * analyze the performance of the approach according to a set of metrics.

CS221: Artificial Intelligence: Principles and Techniques

Stanford / Autumn 2022-2023

[Calendar] [Modules] [Coursework] [Schedule]



- Lectures: Mon/Wed 1:30-2:50pm in NVIDIA Auditorium.
- Problem sessions: Fridays 1:30-2:20pm in Huang 018.
- Office hours, homework parties: see the [Calendar](#).
- To contact the teaching staff, please use [Ed](#); for more personal/sensitive matters, email cs221-aut22-23-lead-staff@lists.stanford.edu.

Teaching Staff


Percy Liang
Instructor

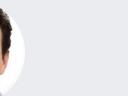

Dorsa Sadigh
Instructor

CS229: Machine Learning

Instructors


Andrew Ng


Moses Charikar

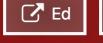

Carlos Guestrin

Course Description This course provides a broad introduction to machine learning and statistical pattern recognition. Topics include: supervised learning (generative/discriminative learning, parametric/non-parametric learning, neural networks, support vector machines); unsupervised learning (clustering, dimensionality reduction, kernel methods); learning theory



CS 228 - Probabilistic Graphical Models

Winter 2021-22

 Ed  Calendar  Course Notes

[Logistics](#) | [Course Info](#) | [Syllabus](#) | [Other Resources](#)

Logistics

- Lectures: Tue, Thu, 9:45am-11:15am, Nvidia Auditorium
- Office Hours and Sections: [Google Calendar](#)

Statistics 200: Introduction to Statistical Inference

Zhou Fan, Stanford University, Autumn 2016

[Home](#)

[Lectures](#)

[Homework](#)

[Grades](#)

[Piazza](#)

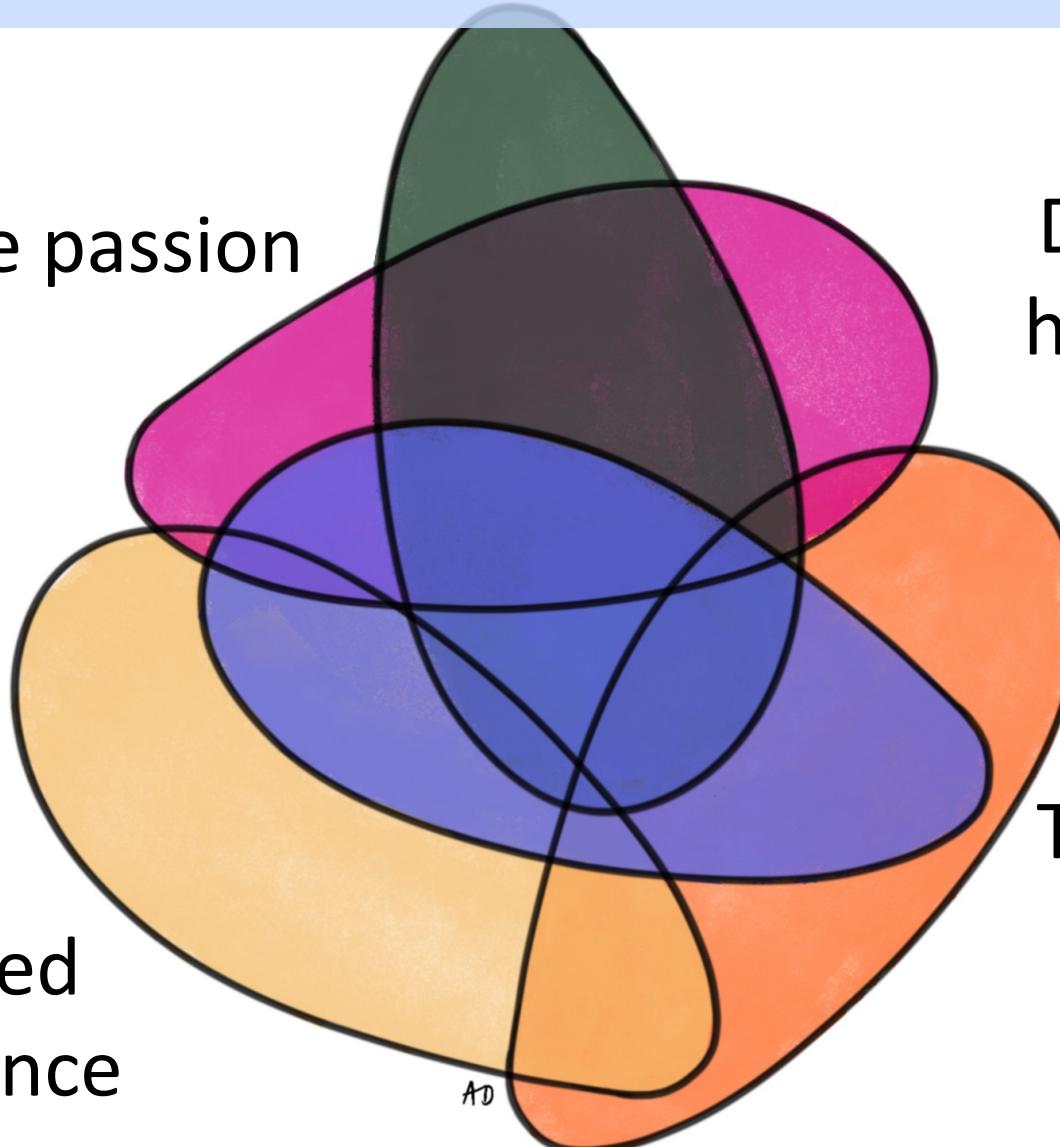
Think about intersectionality

Your side passion

Your lived
experience

Data that you
have access to

Thompson
sampling



Last Class...

CS109

AI

Uncertainty Theory

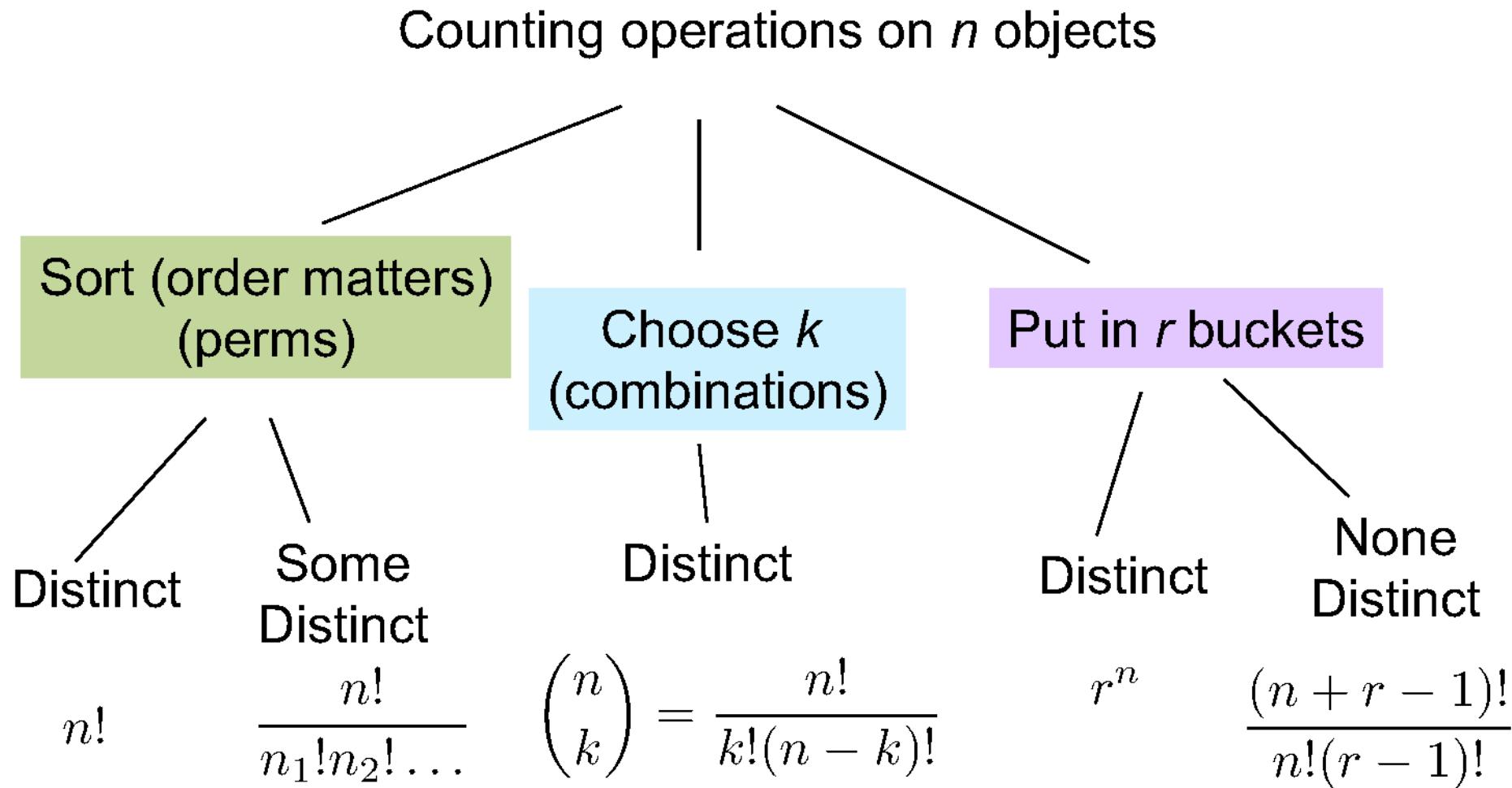
Single Random
Variables

Probabilistic
Models

Counting

Probability Fundamentals

Counting Rules



Counting



Ayesha



Tim



Irina



Joey



Waddie



COURSE VALUES

Everyone is welcome.
Intellectual joy. Be kind. Be
humane. Social connection.
Learn by doing. Thrill of
building. Adapt to new
contexts. **Especially in a**
hard time ☀️.



WHEN YOU MEET YOUR BEST FRIEND

Somewhere you didn't expect to.



Trailing the dovetail shuffle to it's lair – Persi Diaconosis

What is a Probability?

$$P(E) = \lim_{n \rightarrow \infty} \frac{n(E)}{n}$$

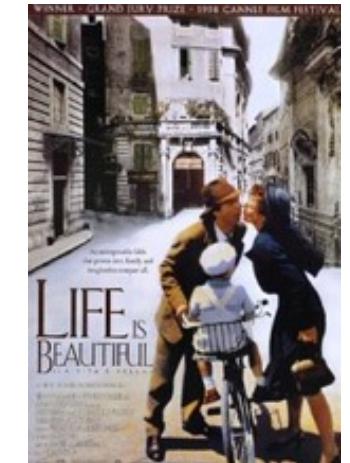
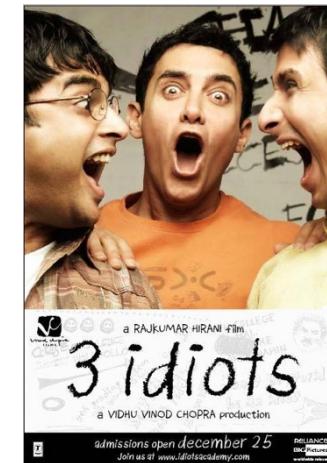
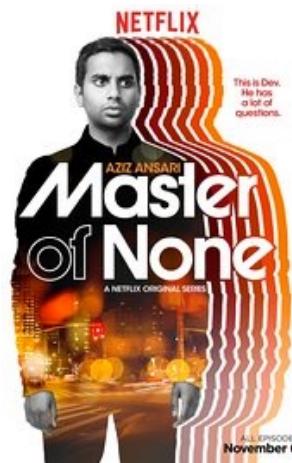


$$P(E|F) = \frac{P(EF)}{P(F)}$$

Definition of
Cond. Probability

Netflix and Learn

- Let E be the event that a user watches the given movie.
- Let F be the event that the same user watches CODA (2021).



$$P(E) = 0.19$$

$$P(E) = 0.32$$

$$P(E) = 0.20$$

$$P(E) = 0.09$$

$$P(E) = 0.20$$

$$P(E|F) = 0.14$$

$$P(E|F) = 0.35$$

$$P(E|F) = 0.20$$

$$P(E|F) = 0.72$$

$$P(E|F) = 0.42$$

Monty Hall Problem

and Wayne Brady



Marilyn discovers the
Probability Bug

Zika Test



Positive Zika.

What is the probability of zika?

-
- *0.1% of people have zika*
 - *90% positive rate for people with zika*
 - *7% positive rate for people without zika*

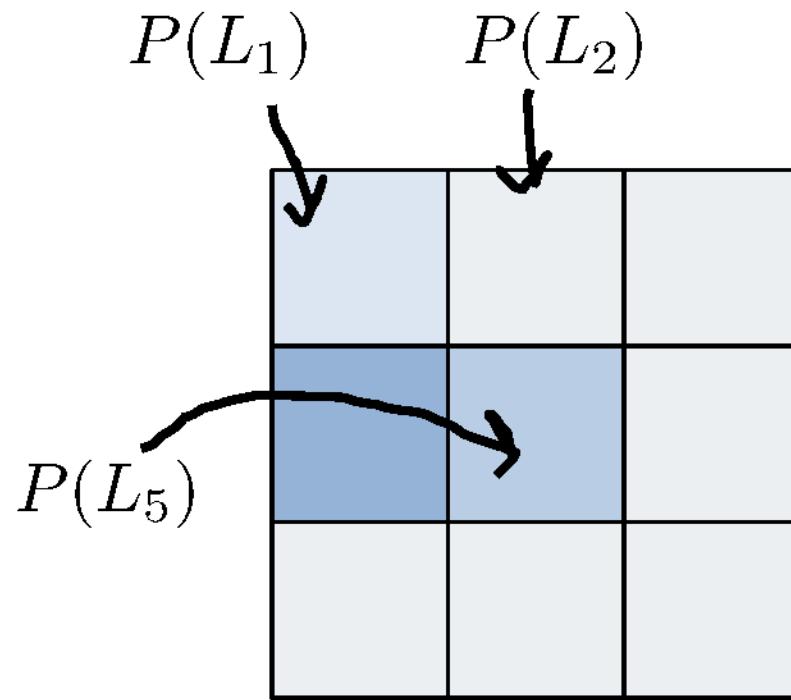
The right answer is 1%



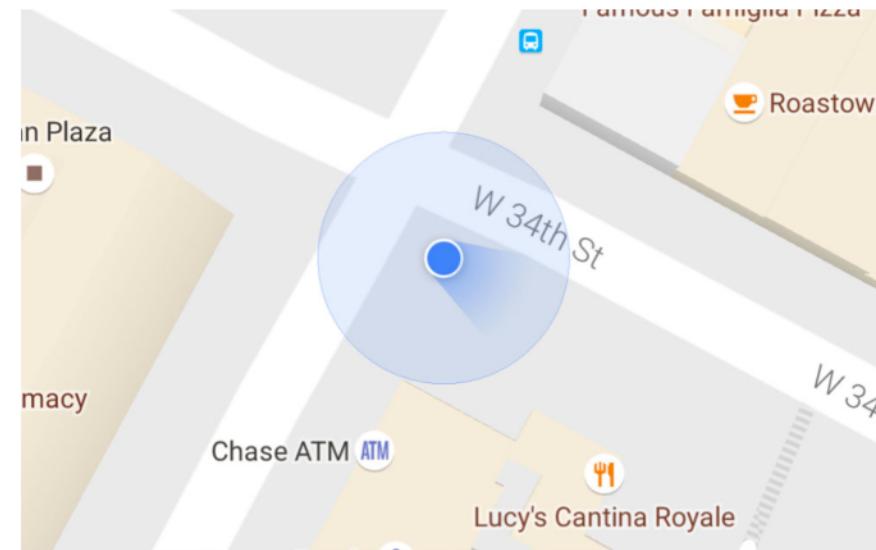
Bayes Theorem Intuition



Update Belief

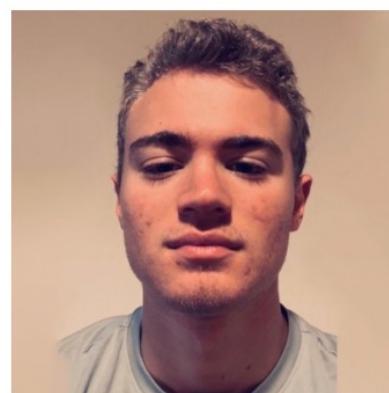
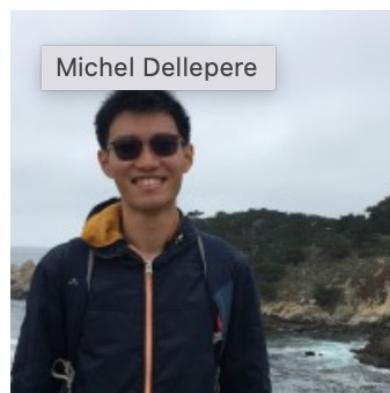
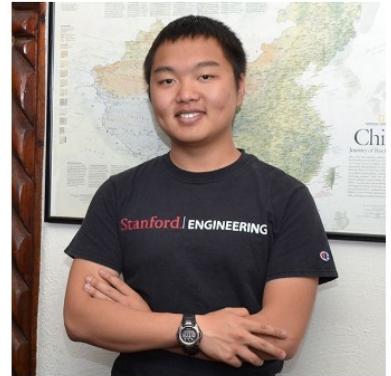


Before Observation



Recall our Ebola Bats



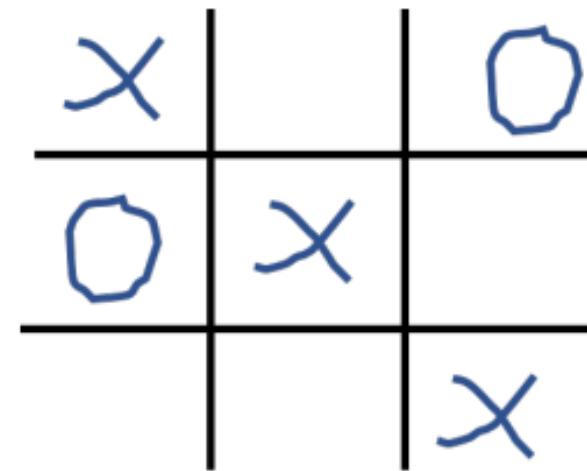


Third Year of Sections



I'm not a robot

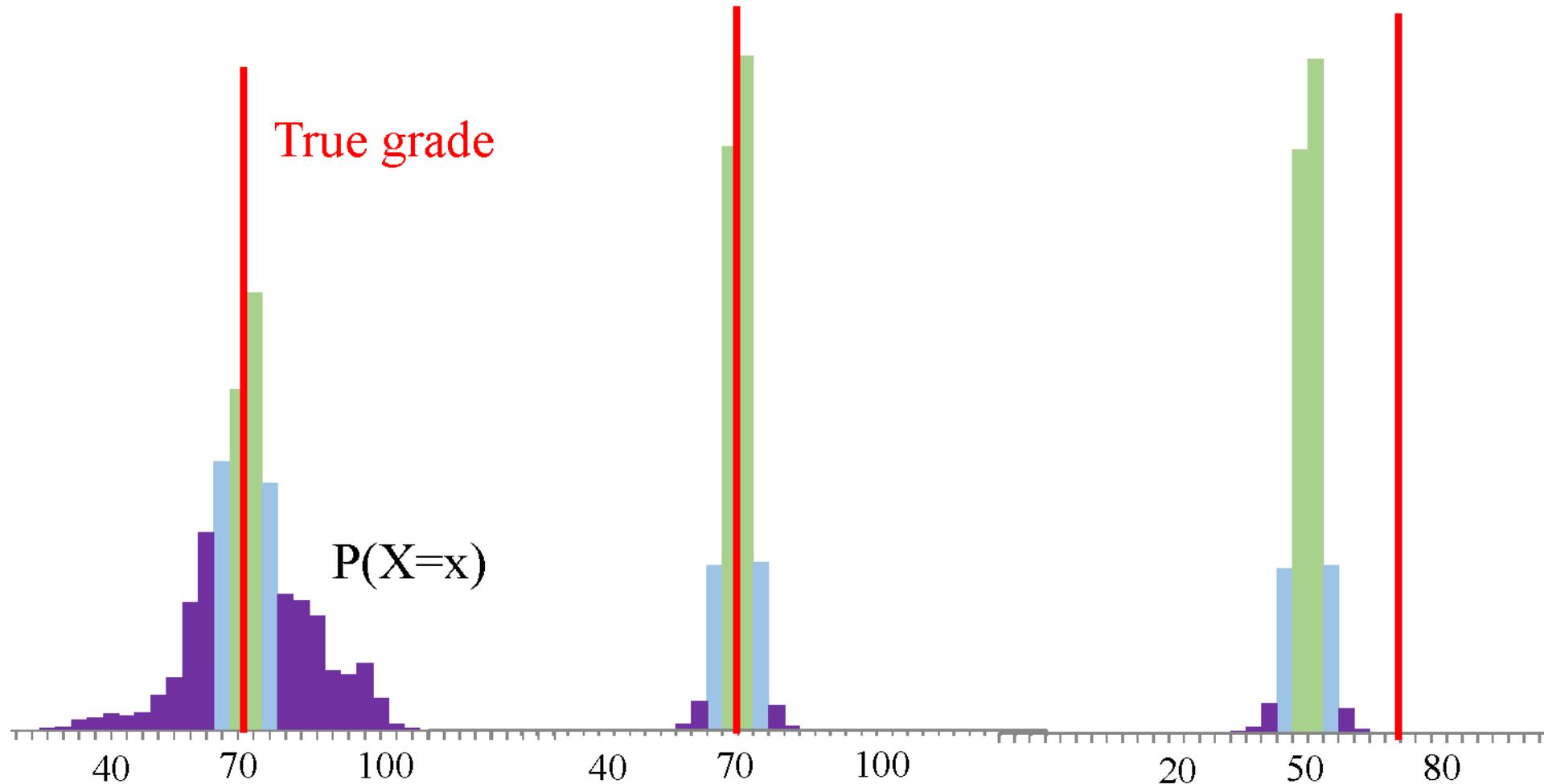
reCAPTCHA
Privacy - Terms

A standard reCAPTCHA interface. It features a checkbox labeled "I'm not a robot". To the right is the reCAPTCHA logo, which is a stylized "C" with a blue and grey swirl. Below the logo is the word "reCAPTCHA" and a link to "Privacy - Terms".



Random Variables

X is the score a peer grader gives to an assignment submission

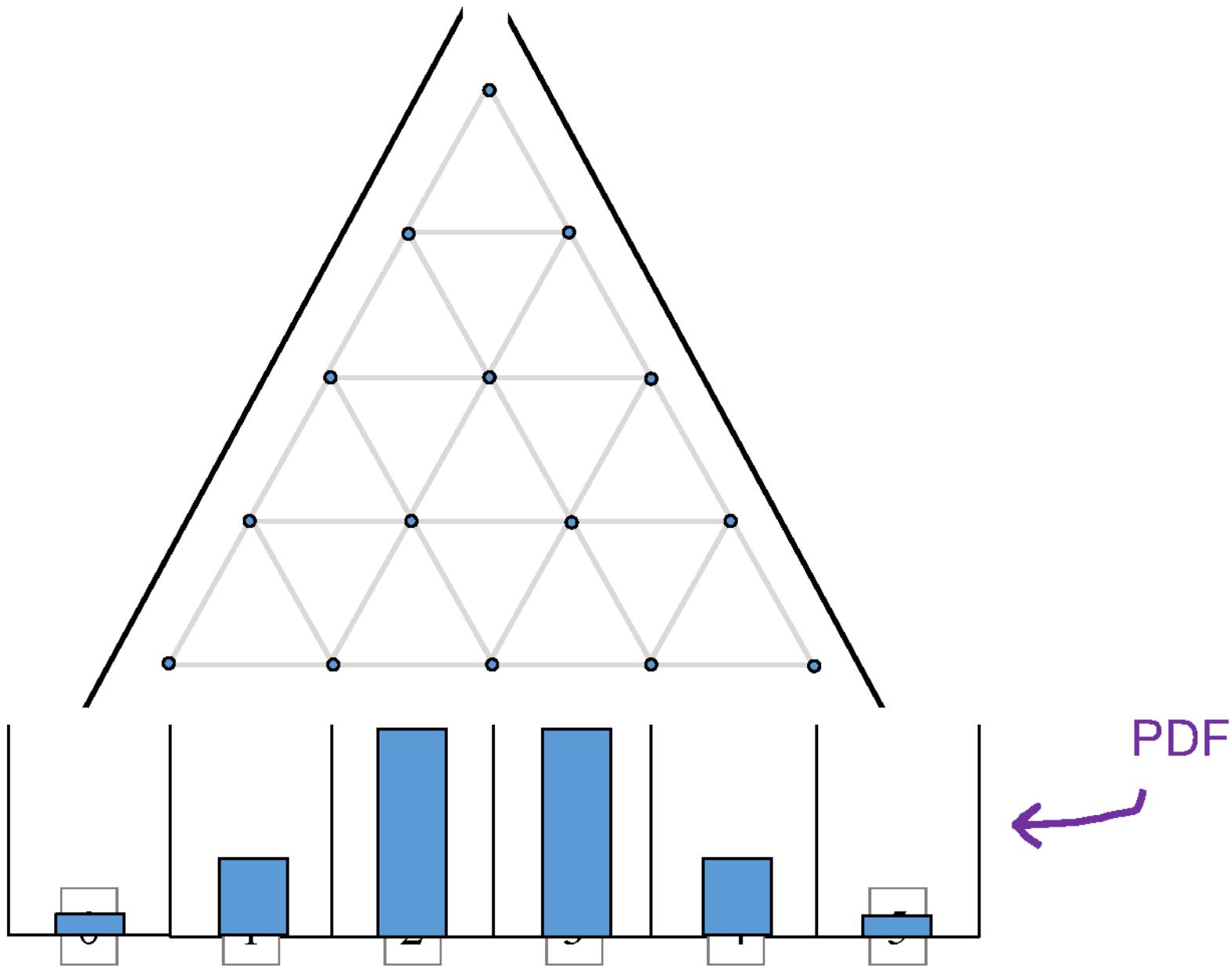


A

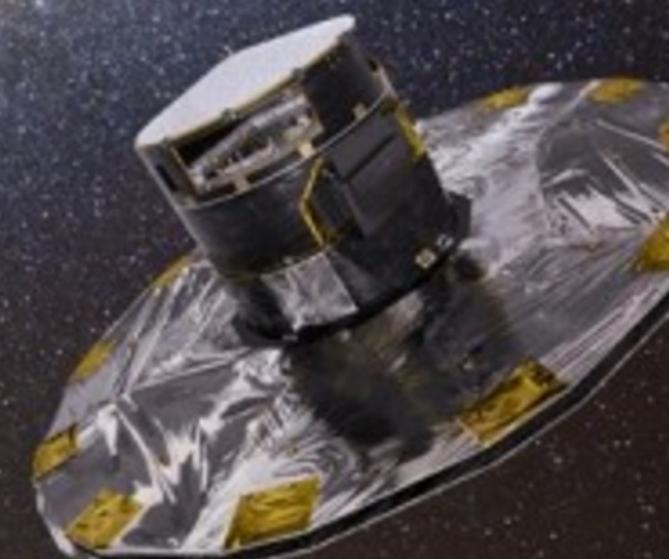
B

C

Binomial



1001

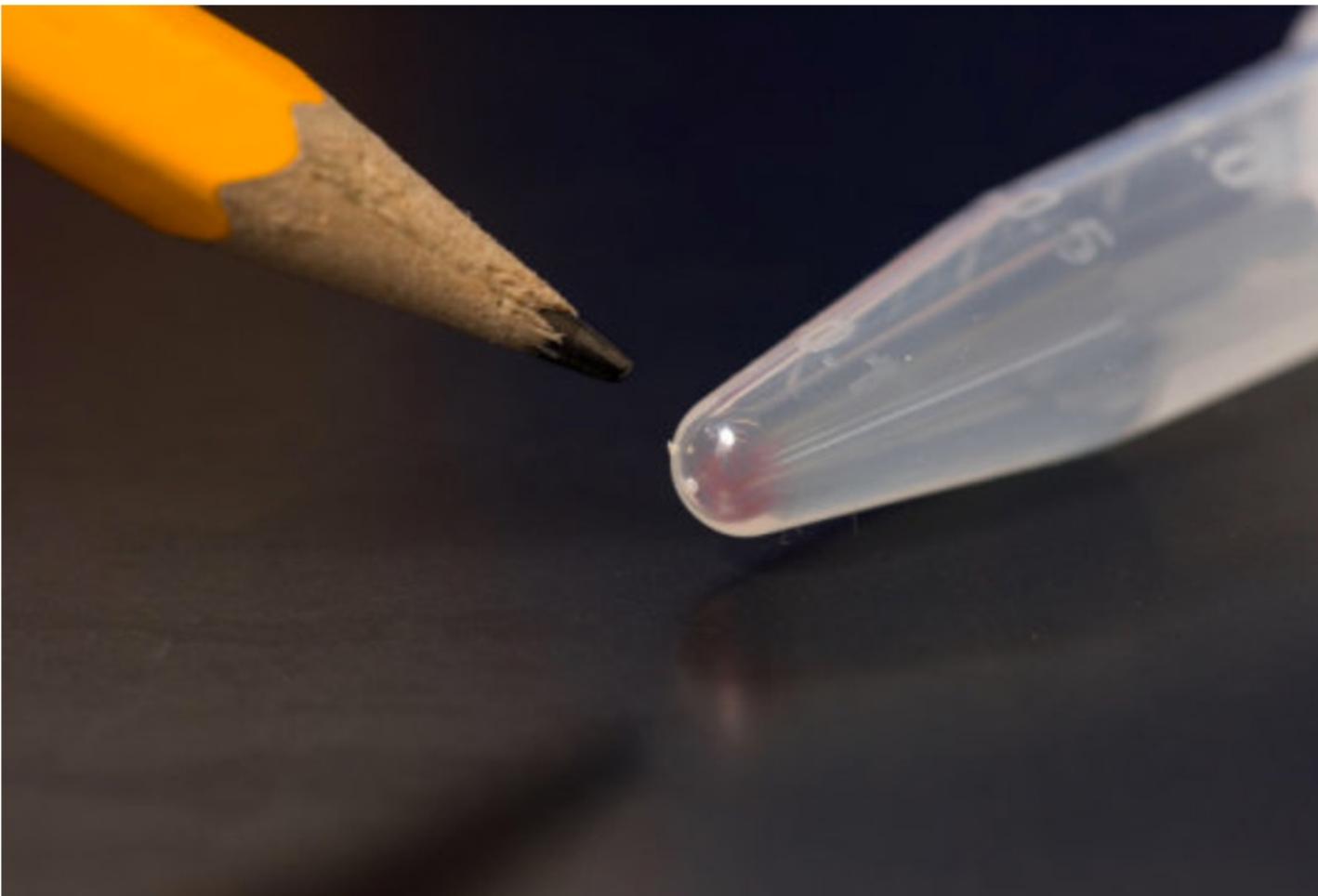


Geometric

Sequence 1:

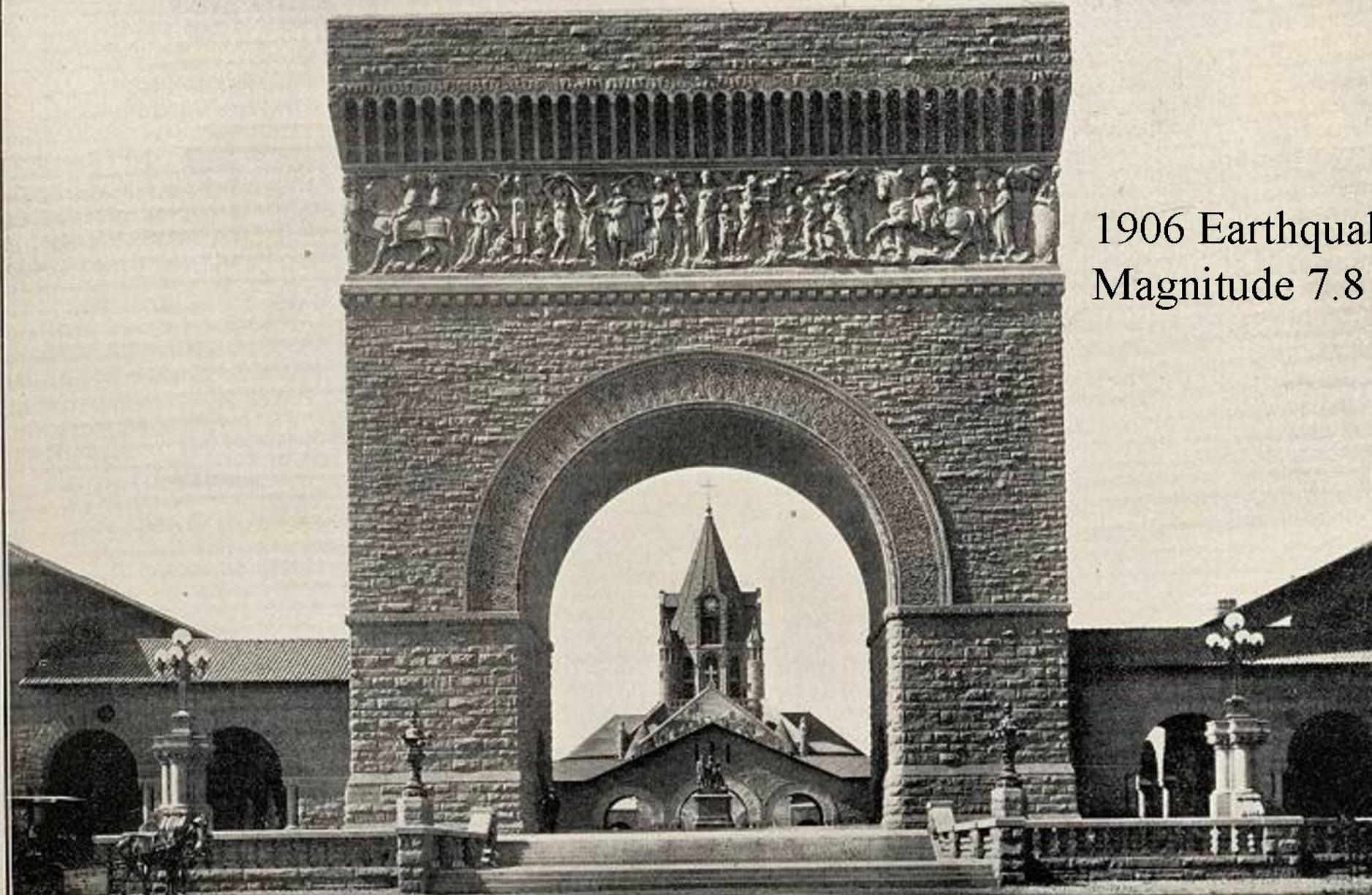
TTTHHTHTTHTTTHTTTHTTTHTHTHHTH
HTTHHTTHHHHTHHTHTTHTHTHTHHTHHTH
HTHHTHHHTHTHTTHTHTHHTHTHTHTHTH
TTTHHTHTTHTHTHTHTHTHTHHTHHTHHTH
TTTHHTHTHTHTHHTHHTHHTHHTHHTHHTH

Storing Data on DNA



All the movies, images, emails and other digital data from more than 600 smartphones (10,000 gigabytes) can be stored in the faint pink smear of DNA at the end of this test tube.

1906 Earthquake
Magnitude 7.8



ILL. No. 65. MEMORIAL ARCH, WITH CHURCH IN BACKGROUND, STANFORD UNIVERSITY, SHOWING TYPES OF CARVED WORK WITH THE SANDSTONE.



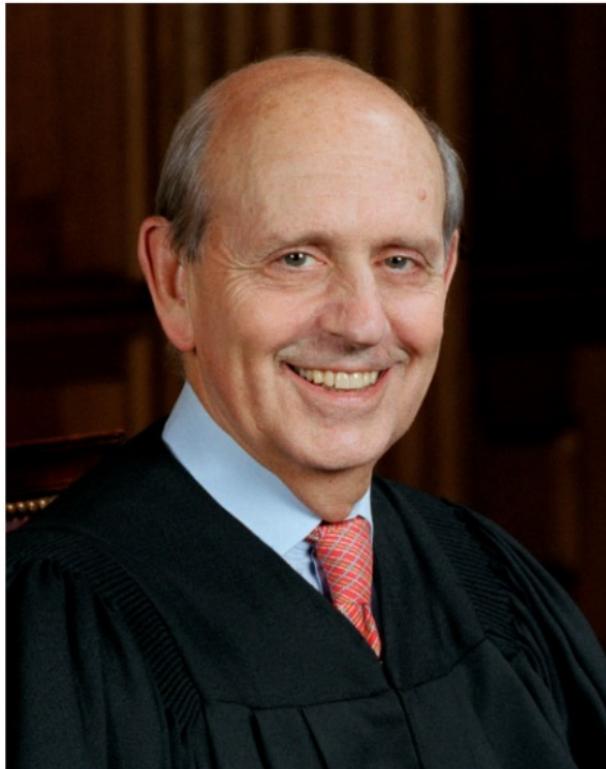
Probability for Extreme Weather?

Bit Coin Mining

You “mine a bitcoin” if, for given data D , you find a number N such that $\text{Hash}(D, N)$ produces a string that starts with g zeroes.



Representative Juries



Simulate

Simulation:

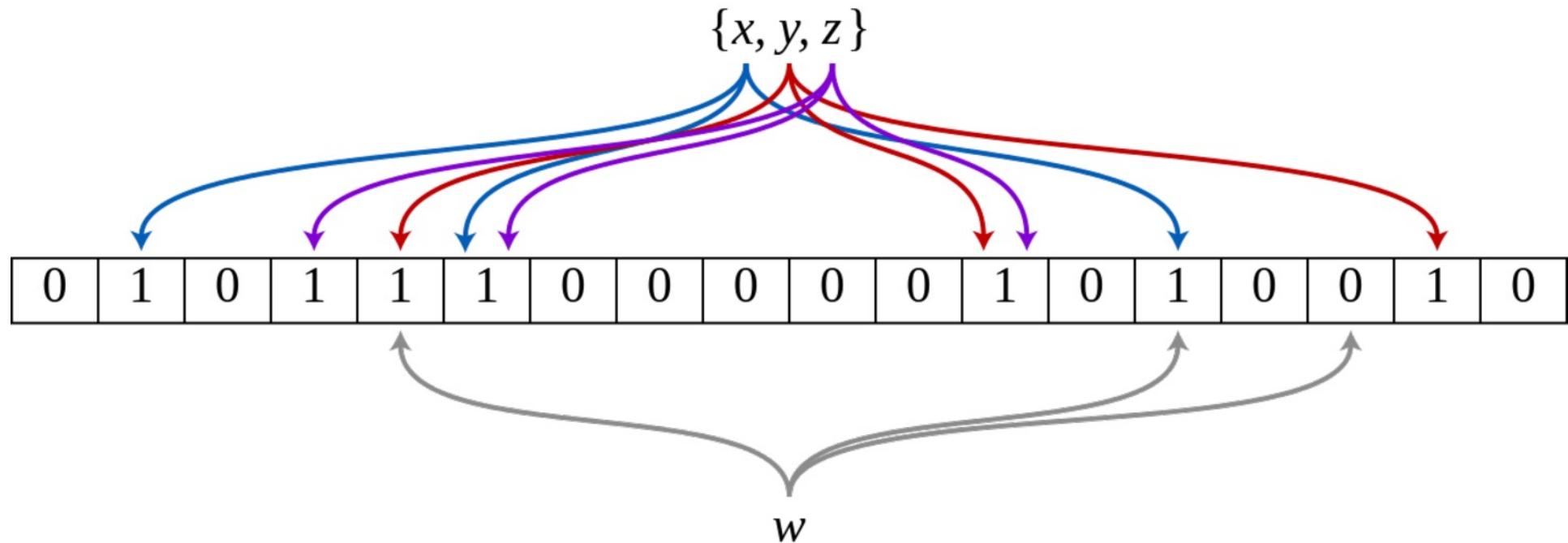


Dating at Stanford

Each person you date has a 0.2 probability of being someone you spend your life with. What is the average number of people one will date? What is the standard deviation?



Bloom Filter



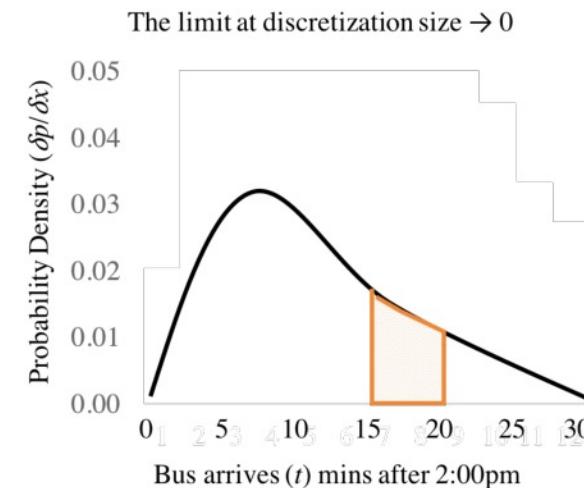
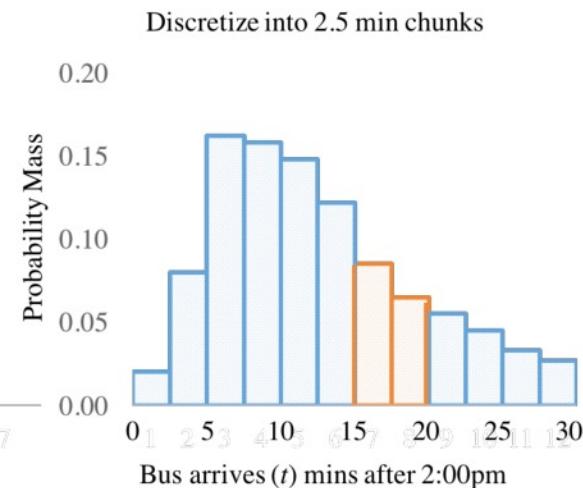
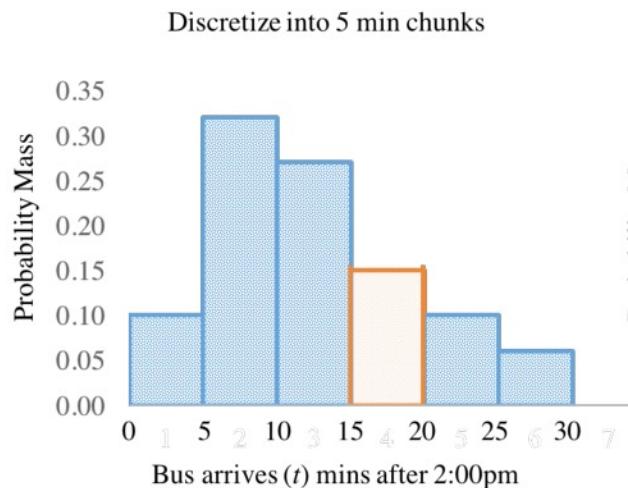
random() ?

Riding the Marguerite

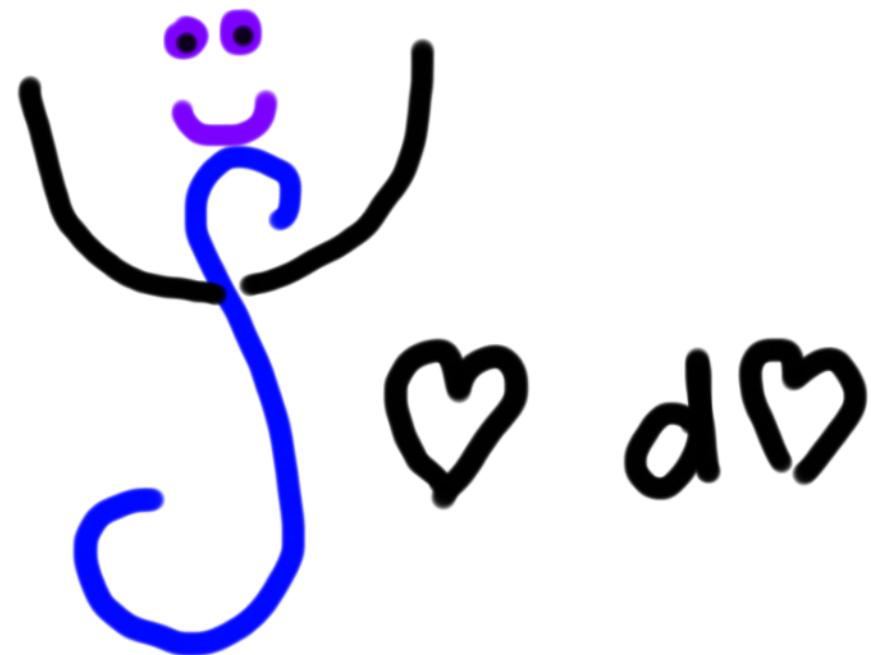


*You are running to the bus stop.
You don't know exactly when
the bus arrives. You arrive at
2:20pm.*

What is $P(\text{wait} < 5 \text{ min})$?



Integrals

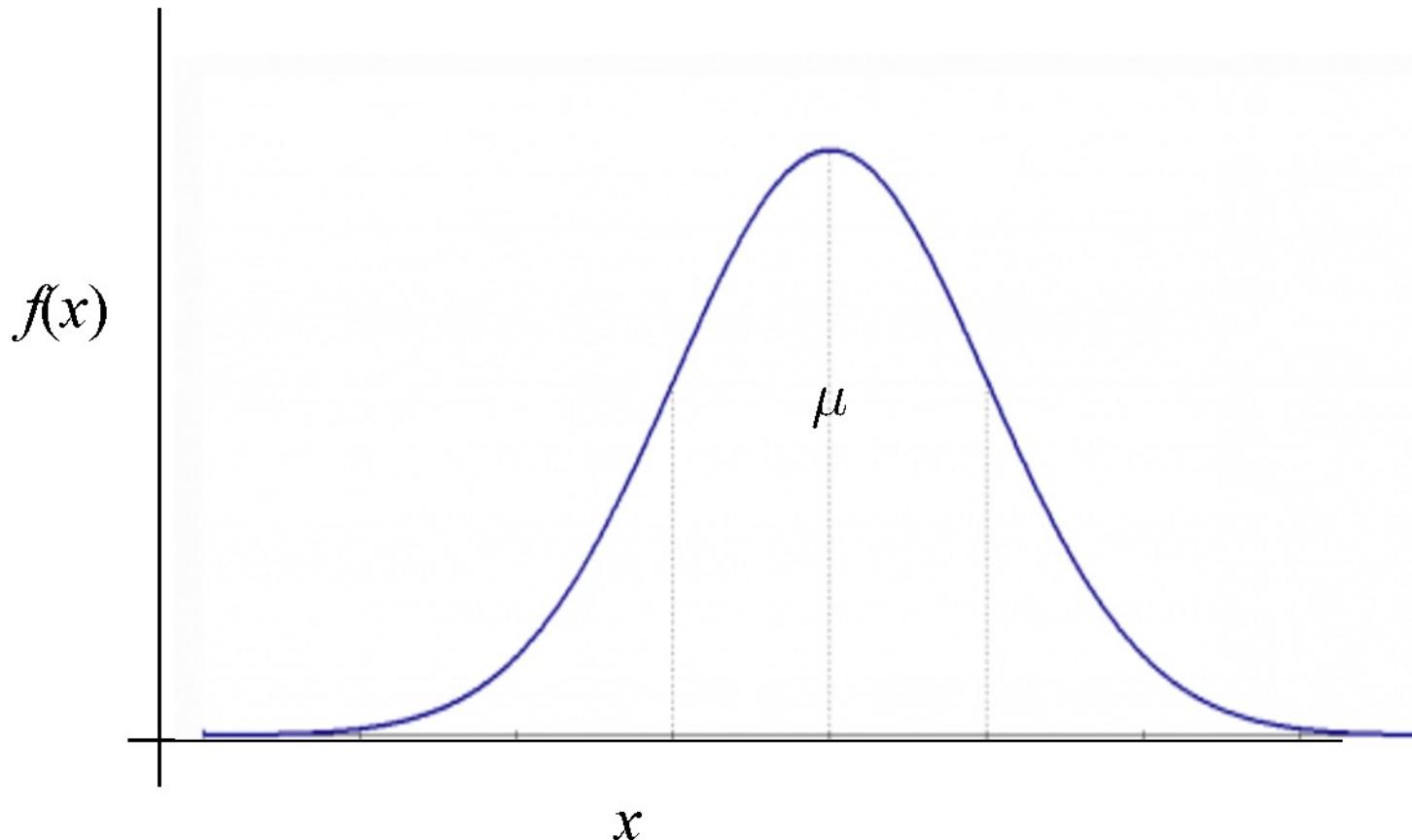


*loving, not scary

Probability Density Function

$$\mathcal{N}(\mu, \sigma^2)$$

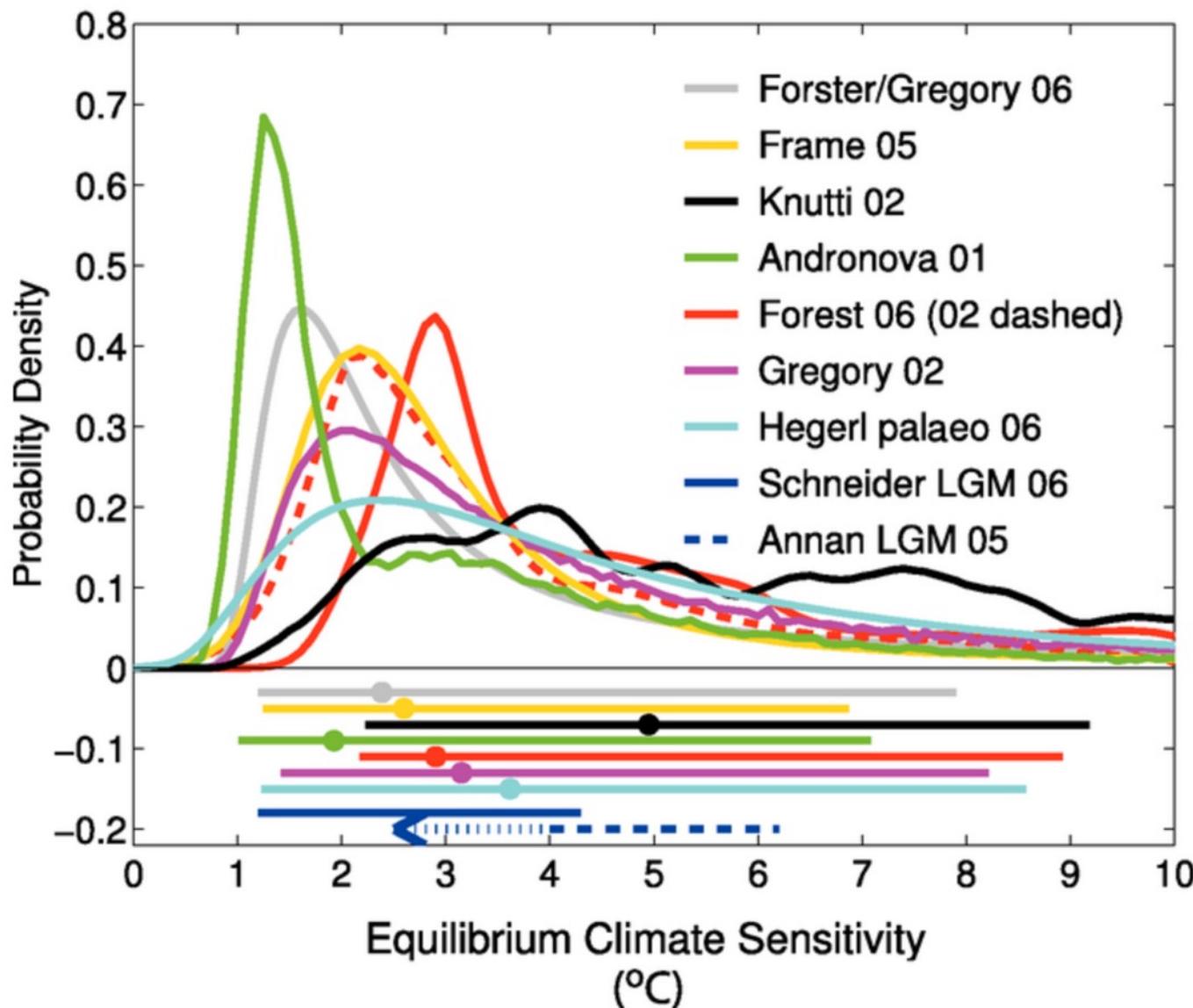
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$



What do you get if you
integrate over a
probability *density* function?

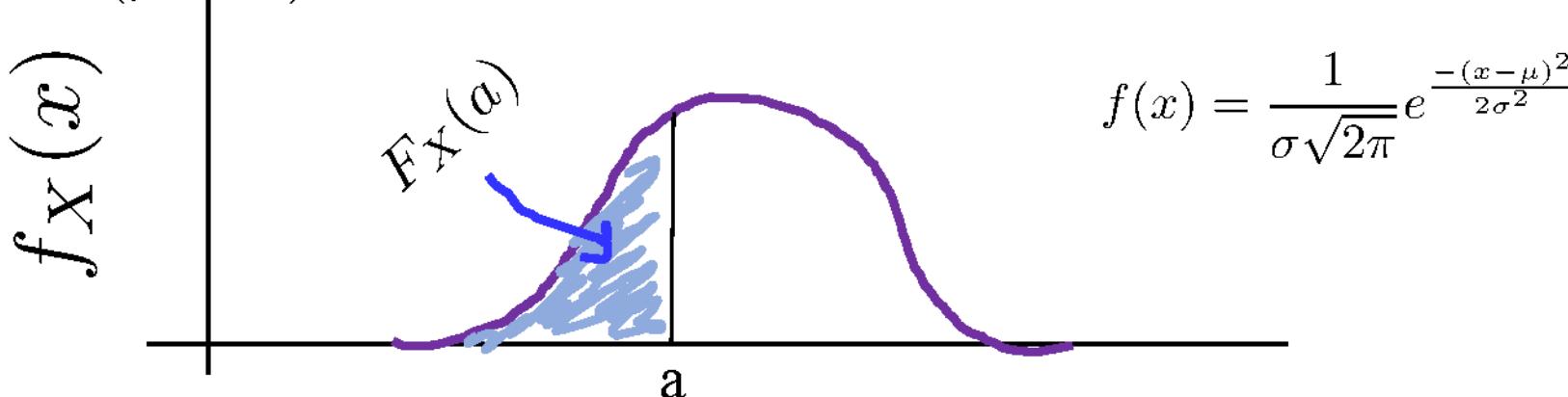
A probability!

Climate Sensitivity

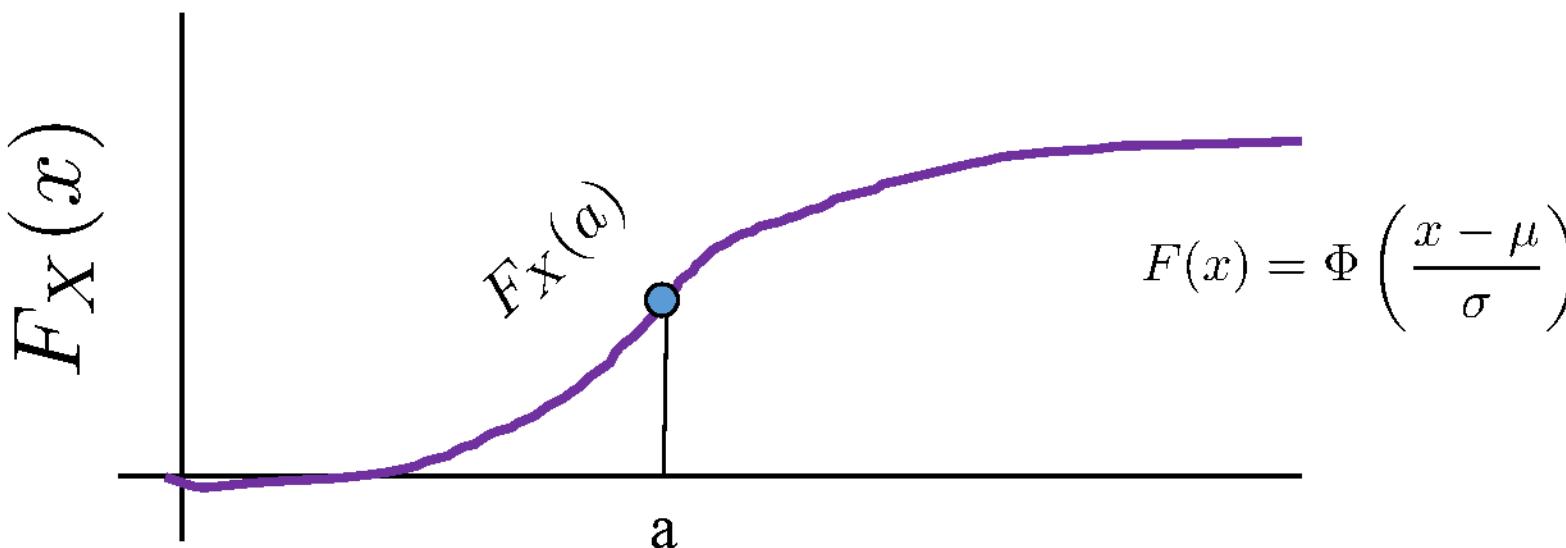


PDF and CDF of a Normal

$$X \sim N(\mu, \sigma^2)$$



$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



$$F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

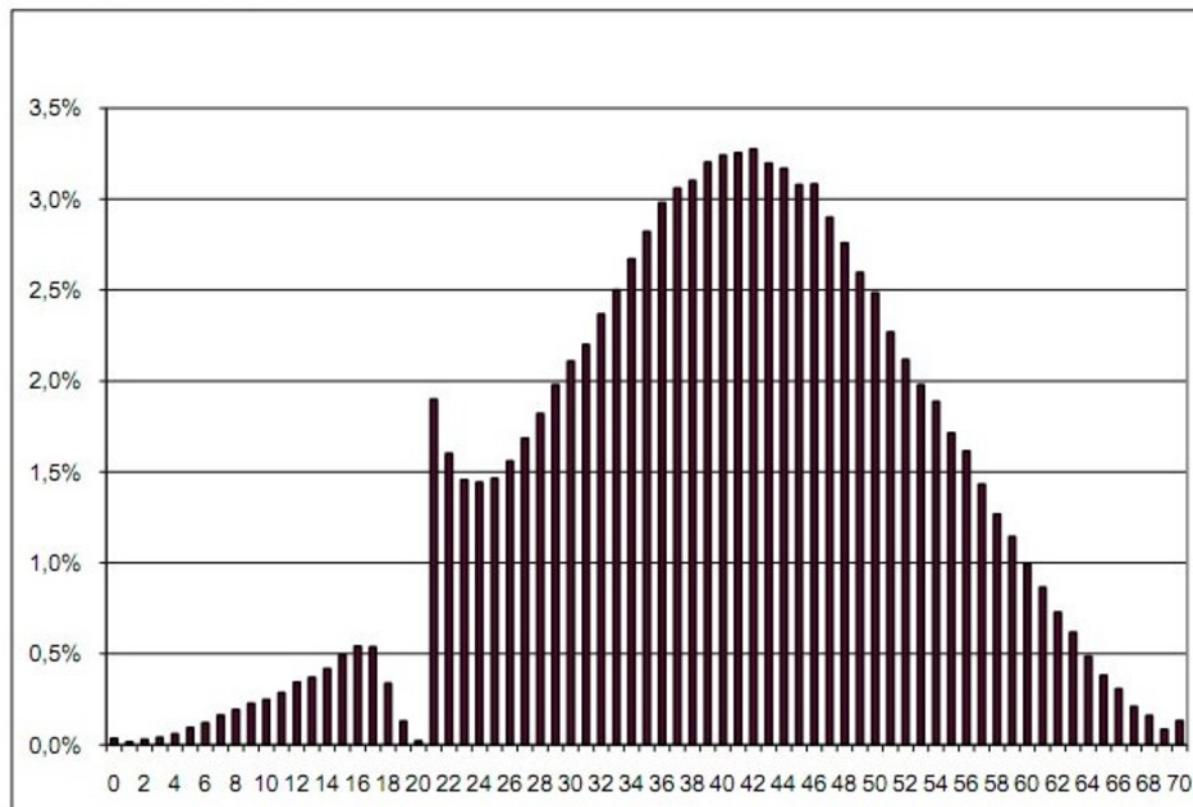
A CDF is the integral from $-\infty$ to x of the PDF

Altruism?

Scores for a standardized test that students in Poland are required to pass before moving on in school

See if you can guess the minimum score to pass the test.

2.1. Poziom podstawowy



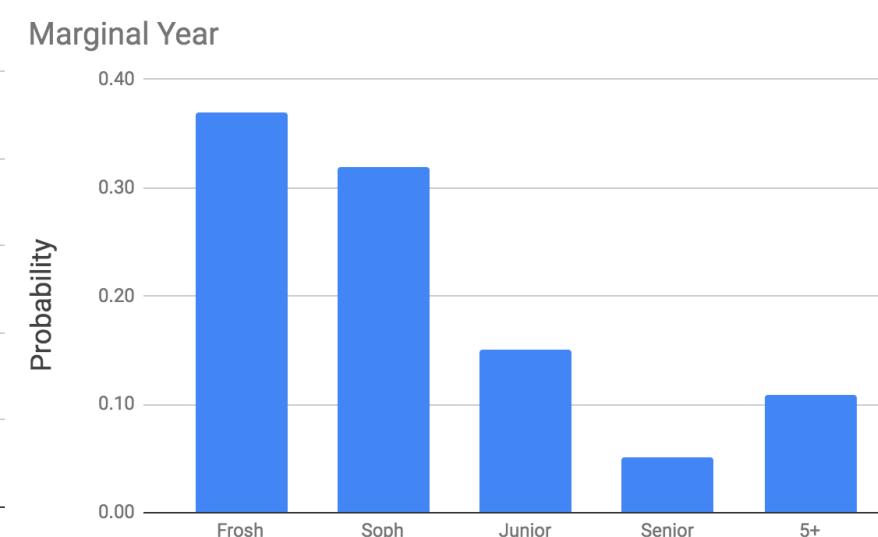
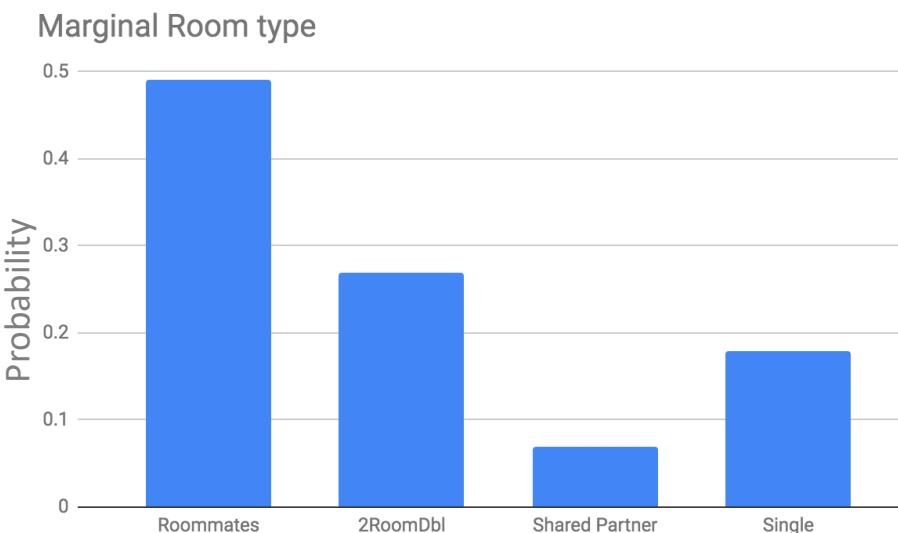
Wykres 1. Rozkład wyników na poziomie podstawowym

Probabilistic Models



Joint Probability Table

	Roommates	2RoomDbl	Shared Partner	Single	
Frosh	0.30	0.07	0.00	0.00	0.37
Soph	0.12	0.18	0.00	0.03	0.32
Junior	0.04	0.01	0.00	0.10	0.15
Senior	0.01	0.02	0.02	0.01	0.05
5+	0.02	0.00	0.05	0.04	0.11
	0.49	0.27	0.07	0.18	1.00



Multinomial

Example document:

“Pay for Viagra with a credit-card. Viagra is great.
So are credit-cards. Risk free Viagra. Click for free.”

$$n = 18$$

$$P \left(\begin{array}{c} \text{Viagra} = 2 \\ \text{Free} = 2 \\ \text{Risk} = 1 \\ \text{Credit-card: } 2 \\ \dots \\ \text{For} = 2 \end{array} \mid \text{spam} \right) = \frac{n!}{2!2!\dots2!} p_{\text{viagra}}^2 p_{\text{free}}^2 \dots p_{\text{for}}^2$$

Probability of seeing
this document | spam

It's a Multinomial!

The probability of a word in
spam email being viagra



General “Inference”



General “Inference”

WebMD® Symptom Checker BETA

INFO SYMPTOMS > QUESTIONS CONDITIONS DETAILS TREATMENT

Add more symptoms

Type your main symptom here

or Choose common symptoms

bloating cough diarrhea dizziness fatigue
fever headache muscle cramp nausea
throat irritation

AGE 30 GENDER Male

MY SYMPTOMS

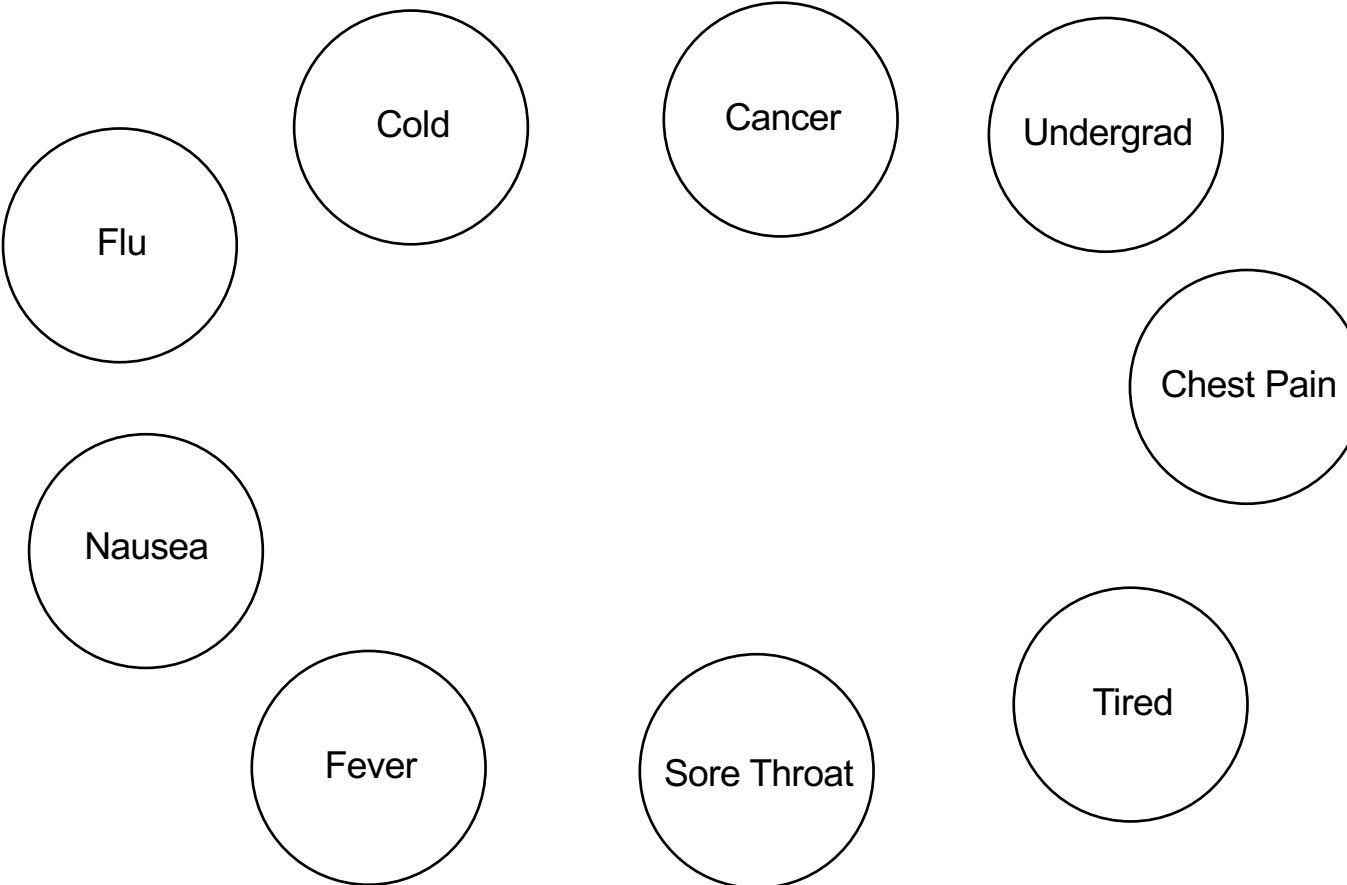
cough × throat irritation ×
sneezing ×

Results Strength: MODERATE

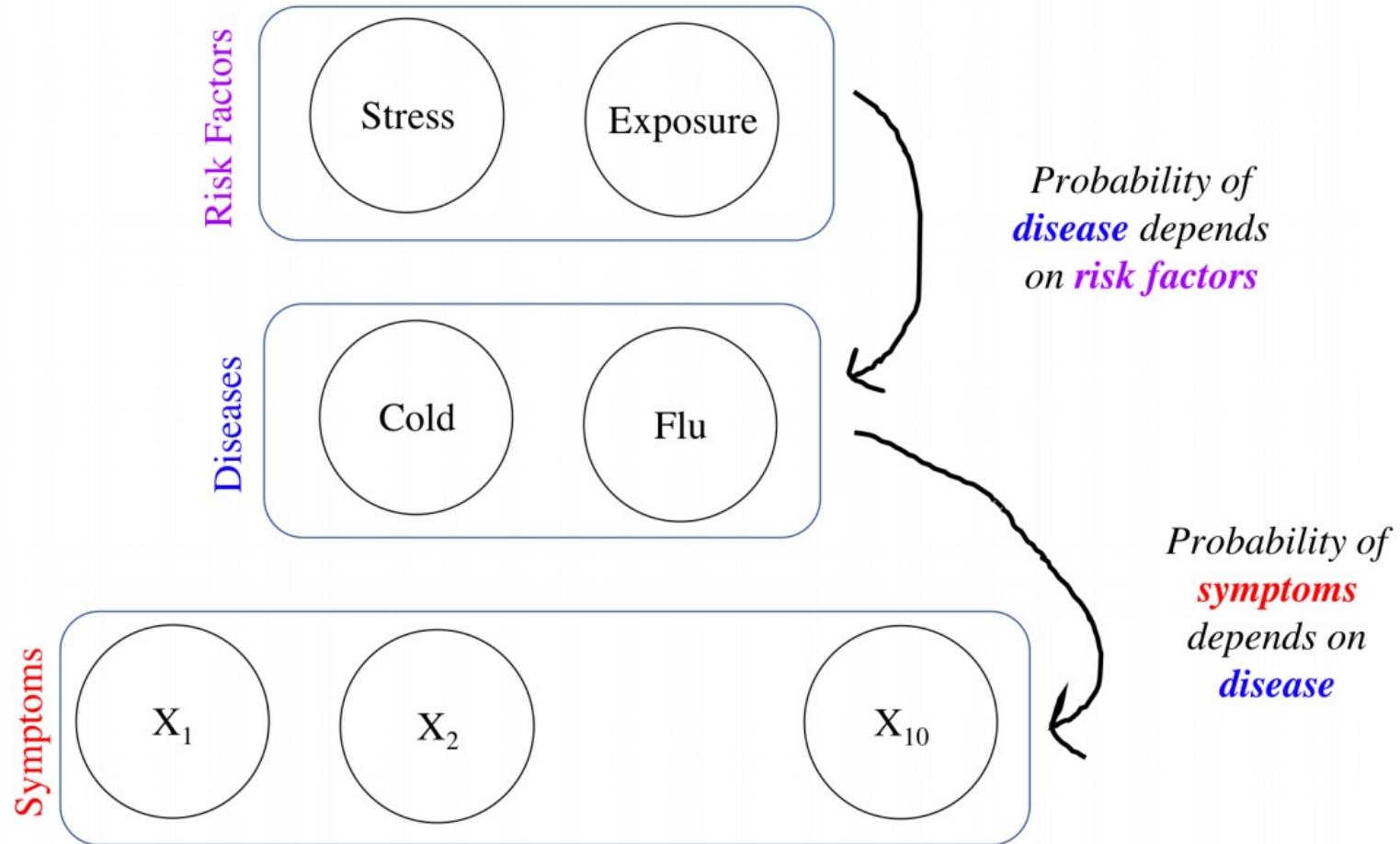


< Previous Info Continue >

Lots of Random Vars?



Bayes Nets!



Alg #1: Rejection Sampling

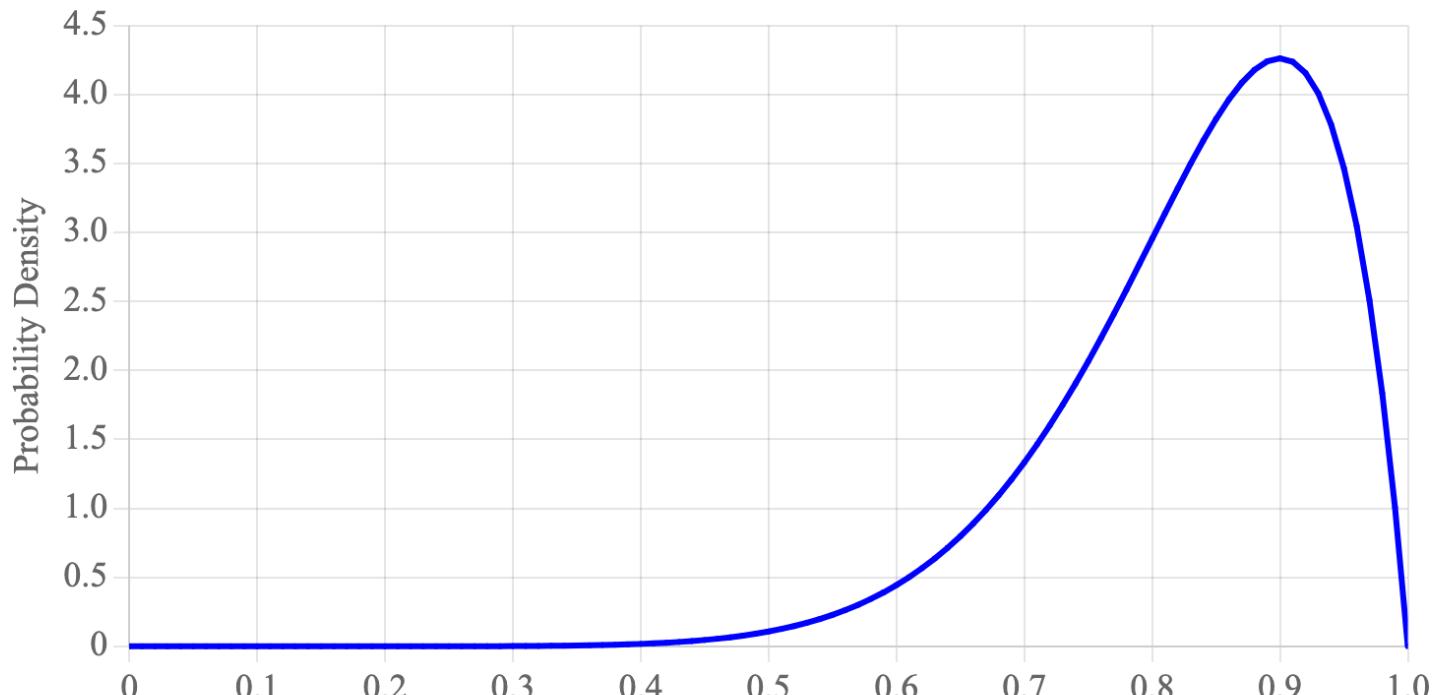
```
3 N_SAMPLES = 100000
4
5 # Program: Joint Sa
6 # -----
7 # we can answer any
8 # with multivariate
9 # where conditioned
10 def main():
11     obs = getObservation()
12     print 'Observation: ', obs
13
14     samples = sampleN(100000)
15     prob = probFlu(samples)
16     print 'Pr(Flu | Obs) = ', prob
```

Uncertainty Theory



$$\begin{aligned}
f(X = x | H = 9, T = 1) &= \frac{P(H = 9, T = 1 | X = x) \cdot f(X = x)}{P(H = 9, T = 1)} && \text{Bayes Theorem} \\
&= \frac{\binom{10}{9} x^9 (1-x)^1 \cdot f(X = x)}{P(H = 9, T = 1)} && \text{Binomial PMF} \\
&= \frac{\binom{10}{9} x^9 (1-x)^1 \cdot 1}{P(H = 9, T = 1)} && \text{Uniform PDF} \\
&= \frac{\binom{10}{9}}{P(H = 9, T = 1)} x^9 (1-x)^1 && \text{Constants to front} \\
&= K \cdot x^9 (1-x)^1 && \text{Rename constant}
\end{aligned}$$

Lets take a look at that function. For now we can let $K = \frac{1}{110}$. Regardless of K we will get the same shape, just scaled:



Lets Play!

Drug A

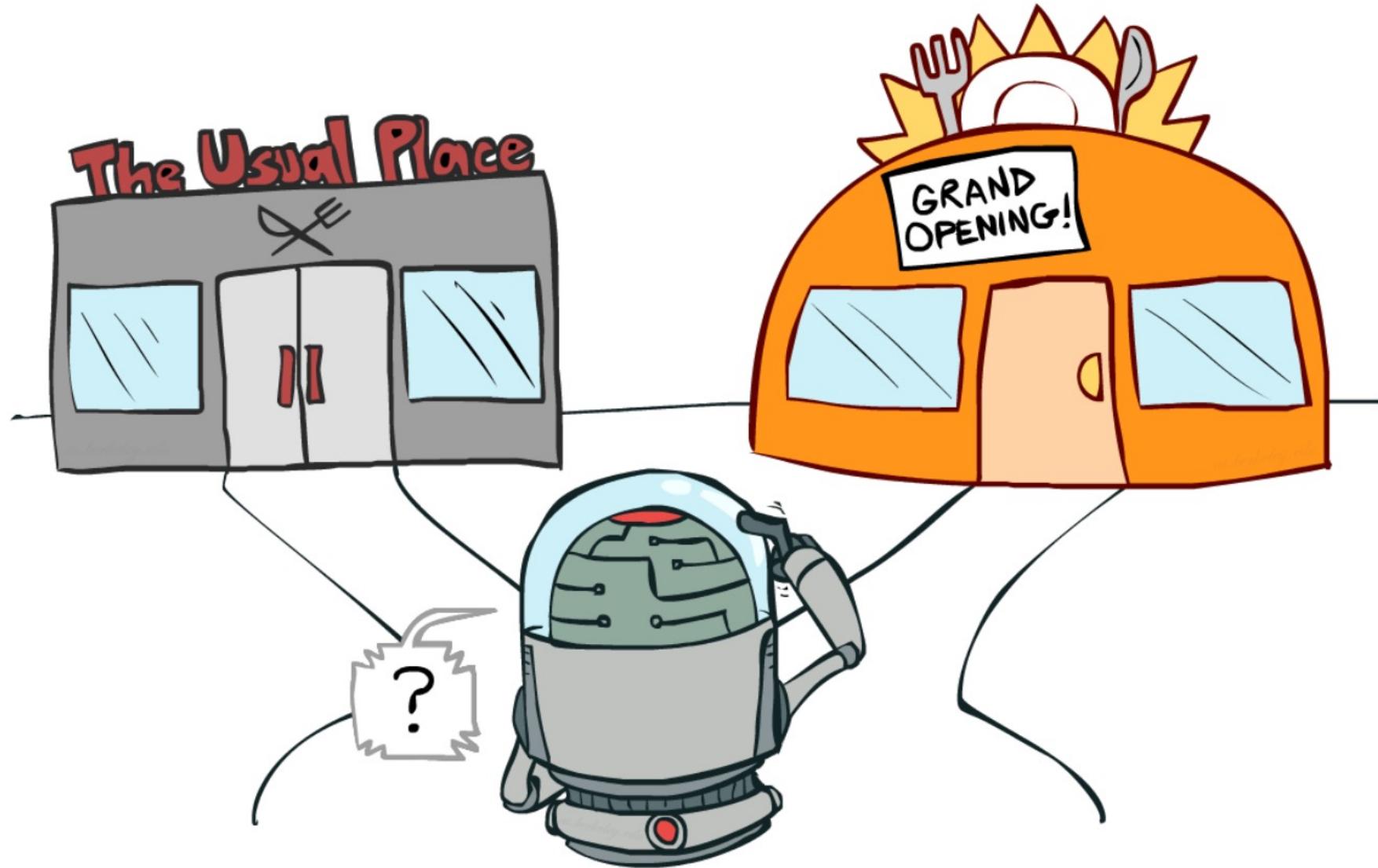


Drug B

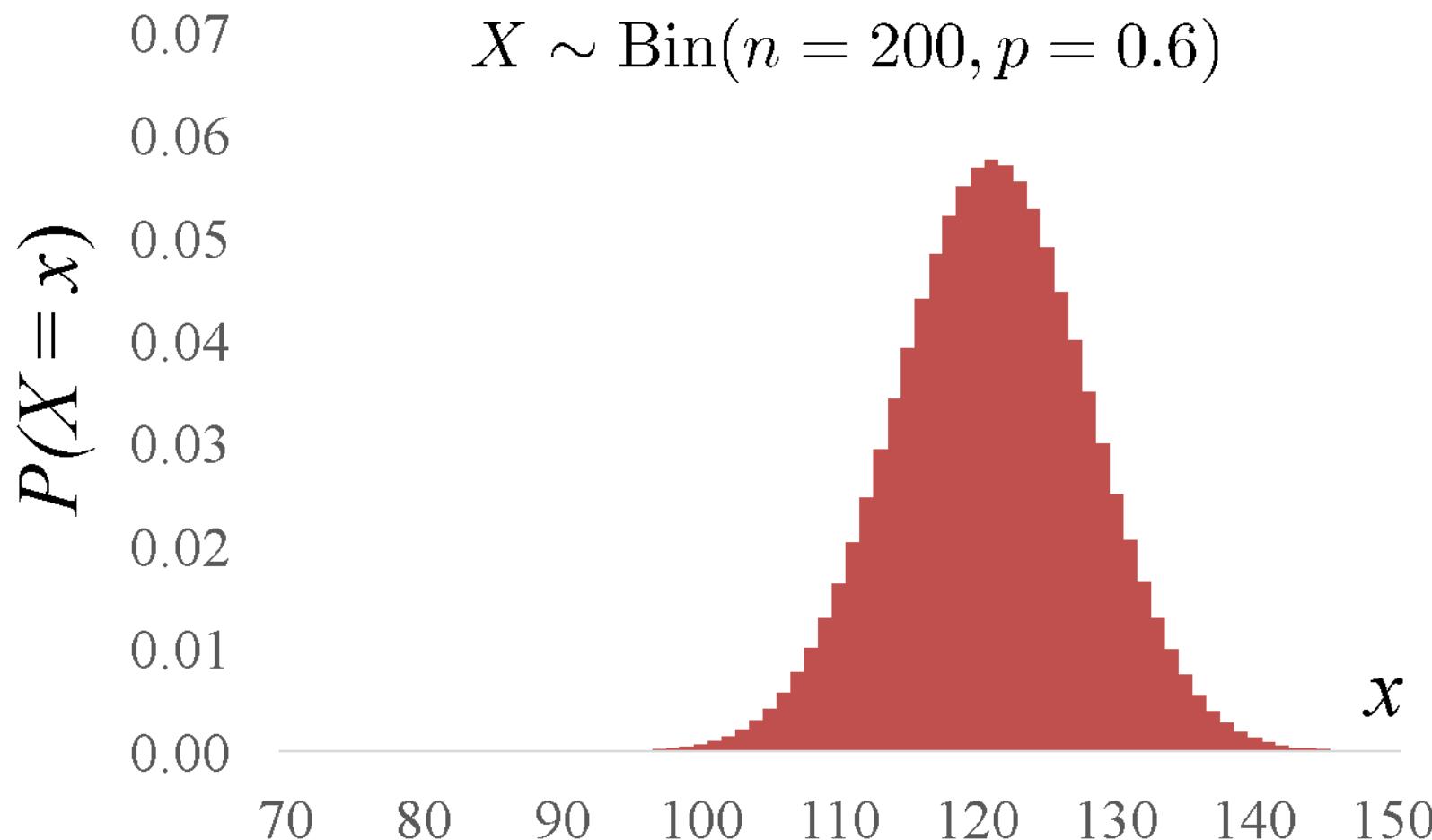


Which one do you give to a patient?

Thompson Sampling



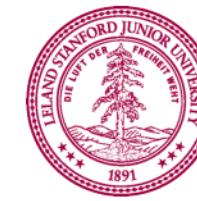
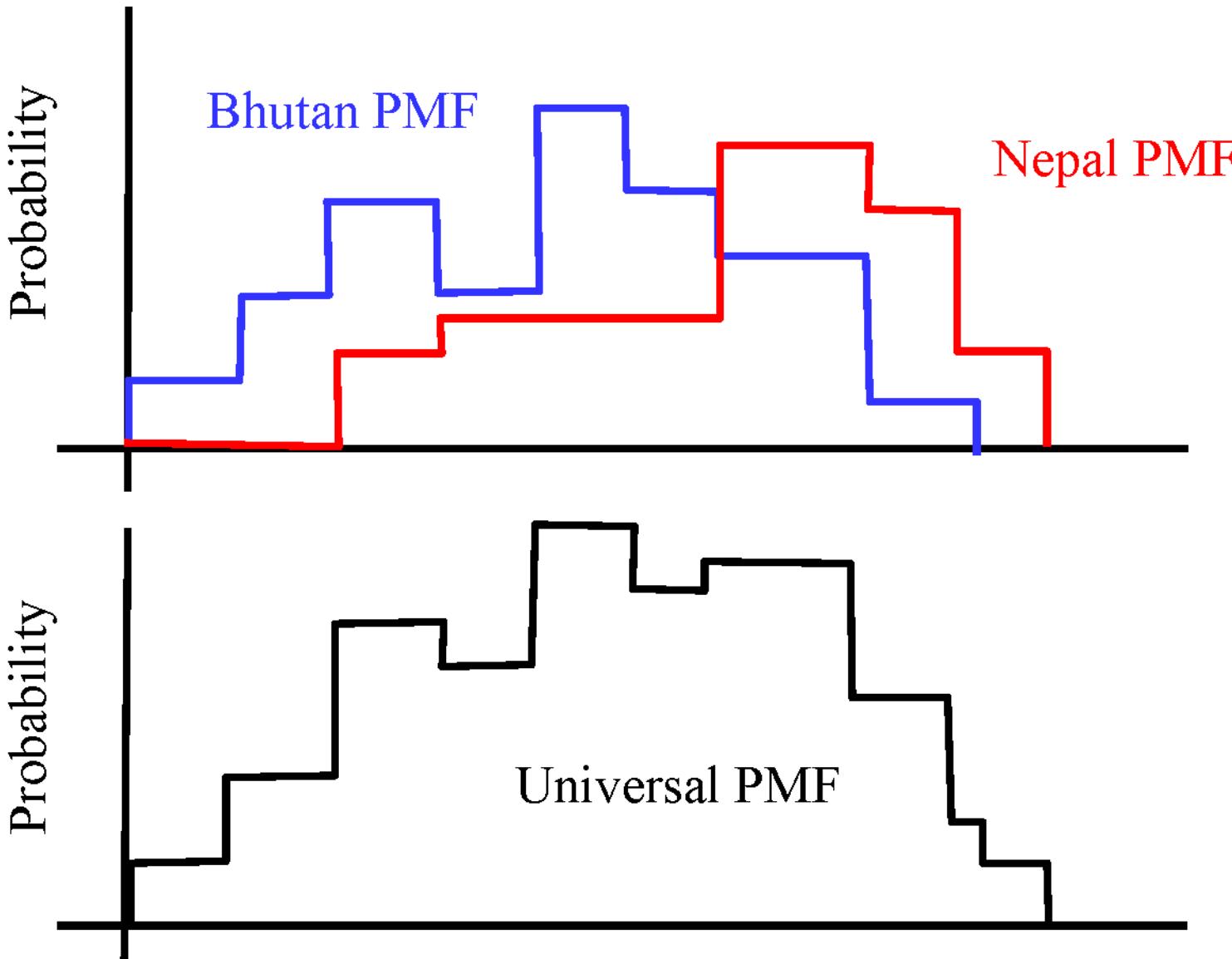
C.L.T. Explains This



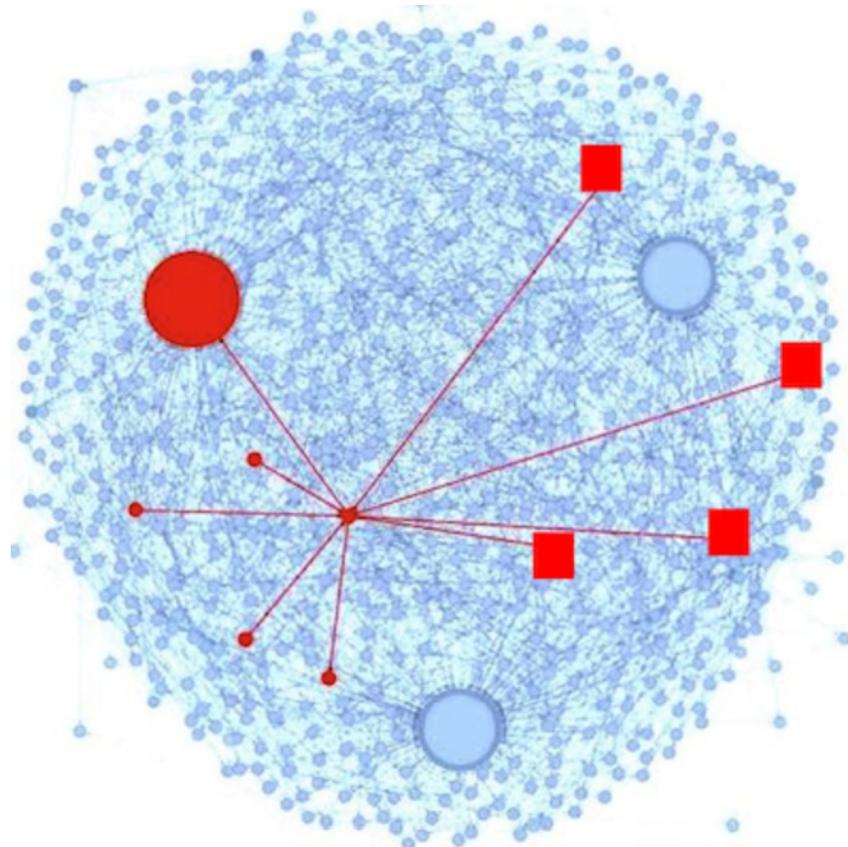


Bootstrap

Universal Sample



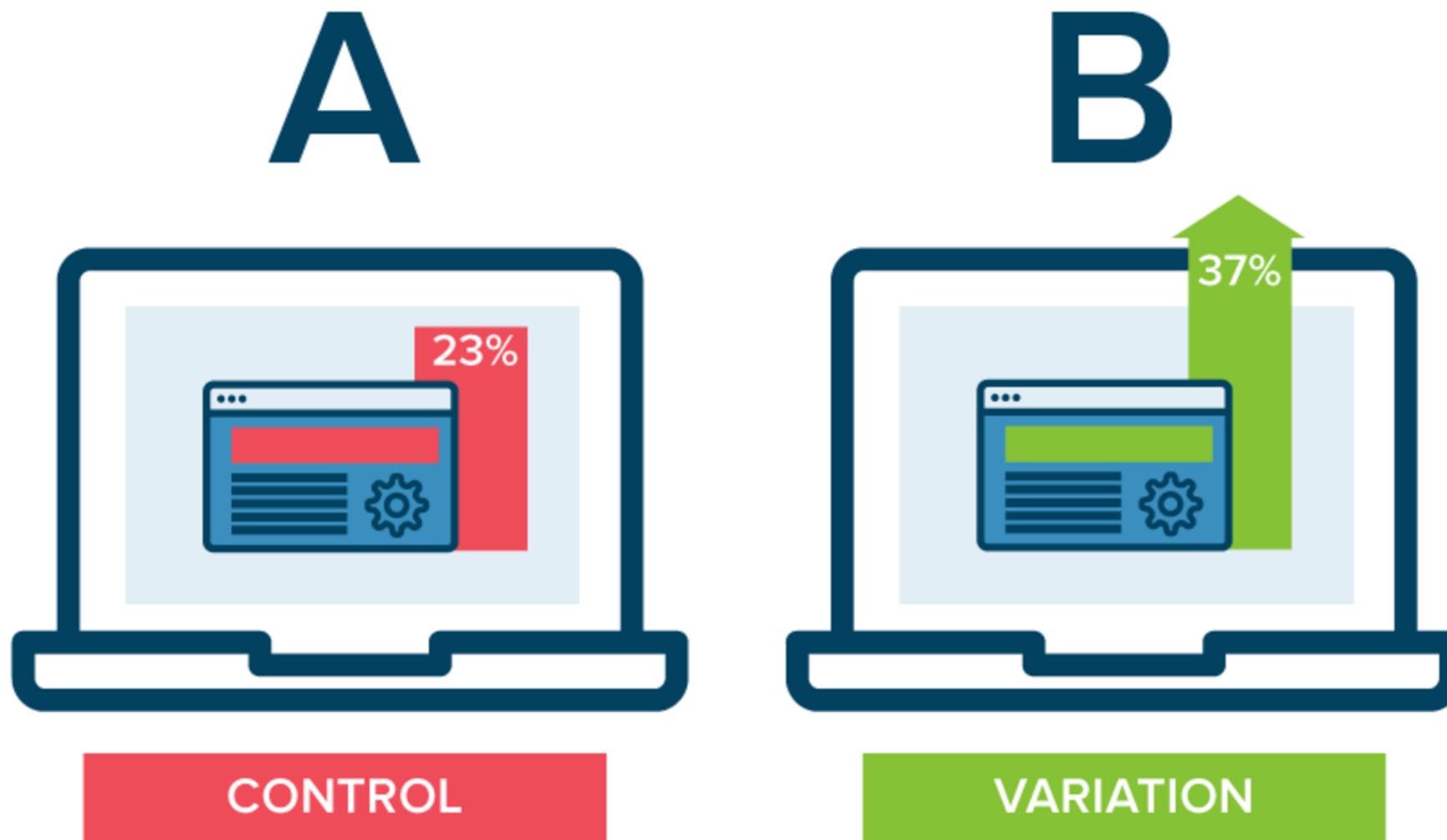
Peer Grading



Peer Grading on Coursera
HCI.

31,067 peer grades for
3,607 students.

A/B Testing

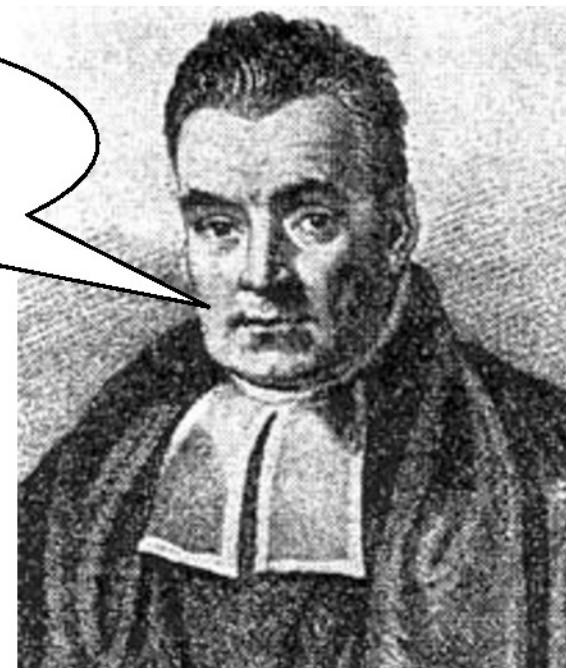


Machine Learning



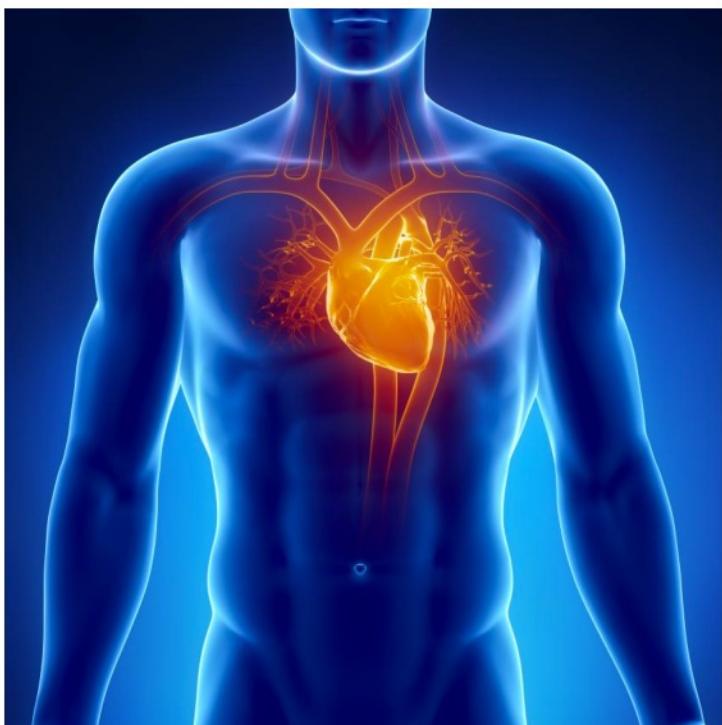
MAP: Most Probable Parameter

So good to see
you again!



Machine Learning

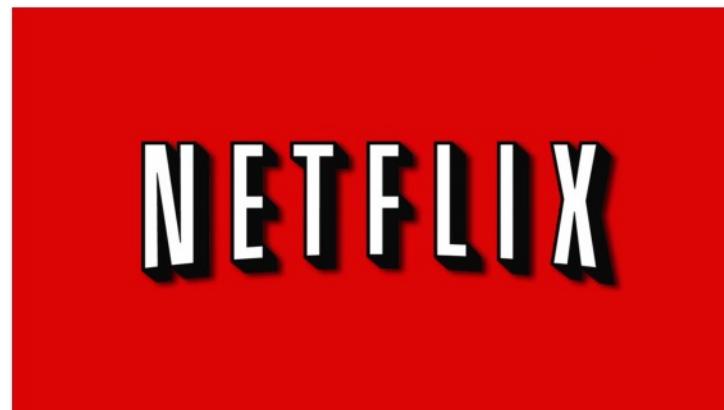
Heart



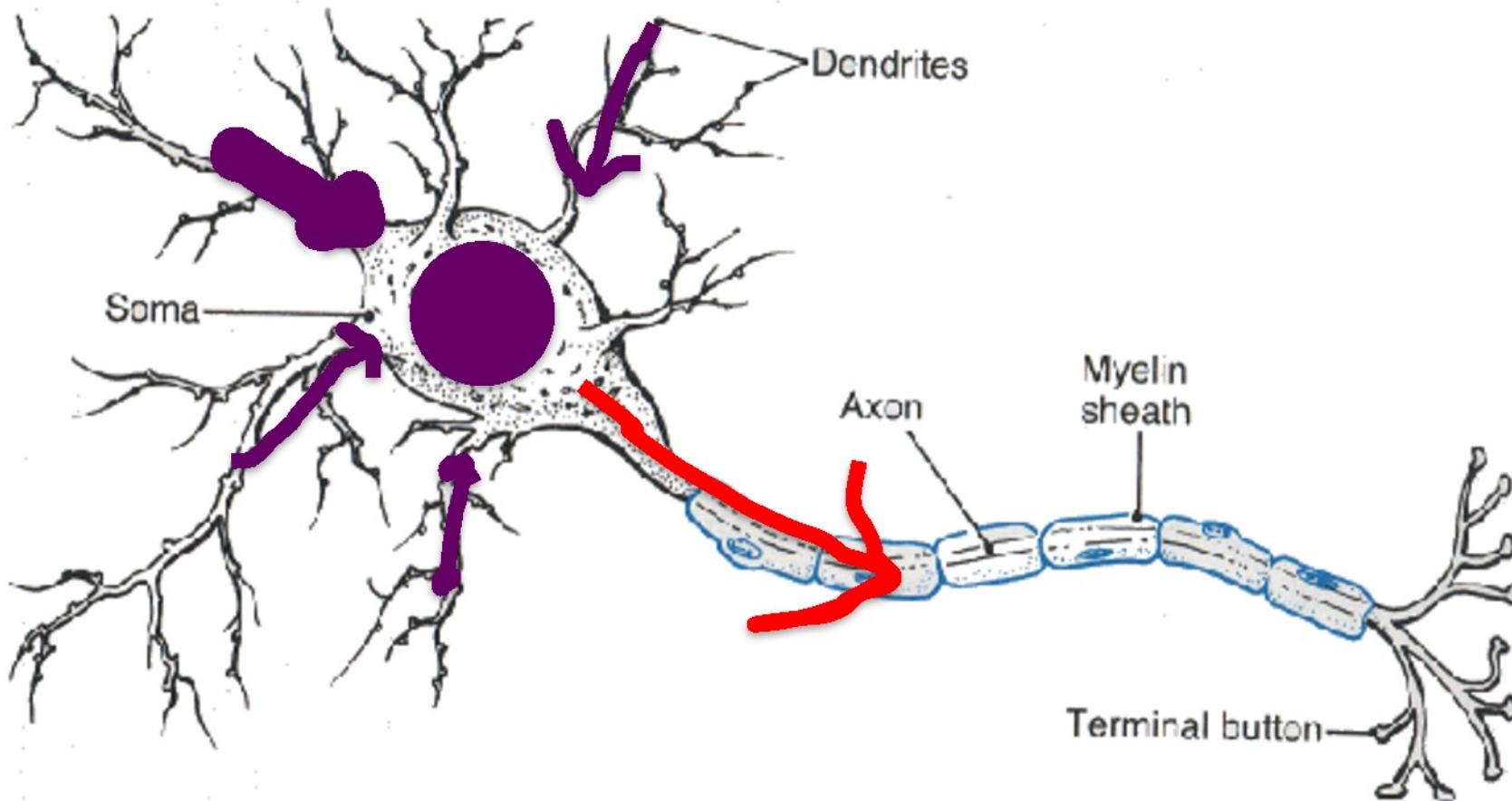
Ancestry



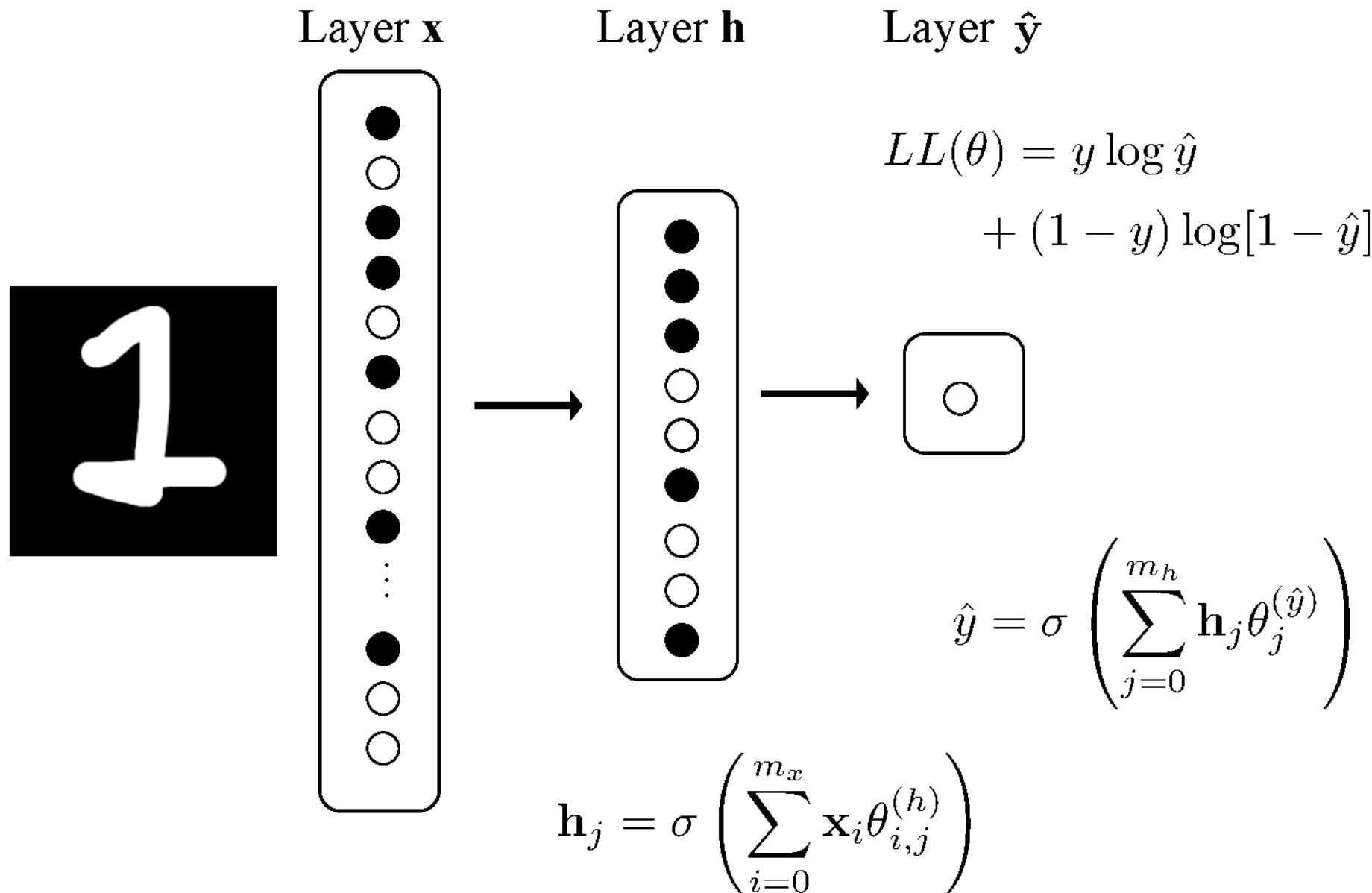
Netflix



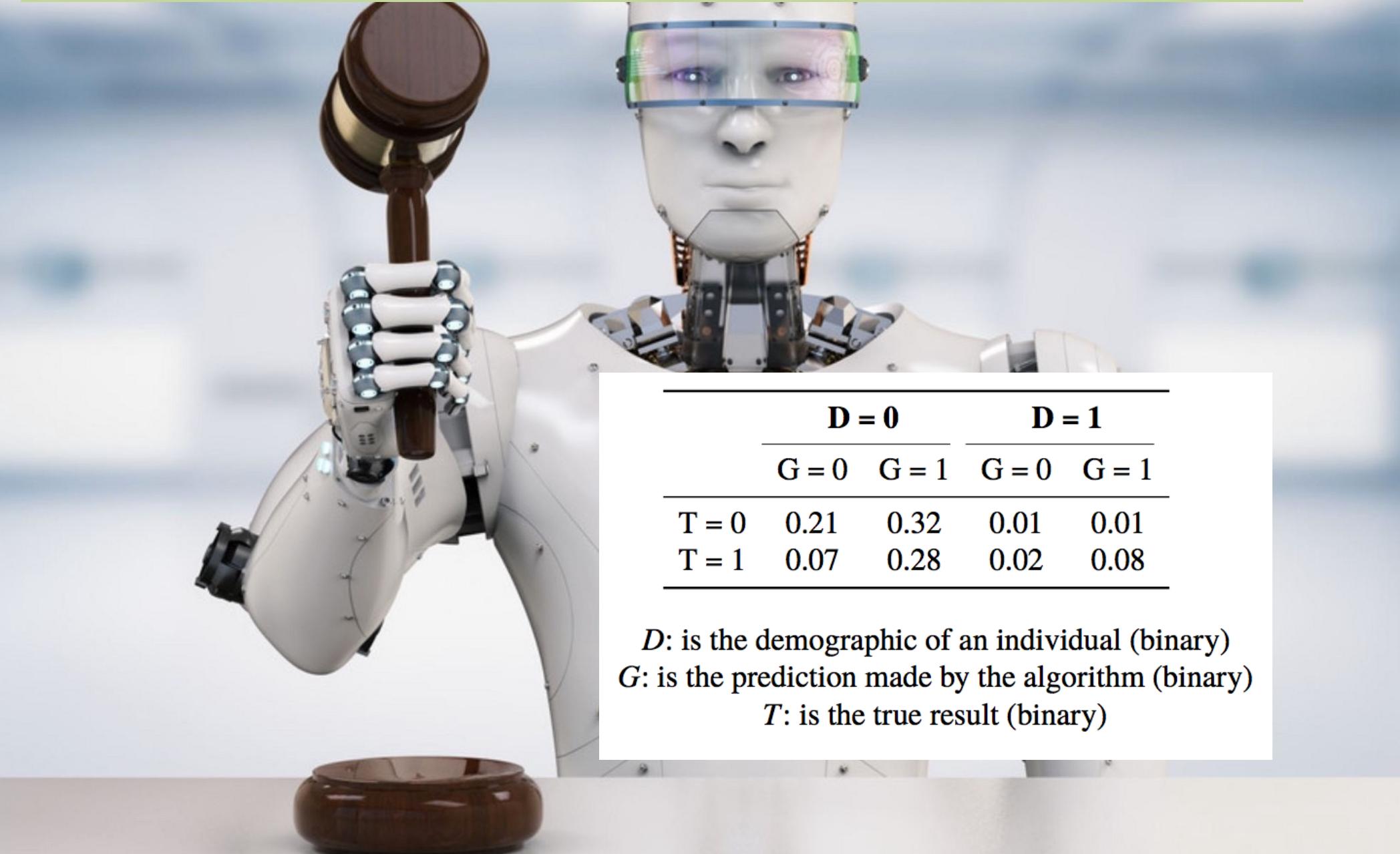
Logistic Regression



Deep Learning



Algorithmic Fairness



		D = 0		D = 1	
		G = 0	G = 1	G = 0	G = 1
T = 0	T = 0	0.21	0.32	0.01	0.01
	T = 1	0.07	0.28	0.02	0.08

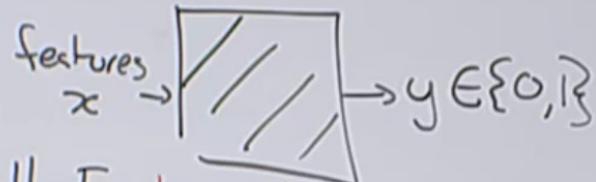
D: is the demographic of an individual (binary)

G: is the prediction made by the algorithm (binary)

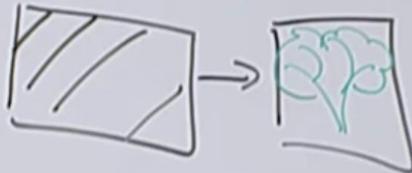
T: is the true result (binary)

DallE and GPT

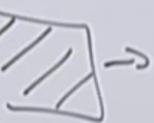
Pset 6:

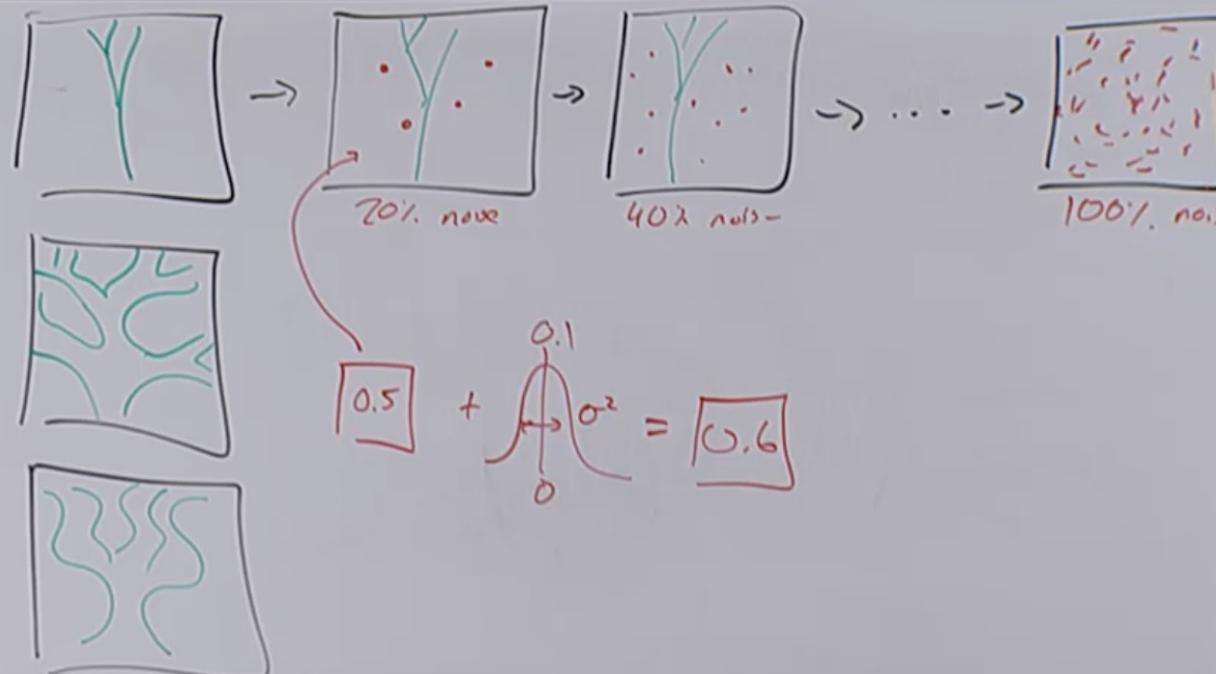


Dall E tree block ?: white:



GPT:

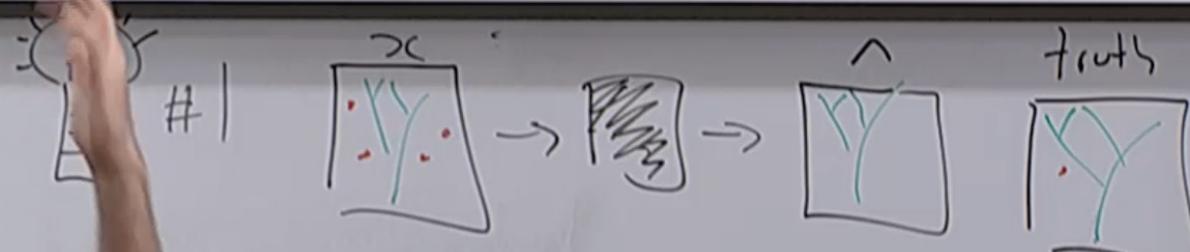
"I love" \rightarrow  \rightarrow "CS109 and pancakes"



reverse
a classification
predictor

scope
Google

how
to
predict



hey GPT can you help me with my homework?

Of course! I'd be happy to help you with your homework. Can you provide some details about what you need help with?

A fair 6-sided die is repeatedly rolled until the total sum of all the rolls exceeds 300. Approximate the probability that at least 80 rolls are necessary to reach a sum that exceeds 300.

To solve this problem, we need to use the geometric distribution. This distribution is used to model the number of Bernoulli trials (i.e. experiments with only two possible outcomes, such as rolling a die) that must be conducted in order to obtain a success. In this case, rolling a 6 on the die is considered a success because it contributes the most to the total sum.

The probability of success on each roll is $1/6$, since there is only one favorable outcome (rolling a 6) out of a total of 6 possible outcomes. This means that the probability of failure on each roll is $5/6$.

The probability of obtaining at least 80 successes before the first failure is given by:

$$P(X \geq 80) = 1 - P(X < 80)$$

where X is the random variable representing the number of successes before the first failure.

Free Research Preview: ChatGPT is optimized for dialogue. Our goal is to make AI systems more natural to interact with, and your feedback will help us improve our systems and make them safer.

This is the probability that at least 80 rolls are necessary to reach a sum that exceeds 300. I hope this helps! Let me know if you have any other questions.

hmm that doesn't seem correct. I am pretty sure you are supposed to use the central limit theorem

You are correct that the central limit theorem can be used to approximate the probability in this problem. The central limit theorem states that the sum of a large number of independent and identically distributed (i.i.d.) random variables will be approximately normally distributed, even if the individual random variables are not normally distributed.

In this problem, we are summing the outcomes of a large number of independent rolls of a fair 6-sided die. This means that the central limit theorem can be applied to approximate the distribution of the sum.

To use the central limit theorem, we first need to compute the mean and standard deviation of the sum. The mean of the sum is simply the mean of a single roll multiplied by the number of rolls. Since each roll has a mean of 3.5 (obtained by adding the possible outcomes and dividing by 6), the mean of the sum is $3.5 * 80 = 280$.

The standard deviation of the sum is equal to the standard deviation of a single roll multiplied by the square root of the number of rolls. Since the standard deviation of a single roll is 1.7 (obtained by using the formula $\sqrt{1/6 * (5/6)} = 1.7$), the standard deviation of the sum is $1.7 * \sqrt{80} = 12.4$.

Free Research Preview: ChatGPT is optimized for dialogue. Our goal is to make AI systems more natural to interact with, and your feedback will help us improve our systems and make them safer.

By the numbers

~ 30 Major Keys



Naïve Bayes Assumption:

$$P(\mathbf{x}|y) = \prod_i P(x_i|y)$$

64 Course Reader Chapters

Mac OS X browser window showing the "Name to Age" demo.

The URL in the address bar is <https://chrispiech.github.io/probabilityForComputerScientists/en/examples/name2age/>.

The sidebar contains a list of course reader chapters:

- 100 Binomial Problems
- Jury Selection
- Grading Eye Inflammation
- Gaussian CDF Calculator
- Grades are Not Normal
- Curse of Dimensionality
- Probability of Baby Delivery

Part 3: Probabilistic Models

- Joint Probability
- Multinomial
- Continuous Joint
- Inference
- Bayesian Networks
- Independence in Variables
- Correlation
- General Inference
- Worked Examples
 - CS109 Logo
 - Fairness in AI
 - Federalist Paper Authorship
 - Name to Age
 - Bridge Distribution
 - Tracking in 2D

Part 4: Uncertainty Theory

- Beta Distribution
- Adding Random Variables
- Central Limit Theorem
- Sampling
- Bootstrapping
- Algorithmic Analysis
- Worked Examples

The main content area shows the following text:

Here are the assumptions that our demo makes:

1. This data only is accurate for names of people in the US. The probability of age given names could be very different in other countries.
2. The US census is not perfect. It does not capture all people who are resident in the US, and there are demographics which are underrepresented. This will also skew our results.

Demo

Query Name: Katherine ✓

A line graph showing the probability distribution of ages for the name "Katherine". The x-axis is labeled "Age" and ranges from 0 to 120. The y-axis is labeled "Probability" and ranges from 0.002 to 0.020. The distribution is bimodal, with peaks around age 30 and age 100.

Age	Probability
10	0.006
20	0.013
30	0.019
40	0.017
50	0.008
60	0.009
70	0.012
80	0.005
90	0.004
100	0.009
110	0.007

Records with name: 589753

This demo is based on real data from US Social Security applications between 1914 and 2014. Thank you to <https://www.kaggle.com/kaggle/us-baby-names> for compiling the data. [Download Data](#).

Names that Give Away Your Age

Some names have certain years where they were exceptionally popular. These names provide quite a lot of information about birth year. Let's look at some of the names with the highest max probability.

1 Personal Challenge



1 New Counter



thank you