Impossibilities in Computing

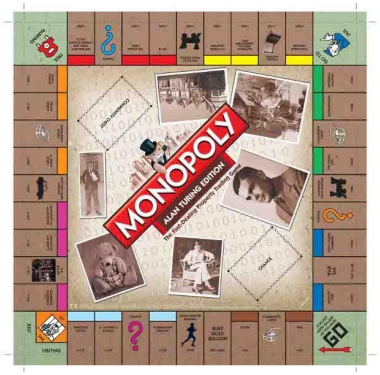
Lecture 3

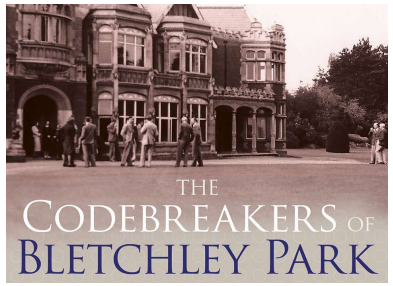
Unit 2

ML and Society (Spring 2024)

Atri Rudra

Pass Phrase: **Alan Turing**



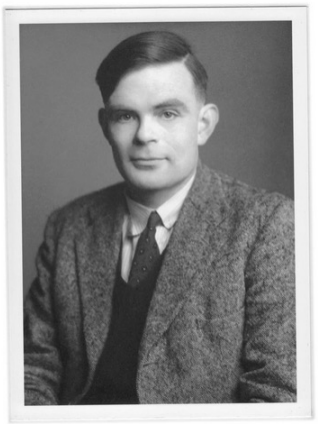


The New York Times

Overlooked No More: Alan Turing, Condemned Code Breaker and Computer Visionary

His ideas led to early versions of modern computing and helped win World War II. Yet he died as a criminal for his homosexuality.

Share full article 143



Alan Turing in 1951. Though he is regarded today as one the most innovative thinkers of the 20th century, at his death many of his wartime accomplishments were classified. Godrey Argent Studio, via The Royal Society

VOL. LIX. NO. 236.] [October, 1950

MIND

A QUARTERLY REVIEW

OF

PSYCHOLOGY AND PHILOSOPHY

I.—COMPUTING MACHINERY AND INTELLIGENCE

BY A. M. TURING

1. *The Imitation Game.*

I PROPOSE to consider the question, 'Can machines think?' This should begin with definitions of the meaning of the terms 'machine' and 'think'. The definitions might be framed so as to reflect so far as possible the normal use of the words, but this attitude is dangerous. If the meaning of the words 'machine' and 'think' are to be found by examining how they are commonly used it is difficult to escape the conclusion that the meaning and the answer to the question, 'Can machines think?' is to be sought in a statistical survey such as a Gallup poll. But this is absurd. Instead of attempting such a definition I shall replace the question by another, which is closely related to it and is expressed in relatively unambiguous words.

The new form of the problem can be described in terms of a game which we call the 'imitation game'. It is played with three people, a man (A), a woman (B), and an interrogator (C) who may be of either sex. The interrogator stays in a room apart from the other two. The object of the game for the interrogator is to determine which of the other two is the man and which is the woman. He knows them by labels X and Y, and at the end of the game he says either 'X is A and Y is B' or 'X is B and Y is A'. The interrogator is allowed to put questions to A and B thus:

C: Will X please tell me the length of his or her hair?

1. Benedict Cumberbatch played Alan Turing in the 2014 movie The Imitation Game https://www.imdb.com/title/tt2084970/
2. Turing was an avid Monopoly player https://www.theguardian.com/technology/2012/sep/10/alan-turing-monopoly-board-google
3. Turing was 5th in the British marathon trials for the 1948 Olympics https://kottke.org/18/04/alan-turing-was-an-excellent-runner
4. Turing led the effort to break Nazi code ("Enigma") at Bletchley Park https://www.nationalww2museum.org/war/articles/alan-turing-betchley-park
5. Turing was a gay man who given a choice between "chemical castration" and imprisonment for homosexual acts https://www.nytimes.com/2019/06/05/obituaries/alan-turing-overlooked.html
6. Turing write what is the considered the first major paper on AI https://academic.oup.com/mind/article/LIX/236/433/986238

Unit 1 grading

note @49 stop following 7 views

Actions

[Combined] Incoming emails/Autolab (for ML&Soc) grades for Unit 1 Group Submission

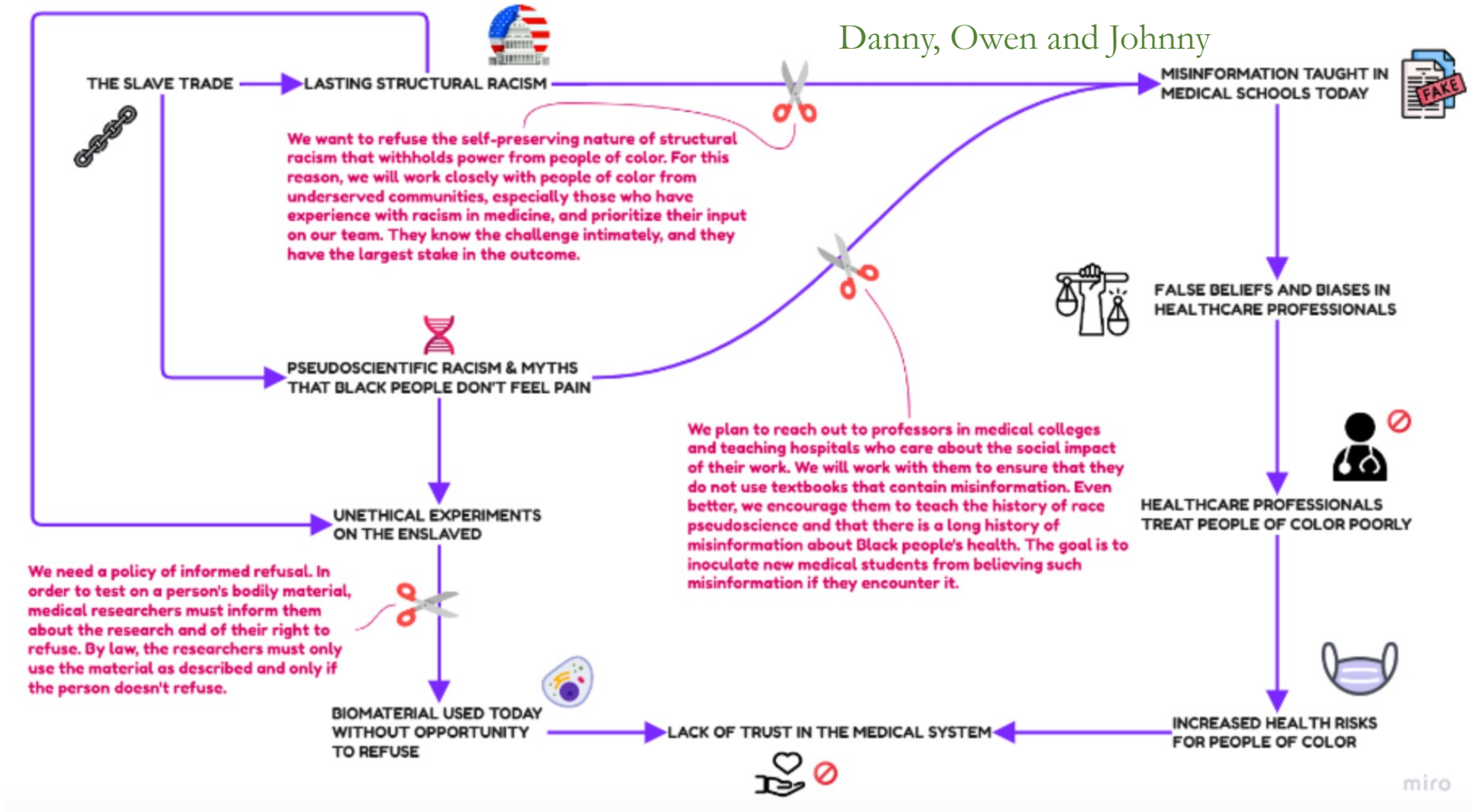
Hey all-

At some point tonight you'll receive an email with comments on your Unit 1 Group Submission. ML&Soc students will also see grades released on Autolab later tonight. Some details on this:

- Atri and Kenny graded on this rubric, per question: Unit_1_Group_Submission_Rubric.pdf. Dalia gave a single overall score. The final score was an average of these two.
- Your email will just contain a table with three columns. The first is the question (e.g. Problem, Current World, etc.). The third column contains three rows per question - one for Dr. Muller's comments, one for Atri/Kenny's comments, and one with the level grading from the above rubric (only used by Atri/Kenny). The second column tells you which is which. Hopefully this is clear when you get the email!
- This is only for the group score for unit 1.
    - We'll release details on the peer survey scores by early next week.
- If you have any questions, please let us know, and/or feel free to set up a meeting with us!

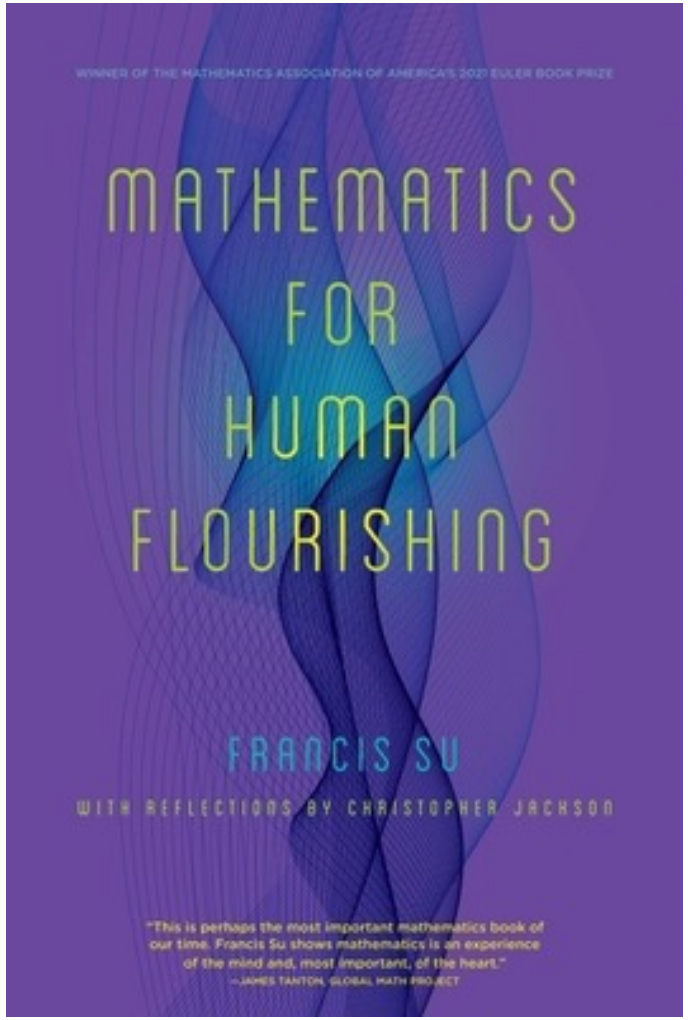
unit1 combined

Surprising for us: Lack of causal diagrams



TQE


IN DETAIL
To predict and serve?
Predictive policing systems are used increasingly by law enforcement to try to prevent crime before it occurs. But what happens when these systems are trained using biased data? **Kristian Lum** and **William Isaac** consider the evidence – and the social consequences
Downloaded from


WINNER OF THE MATHEMATICS ASSOCIATION OF AMERICA'S 2021 EULER BOOK PRIZE
MATHEMATICS FOR HUMAN FLOURISHING
FRANCIS SU
WITH REFLECTIONS BY CHRISTOPHER JACKSON
"This is perhaps the most important mathematics book of our time. Francis Su shows mathematics is an experience of the mind and, most important, of the heart."
—JAMES TANTON, GLOBAL MATH PROJECT

Process

- Form groups
- I'll start off with three things that stood out
- Reflect on two Questions (10 minutes, ish)
- Share back on your group's reflections (10 minutes)
- Discuss **one question the class should talk about** (5 minutes, ish)
- Share back (5 minutes)
- Pick 2 to discuss in small groups (5 minutes)
- Share back (5 minutes)
- Potentially, rinse and repeat
- Break

Three things that stood out in the responses

Many of y'all

Identified with the parts on imposter syndrome

Said we should give more access to (math) education to students from marginalized groups

Will educating folks be enough to end white supremacy? If not, what else will be needed?

Wondered (how deep math will be useful to) build good algorithms

If somehow I guarantee you a perfect algorithm for the problem you want to solve, will that be enough?

TQEs

Two TQEs to discuss with your group

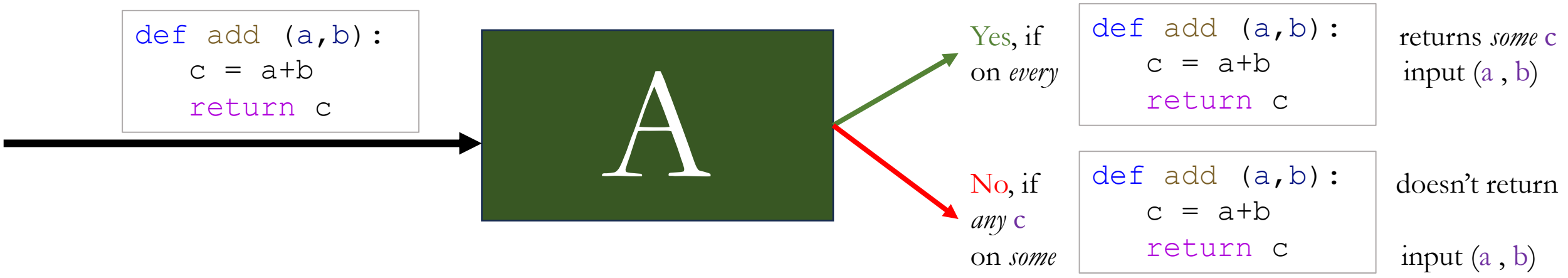
Break!

Mon recap: Halting Problem

Input: A program P

Output: Yes if P terminates on all possible inputs
No otherwise

Let A be a program that solves the Halting problem on all inputs

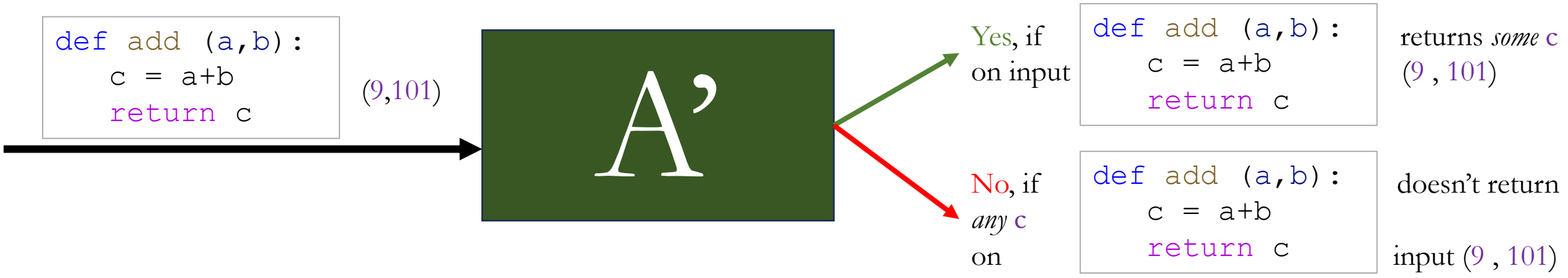


Meta Q: Halting Problem (ver 2)

Input: A program P and input I

Output: Yes if P terminates on I
No otherwise

Let A' be a program that solves the Halting problem (ver 2)



Halting Problem: ver 1 vs. ver 2

Input: A program P

Output: Yes if P terminates on all possible inputs
No otherwise

Input: A program P and input I

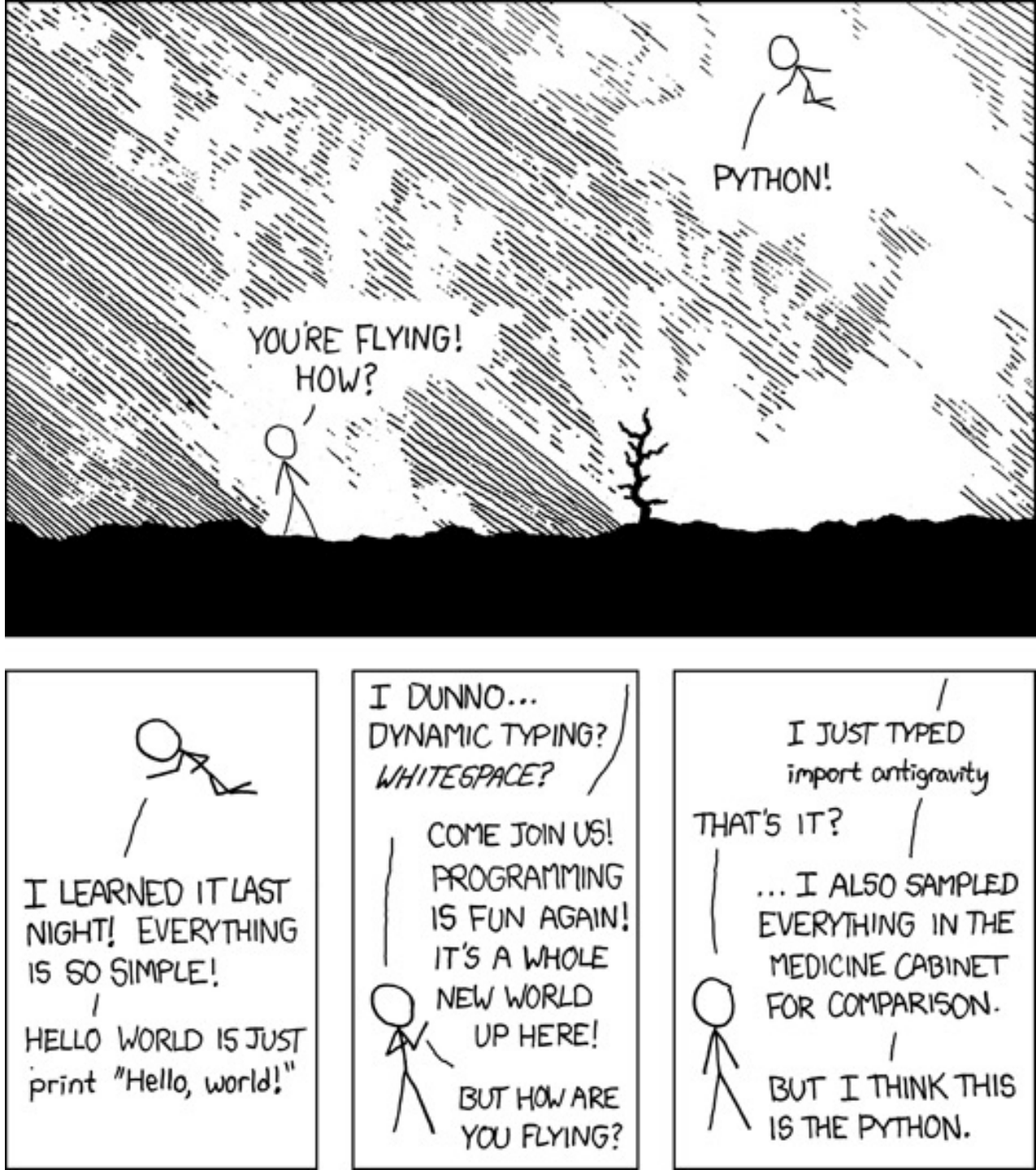
Output: Yes if P terminates on I
No otherwise

Ver 1 is harder than ver 2 (can *prove* this!)

Theorem: There is no magic box A' for ver 2 (as long as A' terminates)

Notebook detour

Let's setup stuff in Python



https://xkcd.com/353/

Questions?

Next up:

Input: A program P and input I

Output: Yes if P terminates on I
No otherwise

Theorem: There is no magic box A' for ver 2 (as long as A' terminates)

Let us assume a `magic_box` for A' exists!

```
def magic_box ( P, I):
# This is a magic box so there is no real code here!
''' Return True if P halts on I and False otherwise'''
```

Assume that

1. `magic_box` terminates on all inputs
2. `magic_box` ALWAYS correctly decides if P halts on I or not

A new function `contradiction`

```
def contradiction ( P ): # This function takes a program as an input

#Run magic_box on (P,P)
if magic_box (P,P): # Use an UTM to make this call
   while True:
      pass # Do nothing

return # Just terminate if magic_box(P,P) returns False
```

Since we assumed `magic_box` exists `contradiction` is well defined!

A function call

```
contradiction (contradiction) # Use an UTM to make this call
```

Since we assumed `contradiction` is well defined, the above is a legit function call!

Question: why is this legit?

```
contradiction (contradiction) # Use an UTM to make this call
```

Since we assumed `contradiction` is well defined, the above is a legit function call!

```
def contradiction ( P ): # This function takes a program as an input

#Run magic_box on (P,P)
if magic_box (P,P): # Use an UTM to make this call
    while True:
        pass # Do nothing

return # Just terminate if magic_box(P,P) returns False
```

1. Since data and programs are the same, we can pass a program as an input
2. A Universal program/TM can run any program on any input
3. We assumed `magic_box` exists
4. If AI chip companies can be a thing, why not this?

Choose ALL options that apply

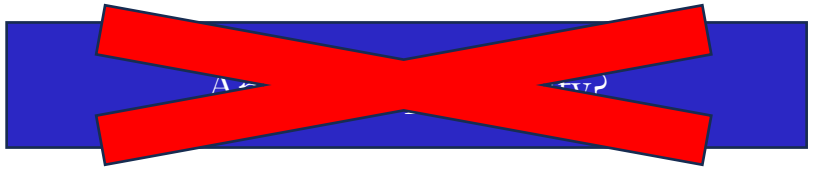
Wait, what?!!!

`contradiction (contradiction)`

What are outcomes of the function call `contradiction (contradiction)`?

It terminates

It does not terminate



Case 1:

`contradiction (contradiction)` terminates

```
def contradiction ( P ): # This function takes a program as an input

#Run magic_box on (P,P)
if magic_box (P,P): # Use an UTM to make this call
    while True:
        pass # Do nothing

return # Just terminate if magic_box(P,P) returns False
```

This does NOT terminate!!

What should this call return?

```
contradiction (contradiction ):

if magic_box (contradiction, contradiction):
    while True:
        pass # Do nothing

return
```

We get into an infinite loop here!

Case 2:

`contradiction (contradiction)` does not terminate

```
def contradiction ( P ): # This function takes a program as an input

#Run magic_box on (P,P)
if magic_box (P,P): # Use an UTM to make this call
    while True:
        pass # Do nothing

return # Just terminate if magic_box(P,P) returns False
```

This DOES terminate!!

What should this call return?

```
contradiction (contradiction ):

if magic_box (contradiction, contradiction):
    while True:
        pass # Do nothing

return
```

We return here!

Questions?

Let's recap: Argue `magic_box` doesn't exist

1. Assume `magic_box` exists

2. Defined a function `contradiction` that uses `magic_box`

3. Looked at the ONLY two possibilities

3.1. `contradiction(contradiction)` terminates

→ `contradiction(contradiction)` does NOT terminate

3.2. `contradiction(contradiction)` does NOT terminate

→ `contradiction(contradiction)` terminates

This clearly is absurd!!

Which specific step in the above "argument" is wrong?

Congrats: You just did your 1st ML&Soc proof!

Questions?