

# **Responsible Data Science**

## **Fairness as Equality of Opportunity**

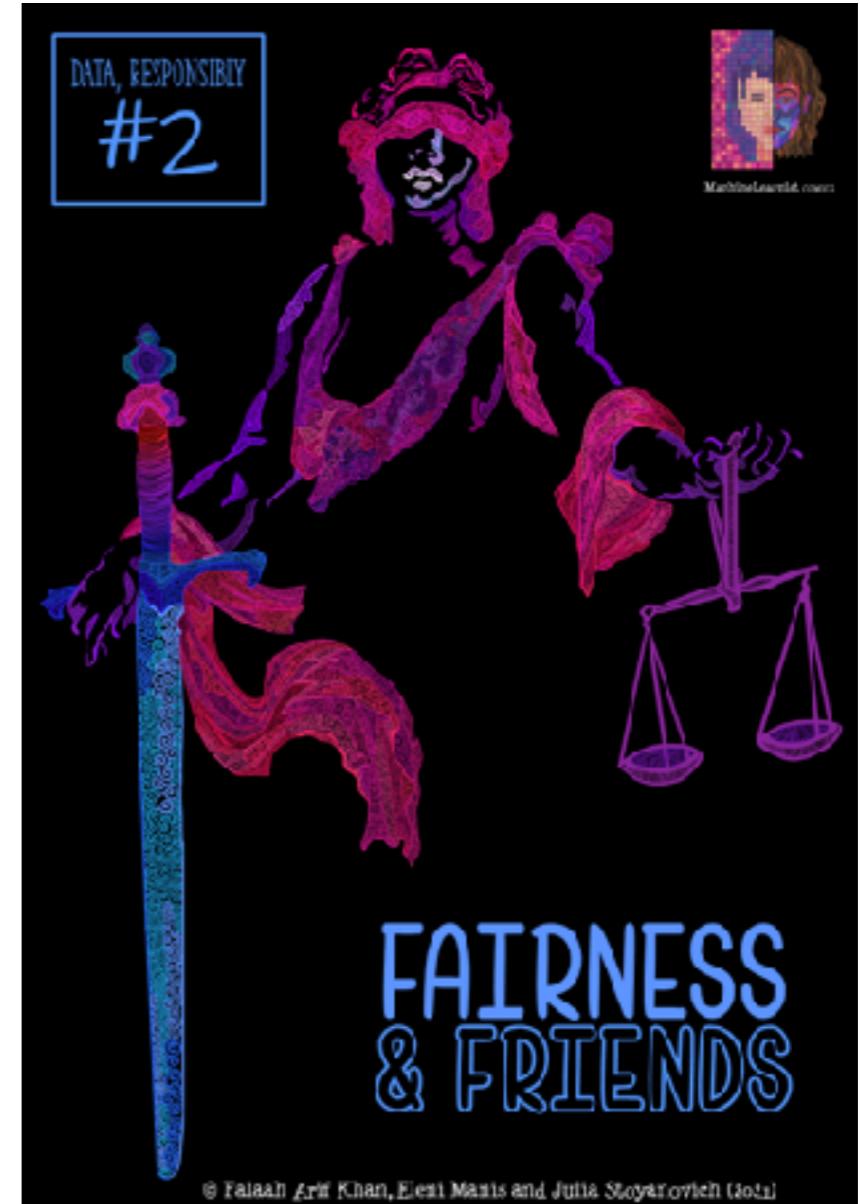
*February 15, 2022*

---

**Prof. Julia Stoyanovich**

Center for Data Science &  
Computer Science and Engineering  
New York University

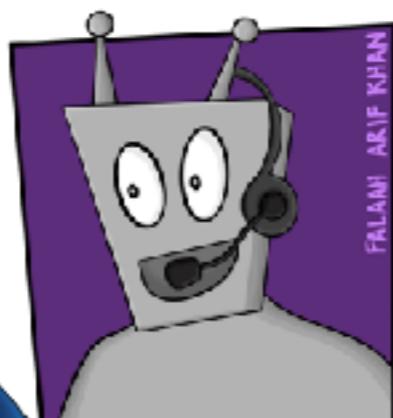
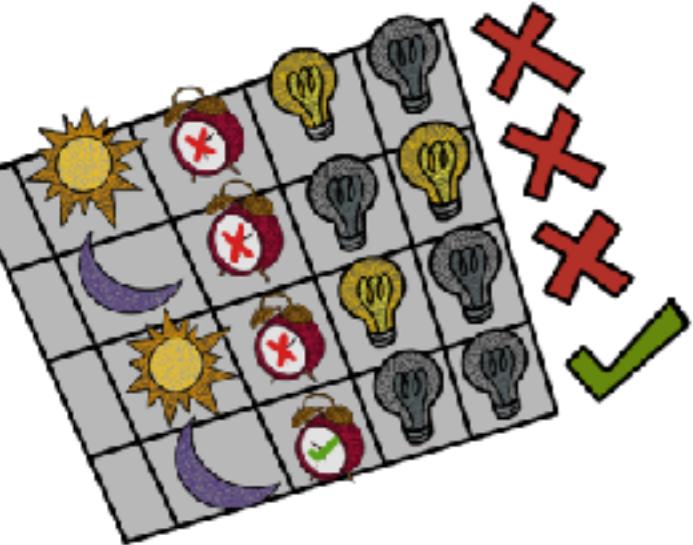
# UPDATE: This week's reading



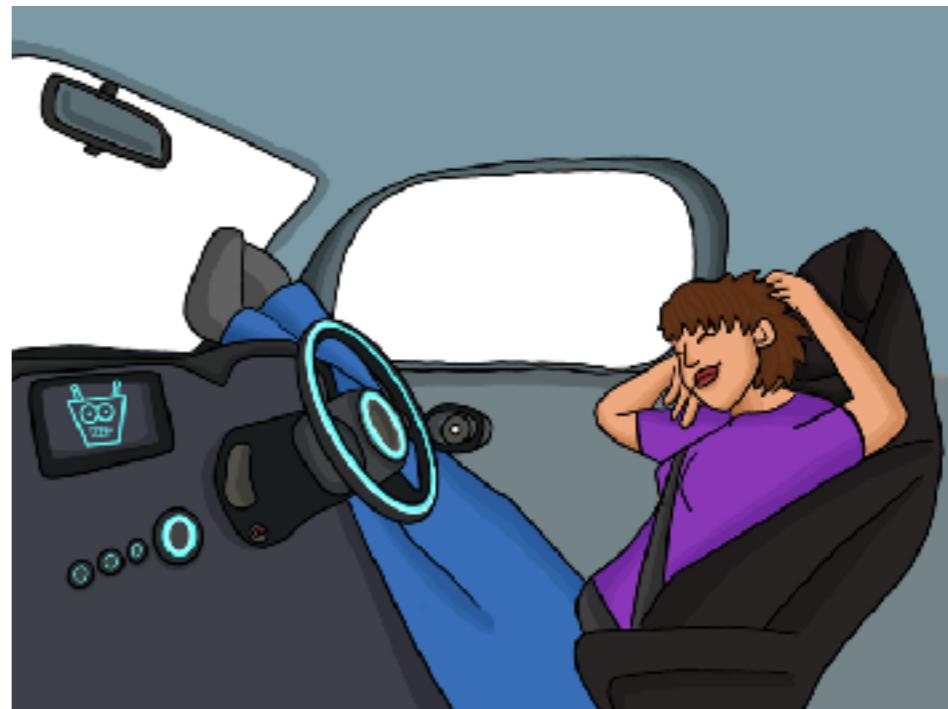
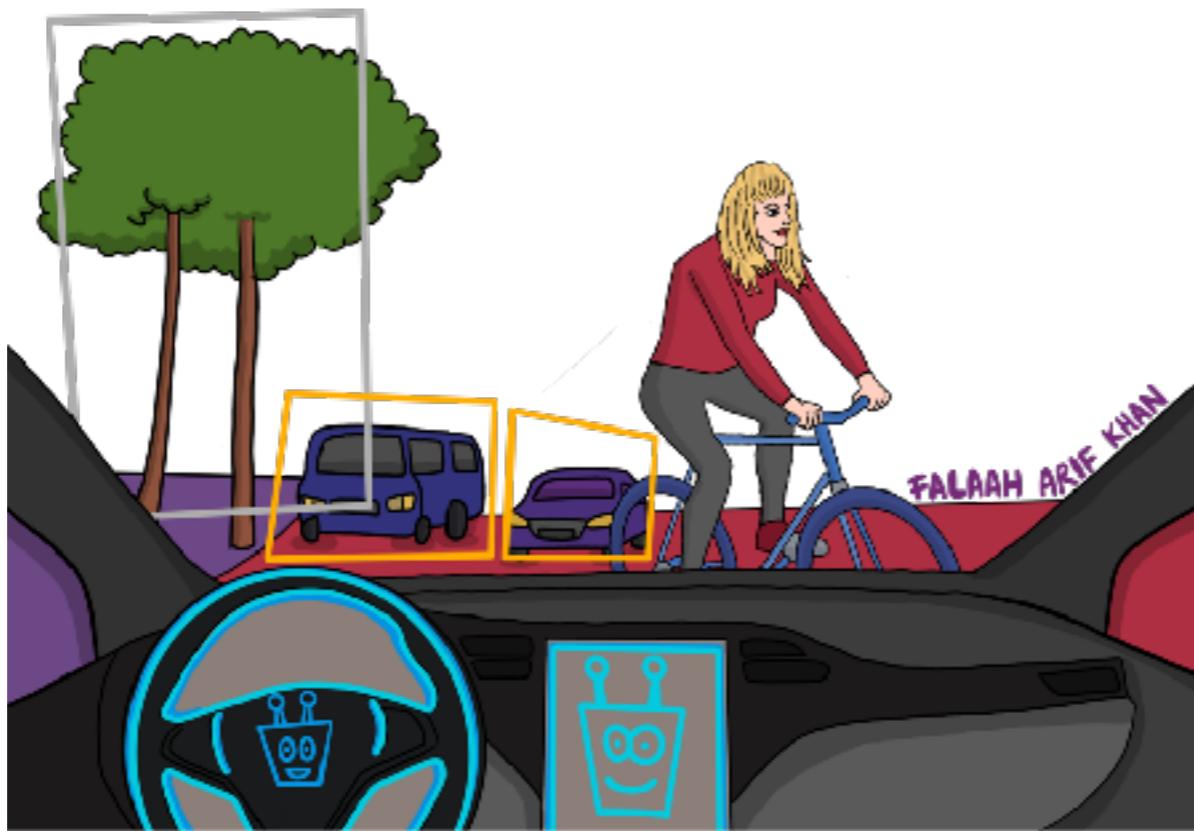
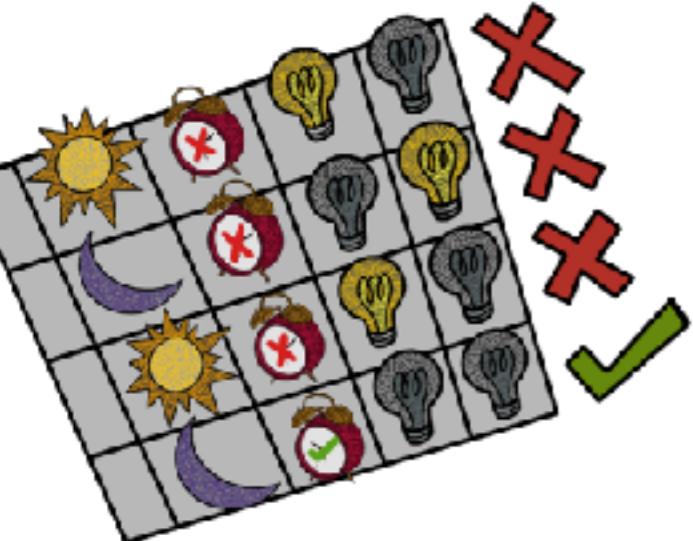


AI ethics teaser

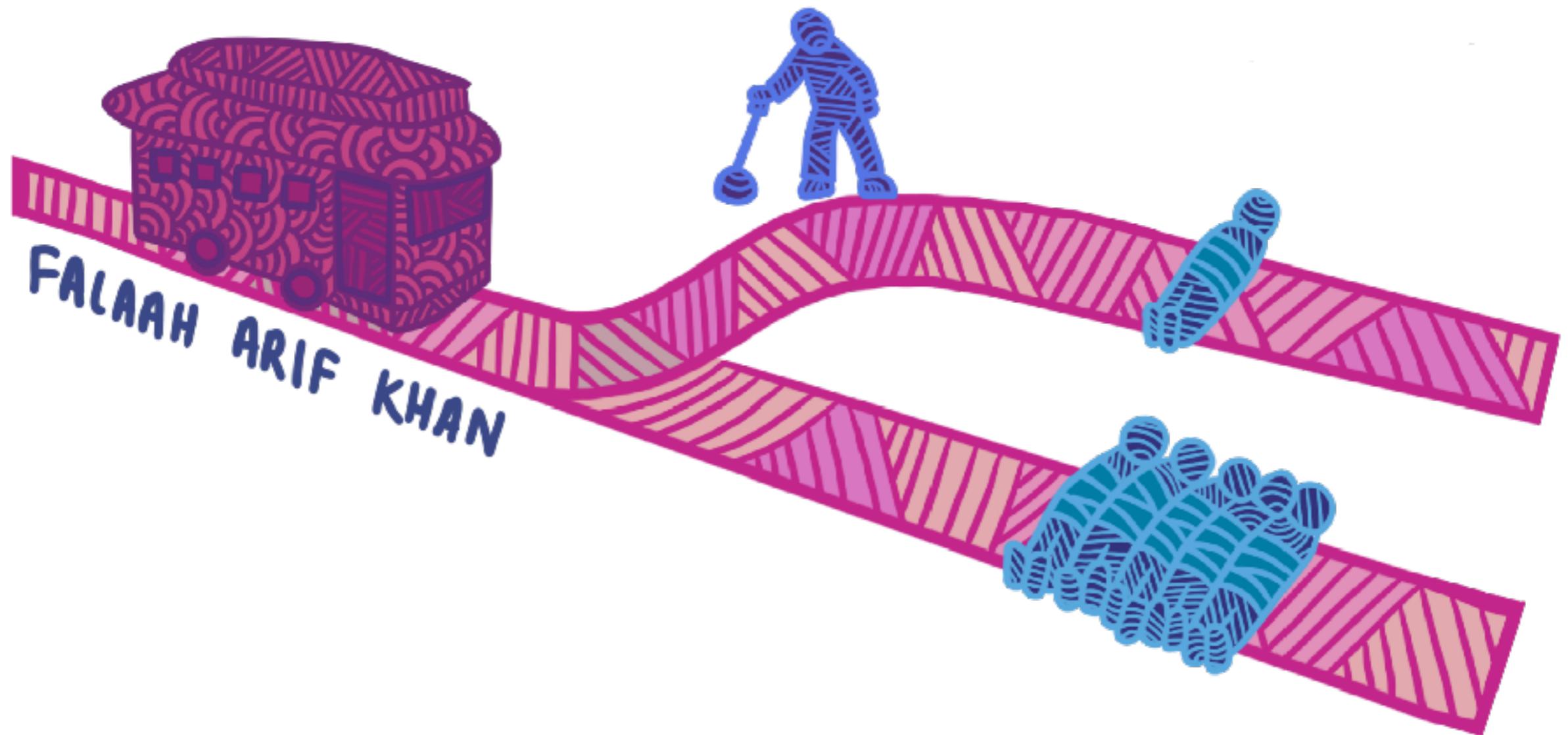
# Mistakes lead to harms



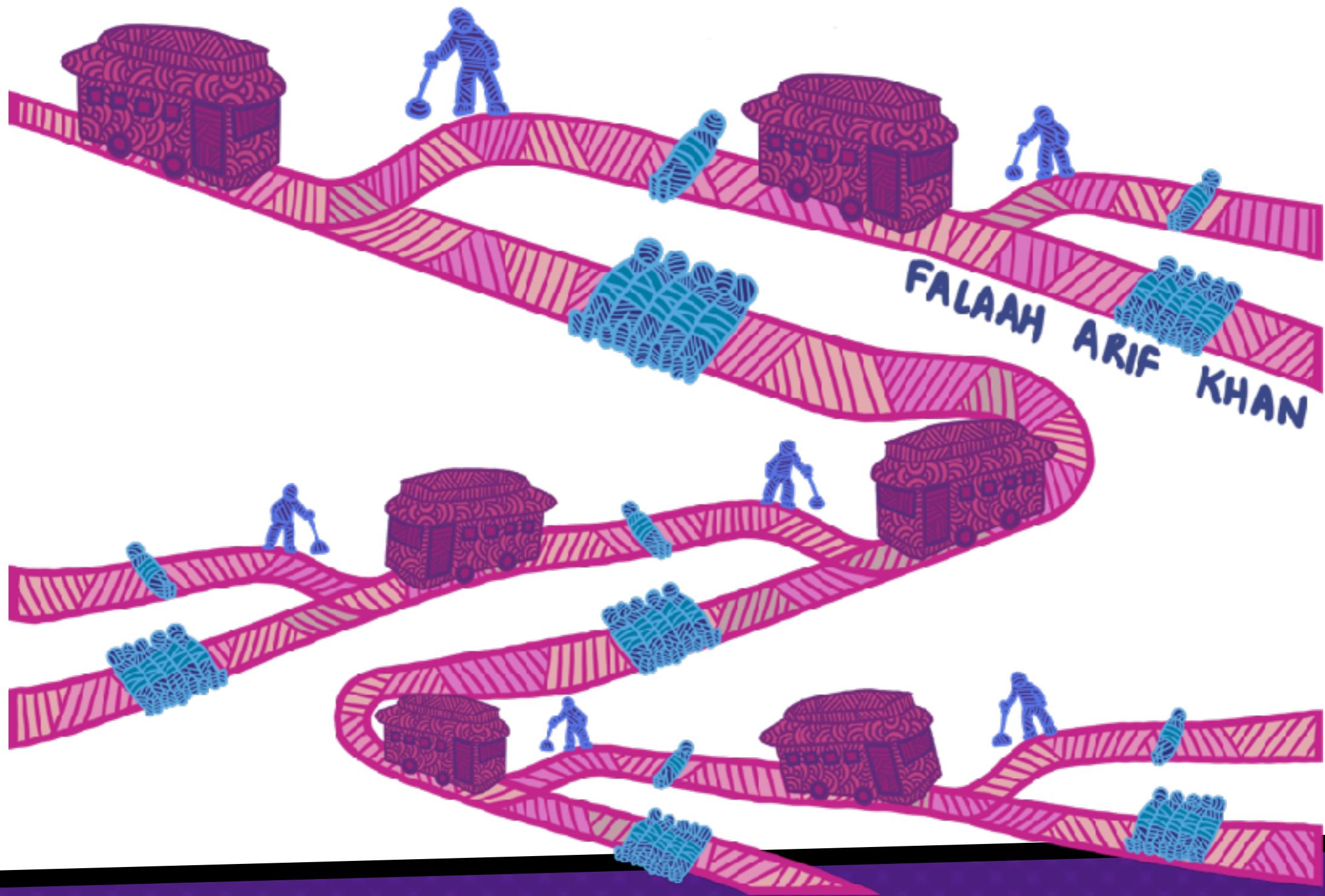
# Mistakes lead to harms



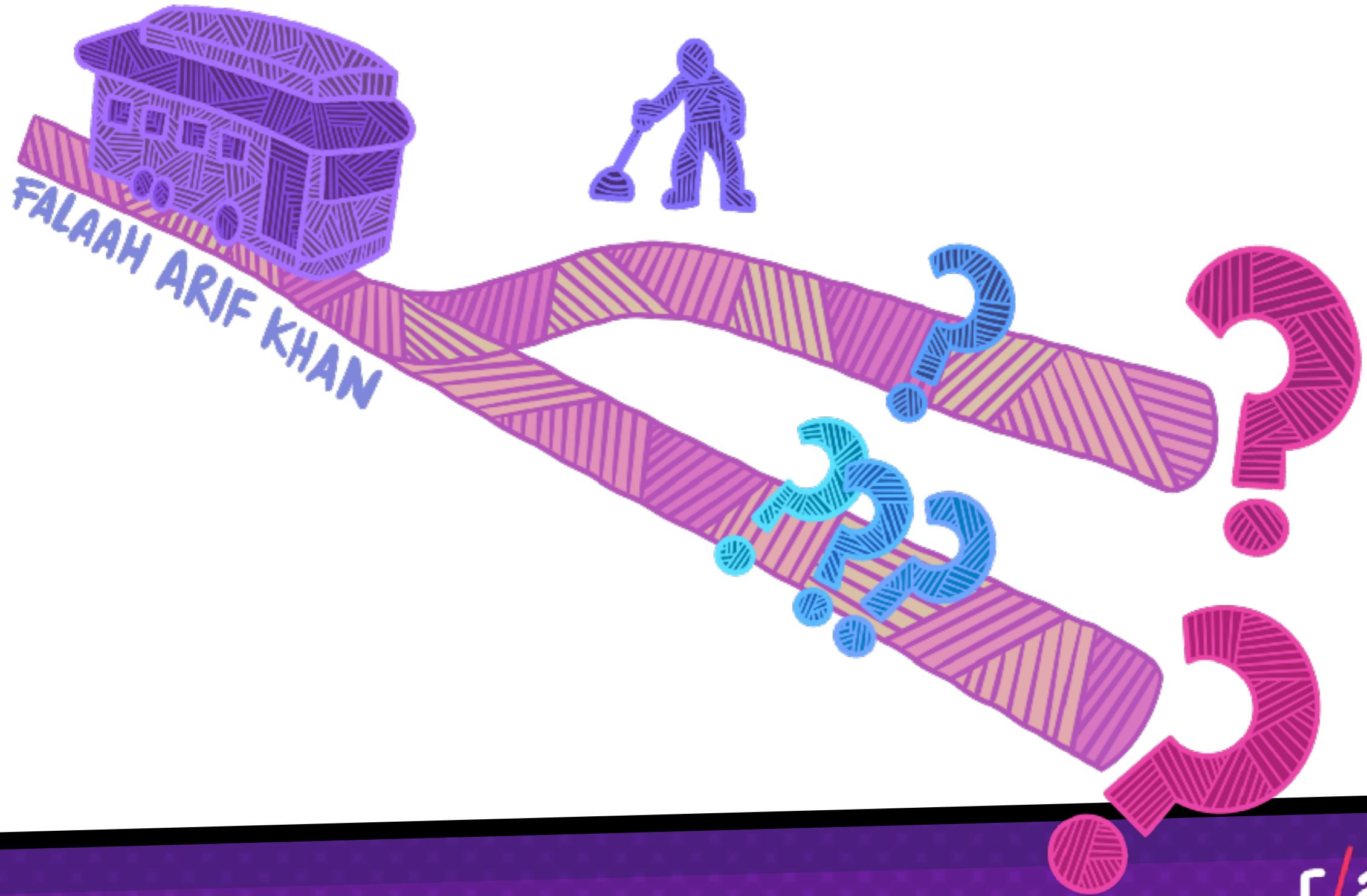
# The trolley problem



# The trolley problem



# Dealing with uncertainty

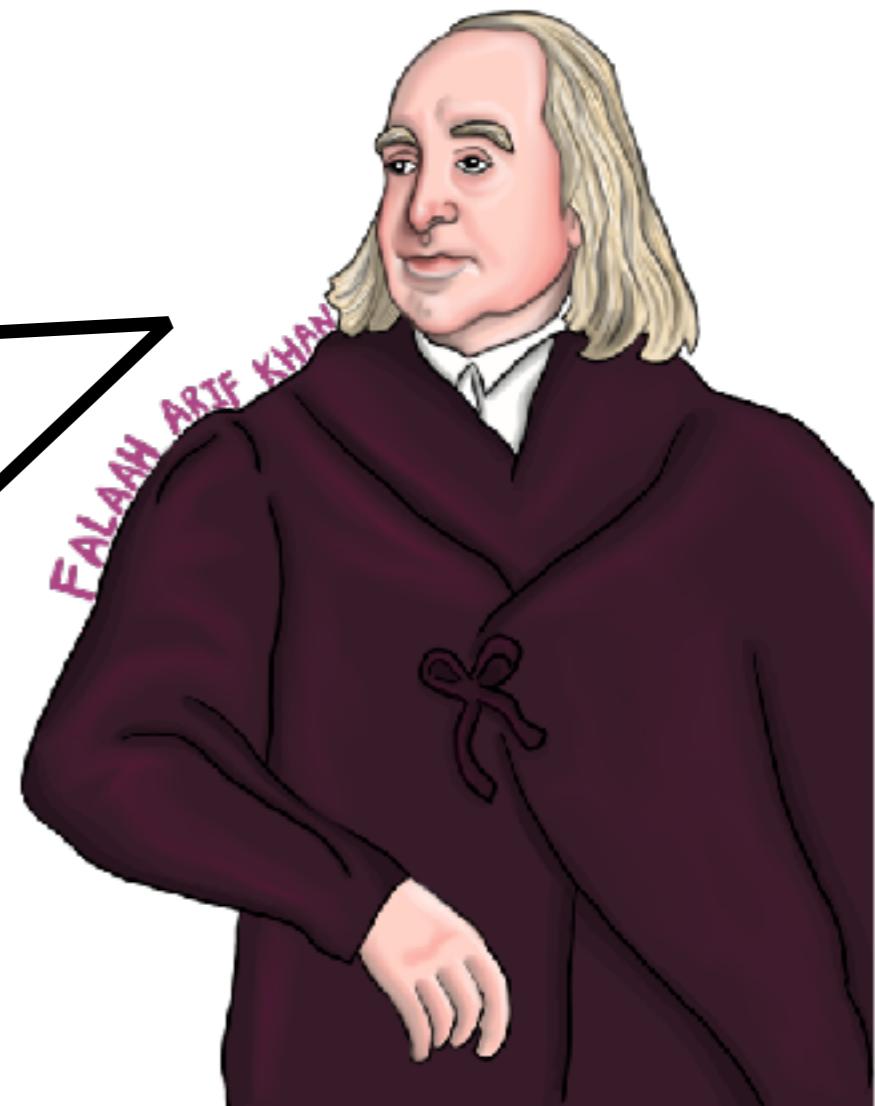


# Utilitarianism



"It is the greatest happiness  
of the greatest number that  
is the measure of right and  
wrong."

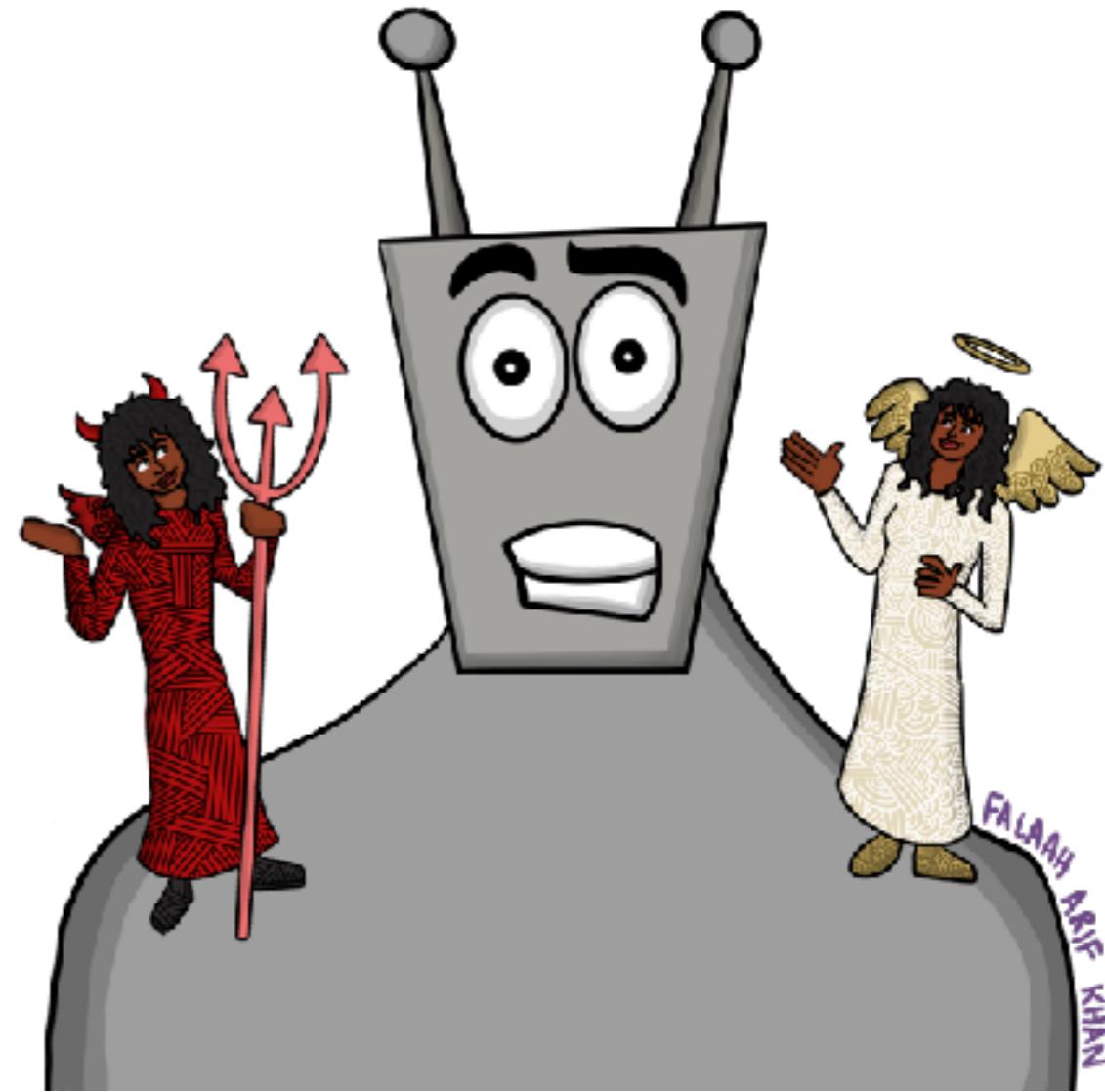
*Jeremy Bentham*



# Algorithmic morality?

## Algorithmic morality

is the act of attributing moral reasoning to algorithmic systems



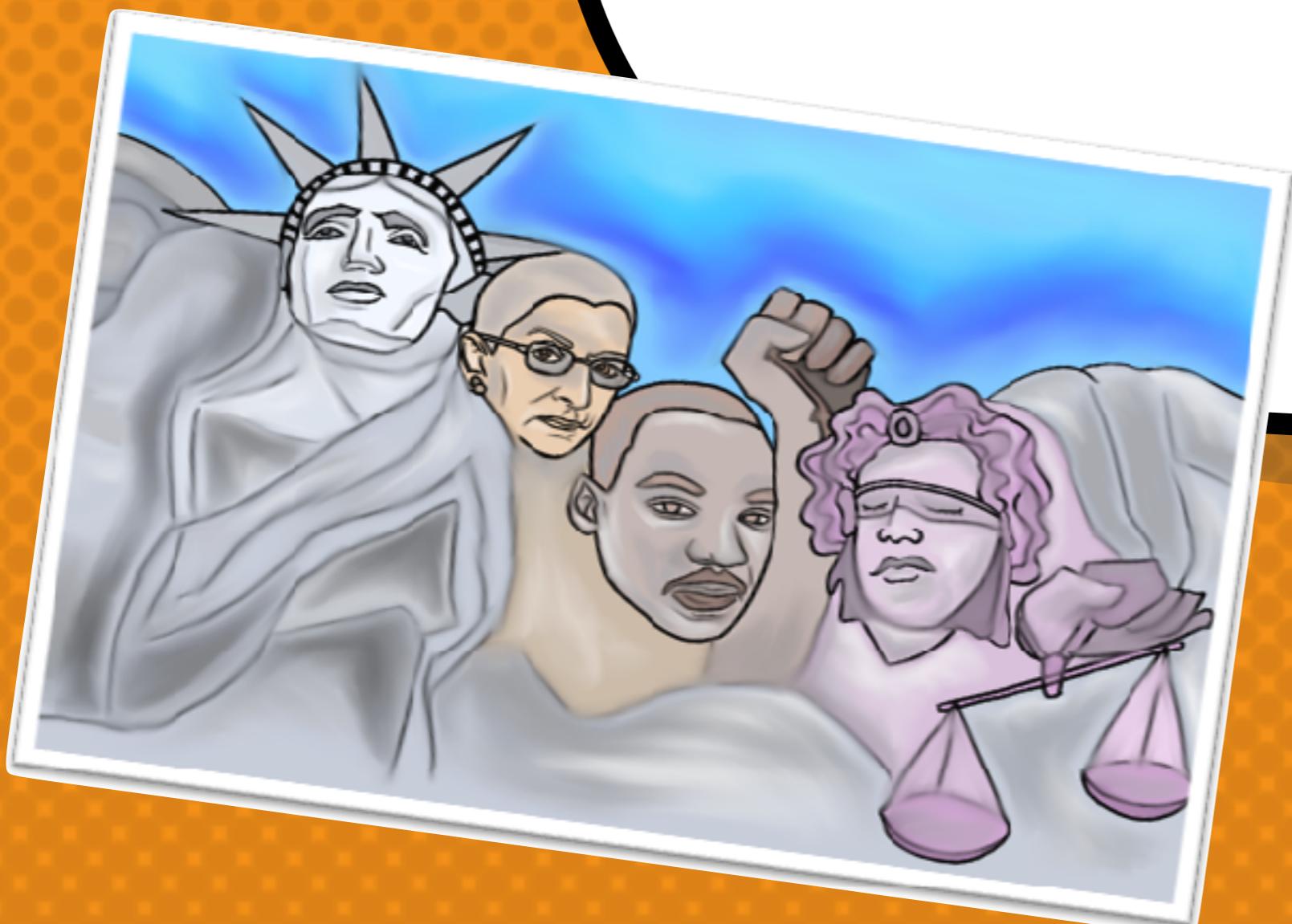
# Algorithmic morality?



# Tech rooted in people



# back to fairness



r/ai

# Fair resource allocation



executive



sous chefs

line chefs

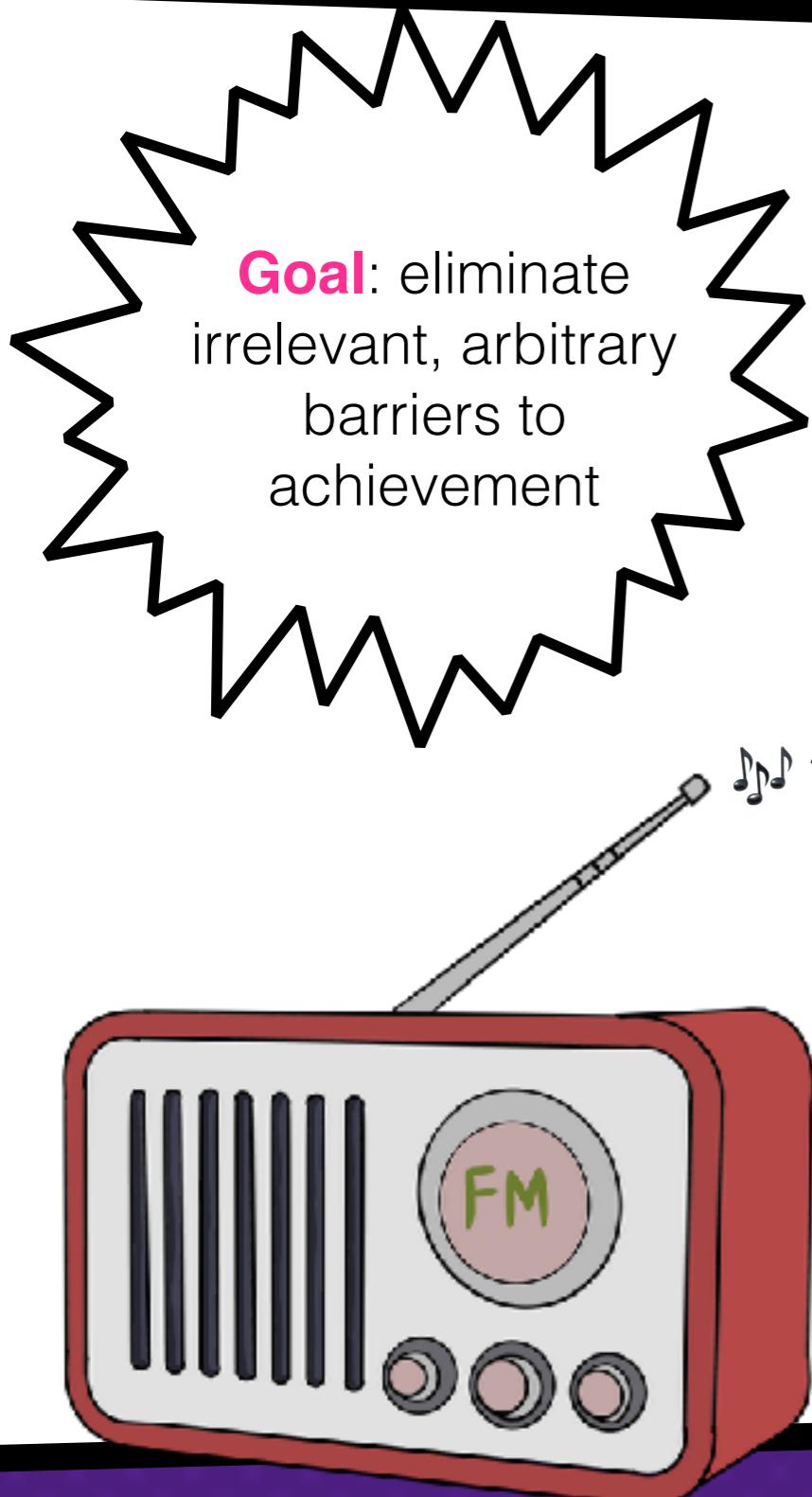
# Fair resource allocation





fairness as  
equality of  
opportunity

# Meet Equality of Opportunity (EO)



♪ Your daddy is rich...  
and your mama's good looking ♪  
...but that won't help you  
in an EO world



# Group fairness as EO



## Group fairness

- Protected group membership is irrelevant to correct or positive classification

## Equality of Opportunity / Substantive

- Irrelevant characteristics (such as group membership) don't affect outcomes



# Individual fairness as EO



## Individual fairness

- Similar treatment of similar individuals
- Only irrelevant characteristics separate similar people

## Equality of Opportunity / Formal

- Irrelevant characteristics don't lead to different treatment of similar people



# The EO Empire

Libertarians now live outside the EO empire



Formal-ville

Rawlsian

luck egalitarian

# Formal EO: Careers open to talents



- In any contest, applicants should only be judged by job-relevant qualifications
- “See nothing irrelevant, speak nothing irrelevant, hear nothing irrelevant”
- Codified as “**fairness through blindness**” with its known weaknesses

# Formal+ EO: Test validity

- A test that systematically under / over estimates people in a way that tracks group membership violates formal EO
- Measures of accuracy or test validity should be broken out by demographic group
- **"Equal opportunity"** [Hardt et al. 2016] codifies formal+ EO



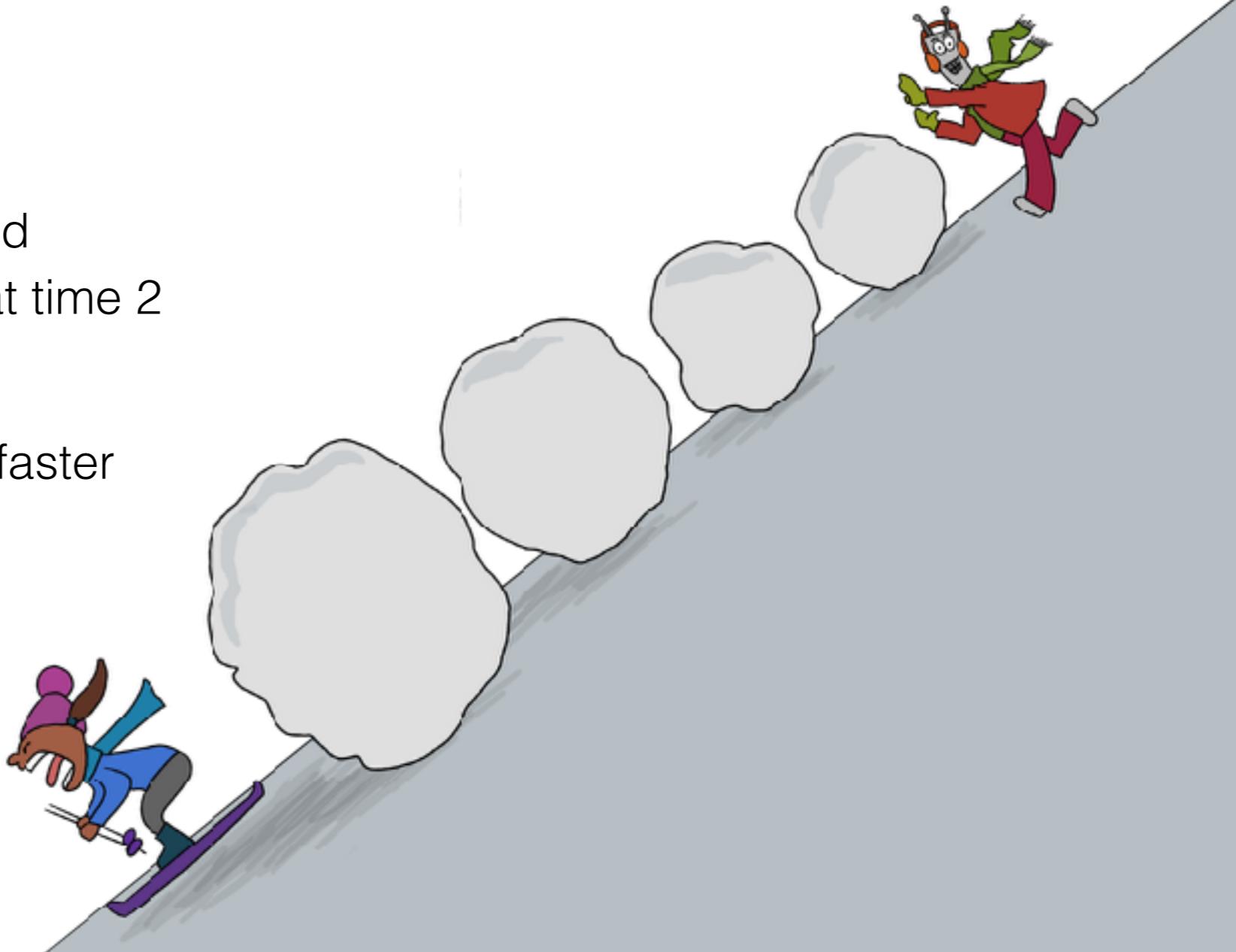
# Formal EO's “before” problem

- Formal EO's appeal: relevant skills in, irrelevant characteristics out
- But OK to use irrelevant privileges before competition
- So privileges affect competition outcomes



# Formal EO's “after” problem

- Winners at time 1 gain improved characteristics for competing at time 2
- Winners win faster, losers lose faster



# “Before” + “after” → discrimination laundering

- Real world discrimination against some leads to privileges for others
- According to formal EO, it's OK to convert privileges to qualifications
- Winning on the basis of qualifications leads to more winning on qualifications
- Discrimination recedes from view...

*“Racial discrimination in on-the-job training is illegal; discrimination on the basis of differences in human capital due to differences in on-the-job training is not”*

(Elizabeth Anderson, The Imperative of Integration)





**Q&A  
discussion**

# The EO Empire

Libertarians now live outside the EO village



Formal-ville

Rawlsian

luck egalitarian

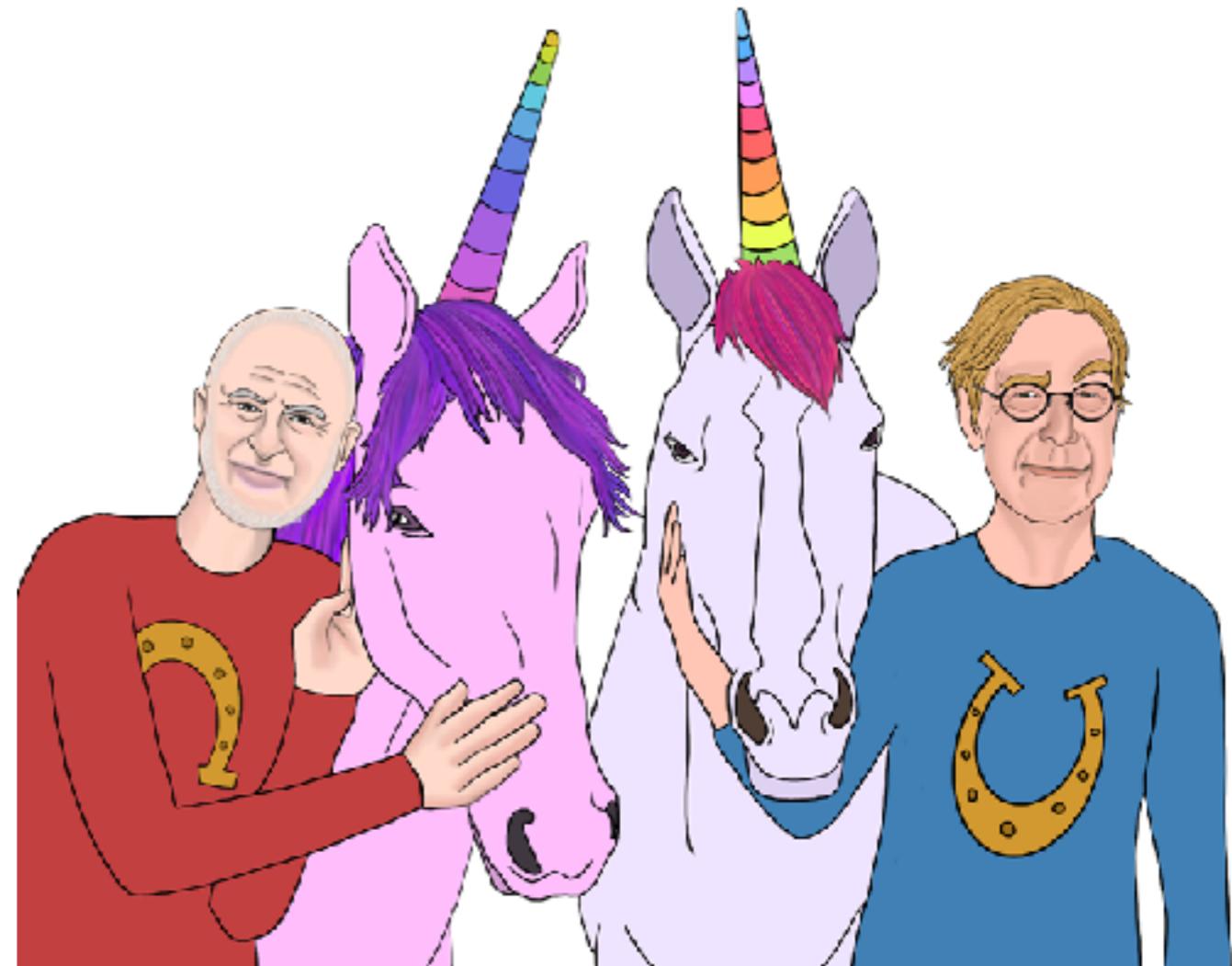
# Substantive EO: Luck egalitarian



The luck egalitarians gather around the communal fire, forsaking all disparities in talent and effort, in favor of unicorns on rainbows!

# Substantive EO: Luck egalitarian

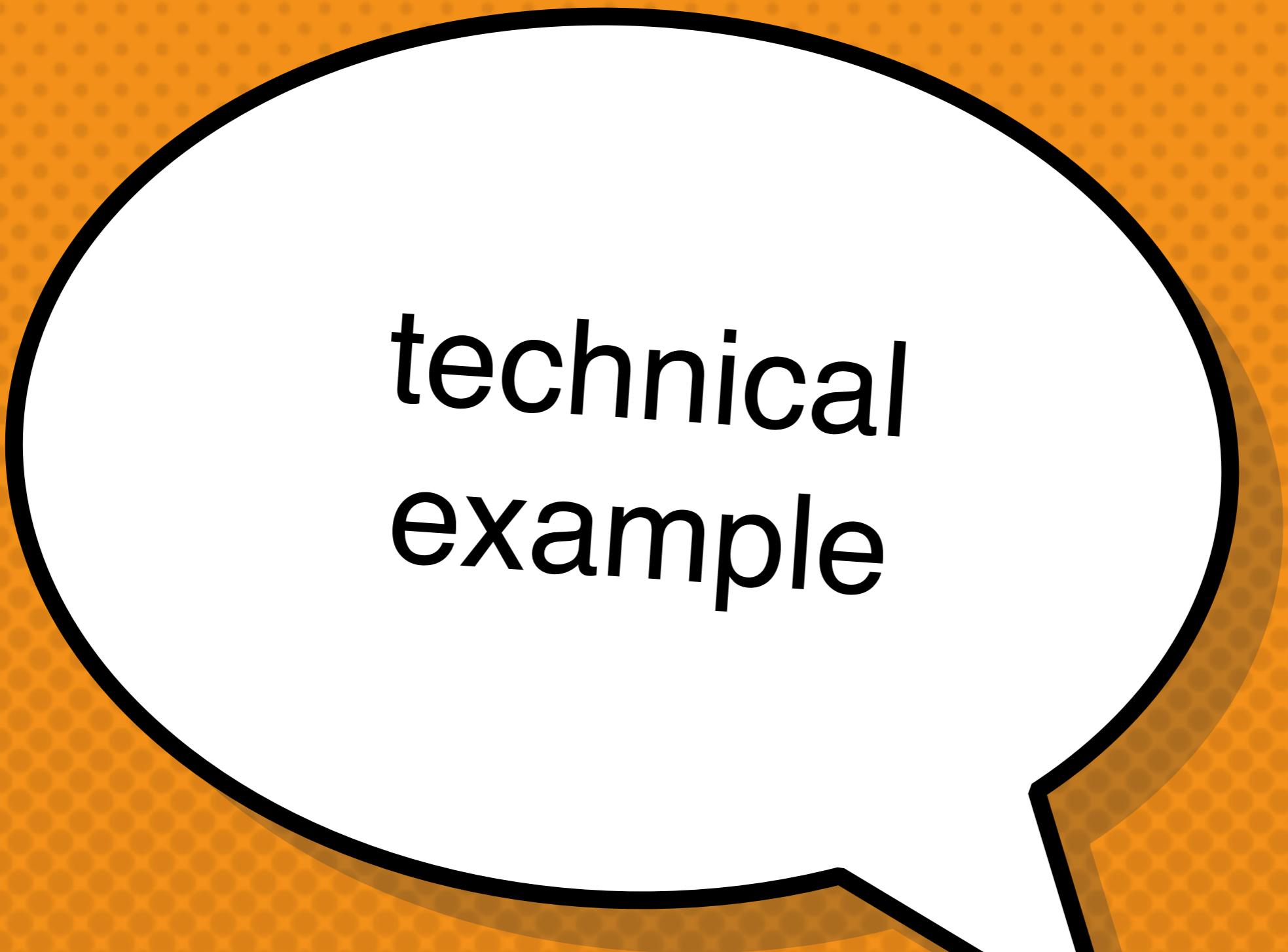
- Outcomes should only be affected by “choice luck” (one’s responsible choices), not by “brute luck”
- But how do we make this separation?



# Substantive EO: Luck egalitarian: Roemer

- No split between responsible effort and irrelevant circumstance
- But there is still an apples and oranges problem





technical  
example

# Diverse balanced ranking

## Goals

**diversity**: pick  $k = 4$  candidates, including 2 of each gender, and at least one per race

**utility**: maximize the total score of selected candidates

	Male		Female	
White	A (99)	B (98)	C (96)	D (95)
Black	E (91)	F (91)	G (90)	H (89)
Asian	I (87)	J (87)	K (86)	L (83)

score = 373

score = 372



## Problem

picked the best White and male candidates (A, B) but did not pick the best Black (E, F), Asian (I, J), or female (C, D) candidates

## Beliefs

scores are more informative within a group than across groups - **effort is relative to circumstance**

it is important to **reward effort**

# From beliefs to interventions

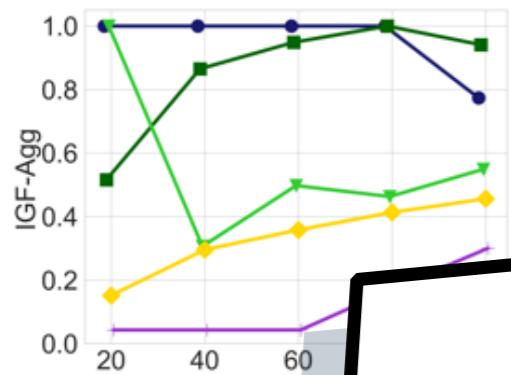
Fairness for female candidates       $83 / 95 = 0.91$

C	D	G	H	K	L
95	95	90	86	83	83

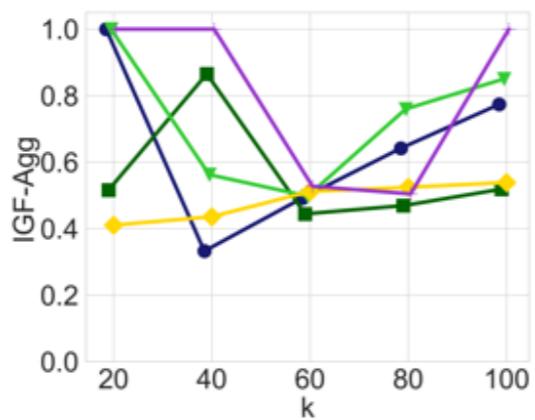
highest-scoring  
skipped

lowest-scoring  
selected

BEFORE: diversity constraints only



AFTER: diversity and fairness constraints



## Beliefs

scores are more informative within a group than across groups -  
**effort is relative to circumstance**

it is important to **reward effort**

# Substantive EO: Rawlsian



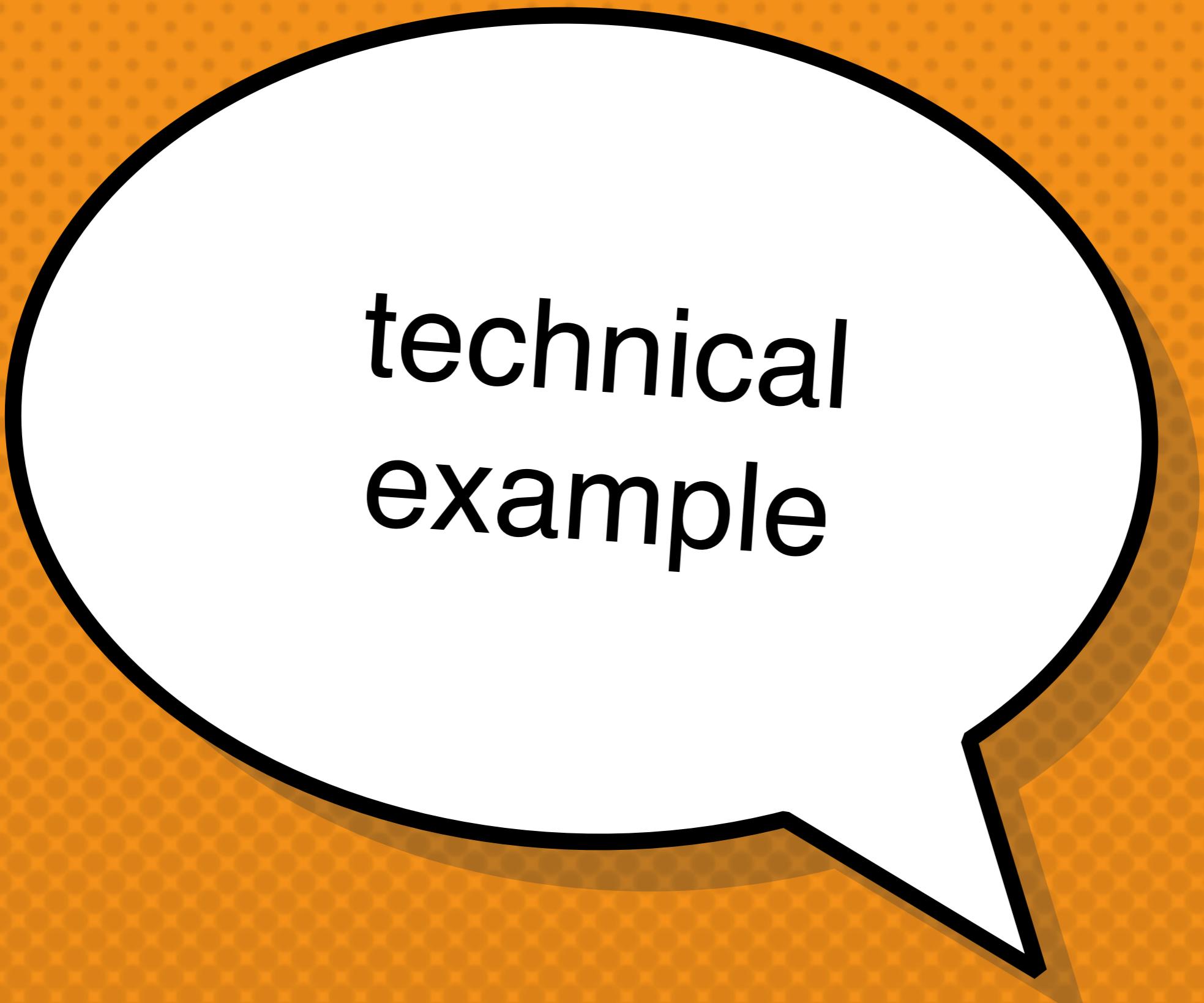
- Equally talented babies must have equal life prospects
- Emphasis is on equality of **developmental opportunities**
- All people - rich or poor - must have the same opportunities to develop their qualifications, so that at the point of competition they are equally likely to succeed

# Misconceptions of Rawls in Fair-ML



- In fair-ML, statistical parity and equality of odds are believed to operationalize Rawlsian fair EO. But this is not so!
- Rawlsian EO is fundamentally about providing developmental opportunities **before** competitions





technical  
example

# Intersectional causal fairness

	gender	race	X	Y
B	m	w	6	12
C	m	<b>b</b>	5	9
D	<b>f</b>	w	6	8
E	m	w	4	7
F	<b>f</b>	<b>b</b>	3	6
K	<b>f</b>	a	5	5
L	m	<b>b</b>	1	3
O	<b>f</b>	w	1	1

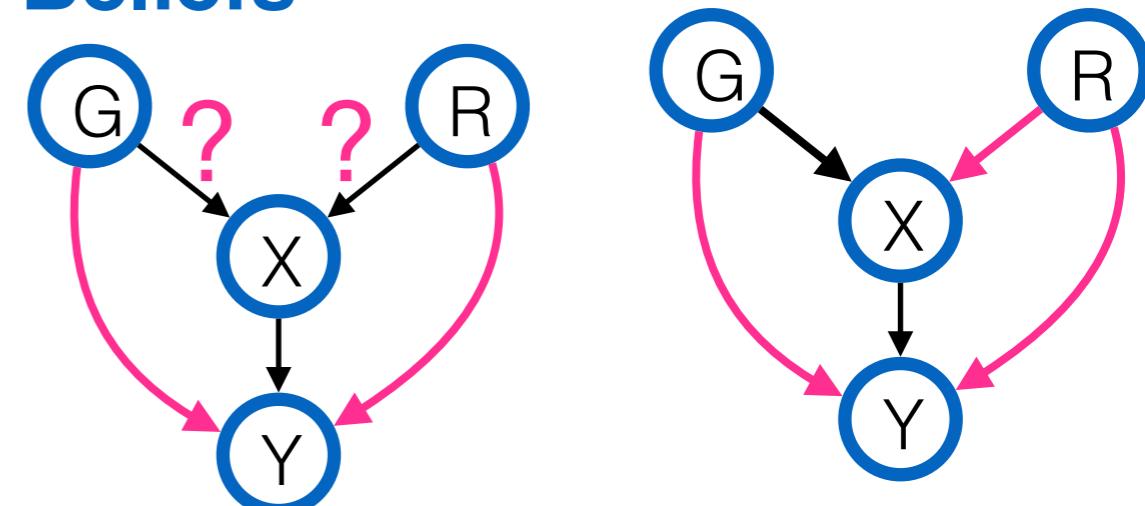
## Problem

weight lifting ability is mapping to qualification score differently depending on gender

## Goal

hire  $k = 4$  best-qualified candidates at a moving company

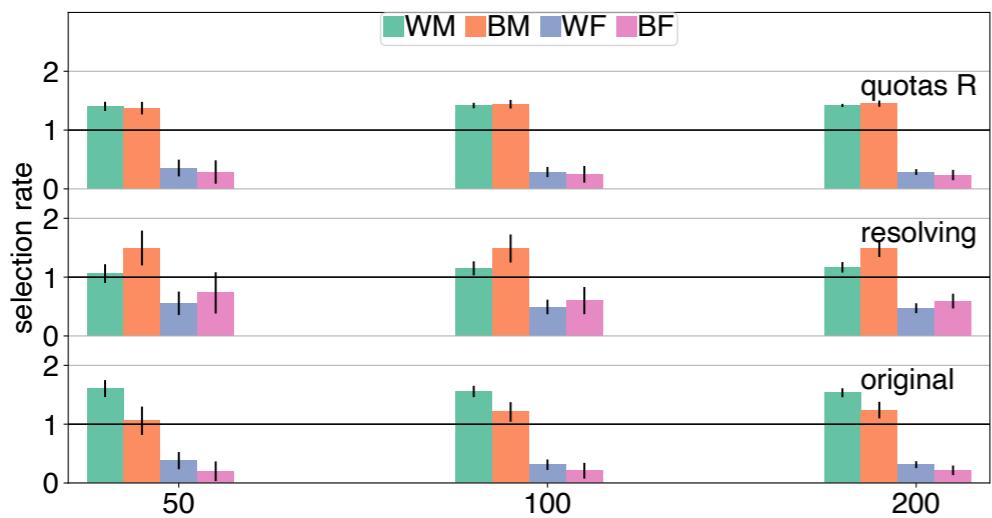
## Beliefs



# From beliefs to interventions

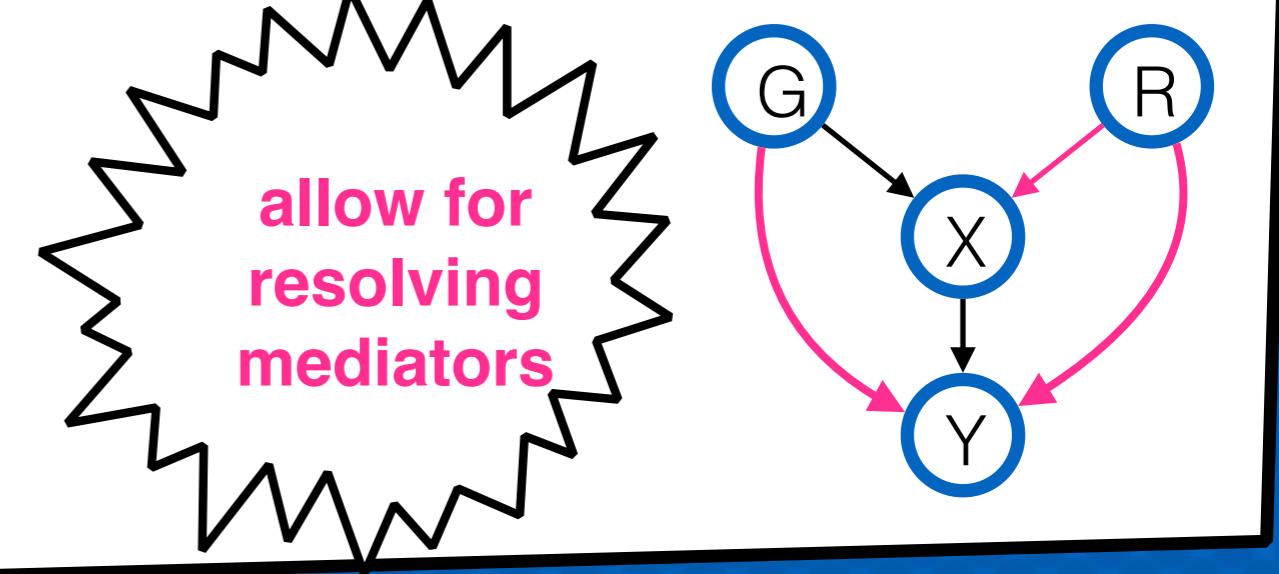
**Idea:** Compute **counterfactual scores**, treating each individual in the sample as though they had belonged to one intersectional group (e.g., Black women). Rank on those scores.

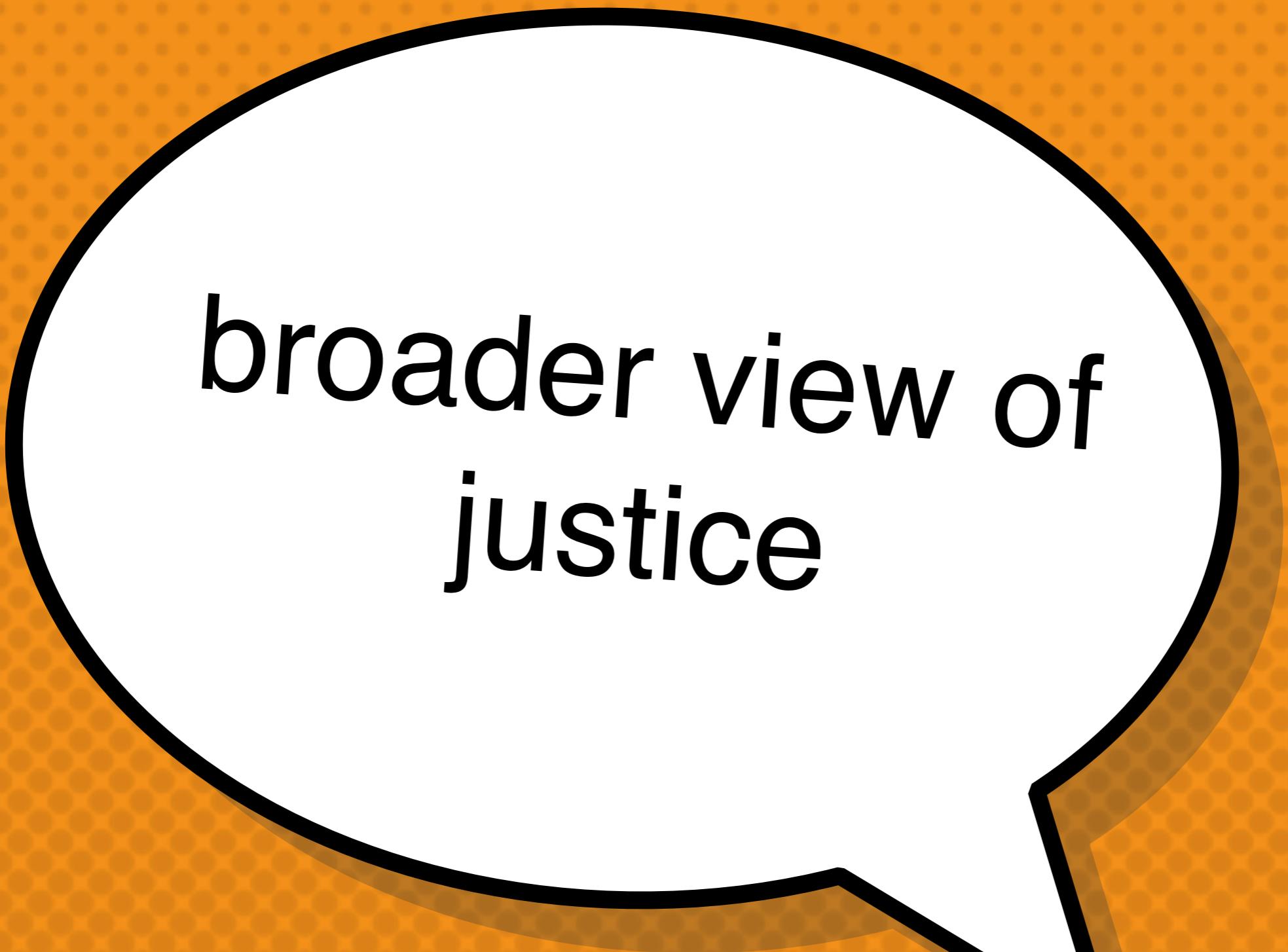
This process produces a **counterfactually fair ranking**.



## Beliefs

allow for  
resolving  
mediators





broader view of  
justice

# EO over a lifetime



- Disadvantages may preclude people from getting a ‘fair’ shot in a competition
- This idea of a Sisyphean struggle for certain candidates motivates us to take a broader look at EO

# Rawls' “natural lottery”

**Natural lottery:** Talents and fortune are distributed arbitrarily.

**Difference principle:** Since we don't deserve our starting points in life, we must work towards a social system that serves everyone.

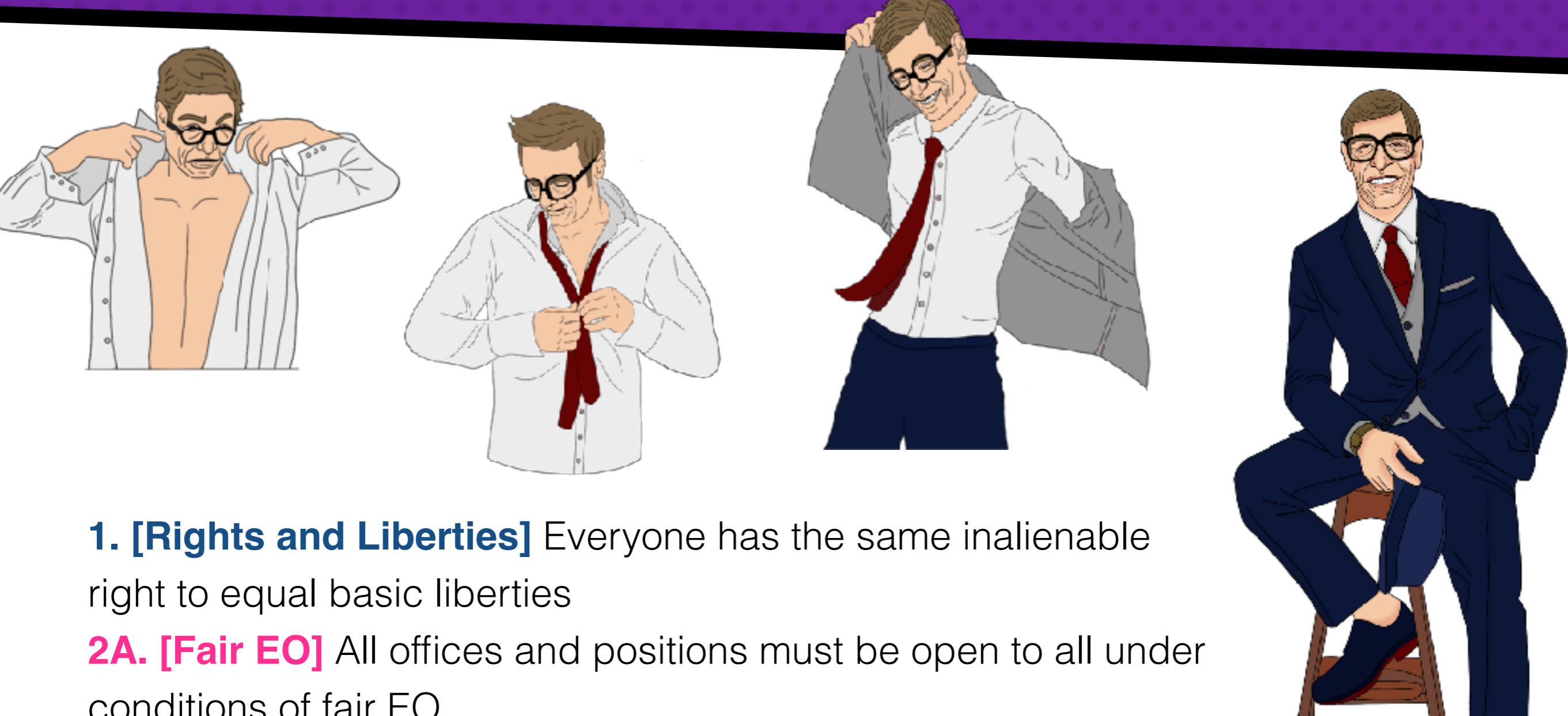


# Rawls' “original position”

**The Veil of Ignorance:** If citizens do not know their race, class, sex, social position (or any other characteristics that might cause them to favor people like themselves), they will advocate for all social positions and their attached privileges to be distributed fairly.



# Broader view of justice



**1. [Rights and Liberties]** Everyone has the same inalienable right to equal basic liberties

**2A. [Fair EO]** All offices and positions must be open to all under conditions of fair EO

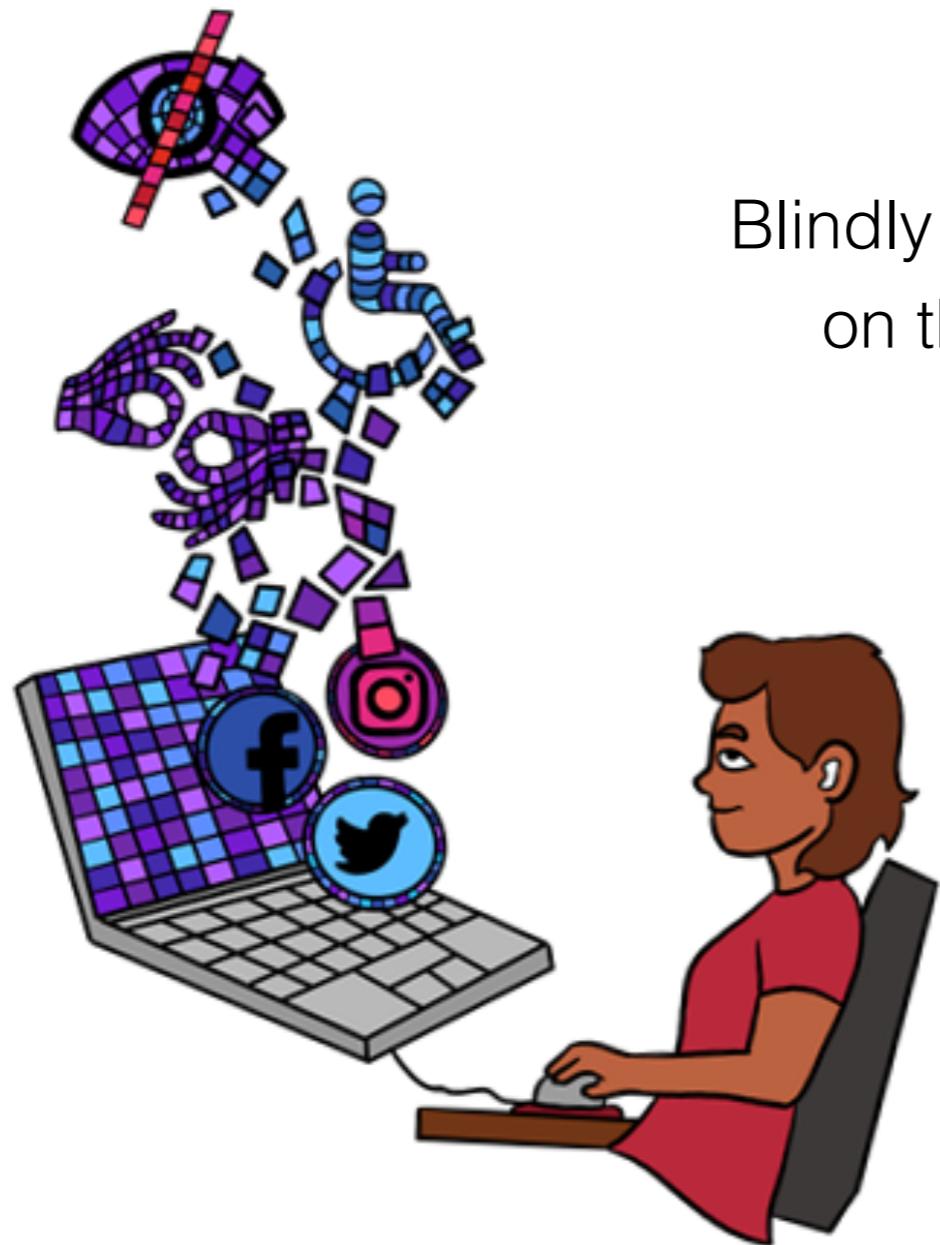
**2B. [Difference Principle]** Social and economic inequalities must be of the greatest benefit to the least advantaged

# Broader view of justice

Blindly satisfying EO may  
infringe on **basic liberties**.



# Broader view of justice



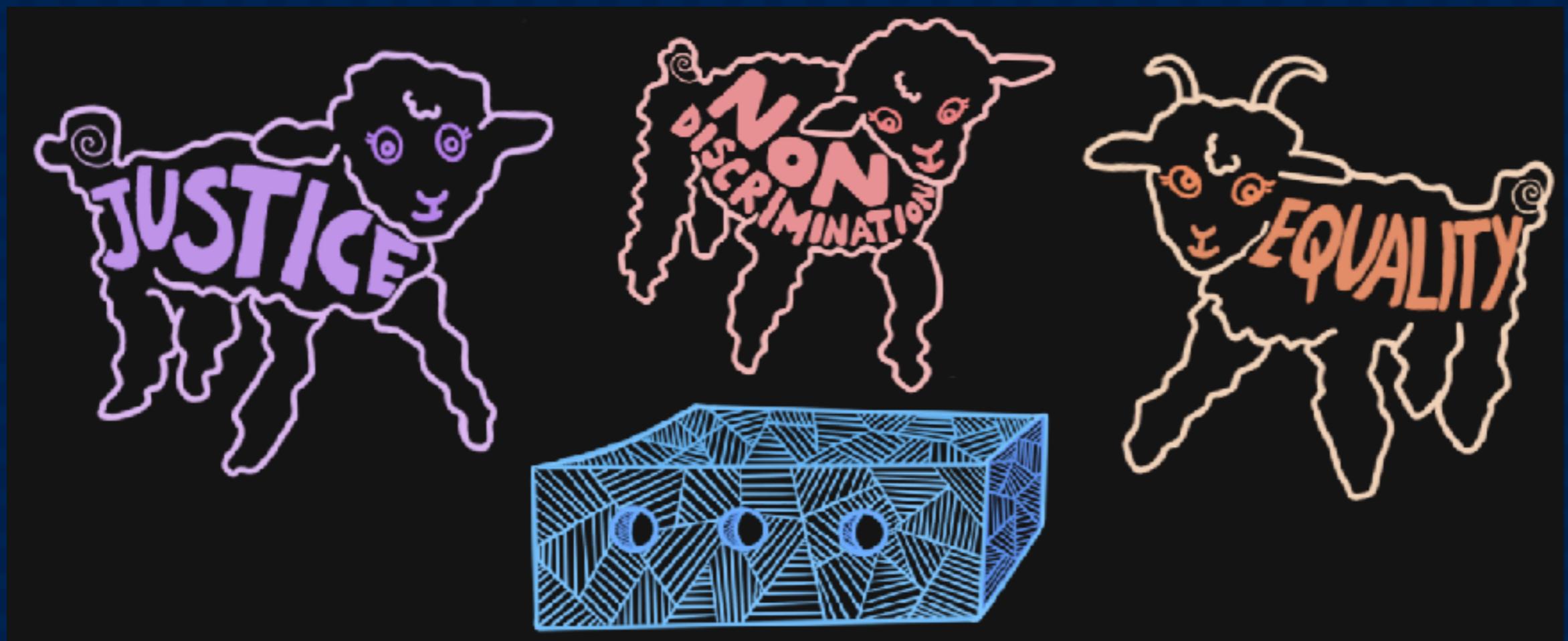
Blindly satisfying EO may infringe  
on the **freedom of speech**.





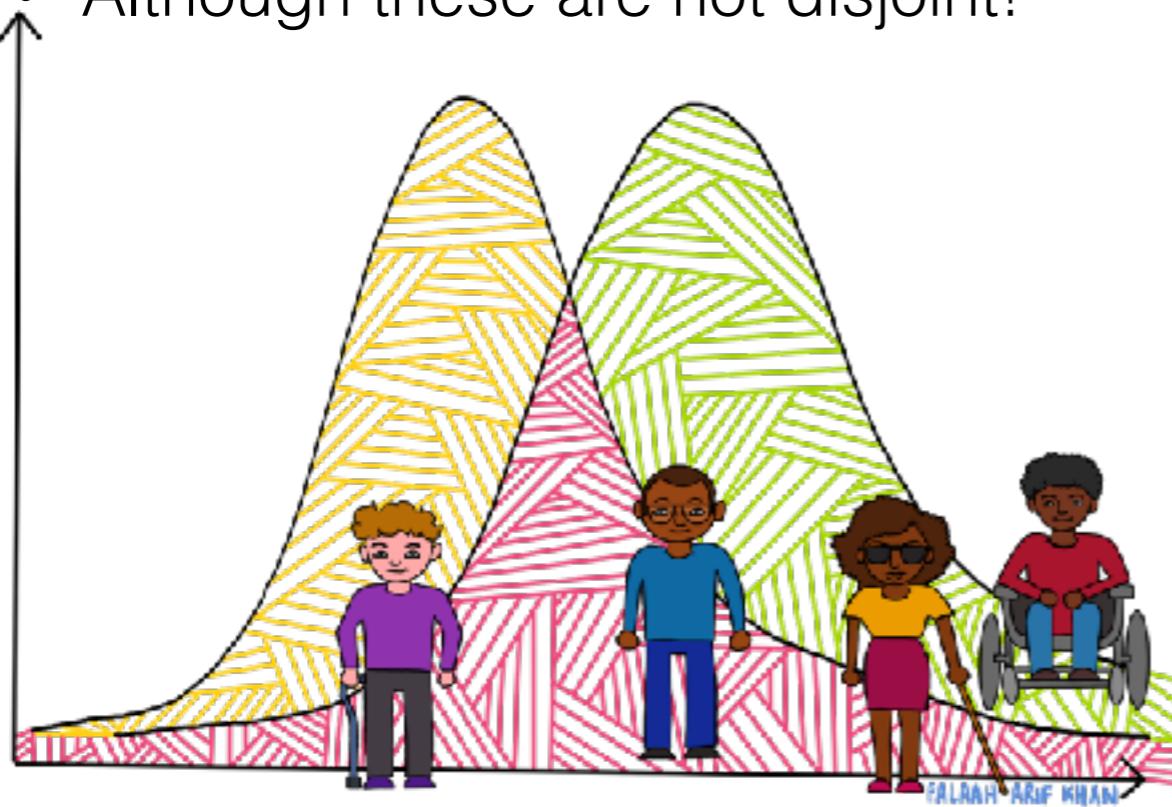
take-aways

Still need interpretability!



# Framing technical interventions

- Normative choices first
- Technical choices later
- Although these are not disjoint!



## Next time

- fairness beyond distributive justice
- the responsible AI lifecycle



**Q&A  
discussion**

# **Responsible Data Science**

## Fairness as Equality of Opportunity

---

**Thank you!**



**NYU**

TANDON SCHOOL  
OF ENGINEERING



**NYU**

Center for  
Data Science

**r/ai**