

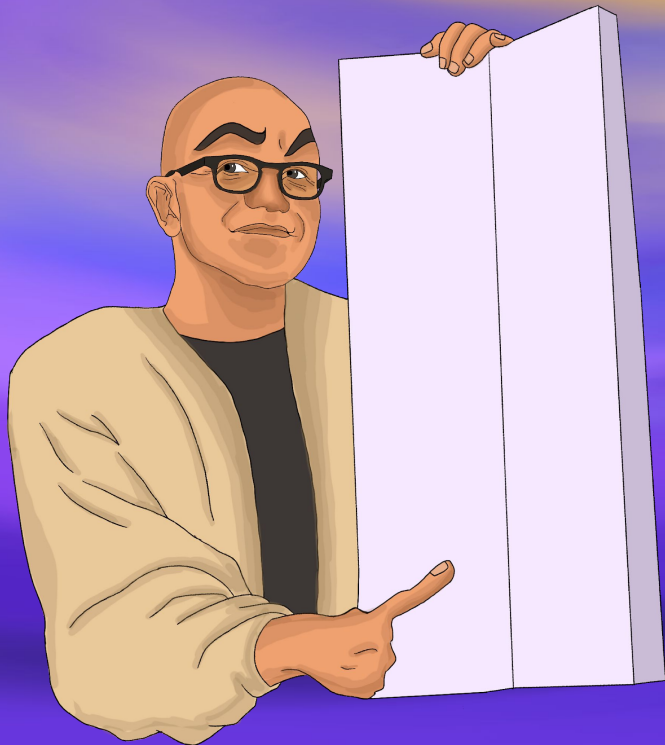


FAIRNESS & FRIENDS

(Falaah Arif Khan, Eleni Manis, and Julia Stoyanovich)

Tenets of Fair-ML

1. Be clear that there is no one correct notion of Fairness, and yet feel free to propose blanket software solutions for all datasets and applications



2. Be clear that ethics research is important insofar as it does not shed any bad light on the company and its products

Tenets of Fair-ML



3. Be clear that ML systems are biased when data is biased. To get an outcome that looks fair, simply train the exact system on de-biased data

4. Be clear that expertise in building unethical AI is a market advantage and can be launched as 'Ethics-As-A-Service'



What is 'Fairness'?



'Fair' Resource Allocation



(EXECUTIVE)

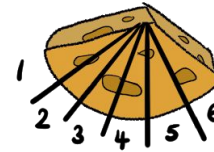


(SOUS CHEFS)



(ROOKIES/LINE CHEFS)

Is it **Equality** in the Distribution
of some commodity/outcome?
(in the economic sense)



Is it some notion of **Distributive Justice**?
(from political philosophy)



Is it **Non-Discrimination**?

(from legal doctrines)



Which 'Fairness' metric is suitable for our particular context?

Is it **Equality in the Distribution** of some commodity/outcome?
(in the economic sense)

Is it some notion of **Distributive Justice**?
(from political philosophy)

Is it **Non-Discrimination**?
(from legal doctrines)



Maybe we can get some guidance from political philosophy!

Meet EOP (Equality of Opportunity)!

- Eliminates irrelevant, arbitrary barriers to achievement
- Irrelevant personal characteristics don't help or hurt access to desirable positions, outcomes



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



EOP as group fairness



Statistical fairness:

- ID protected groups
- Membership irrelevant to correct or positive classification

EOP:

- Irrelevant characteristics also don't affect outcomes

EOP as individual fairness



Individual fairness:

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

EOP:

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

It's the....

Age of Equality Of Opportunity! (EOP)



Age of EOP!

Libertarian





Libertarian

Any holding acquired without cheating, is claimed 'fairly', even if some end up with significantly more than others



Libertarian

“You do
you!”

- Emphasis is on people's rights and liberties
- The libertarian will object to any 'ill-gotten gains' - no cheating or defrauding allowed

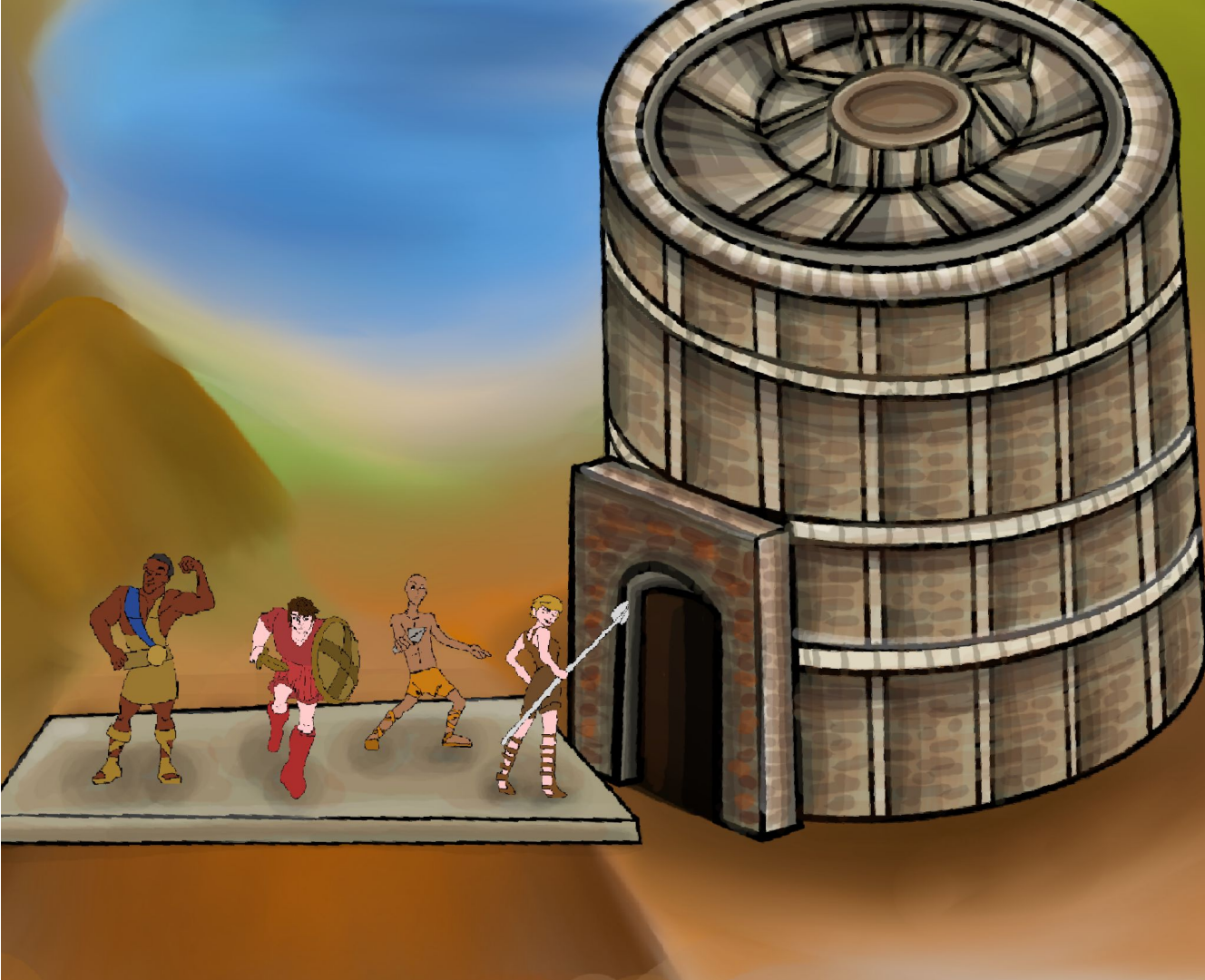
Age of EOP!



Formal EOP

Formal EOP

The theatre of formal EOP is open to all talents. You fight with what you have - no special treatment once you're in.



Formal EOP

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



Fairness through Blindness can pack a punch

- Some characteristics can be excluded
- “Ban the box”: blindness to criminal history during resume screening
- Blindness to credit history



Formal EOP- not just fairness as blindness

A test that systematically under/overestimates people in a way that tracks group membership violates formal EOP

Measures of accuracy or test validity broken out by demographic group



Age of EOP!



Formal EOP

- Strong fairness guarantees
- Fairness through blindness and measures of test accuracy & validity broken out by group
- But formal EOP doesn't satisfy ALL fairness cravings

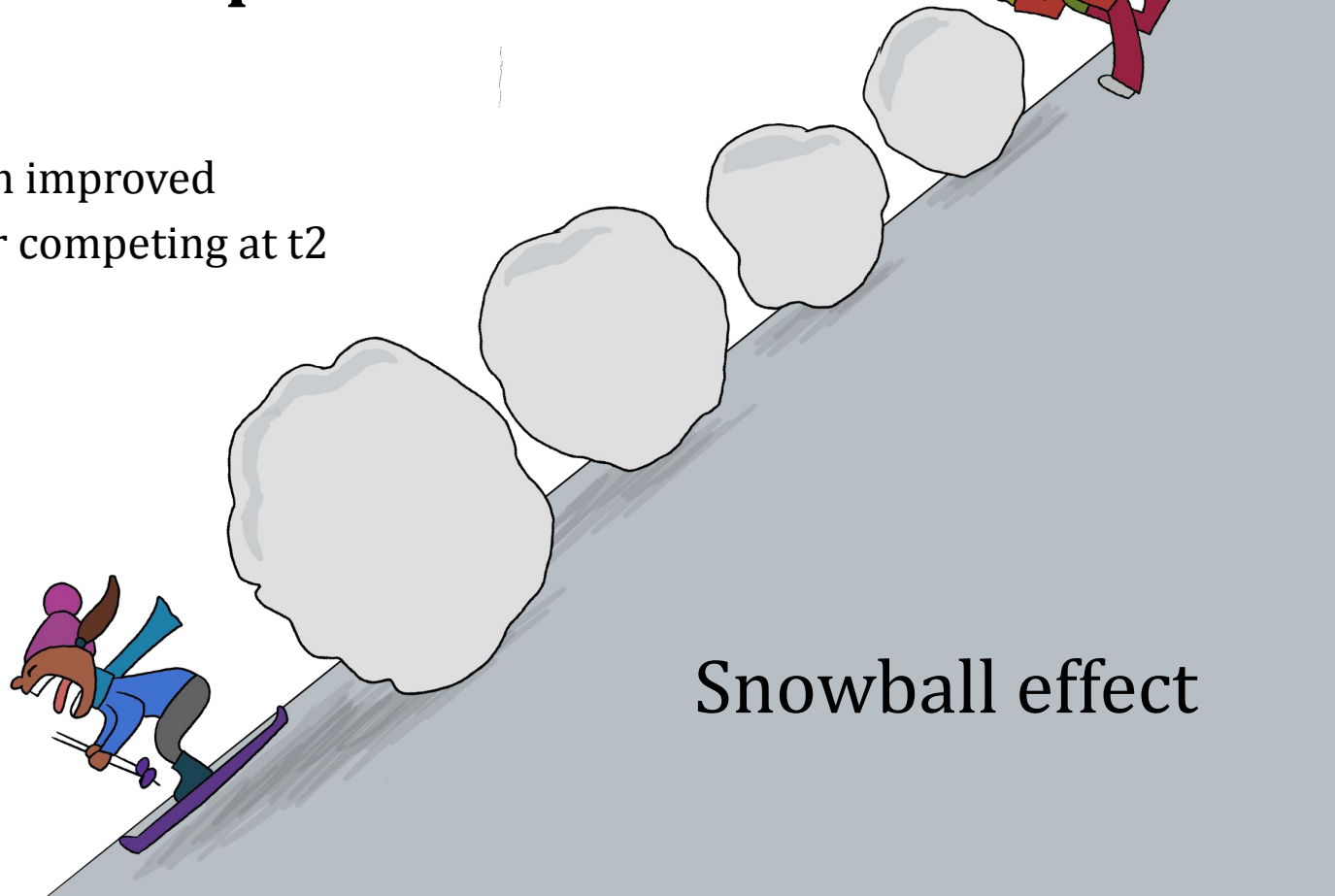
Formal EOP's “before” problem

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



Formal EOP's "after" problem

- Winners at t_1 gain improved characteristics for competing at t_2
- Losers lose faster



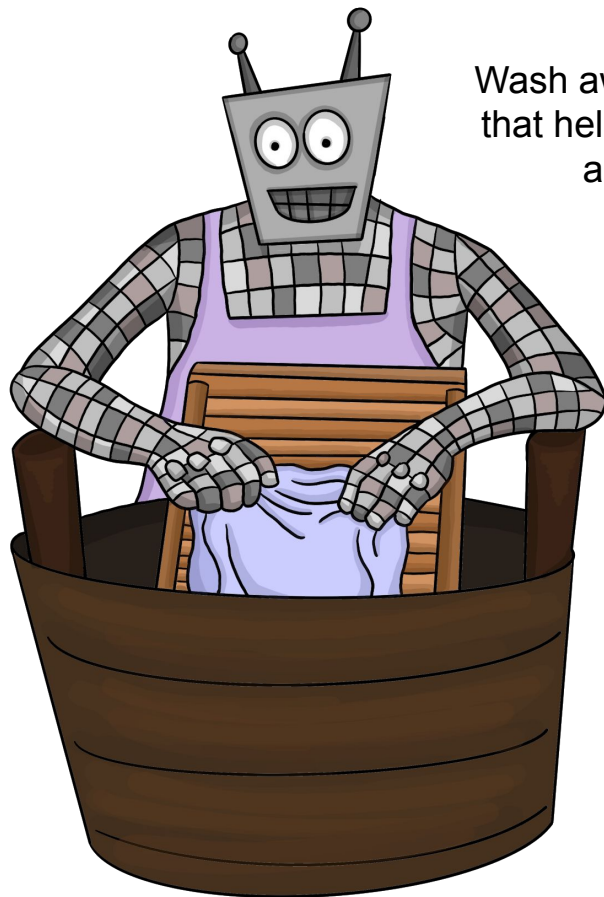
Snowball effect

Before + after problems → discrimination laundering

- Real world discrimination → privileges
- 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*)



Questions?



Age of EOP!

Substantive:
Rawlsian





Substantive EOP: Rawlsian

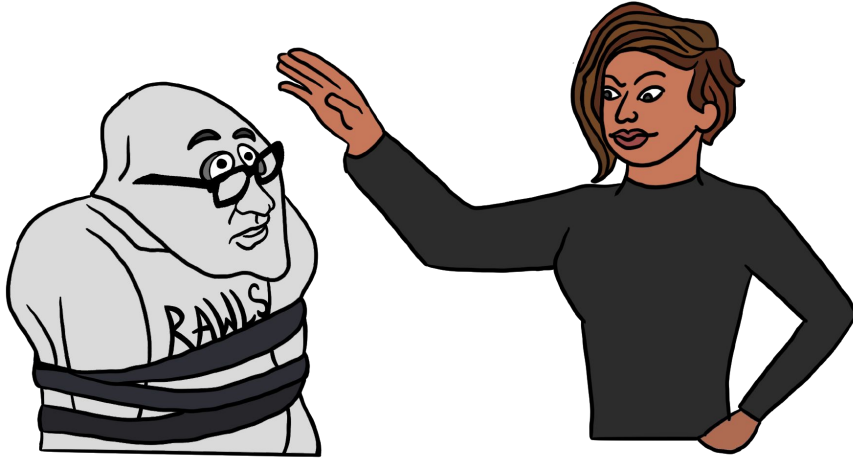
The bouncy castle of social security
has strategically placed trampolines
to propel individuals toward the
opportunities they seek

Substantive EOP: Rawlsian



- 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

Substantive EOP: Rawlsian

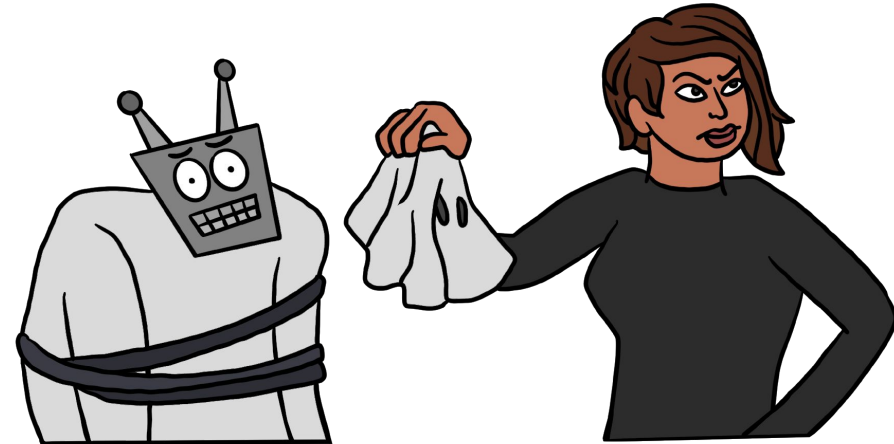


Fair-ML formulations of Rawls's Fair EOP are statistical parity and equality of odds.

BUT, these measures distort Rawlsian EOP!

Rawlsian EOP is fundamentally concerned with providing developmental opportunities **before** competitions

And not to measure how equitably a competition distributes developmental opportunities in **advance of later competitions!** (Fair-ML's focus)



Age of EOP!



**Substantive:
Luck-Egalitarian**

Substantive EOP: Luck Egalitarian

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



Substantive EOP: Luck Egalitarian

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

“Nothing that you didn't choose for yourself should affect your life prospects”



Substantive EOP: Roemer

- No split between responsible effort and irrelevant circumstances
- But how to compare apples and oranges?



Substantive EOP without effort/circumstances split

- Effort and circumstances can't be disentangled
- That's OK: circus families→circus kids
- Not OK: privileged family→access to wide range of desirable positions



Substantive EOP without effort/circumstances split

- Why not attend only to responsible effort?
- To avoid putting fairness and good decision-making at odds
- Good hiring decisions depend on considering all of a person's skill



Substantive EOP, take 2

Formal:

Selection process rewards
existing qualifications

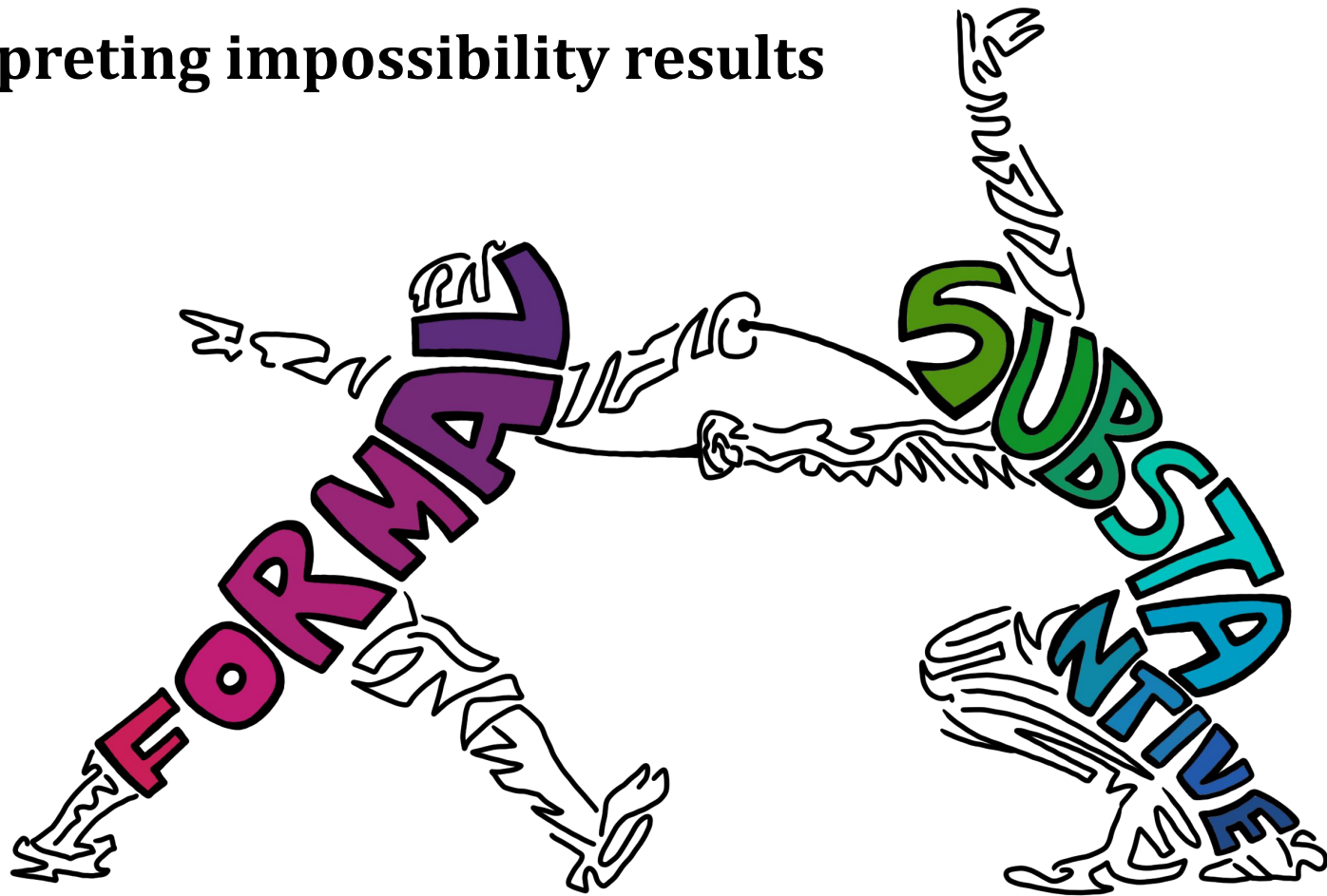


Substantive:

Selection process distributes
opportunities to develop
qualifications



Re-interpreting impossibility results



Normative guidance



When: qualifications are not
affected by circumstances of birth

Normative guidance



When: judges might be swayed by
irrelevant traits like appearance

Normative guidance



When: blindness to 'irrelevant' traits can boost the prospects of candidates

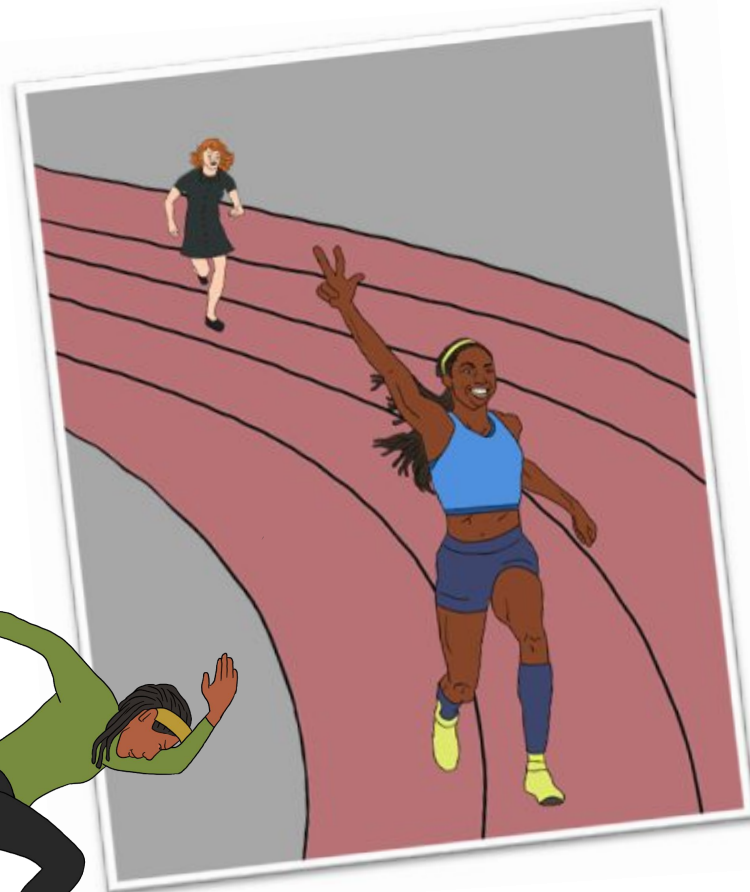
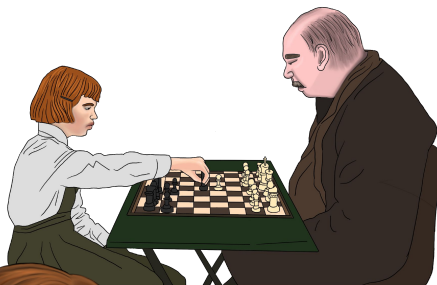
Normative guidance

When: disadvantages preclude people from getting a 'fair' shot



EOP over a lifetime

From the POV of Justice:
People should have genuine opportunities to
realistically achieve goals



Not merely formal opportunities to
compete for jobs!

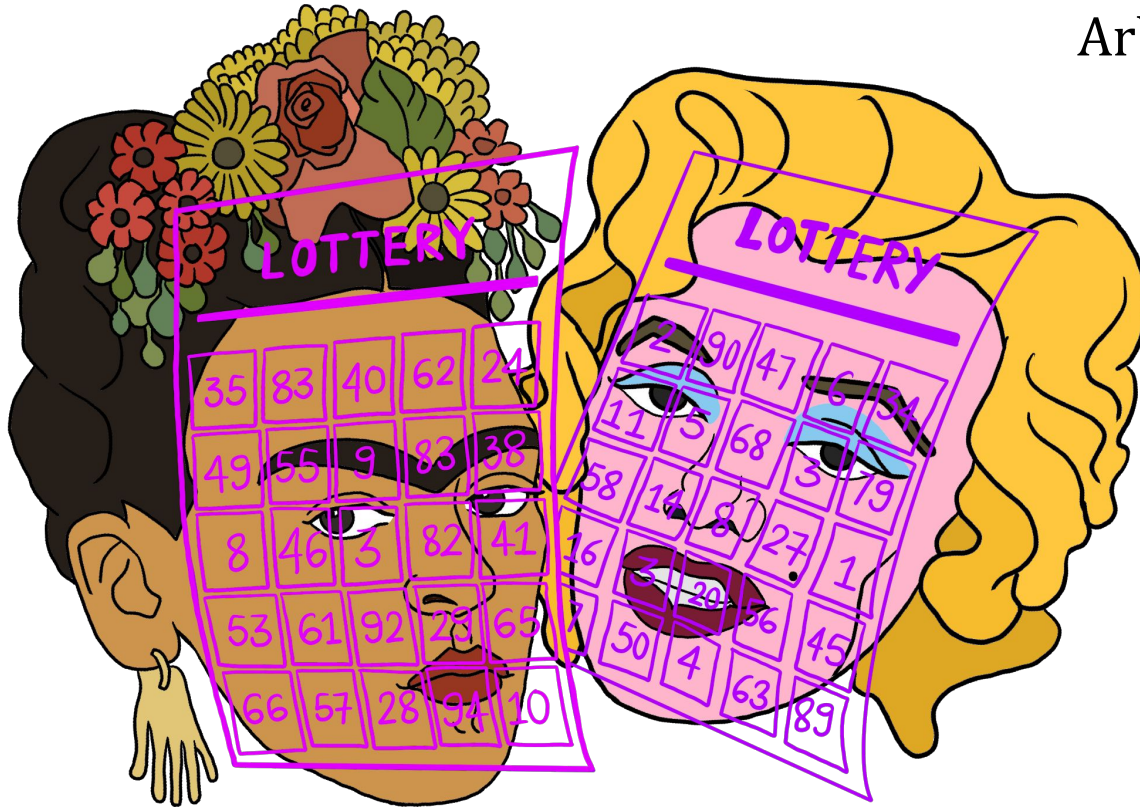
Broader view of justice



Original Position-
“Veil of Ignorance”

Broader view of justice

Natural Lottery-
Arbitrary distribution of
talents and fortune



Broader view of justice



1. [Rights and Liberties] Everyone has the same inalienable right to equal basic liberties
2. (a) [Rawlsian Fair EOP] All offices and positions must be open to all under conditions of fair equality of opportunity
2. (b) [Difference Principle] Social and economic inequalities must be of the greatest benefit to the least advantaged

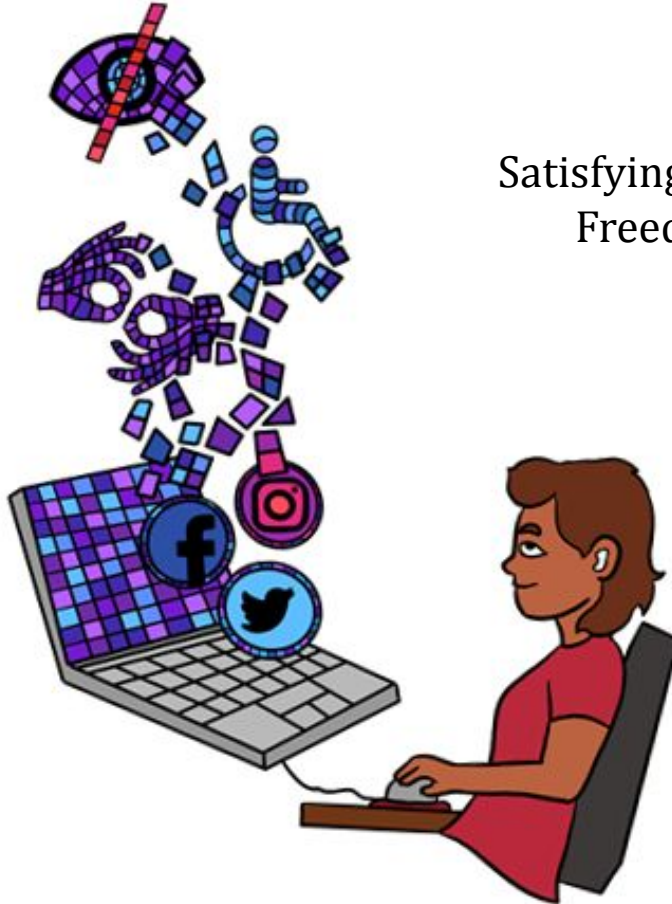
Broader view of justice

Satisfying EOP infringes on
basic liberties



Rawls (1971)

Broader view of justice

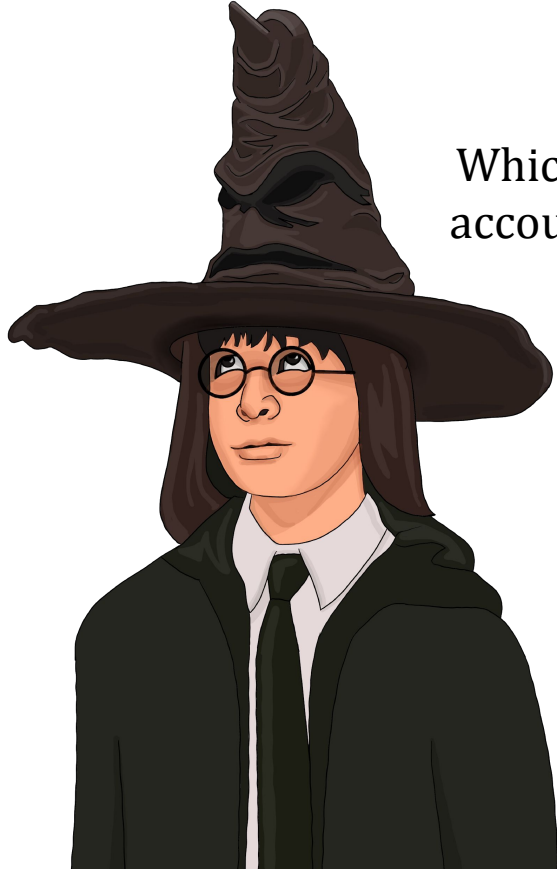


Satisfying EOP infringes on
Freedom of Speech



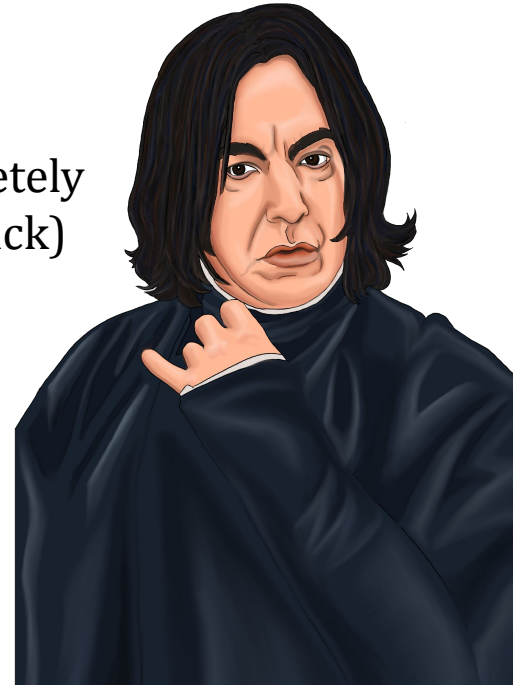
Rawls (1971)

Limitations in guidance - brute luck vs. choice luck

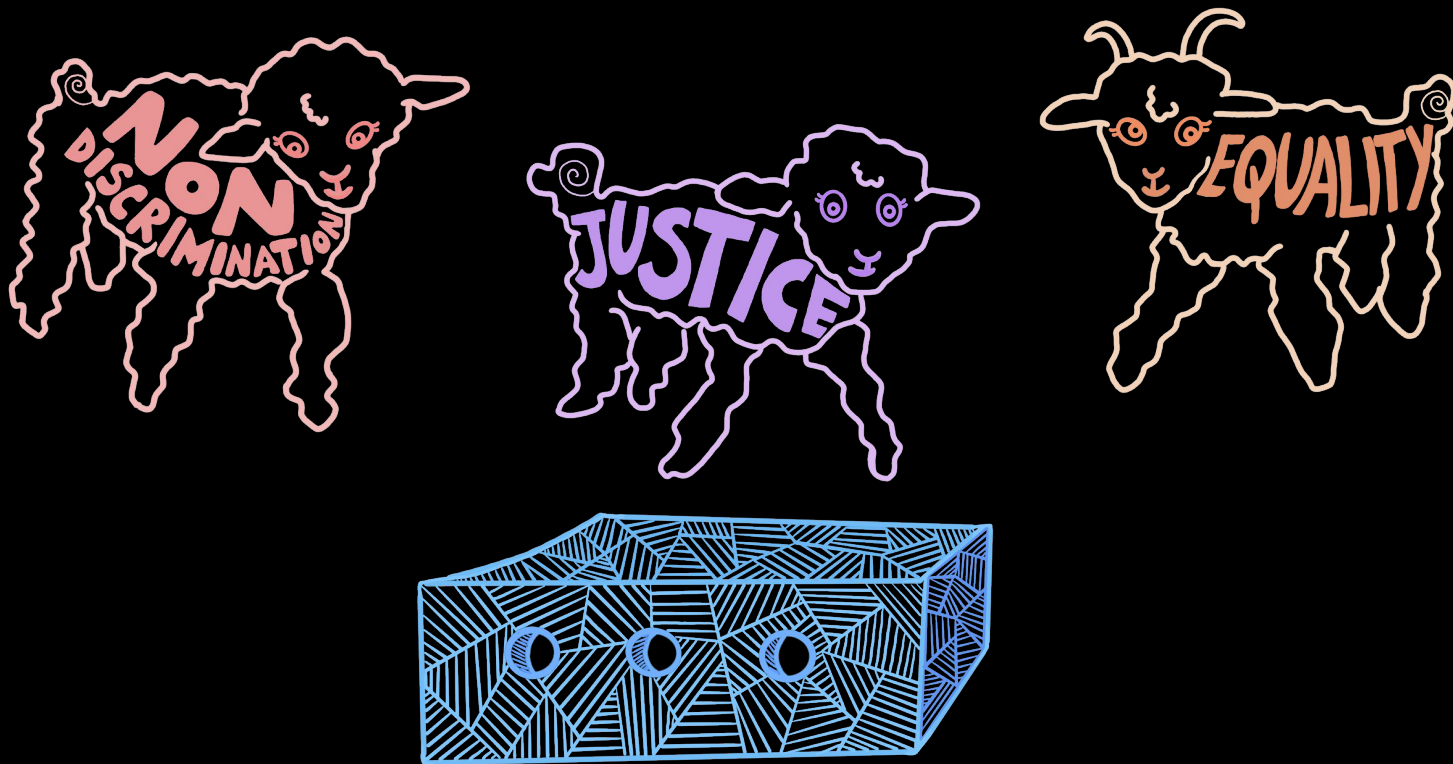


Which characteristics can we hold one accountable for? (Responsible choices)

And which matters are completely out of their control? (Brute luck)

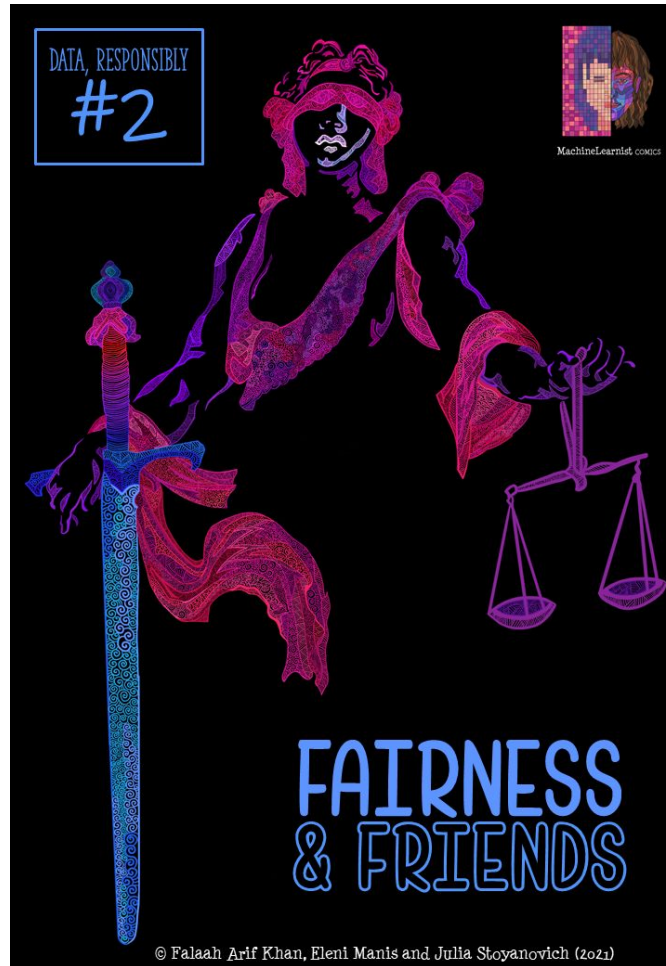


Limitations in guidance - interpretability



The fairness that you asked for is inside this box!

Thank you!



<https://dataresponsibly.github.io/comics>

Questions?

