



FAIRNESS & FRIENDS

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

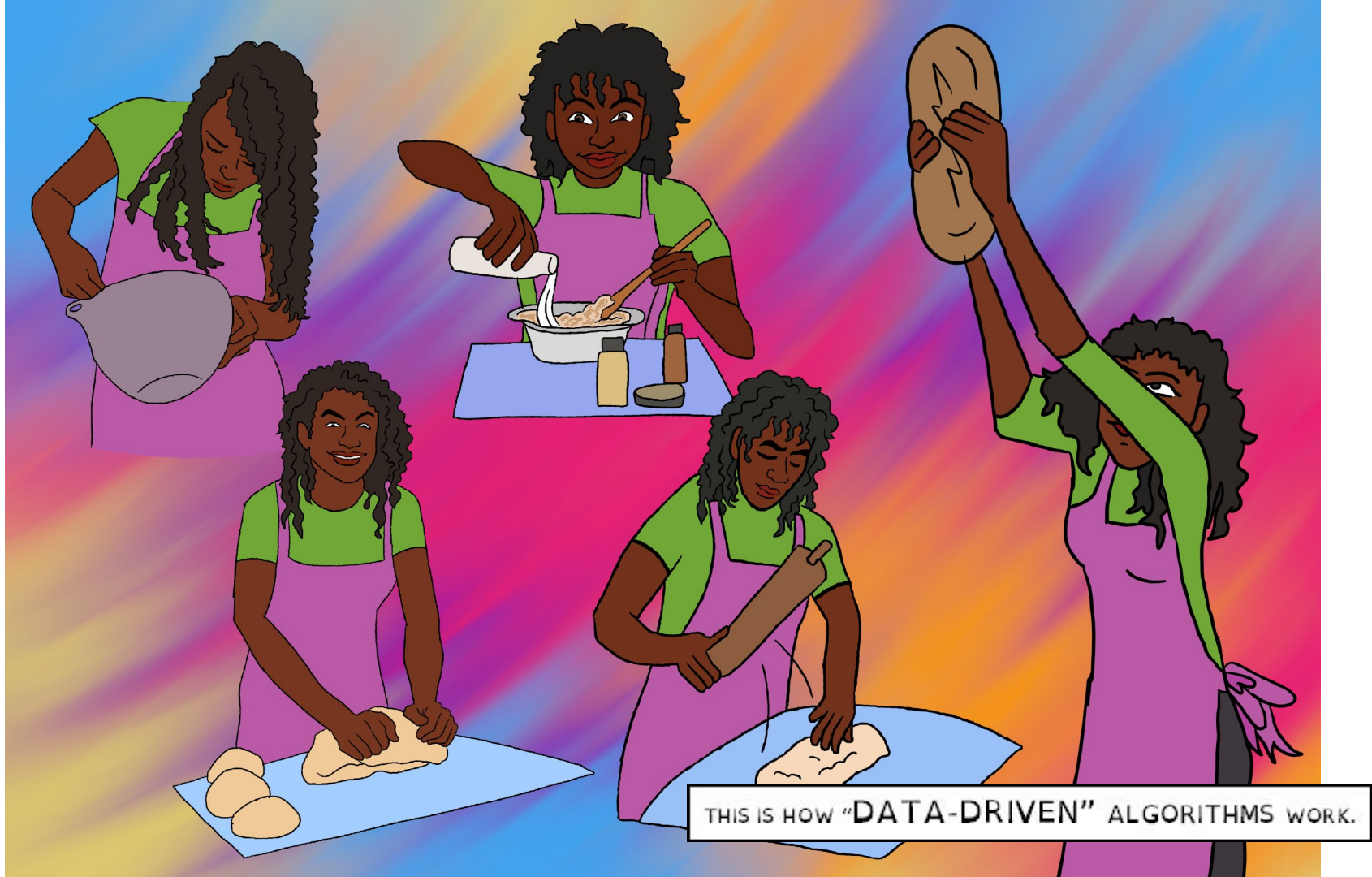


ible

rea



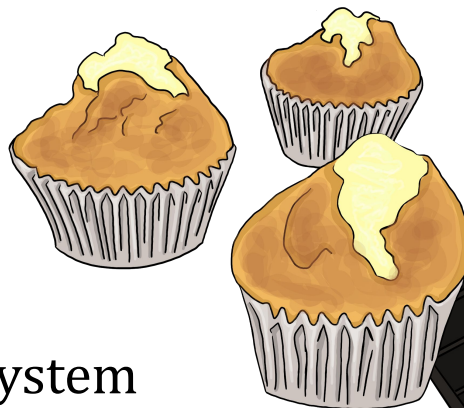
WE CALL SUCH ALGORITHMS "RULE-BASED"



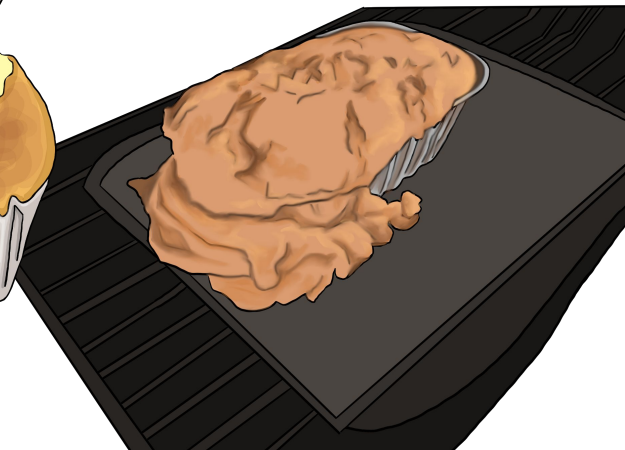
THIS IS HOW "DATA-DRIVEN" ALGORITHMS WORK.

All about that Bias...

Pre-existing - In the data
(Example: Racial and gender stereotypes
in language models)



Technical - In the technical system
(Example: Hardware limitations, incorrect choices
of representation, strong modelling assumptions)

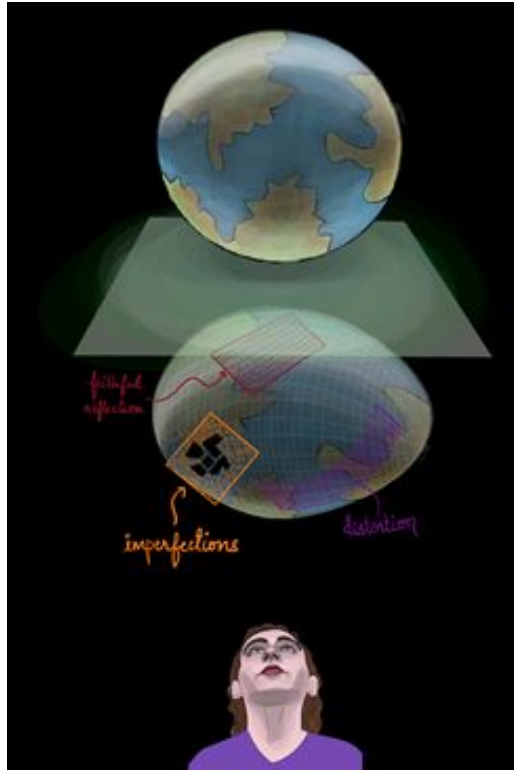


What is bias? (Recap)

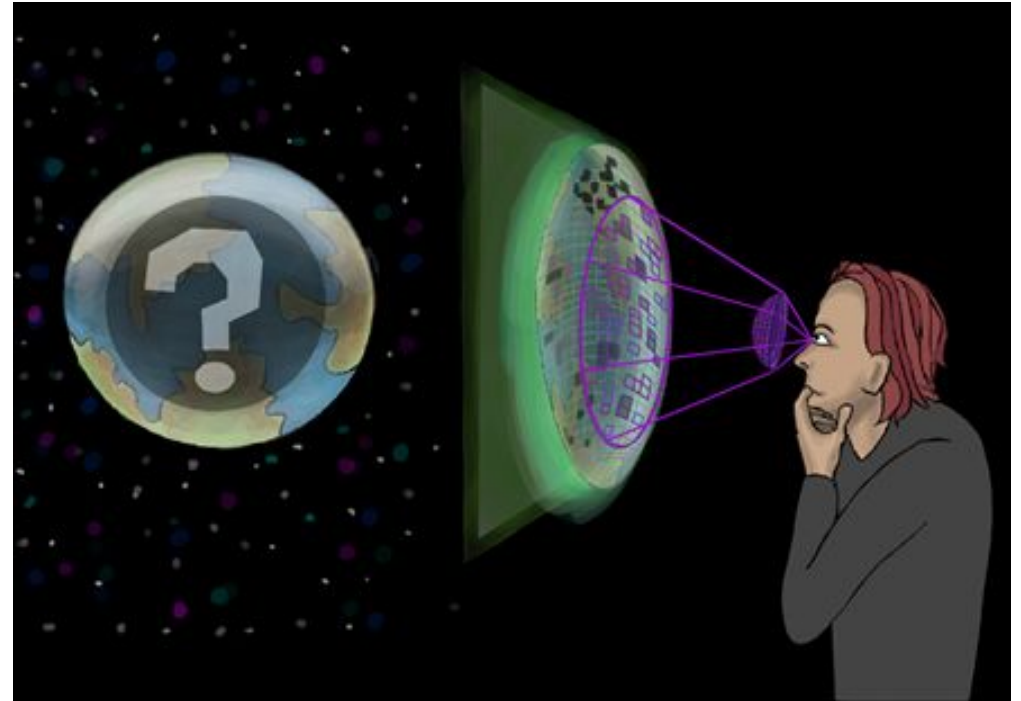
Emergent - due to decisions
(Example: Feedback loops, behavioral changes)



1. Data is a mirror reflection of the world



2. Bias is the data is the distortions in the reflection



3. Mitigation of bias is the corrective lenses!

'Fairness is Impossible'



(EXECUTIVE)

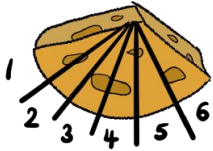


(SOUS CHEFS)



(ROOKIES/LINE CHEFS)

'Fairness is Impossible'



'Fairness is Impossible'



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!

It's the....
Age of
Equality Of
Opportunity!
(EOP)



Age of EOP!

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



Libertarian

- Any welfare/opportunities won 'fair and square' are allowed
- Even if some end up with significantly more than others
- **Not a form of EOP at all** - only enforces a limited notion of **procedural fairness**

Age of EOP!



Formal

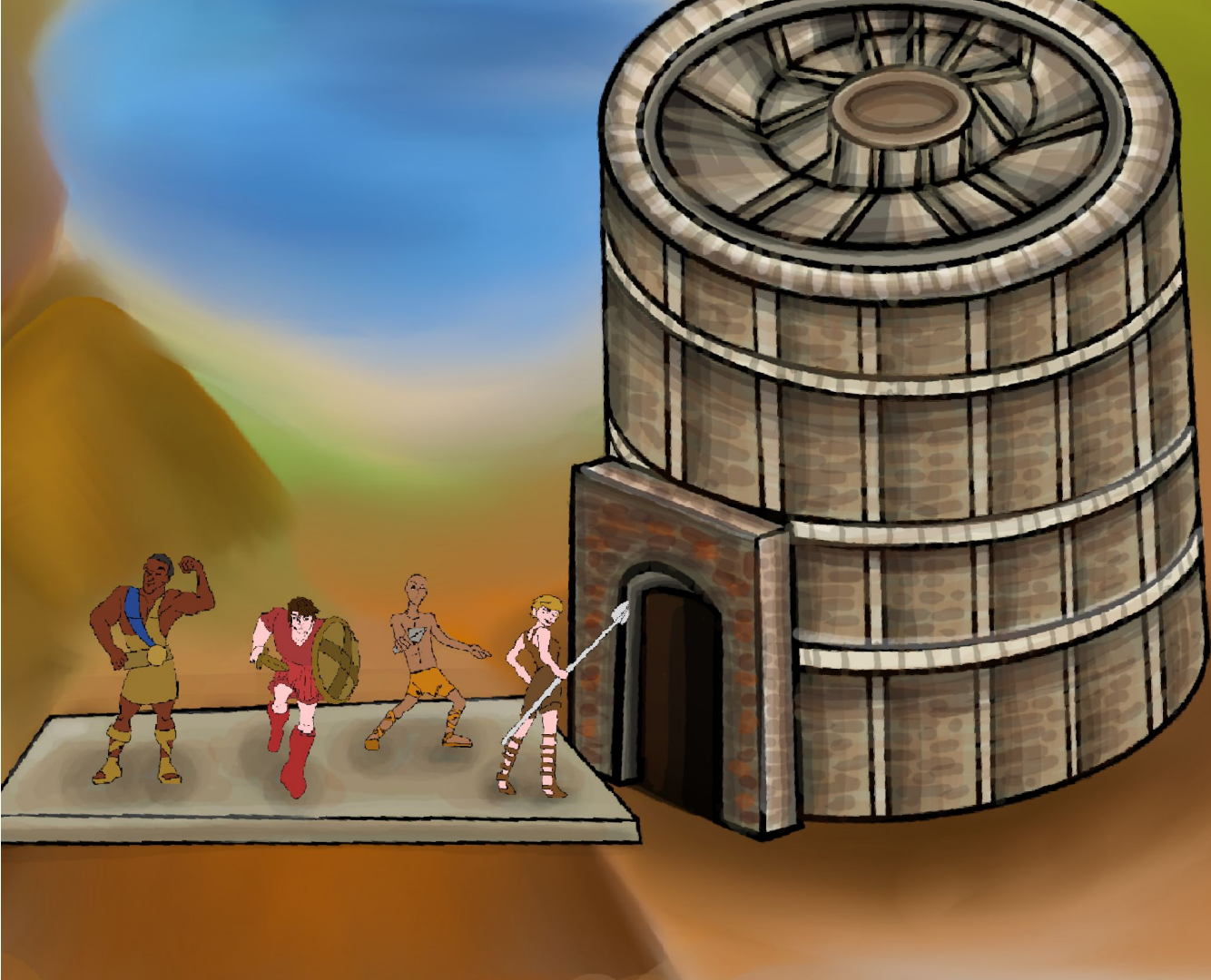
(The theatre of formal EOP is open to all talents - but you fight with what you have, no special treatment once you're in)

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”



Formal EOP

- In any competition, the most qualified person must win
- Formal EOP is codified as “Fairness through Blindness”
- Formal plus - algorithms must not perform systematically worse/better for some demographics

Age of EOP!

Substantive: Rawlsian

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



Substantive EOP: 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.



Substantive EOP: Rawlsian

- Fair-ML formulations include ‘equality of odds’ and ‘statistical parity’
- But! These are not truly Rawlsian - at the point of where the algorithm is making a decision, it’s too late to provide people with opportunities to build qualifications

Age of EOP!



Substantive:
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”

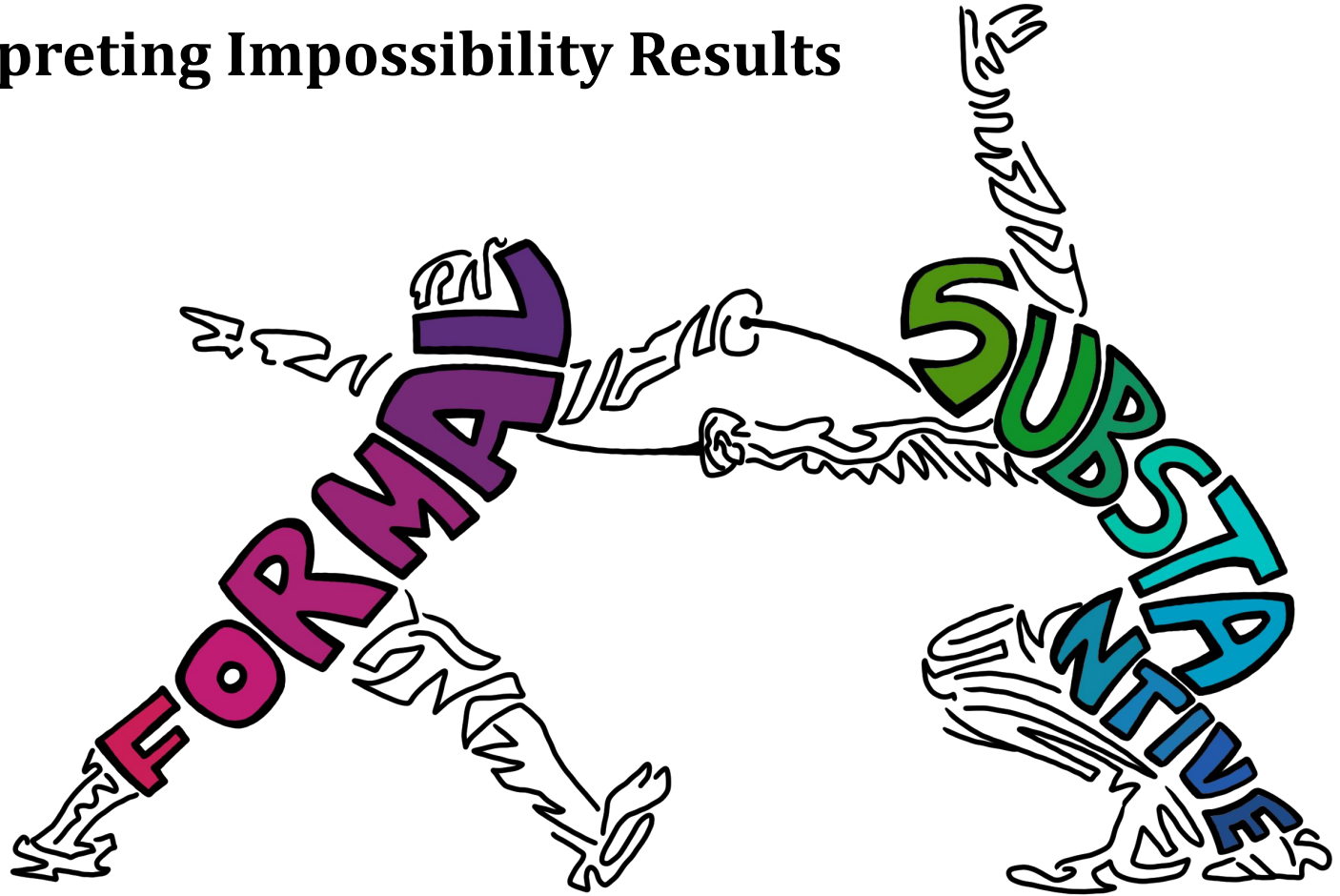




Substantive EOP: Luck Egalitarian

- One popular formulation is Roemer's EOP - only control for certain factors such as sex and race
- Measure a person based on their rank in the effort distribution of their type/circumstance

Re-interpreting Impossibility Results



Re-interpreting Impossibility Results



When: qualifications are not affected by circumstances of birth



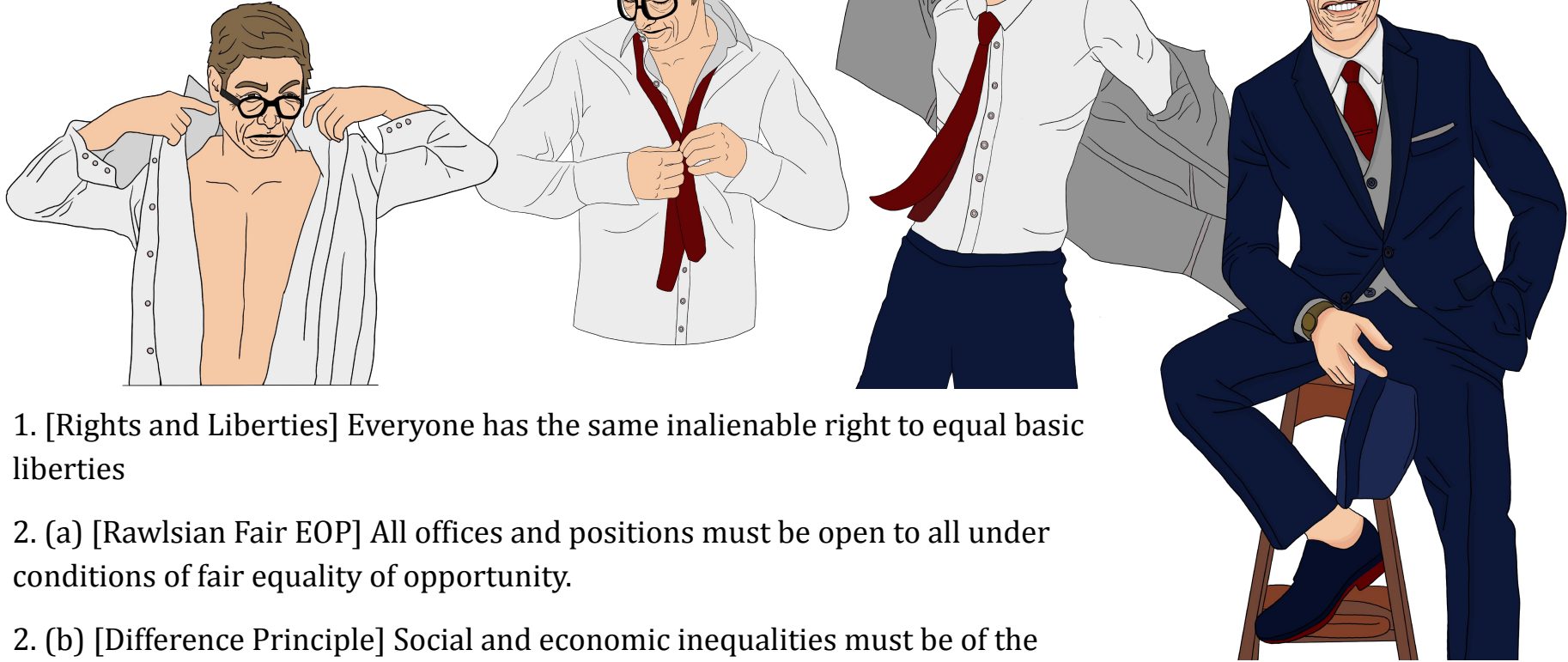
When: judges might be swayed by irrelevant traits like appearance.

Re-interpreting Impossibility Results

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



Broader notions of Justice - Rawls



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

'Fair' hiring of people with disabilities



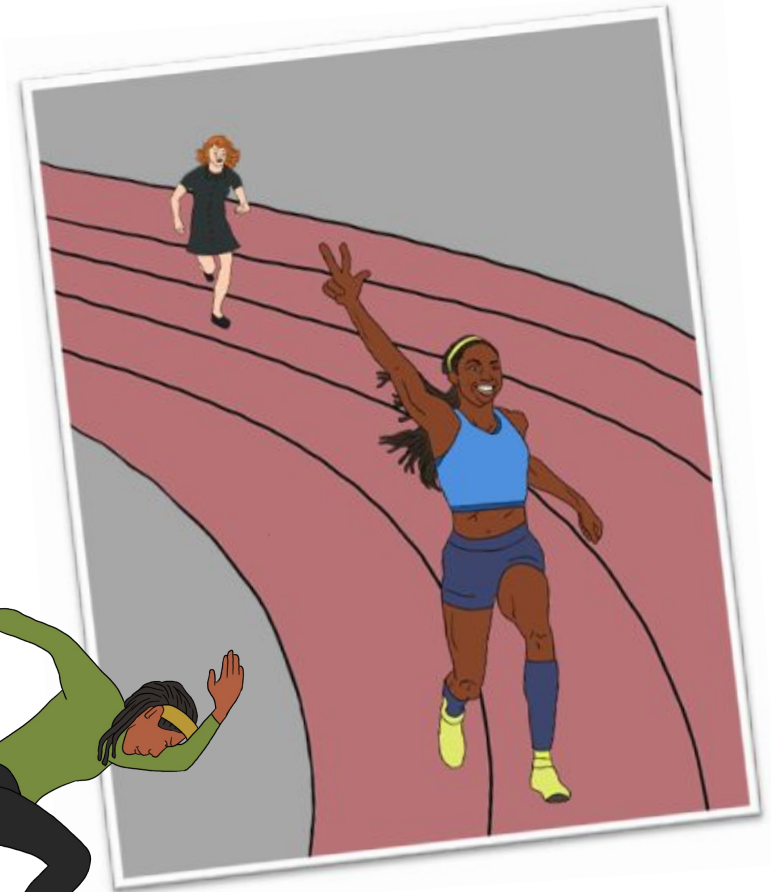
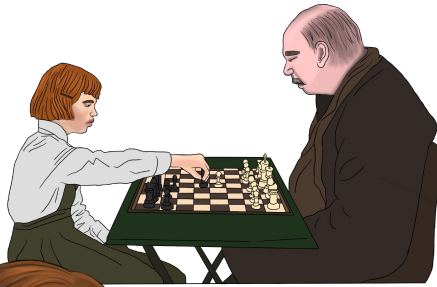
Satisfying Fair EOP, infringes on
Rights and Liberties



Rawls (1971)

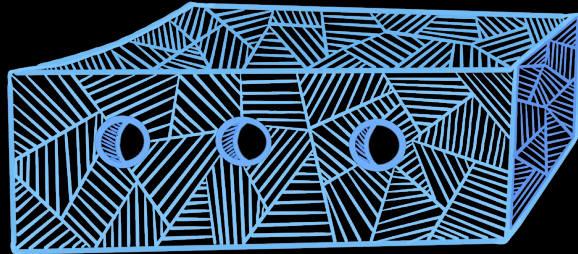
EOP over a lifetime

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



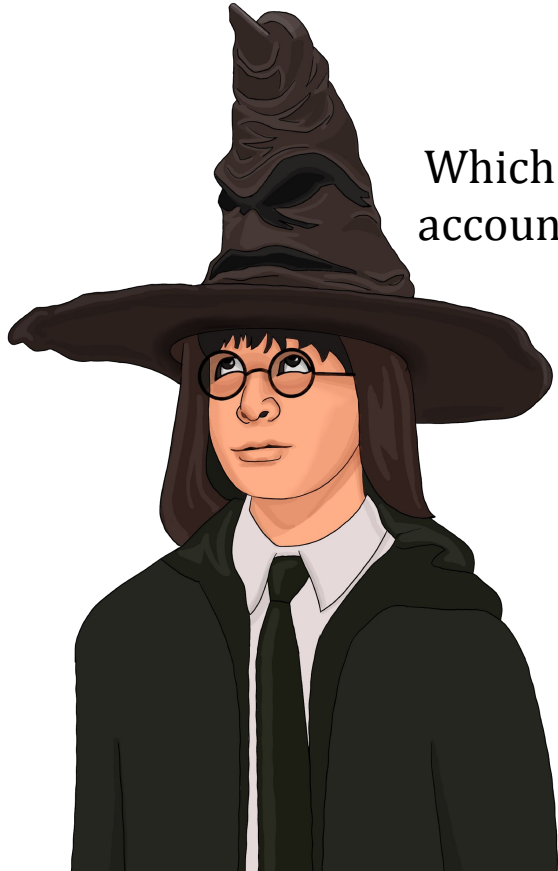
Not merely formal opportunities to
compete for jobs!

Limitations in Guidance - Interpretability



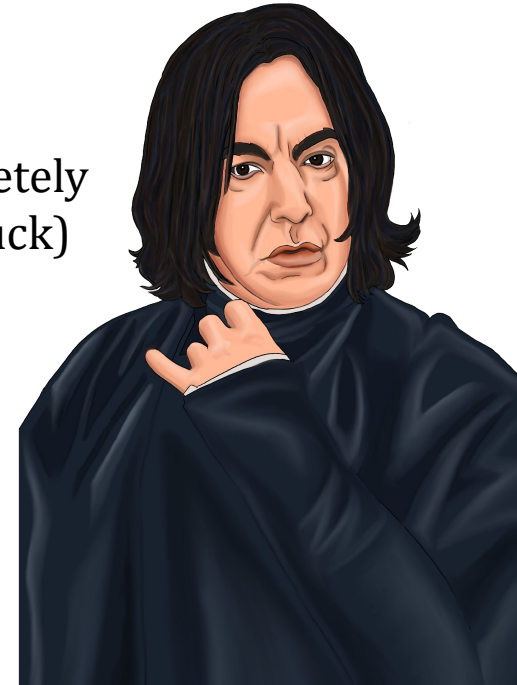
The fairness that you asked for is inside this box!

Limitations in Guidance - Brute luck vs. Choice luck

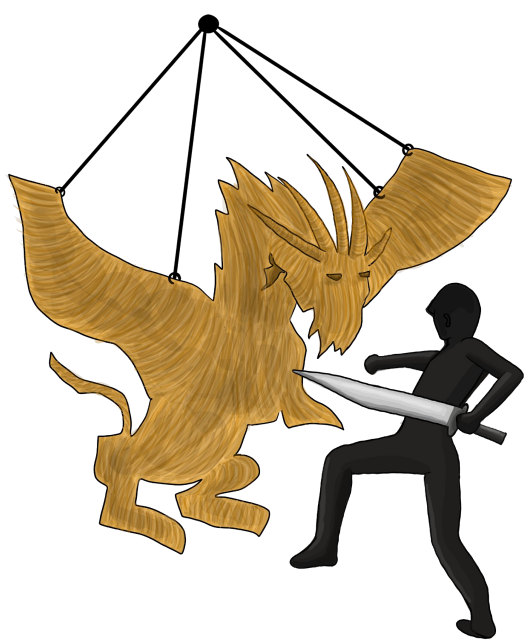


Which characteristics can we hold one accountable for? (Responsible choices)

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



Only technical fixes are not enough!



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'

