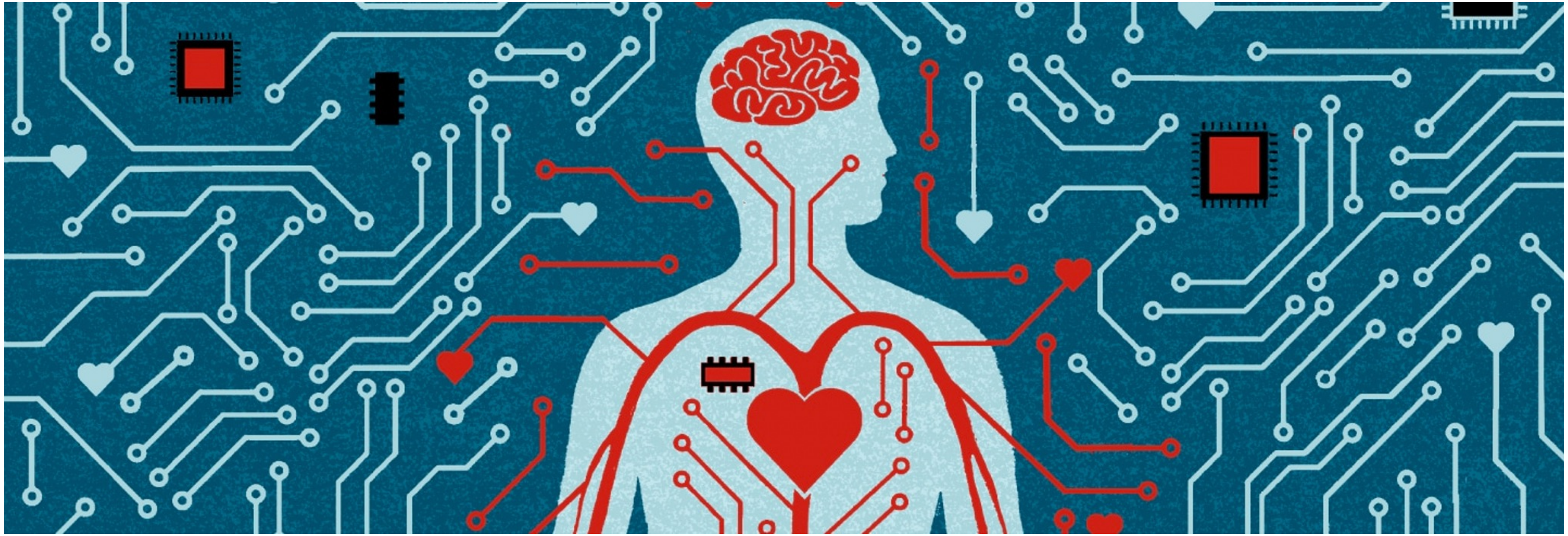
Center for Advanced Study in the Behavioral Sciences at Stanford University

WISCONSIN SCHOOL OF BUSINESS



Actuarial Fairness
principles and perspectives

James Guszczka – Stanford-CASBS
Dani Bauer – University of Wisconsin-Madison

CAS Spring Meeting
May 17, 2022

Actuarial Fairness in Context

Overview of ethical principles

AI ethics principles

AI ethics principles

AI ethics principles

The need to manage tradeoffs

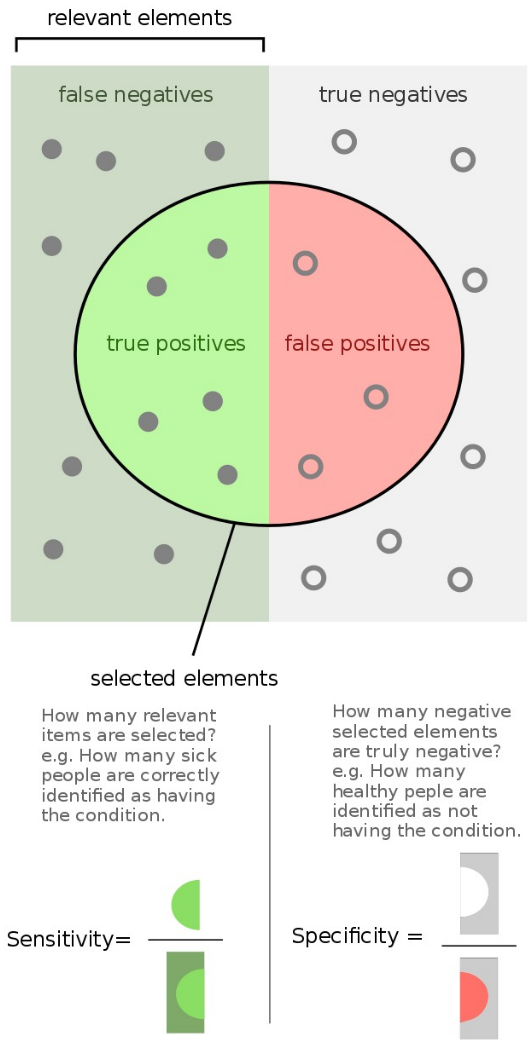
Core ethical principles are **ideals** to be strived for.

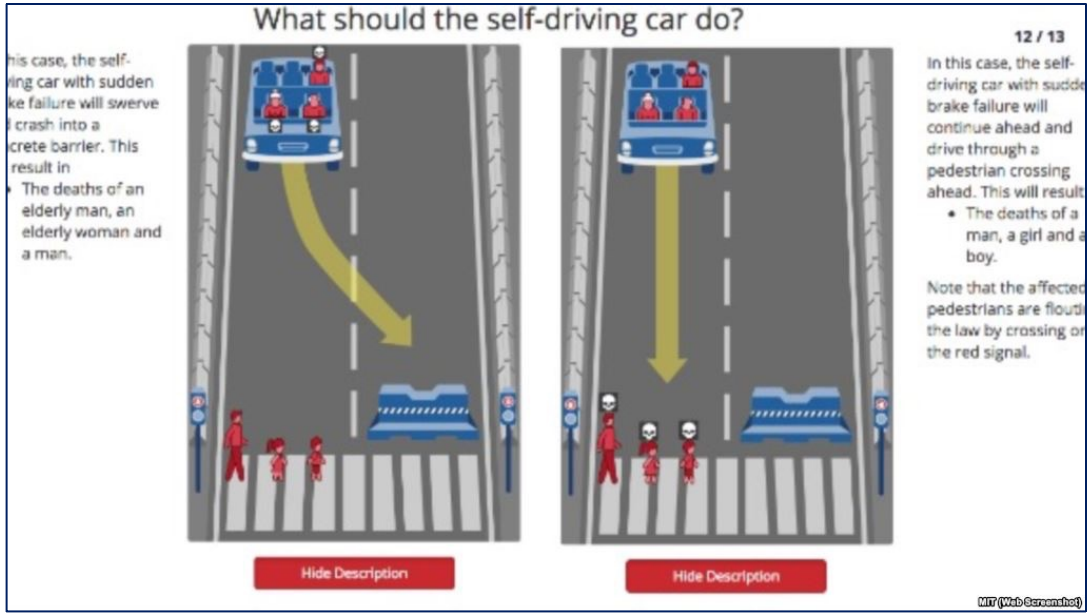
It's often **impossible** to simultaneously satisfy all of them

Trade-offs typically must be deliberated

Innovations can be explored to make tradeoffs less acute

Think of the ethical principles as **design considerations**.



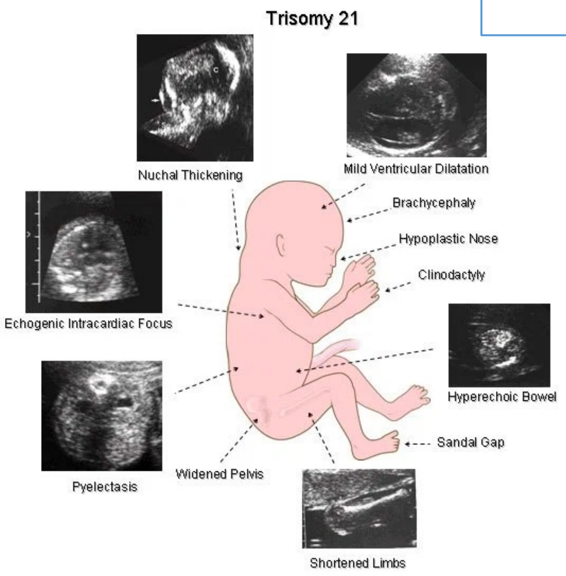


Ethics and quality control

Artificial Intelligence—The Revolution Hasn't Happened Yet

by Michael I. Jordan

> She said, "Ah, that explains why we started seeing an uptick in Down syndrome diagnoses a few years ago. That's when the new machine arrived."



Relationship between <u>ethics</u> and <u>quality control</u>

- Evaluate data provenance
- Ensure operating environment is suitably "regularized"
  (e.g., in the case of autonomous vehicles)
- Ensure end-users are trained and have a good "mental model" of the technology
- Don't neglect the "science" part of data science – need for scientifically informed **judgment** in building and using algorithms

AI and human autonomy

> ***Individual autonomy:*** *The capacity to be one's own person, to live one's life according to reasons and motives that are taken as one's own and not the product of manipulative or distorting external force.*
>
> *— Stanford Encyclopedia of Philosophy*

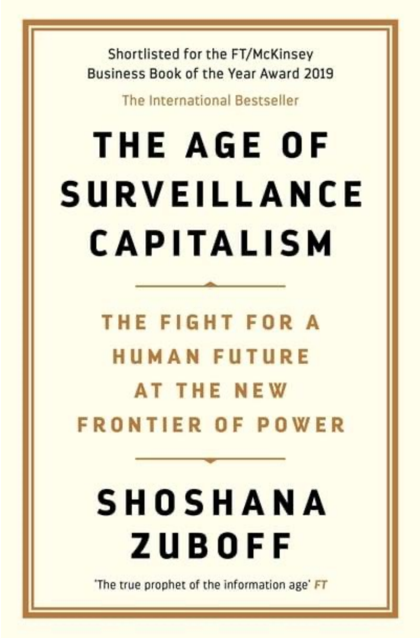
SCIENTIFIC AMERICAN

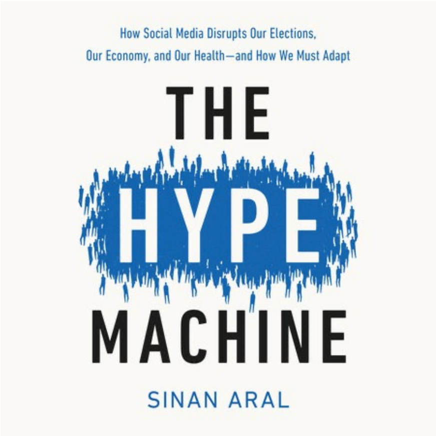
Will Democracy Survive Big Data and Artificial Intelligence?

We are in the middle of a technological upheaval that will transform the way society is organized. We must make the right decisions now

By Dirk Helbing, Bruno S. Frey, Gerd Gigerenzer, Ernst Hafen, Michael Hagner, Yvonne Hofstetter, Jeroen van den Hoven, Roberto V. Zicari, Andrej Zwitter on February 25, 2017

But it won't stop there. Some software platforms are moving towards "persuasive computing." In the future, using sophisticated manipulation technologies, these platforms will be able to steer us through entire courses of action, be it for the execution of complex work processes or to generate free content for Internet platforms, from which corporations earn billions. *The trend goes from programming computers to programming people.*





AI and human autonomy

Choice architecture (“Nudge”) is often criticized as a type of manipulation that undermines human autonomy.

How Uber Uses Psychological Tricks to Push Its Drivers’ Buttons

The company has undertaken an extraordinary experiment in behavioral science to subtly entice an independent work force to maximize its growth.

By NOAM SCHEIBER and graphics by JON HUANG | APRIL 2, 2017

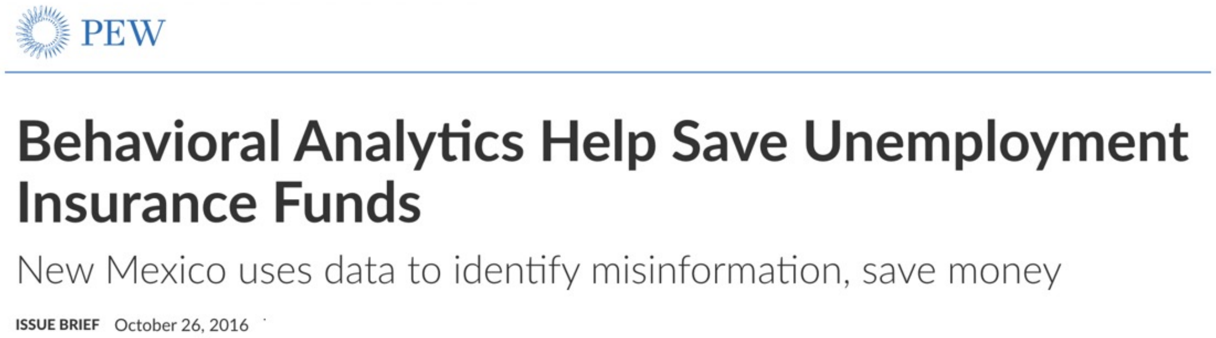
BEHAVIORAL ECONOMICS

Uber Shows How Not to Apply Behavioral Economics

by Francesca Gino

April 13, 2017

Behavioral science and ethical AI

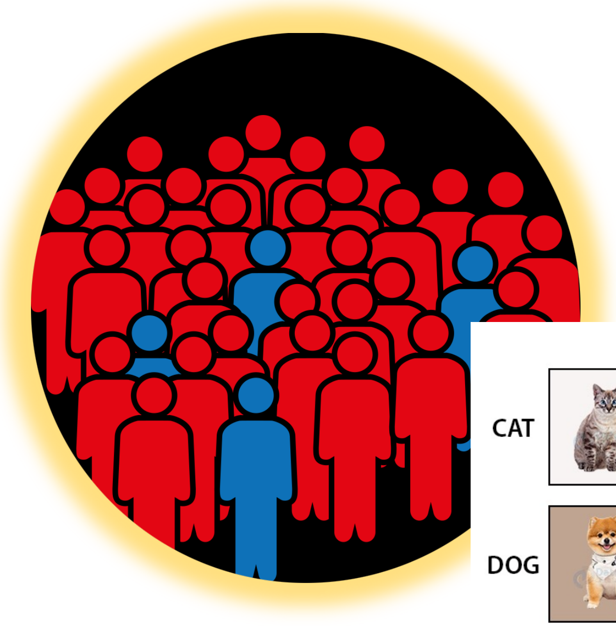

PEW

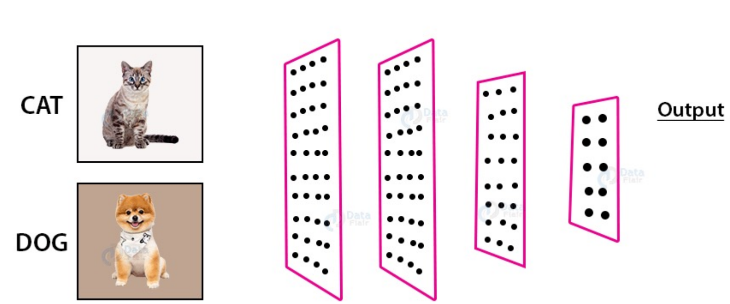
Behavioral Analytics Help Save Unemployment Insurance Funds

New Mexico uses data to identify misinformation, save money

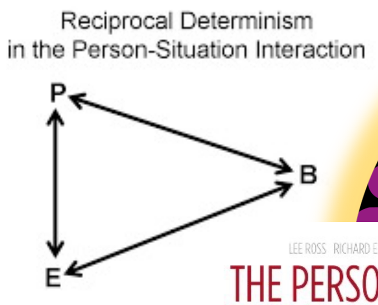
ISSUE BRIEF October 26, 2016

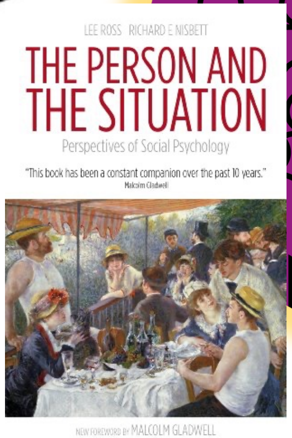
Naïve view



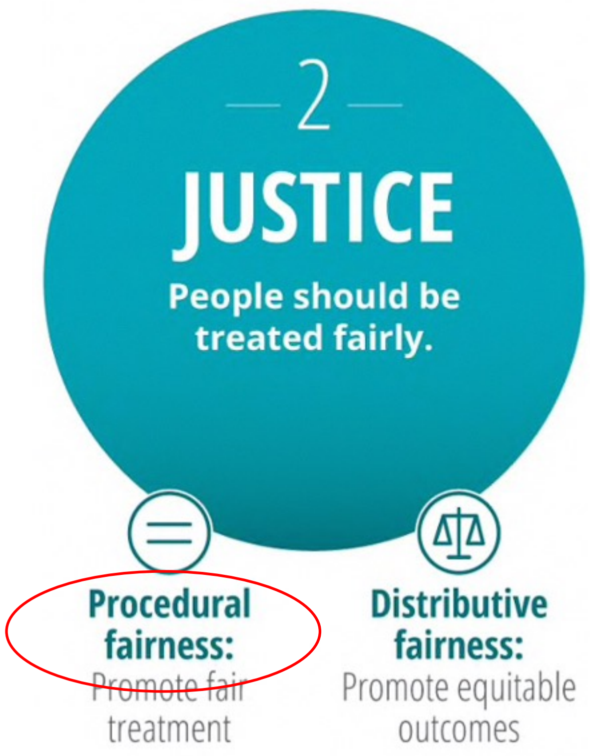


"Nudge" view





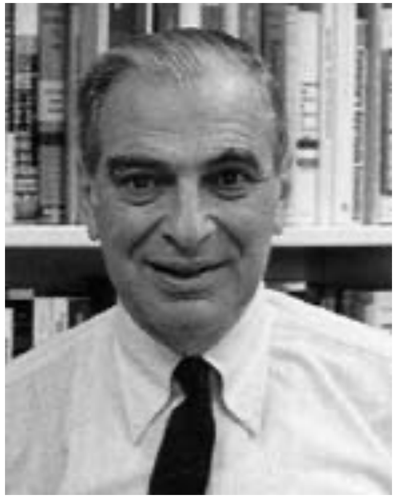
Procedural justice: fairness and transparency of the decision process



- **Perfect procedural justice** – a procedure that is guaranteed to give the desired outcome.
  - E.g.: the person who cuts the cake is the last one to choose a slice
  - We don’t have this in actuarial science because of uncertainty around any estimate of E[loss]
- **Imperfect procedural justice** – the procedure is not guaranteed to give the desired outcome
  - E.g.: a criminal trial. Sometimes guilty go free and vice versa
  - This maps onto actuarial fairness
- **Pure procedural justice** – no criterion for the desired outcome
  - E.g.: gambling
  - Does this map onto price optimization? Two policyholders with the same risk profile could be charged different amounts.

Actuarial Fairness – Kenneth Arrow

"Suppose therefore, an agency, a large insurance company plan, or the government, stands ready to offer insurance against medical costs on an actuarially fair basis; that is, if the costs of medical care are a random variable with mean μ, the company will charge a premium μ, and agree to indemnify the individual for all medical costs. Under these circumstances, the individual will certainly prefer to take out a policy and will have a welfare gain thereby."



Actuarial Fairness – CAS

Principle 1: A *rate* is an estimate of the expected value of future costs.

Ratemaking should provide for all costs so that the insurance system is financially sound.

Principle 2: A rate provides for all costs associated with the transfer of risk.

Ratemaking should provide for the costs of an individual risk transfer so that equity among insureds is maintained. When the experience of an individual risk does not provide a credible basis for estimating these costs, it is appropriate to consider the aggregate experience of similar risks. A rate estimated from such experience is an estimate of the costs of the risk transfer for each individual in the class.

Principle 3: A rate provides for the costs associated with an individual risk transfer.

Ratemaking produces cost estimates that are actuarially sound if the estimation is based on Principles 1, 2, and 3. Such rates comply with four criteria commonly used by actuaries: reasonable, not excessive, not inadequate, and not unfairly discriminatory.

Principle 4: A rate is reasonable and not excessive, inadequate, or unfairly discriminatory if it is an actuarially sound estimate of the expected value of all future costs associated with an individual risk transfer.

Algorithmic fairness beyond insurance

Racial bias skews algorithms widely used to guide care from heart surgery to birth, study finds

TECHNOLOGY

Facial Recognition Is Accurate, if You're a White Guy

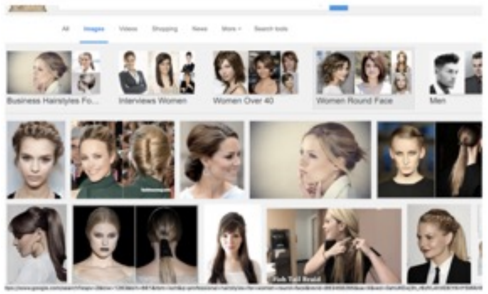
By STEVE LOHR FEB. 9, 2018

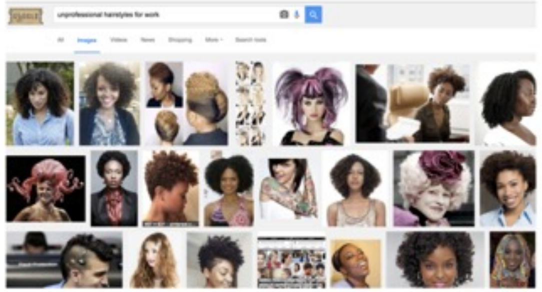
Dissecting racial bias in an algorithm used to manage the health of populations

Ziad Obermeyer[1,2,*], Brian Powers[3], Christine Vogeli[4], Sendhil Mullainathan[5,*,†]

I'll stop calling algorithms racist when you stop anthropomorphizing AI

April 7, 2016 Cathy O'Neil, mathbabe





Ad related to latanya sweeney

Latanya Sweeney Truth
www.instantcheckmate.com/
Looking for Latanya Sweeney? Check Latanya Sweeney's Arrests.

Ads by Google

Latanya Sweeney, Arrested?
1) Enter Name and State. 2) Access Full Background Checks Instantly.
www.instantcheckmate.com/

Latanya Sweeney
Public Records Found For: Latany
www.publicrecords.com/

La Tanya
Search for La Tanya Look Up Fast
www.ask.com/La+Tanya

THE VERGE

Amazon reportedly scraps internal AI recruiting tool that was biased against women

The secret program penalized applications that contained the word "women's"

By James Vincent | @jjvincent | Oct 10, 2018, 7:09am EDT

A well-known algorithmic bias case study

PROPUBLICA

▾ Machine Bias

Machine Bias

There's software used across the country to predict future criminals. And it's biased against blacks.

by Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, ProPublica
May 23, 2016

Prediction Fails Differently for Black Defendants

	WHITE	AFRICAN AMERICAN
Labeled Higher Risk, But Didn't Re-Offend	23.5%	44.9%
Labeled Lower Risk, Yet Did Re-Offend	47.7%	28.0%

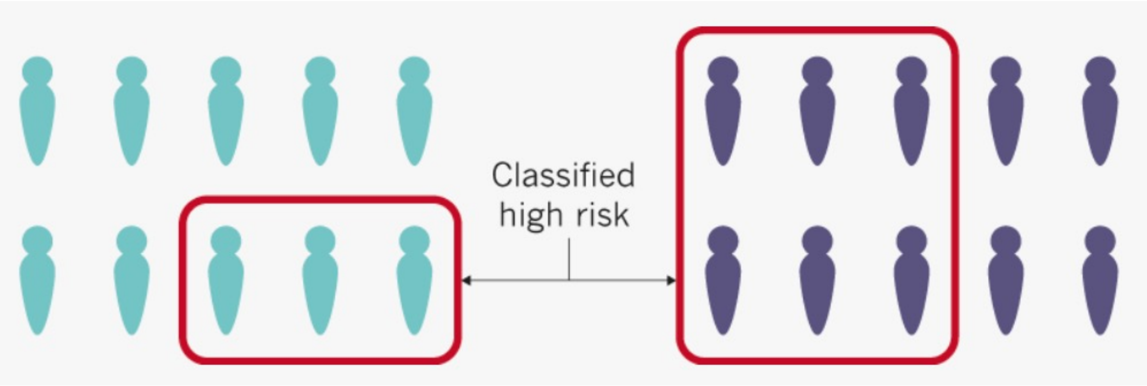
Wisconsin Supreme Court (2016):

- Judges can use risk scores.
- But the scores cannot be a "determinative" factor in whether the defendant is jailed or gets probation.
- Judge must be given a warning about the limits of the algorithm's accuracy.

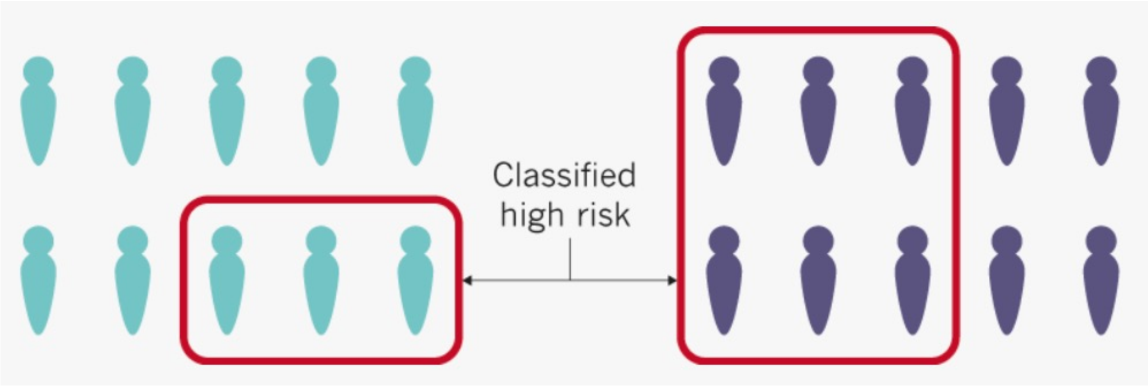
Inherent Trade-Offs in the Fair Determination of Risk Scores

Jon Kleinberg, Sendhil Mullainathan, Manish Raghavan

(Submitted on 19 Sep 2016 (v1), last revised 17 Nov 2016 (this version, v2))



Fact in the world:
Higher base rate for purple than green



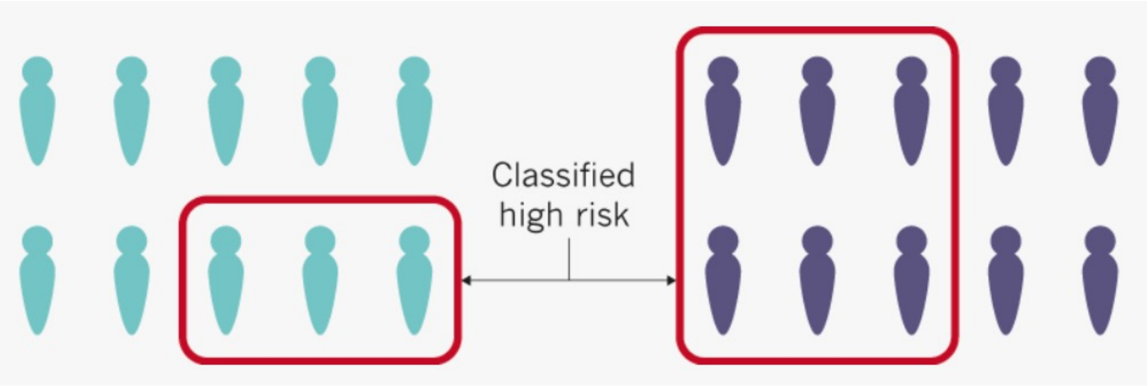
Predictive parity: "high risk" means ***2/3 chance of being re-arrested*** *for each group*

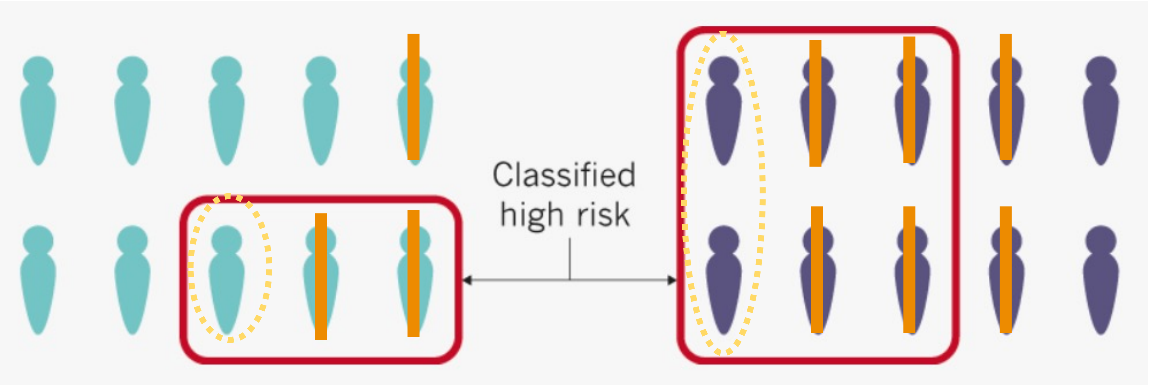
Inherent Trade-Offs in the Fair Determination of Risk Scores

Jon Kleinberg, Sendhil Mullainathan, Manish Raghavan

(Submitted on 19 Sep 2016 (v1), last revised 17 Nov 2016 (this version, v2))

"It turns out
[different false positives rates are]
more or less a statistical artifact"
– Sharad Goel





Fact in the world:
Higher base rate for purple than green

*Predictive parity: "high risk" means **2/3 chance of being re-arrested** for each group*

False positives (1/7 for green; 2/4 for purple):
a mathematical inevitability

The focus is not just be on making the ML model fair but rather on **making the overall system and outcomes fair**

Define (equity) → Measure/ Detect (bias) → Understand (root causes) → Improve (fairness) → Mitigate (the impact) → Monitor & Evaluate

Slide taken from joint presentation with Rayid Ghani

Bias (in outcomes) can come from any of these four components

World → Data → ML Pipeline → Actions

Data

Sample Bias
Measurement Bias
Label Bias

ML Pipeline

System Developers
Complexity or flaws
Design Choices

Slide taken from joint presentation with Rayid Ghani

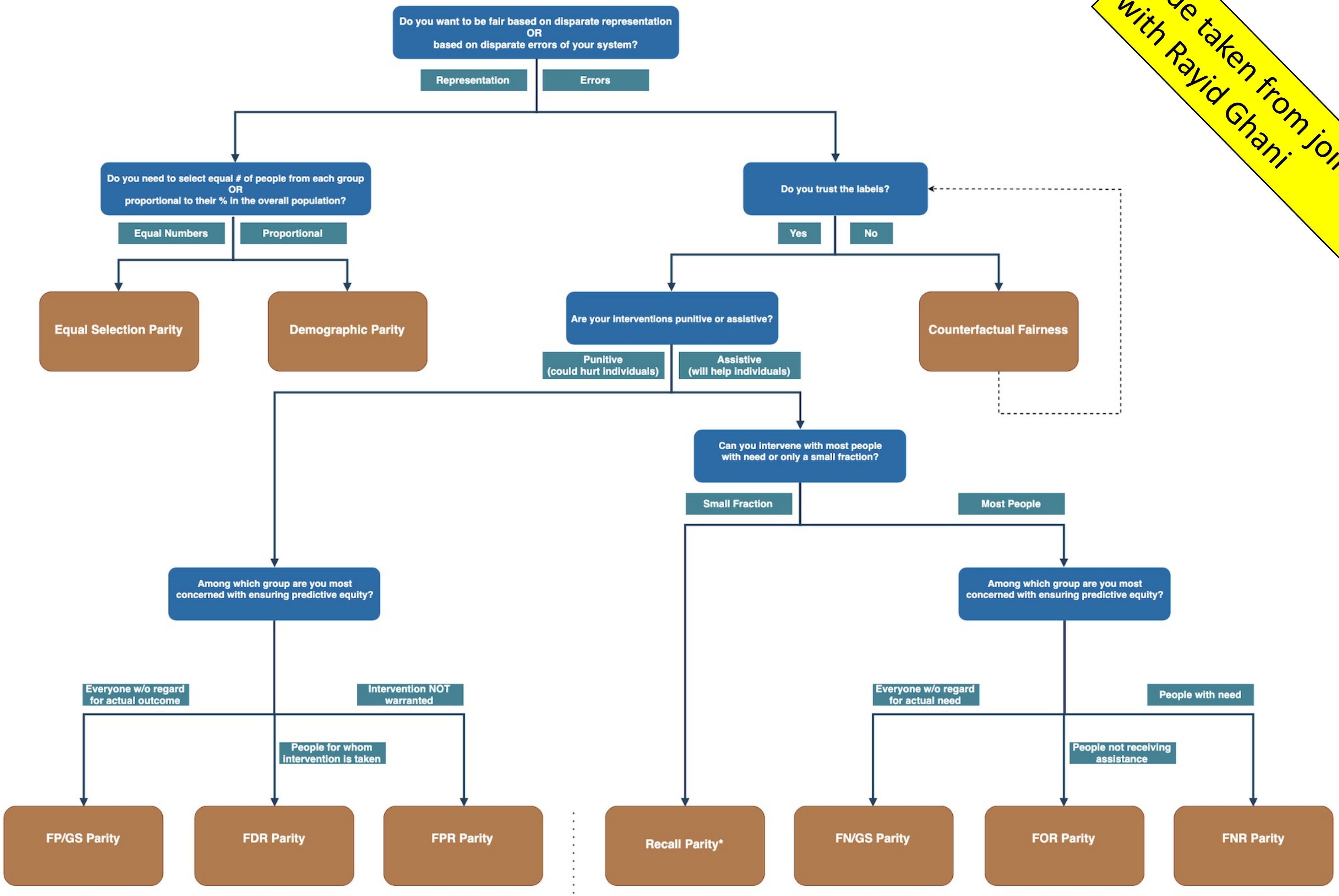
Many Bias Measures: How do we select what we care about?

- Statistical/Demographic Parity
- Impact Parity
- False Discovery Rate Parity
- False Omission Rate Parity
- False Positive Rate Parity
- False Negative Rate Parity
- …

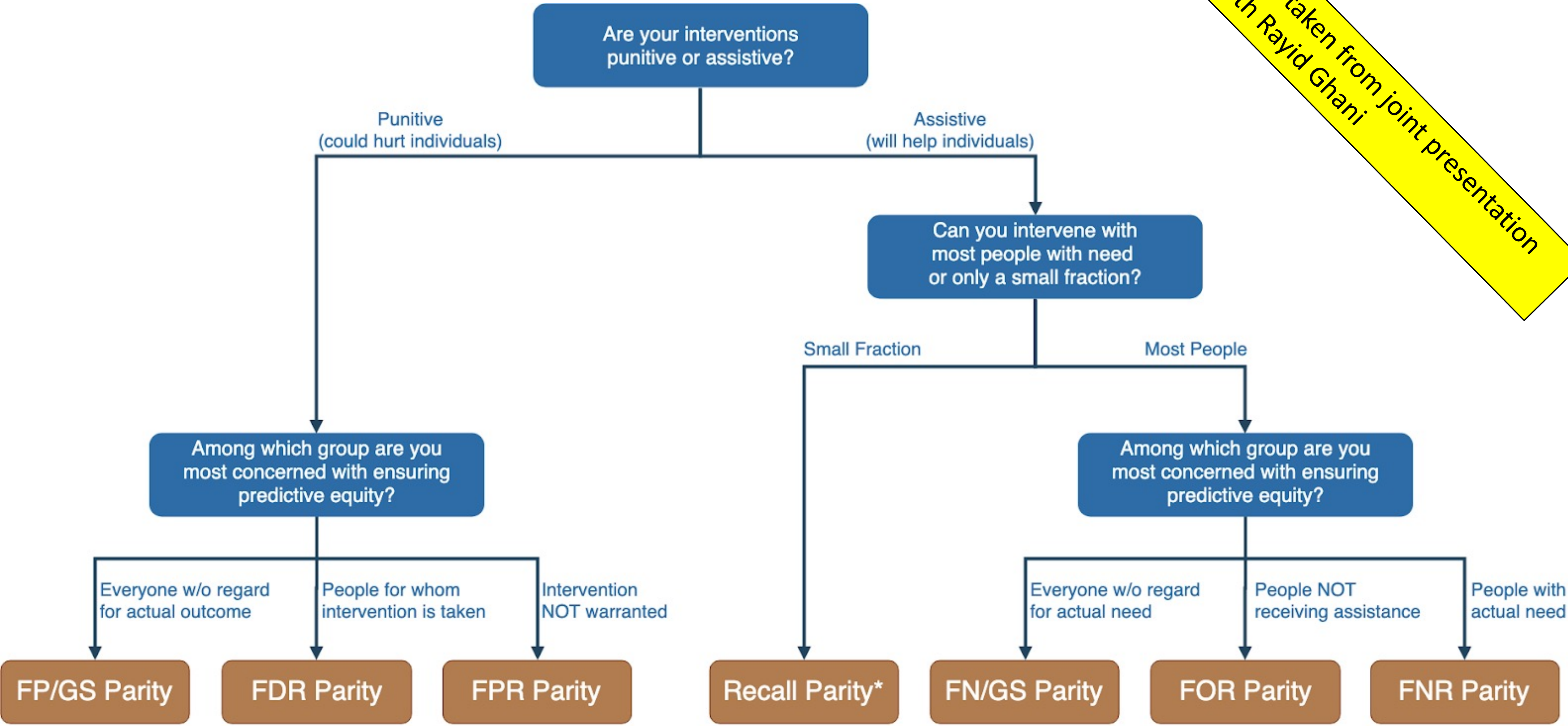
Slide taken from joint presentation with Rayid Ghani

FAIRNESS TREE

Slide taken from joint presentation with Rayid Ghani



Zoomed in Version



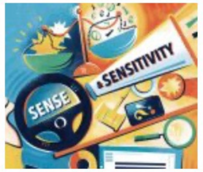
Slide taken from joint presentation with Rayid Ghani

(How) Does this Apply to Insurance?

Fairness and Bias in Actuarial Applications

“During the last two congressional sessions, legislators have introduced bills to eliminate so-called “income proxies” including credit scoring, education level and employment status that could greatly impact how actuaries develop rates. In 2021 three states, Colorado, Michigan and Washington, either enacted legislation or implemented regulation in response to those who insist personal auto insurance rates are unfair or discriminatory.”

actuarialREVIEW CAS



Sense & Sensitivity: Should fairness be a reason to eliminate predictive insurance rating factors?

cover STORY

Sense & Sensitivity: Should fairness be a reason to eliminate predictive insurance rating factors?

BY ANNMARIE GEDDES BARIBEAU

APRIL 5, 2022 BUSINESS AND TECHNOLOGY EMERGING ISSUES

For more than 70 years, insurers and insurance regulators have been sensitive to the issue of potentially discriminatory or unfair rating factors.

Seemingly Obvious Example: Personal Auto Policy sold by a Stock Insurer

- **Possible View**: Insurance pool is a risk sharing device, everyone should pay their expected costs (includes expenses and profits)

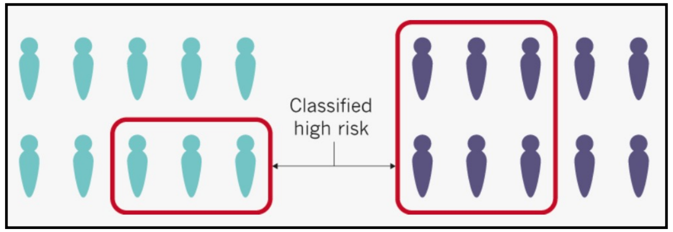
➔Focus on "procedural fairness": Two individuals with the same risk should pay the same premium

➔Procedural Fairness ≈ Actuarial Fairness

➔No need to worry about biases and tradeoffs (?)

- **But: Government mandates car liability coverage, car insurance regulated**

Thought Experiment in Personal Auto



- Risk classes (true – unknown – and imperfectly classified):
  - Low risk
  - High risk
- Two groups of consumers:
  - Protected **Group A**, riskier on average, poorer on average
  - **Group B**, less risky on average, wealthier on average
- Coverage options:
  - (None)
  - Minimum
  - Premium

Who benefits from insurance mandate?

- Without insurance:
  - More likely that member of Group A is at fault
  - More likely that member of Group A can’t pay claim out-of-pocket
  - More likely that member of Group B suffers financial loss (on net)
- With actuarially fair insurance, everyone pays their share

➜Insurance mandate, on net, is a **transfer from Group A to Group B**

…although everyone may be better off because insurance avoids surprises… (“consumption smoothing”)

➜Focus on procedural/actuarial fairness appropriate?

So why not drastically limit risk classification?
(e.g., charge everyone a flat price for coverage)

- Cross subsidization from Group B to Group A – maybe OK?
- Less competition? Limit realized cost savings?

- Issues around **adverse selection** and **moral hazard**:
  - All risky participants will sort into higher coverage
  - Premiums will increase

➜Low risk participants, also and especially from Group A, worse off
  - Possible that even Group A, as a whole, is worse off (“welfare”)
  - Even more pronounced for Group B

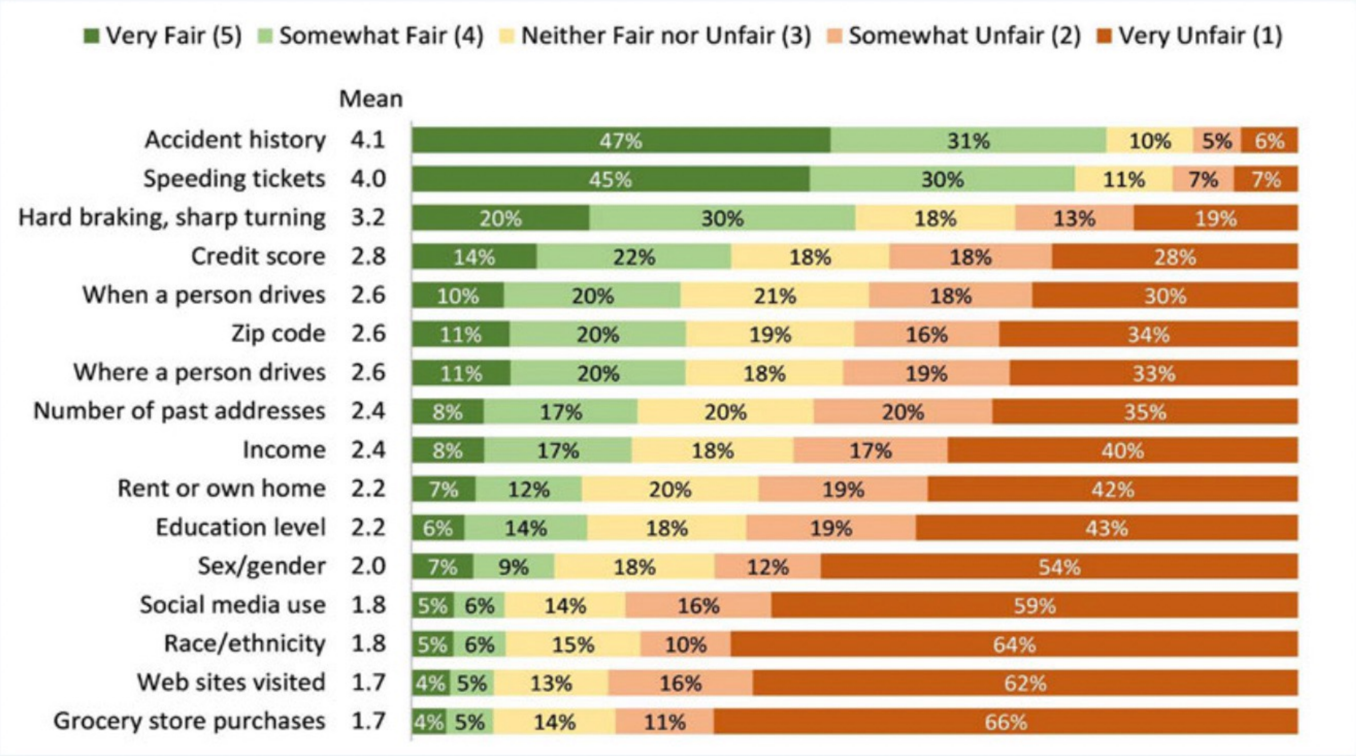
Not simple, (normative) tradeoffs we have to navigate...

What variables to discriminate on (Avraham, 2018; Prince & Schwarcz, 2020):

- If I control the characteristic, it is ok to use it
- If the variable changes over time (mutable, e.g. age), ok to use it (benefit at some point)
- Acceptable if a variable causes an insurance event (cancer in life insurance)
- More acceptable if correlation is higher (better predictors)
- Avoid if variable reinforces existing discrimination
- If inclusion inhibits socially valuable behavior, don’t use

→But there are many grey areas and algorithms are smart (proxy or indirect discrimination)

Figure 1. How Americans rate the fairness of companies using various types of data in car insurance decisions.



Notes: *Survey conducted by YouGov for the author February 11 to 14, 2019. N = 1, 095. Values weighted to be nationally representative.*

Source: *Barbara Kiviat, "Which Data Fairly Differentiate? American Views on the Use of Personal Data in Two Market Settings," Sociological Science 8: 26-47. © 2021.*

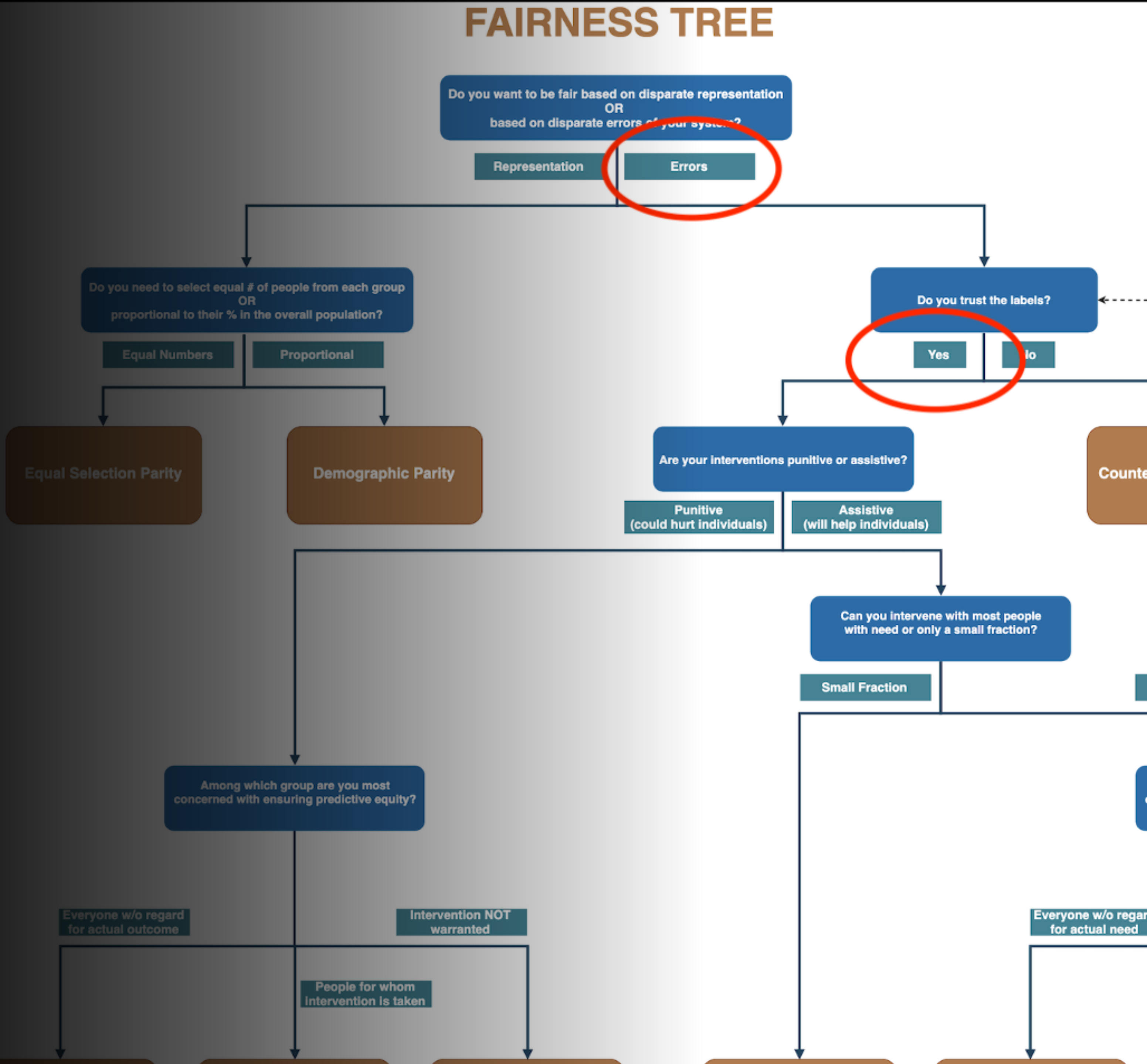
Discriminatory to use location if people in low cost, high crime neighborhood can't move?

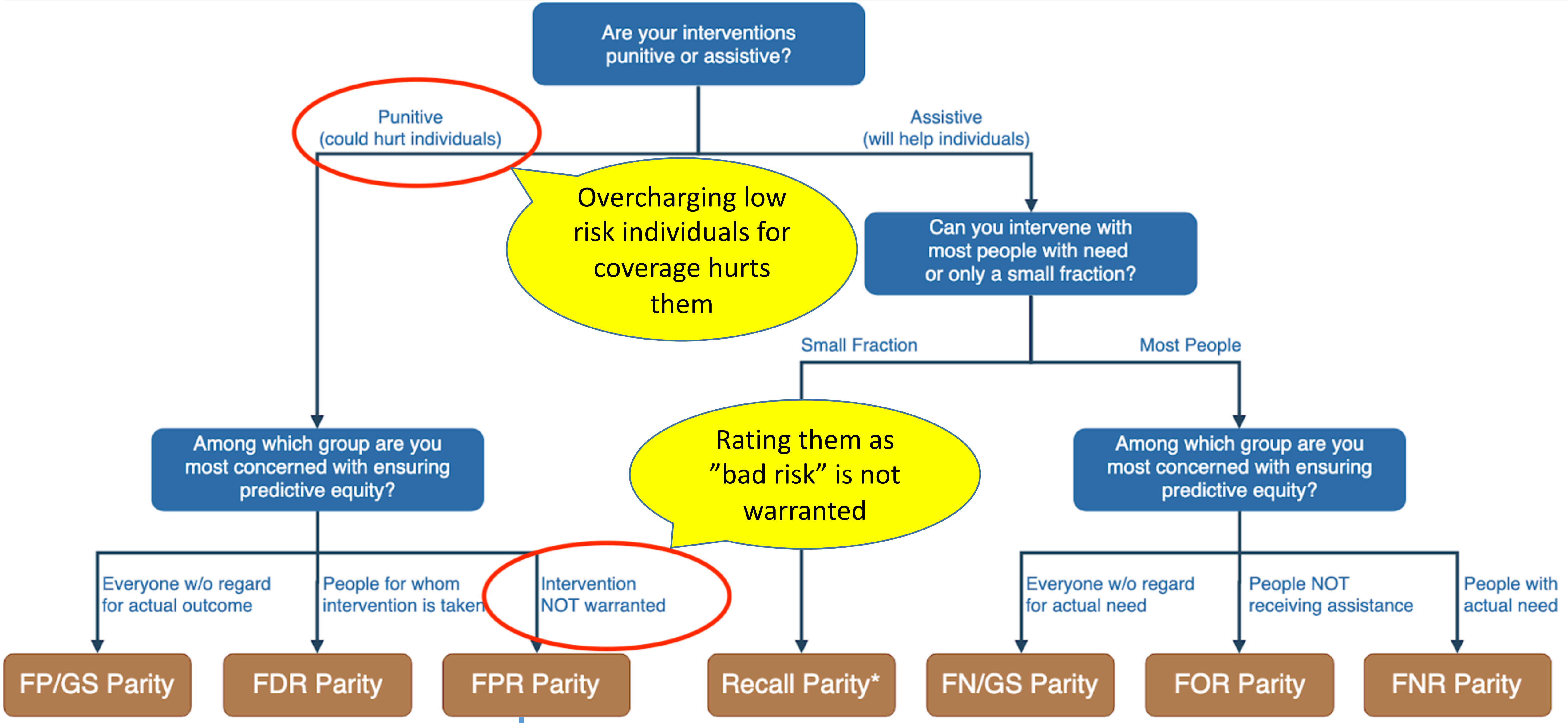
Discriminatory to use location if people in low cost, high crime neighborhood can't move?

Telematics: Tracking time of day unfair to blue collar workers who are more likely to be working at night. When a person drives was considered less fair than credit scoring. (Kiviat study)

OK, let's worry about fairness in ML algorithms – what do we care about?

Rayid Ghani's Tree





Rate of individuals erroneously classified as bad drivers should be the same in Groups A and B (or, in other words, you should not get penalized for being a member of Group A)

How to ensure FPR Parity?

IBM Research Trusted AI | Home | Demo | Reso

AI Fairness 360

This extensible open source toolkit can help you examine, report, and mitigate discrimination and bias in machine learning models throughout the AI application lifecycle. We invite you to use and improve it.

Python API Docs ↗ | Get Python Code ↗ | Get R Code ↗

Not sure what to do first? Start here!

Aequitas

An open source bias audit toolkit for machine learning developers, analysts, and policymakers to audit machine learning models for discrimination and bias, and make informed and equitable decisions around developing and deploying predictive risk-assessment tools.

TRY IT NOW!

```
install.packages('fairness')
library(fairness)
```

arXiv.org > cs > arXiv:1909.05167

Search...
Help | Adv

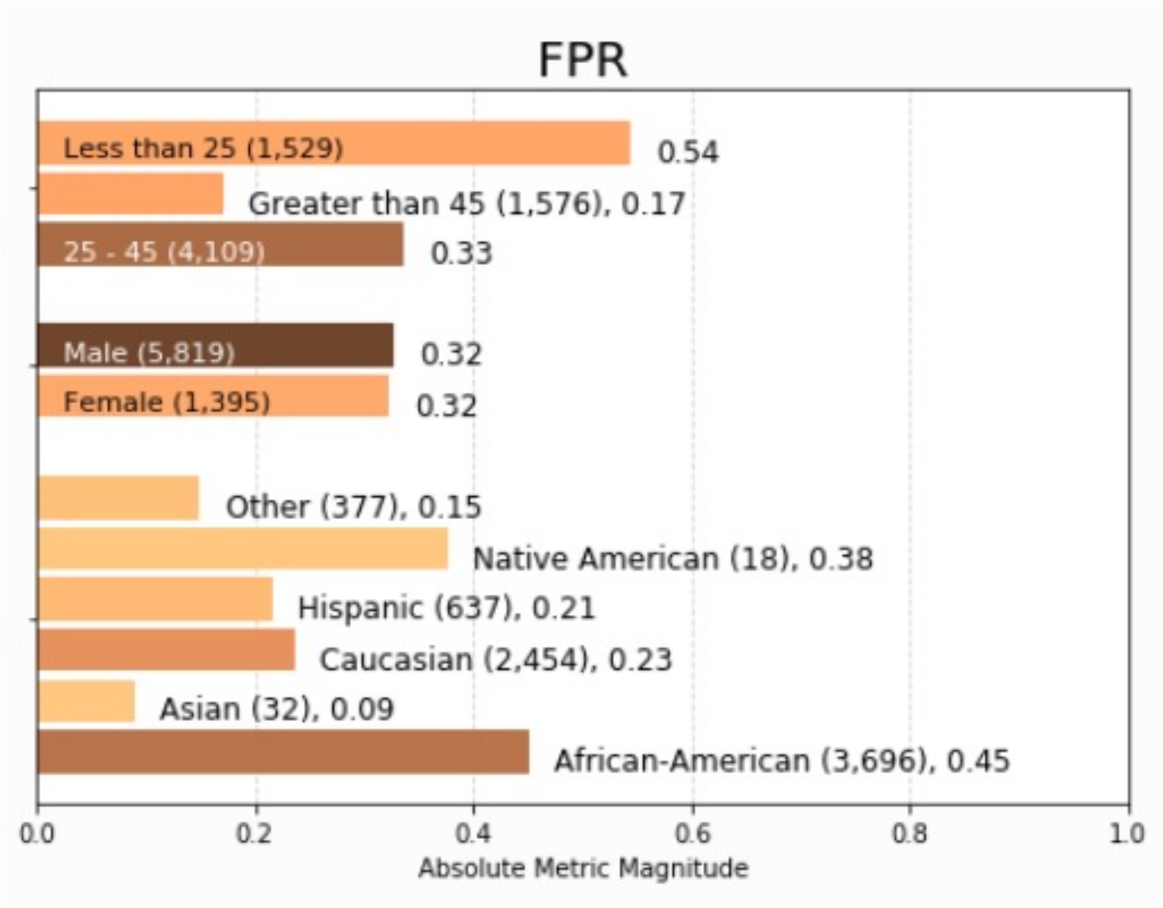
Computer Science > Machine Learning

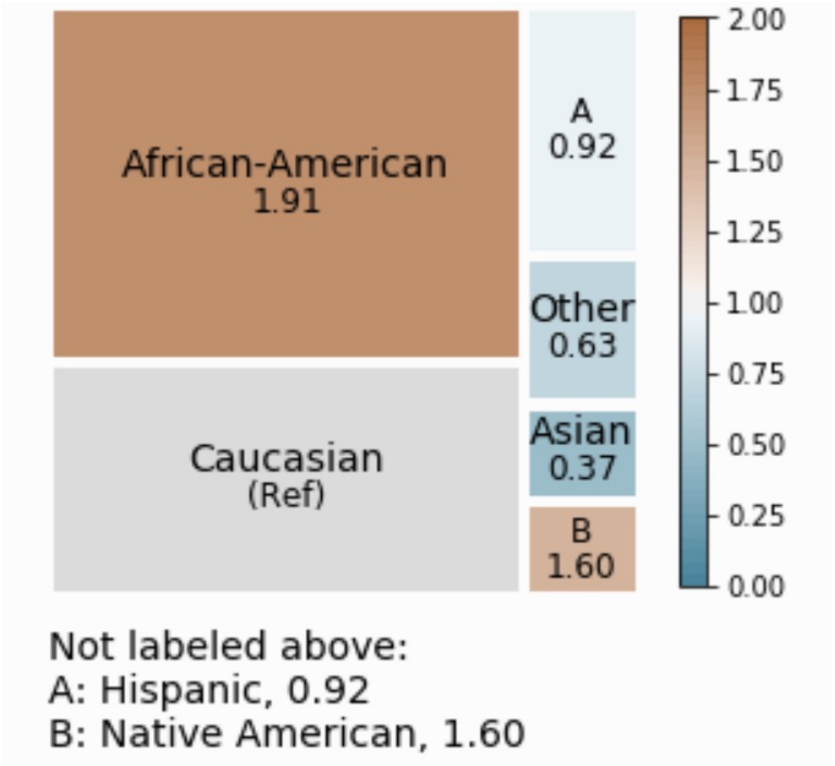
[Submitted on 11 Sep 2019]

FAT Forensics: A Python Toolbox for Algorithmic Fairness, Accountability and Transparency

Kacper Sokol, Raul Santos-Rodriguez, Peter Flach

[Compas Data using Aequitas]

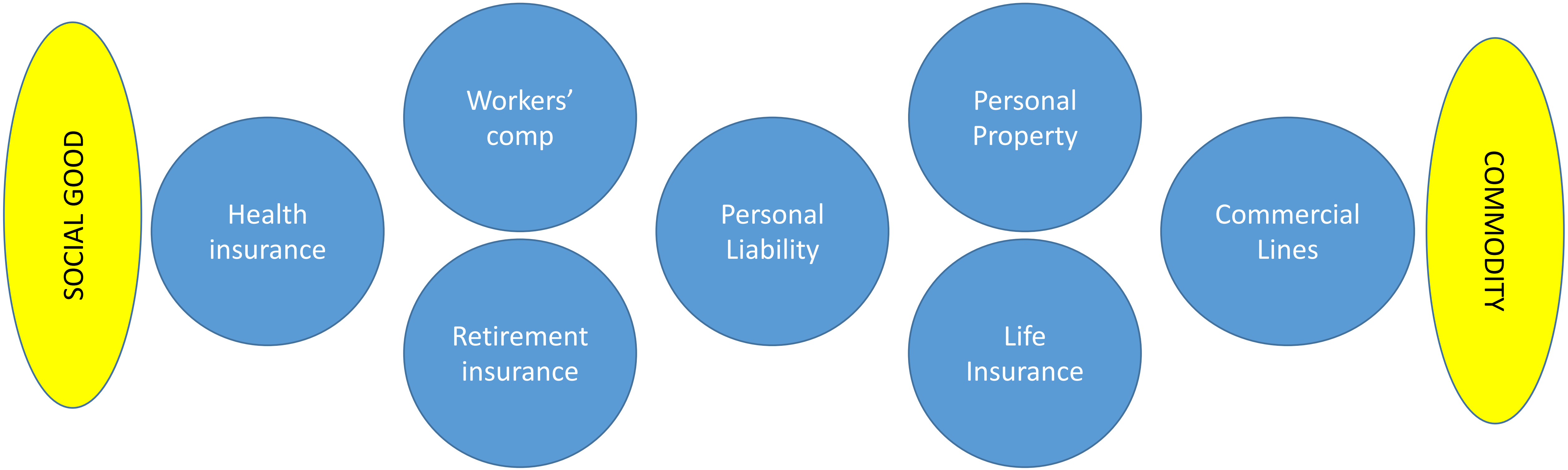




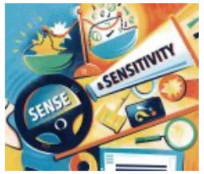
→ Can compare for different models, cutoffs

Is Fairness viable?

- If ascertaining desired fairness possible at <u>low cost</u> regarding accuracy, possibly yes!
  - But what is low cost? And how does one convince competitors?
- Likely depends on type of insurance:



actuarialREV



Sense & Sensitivity
fairness be a reasor
predictive insurance
factors?

“Nobody knows what constitutes an acceptable balance of correlation to a protected class versus correlation to a business operation [...] And there isn’t even data for many of the protected classes to even begin the analysis [...Laws...] will have a negative impact on all companies and especially smaller companies who would have to comply with the law.” (Dave Snyder, APCIA)

cover STORY

“While assuring fairness to everyone’s satisfaction is a laudable objective worthy of pursuit, it is elusive by its very nature. Fairness, or impartiality, can be a matter of perception.”

Sense & Sensitivity: Should fairness be a reason to eliminate predictive insurance rating factors?

“Developing fair rates requires a sensitive balance between multiple rating factors to assure fairness to policyholders while helping insurers achieve business goals.”

BY ANNMARIE GEDDES BARIBEAU

APRIL 5, 2022 BUSINESS AND TECHNOLO

For more than 70 years, insurers and insurance regulators have been sensitive to the issue of potentially discriminatory or unfair rating factors.

Methods for Quantifying Discriminatory Effects on Protected Classes in Insurance

By Roosevelt Mosley, FCAS, CSPA and Radost Wenman, FCAS

As the insurance industry focuses attention on potential racial bias across all practice areas, this paper examines three approaches to defining and measuring fairness in predictive models. It also provides an overview of several bias mitigation techniques that can be performed during the input, modeling, or output phase of a model once a set of fairness criteria has been adopted.

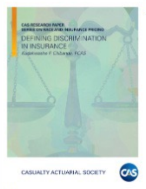
Read More

Approaches to Address Racial Bias in Financial Services: Lessons for the Insurance Industry

By Members of the 2021 CAS Race and Insurance Research Task Force

This paper examines issues of racial bias in lending practice for mortgages, personal and commercial lending, as well as credit-scoring. It looks at these four areas and describes solutions intended to address any potential bias, which may include government intervention, internal bias testing and monitoring measures, and development of new products to mitigate bias.

Read More



Defining Discrimination in Insurance

By Kudakwashe F. Chibanda, FCAS

This paper defines several terms that are currently being used in discussions around potential discrimination in insurance – protected class, unfair discrimination, proxy discrimination, disparate impact, disparate treatment, and disproportionate impact – and provides historical and practical context for them. It also illustrates the inconsistencies in how different stakeholders define these terms.

Read More

Understanding Potential Influences of Racial Bias on P&C Insurance: Four Rating Factors Explored

By Members of the 2021 CAS Race and Insurance Research Task Force

This paper examines four commonly used rating factors in personal lines insurance – credit-based insurance score, geographic location, home ownership, and motor vehicle record – to understand how the data underlying insurance pricing models may be impacted by racially biased policies and practices outside of the system of insurance.

Read More

https://www.casact.org/publications-research/research/research-paper-series-race-and-insurance-pricing

North American Actuarial Journal ›
Latest Articles

Submit an article | Journal homepage

Enter keywords, authors

Listen

Feature Articles

The Discriminating (Pricing) Actuary

Edward W. (Jed) Frees & Fei Huang

Published online: 06 Aug 2021

Download citation https://doi.org/10.1080/10920277.2021.1951296 Check for updates

DISCRIMINATION-FREE INSURANCE PRICING

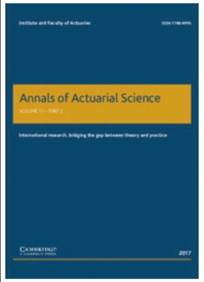
Published online by Cambridge University Press: **07 October 2021**

M. Lindholm, R. Richman, A. Tsanakas and M.V. Wüthrich

Show author details

Article Figures Metrics

Save PDF | Share | Cite | Rights & Permissions



AI: Coming of age?

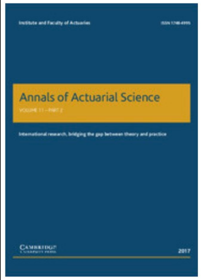
Published online by Cambridge University Press: **19 January 2022**

Trevor Maynard, Luca Baldassarre, Yves-Alexandre de Montjoye, Liz McFall and María Óskarsdóttir

Show author details

Article Metrics

Save PDF | Share | Cite | Rights & Permissions



Multidisciplinary collaboration on discrimination – not just "Nice to Have"

Published online by Cambridge University Press: **01 November 2021**

Chris Dolman, Edward (Jed) Frees and Fei Huang

Show author details

Article Metrics

The need for AI governance and auditing

An emerging data science sub-profession: the algorithm auditor.

- Algorithm auditing should be founded on **more than machine learning**. Social science methodology, ethics, regulation, human-centered design should be brought to bear.
- Often the goal is **to identify tradeoffs** that must be deliberated at societal levels. (e.g., sensitivity/specificity; tradeoffs in different conceptions of "fairness")
- "Since actuaries are intimately acquainted with rating factors and the data behind them and are required to uphold the highest **standards of professional independence**, they should have a greater voice in the rating variable conversation." (A. Baribeau)
- Algorithm auditing should ultimately become the purview of **a learned (data science) profession** with proper credentialing, standards of practice, disciplinary procedures, ties to academia, continuing education, training in ethics, regulation, and professionalism

Harvard Business Review

ECONOMICS & SOCIETY

Why We Need to Audit Algorithms

by James Guszcza , Iyad Rahwan , Will Bible , Manuel Cebrian and Vic Katyal

November 28, 2018