Ethics, Public Policy, and Technological Change

Rob Reich

Mehran Sahami

Head TA: Roberta Fischli

Housekeeping

- Friday sections will discuss the algorithmic decision making case study
- Technical assignment due Tuesday, Jan 27.
- Philosophy paper assignment released next week, due Feb. 12

Today's Agenda:
Algorithmic Decision-Making: Rights and Responsibilities

1. What is Fairness? Rawls's Theory of Justice as Fairness
2. What's the point of criminal justice?
3. How to decide whether to use an algorithm
   - Simulating outcomes
   - Implementing and measuring outcomes
4. Who should decide?

A short philosophical interlude and recommendation:

What is Socrates most famous saying?

The unexamined life is not worth living.

The unlived life is not worth examining.

What did you do over the long MLK holiday weekend?


Pfeiffer Beach, Big Sur

Today's Agenda:
Algorithmic Decision-Making: Rights and Responsibilities

1. What is Fairness? Rawls's Theory of Justice as Fairness
2. What's the point of criminal justice?
3. How to decide whether to use an algorithm
   - Simulating outcomes
   - Implementing and measuring outcomes
4. Who should decide?

Rawls's Justice as Fairness: Basic Framework

John Rawls's groundbreaking theory addresses **domestic justice** within a single society, not global justice across nations. His framework treats society as a cooperative venture designed for mutual advantage among its members.

Distributive Justice

Focuses on fair distribution of rights, opportunities, and resources among society's members.

Pure Procedural Justice

Principles of justice emerges from a fair procedure rather than predetermined outcomes or substantive criteria.

Alternative to Utilitarianism

Challenges utilitarian ethics by respecting individual separateness—utilitarianism "fails to take seriously the distinction between persons."

Society as a cooperative venture for mutual advantage?

From: Steve Jobs, sjobs@apple.com
To: Steve Jobs, sjobs@apple.com
Date: Thursday, September 2, 2010 at 11:08PM

I grow little of the food I eat, and of the little I do grow
I did not breed or perfect the seeds.

I do not make any of my own clothing.

I speak a language I did not invent or refine.

I did not discover the mathematics I use.

I am protected by freedoms and laws I did not conceive
of or legislate, and do not enforce or adjudicate.

I am moved by music I did not create myself.

When I needed medical attention, I was helpless
to help myself survive.

I did not invent the transistor, the microprocessor,
object oriented programming, or most of the technology
I work with.

I love and admire my species, living and dead, and am
totally dependent on them for my life and well being.

How to Select Principles of Justice?

Rawls proposes a social contract approach: the correct principles of justice are those that **all persons would rationally agree to** under fair conditions.

> "Principles that free and rational persons concerned to further their own interests would accept in an initial position of equality as defining the fundamental terms of their association."
>
> —Rawls, *A Theory of Justice*, p. 10

This agreement must occur in conditions of genuine equality, where no party has unfair advantages in negotiating the social contract.

Problem: Designing the Choice Situation

The Challenge

How can we create conditions that allow rational self-interest to operate while neutralizing underlying inequalities and differential bargaining power?

Sources of Inequality

Natural Talents

People possess different innate abilities, intelligence, and physical capacities through no merit of their own.

Inherited Positions

Individuals are born into different social classes, economic circumstances, and family situations.

Bargaining Power

These morally arbitrary factors translate into unequal influence in determining society's fundamental principles.

Rawls's Answer: The Original Position

> **"The idea of the original position is to set up a fair procedure so that any principles agreed to will be just."**
>
> —John Rawls, *A Theory of Justice*, p. 118

This foundational concept explains why Rawls calls his theory **Justice as Fairness**. The fairness of the procedure guarantees the justice of the resulting principles—if the process is fair, the outcome must be just.

The Veil of Ignorance

In the Original Position, rational individuals choose principles of justice without knowing their personal circumstances. This **veil of ignorance** ensures impartial reasoning.

Social Position

Your place in society remains unknown

Natural Abilities

Your talents, intelligence, and strength are concealed

Class Status

Your economic class and social status are hidden

Personal Identity

You don't know your race, sex, or gender identity

Life Goals

Your conception of the good and what ends you'll pursue remain unknown

Achieving a Fair Choice Situation

> "This ensures that no one is advantaged or disadvantaged in the choice of principles by the outcome of natural chance or the contingency of social circumstance. Since all are similarly situated and no one is able to design principles to favor his particular condition, the principles of justice are the outcome of a fair agreement."

—Rawls, §3, p. 11

Veil of Ignorance

Eliminates knowledge of personal circumstances

Equal Situation

All parties share identical conditions and information

Fair Agreement

Principles emerge from genuine equality and freedom

Under these carefully constructed conditions, people are genuinely free and equal, ensuring any agreement reflects justice rather than power.

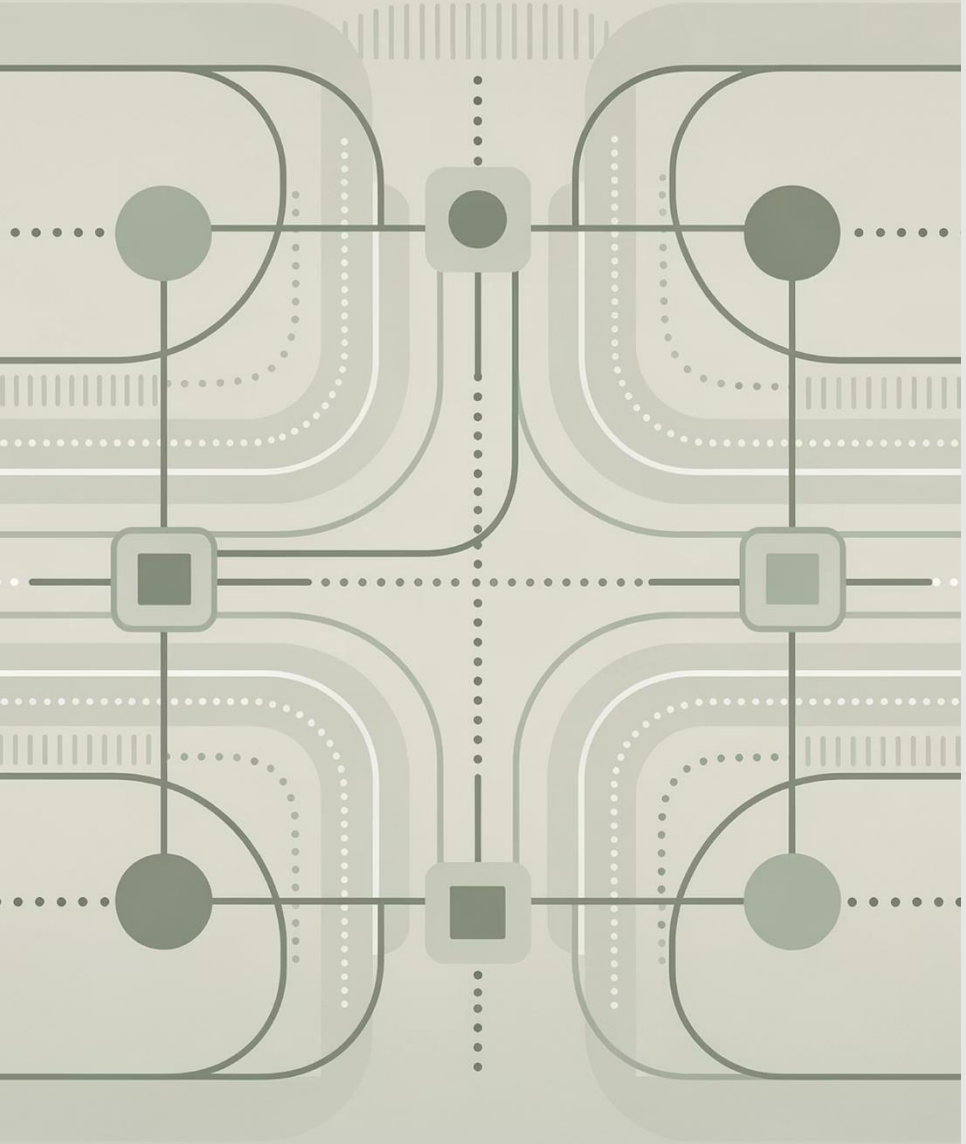
Rawls's Two Principles of Justice

1. The Liberty Principle

Equal right to the most extensive basic liberties for all, compatible with similar liberties for others.

2. The Difference Principle

- Social and economic inequalities must benefit the least advantaged.
- Positions and offices must be open to all under fair equality of opportunity.



Concluding Questions: Algorithmic Fairness

Rawls's framework raises critical questions for our technological age. How should we design algorithmic systems to promote fairness and avoid bias?

Substance vs. Process

Should algorithmic fairness focus on fair outcomes or fair procedures? Individual rights or group equality?

Competing Values

What happens when predictive accuracy conflicts with fairness, transparency, privacy, explainability, or due process?

Power Dynamics

How does algorithmic decision-making reinforce or challenge existing structures of power and inequality in society?

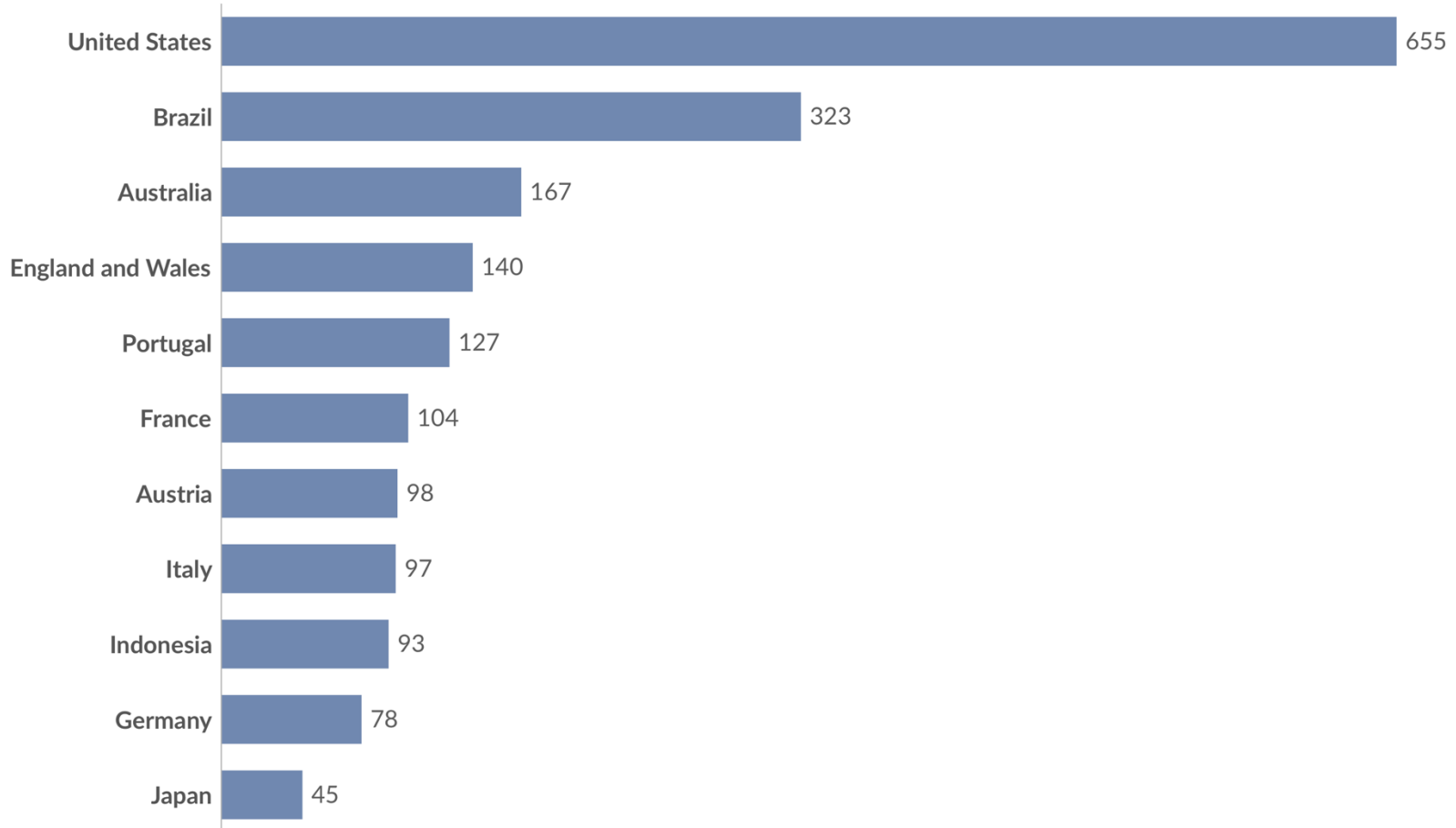
Today's Agenda:
Algorithmic Decision-Making: Rights and Responsibilities

1. What is Fairness? Rawls's Theory of Justice as Fairness
2. What's the point of criminal justice?
3. How to decide whether to use an algorithm
   - Simulating outcomes
   - Implementing and measuring outcomes
4. Who should decide?

Background on Criminal Justice

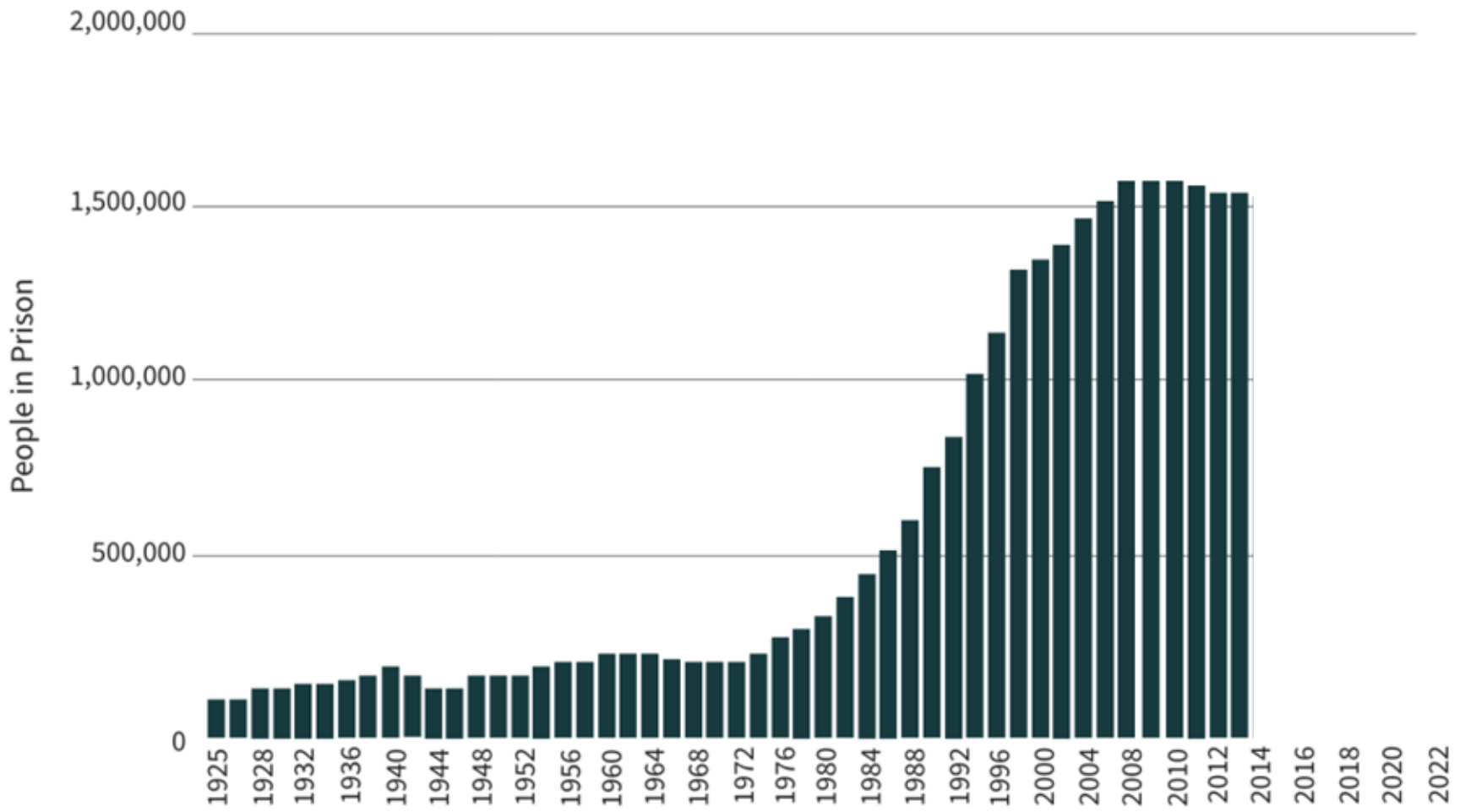
Prison population rate

Number of prisoners per 100,000 people. Shown is data for 2018 or the latest available data before 2018.



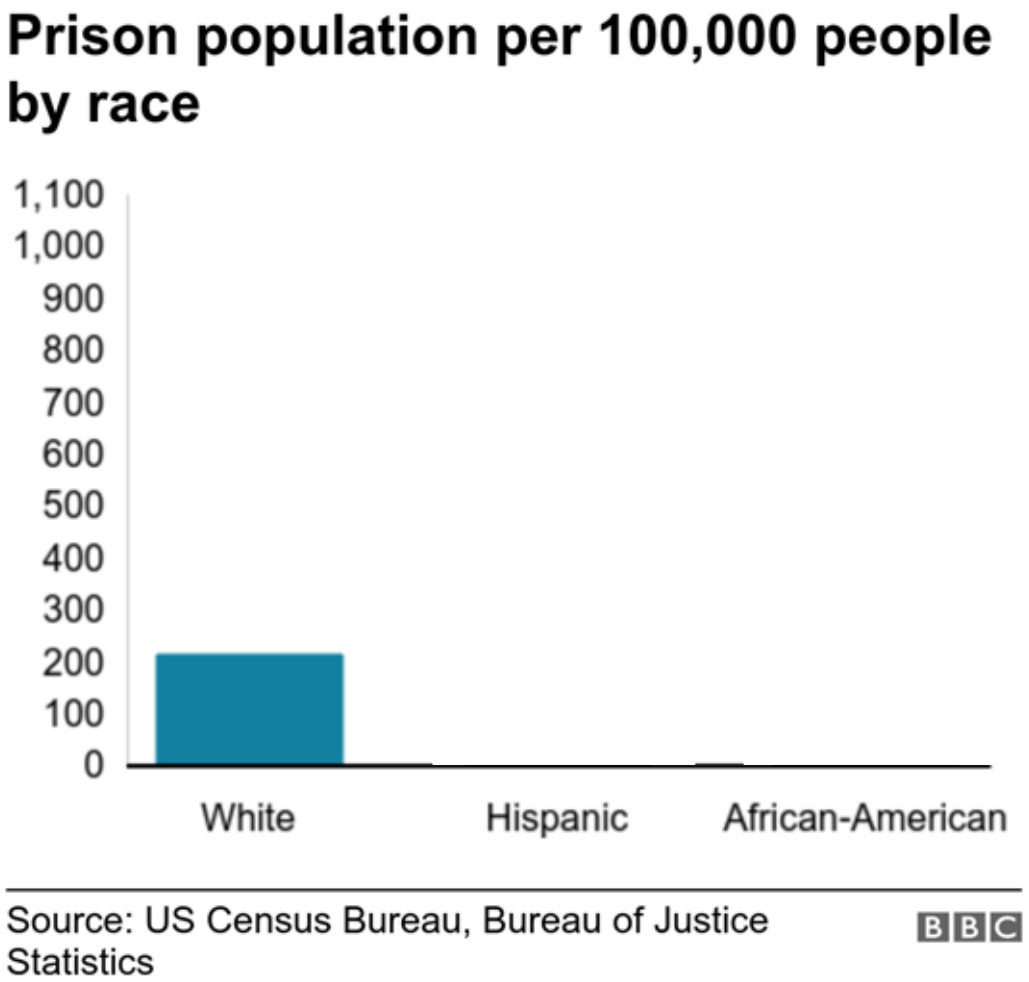
Data source: World Prison Brief (2018)

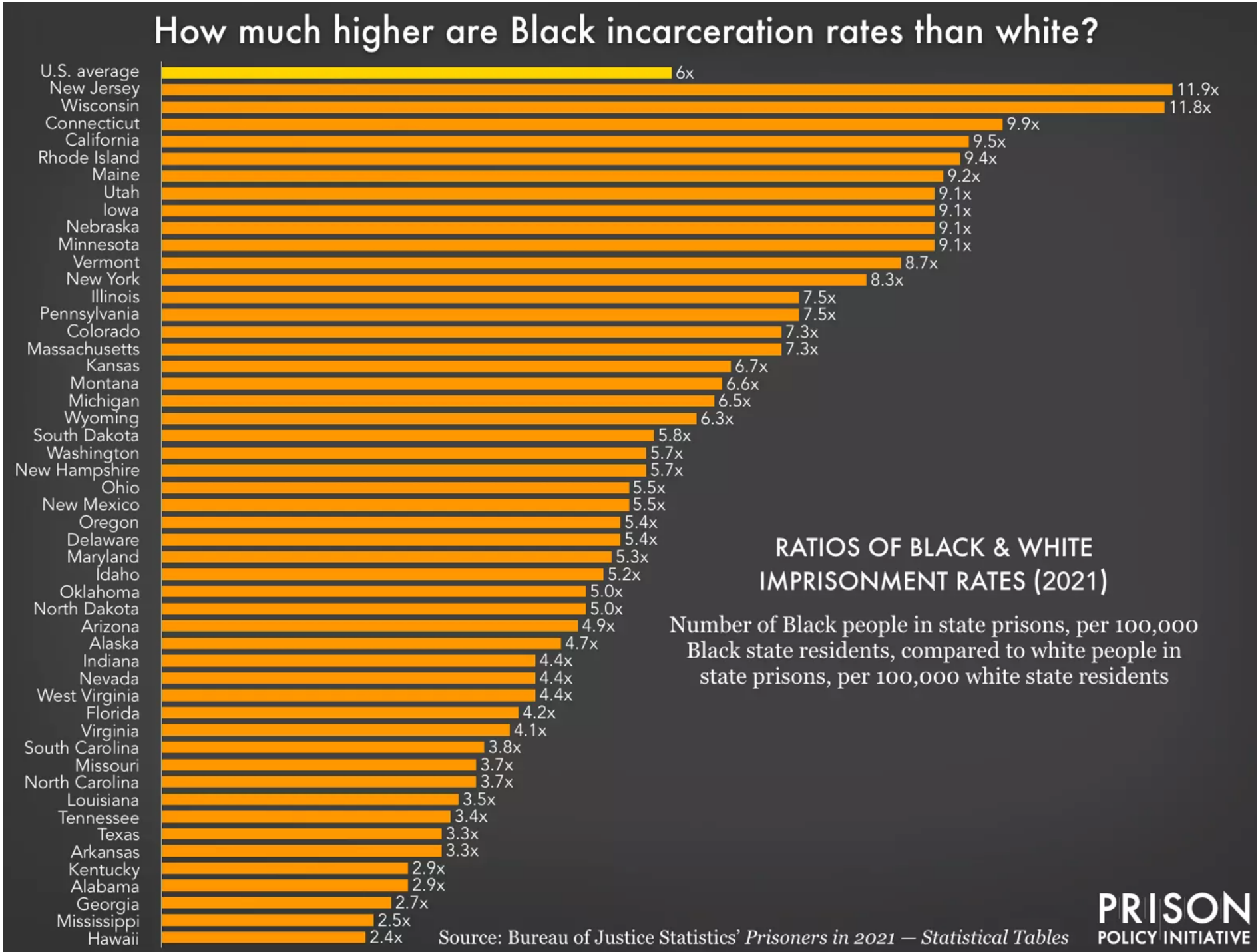
Mass Incarceration



Source: Sentencing Project

Incarceration by Race




How much higher are Black incarceration rates than white?
U.S. average 6x
New Jersey 11.9x
Wisconsin 11.8x
Connecticut 9.9x
California 9.5x
Rhode Island 9.4x
Maine 9.2x
Utah 9.1x
Iowa 9.1x
Nebraska 9.1x
Minnesota 9.1x
Vermont 8.7x
New York 8.3x
Illinois 7.5x
Pennsylvania 7.5x
Colorado 7.3x
Massachusetts 7.3x
Kansas 6.7x
Montana 6.6x
Michigan 6.5x
Wyoming 6.3x
South Dakota 5.8x
Washington 5.7x
New Hampshire 5.7x
Ohio 5.5x
New Mexico 5.5x
Oregon 5.4x
Delaware 5.4x
Maryland 5.3x
Idaho 5.2x
Oklahoma 5.0x
North Dakota 5.0x
Arizona 4.9x
Alaska 4.7x
Indiana 4.4x
Nevada 4.4x
West Virginia 4.4x
Florida 4.2x
Virginia 4.1x
South Carolina 3.8x
Missouri 3.7x
North Carolina 3.7x
Louisiana 3.5x
Tennessee 3.4x
Texas 3.3x
Arkansas 3.3x
Kentucky 2.9x
Alabama 2.9x
Georgia 2.7x
Mississippi 2.5x
Hawaii 2.4x
RATIOS OF BLACK & WHITE IMPRISONMENT RATES (2021)
Number of Black people in state prisons, per 100,000 Black state residents, compared to white people in state prisons, per 100,000 white state residents
Source: Bureau of Justice Statistics' Prisoners in 2021 — Statistical Tables
PRISON POLICY INITIATIVE

A Moment of Reckoning

Source: Brett Sayles

What's the Purpose of Criminal Justice/Imprisonment?

1. Do we need prisons?
   - Some democracies practice punishment without prisons
2. Deterrence
   - Prevent future crime (by others)
3. Incapacitation
   - Prevent future crime (by the accused)
4. Retribution
   - Hold perpetrators responsible for criminal behavior
5. Rehabilitation ("Correction")
   - Help offenders return to society as law-abiding citizens

More Narrowly, Bail Systems and Decisions

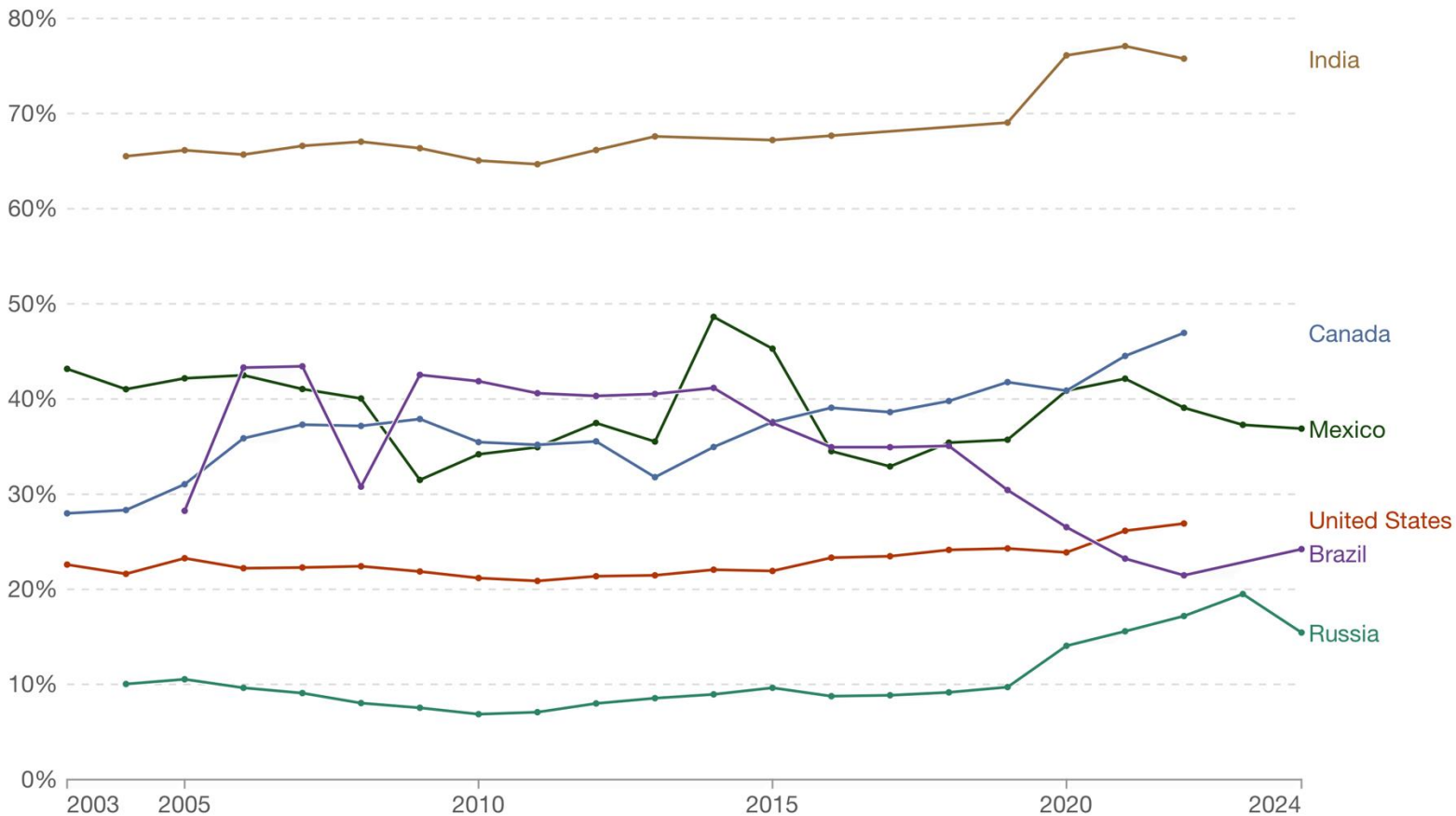
- Bail: release defendant with conditions to ensure court appearance and protect public safety
- Why have bail system at all?

- What criteria would you use if you had to make a decision about whether someone should be released on bail?
- Can algorithms help us make better decisions? What problem are we solving?

Unsentenced detainees as a proportion of overall prison population, 2003 to 2024

Our World in Data

People held in detention who have not yet been sentenced as a percentage of all people held in detention.



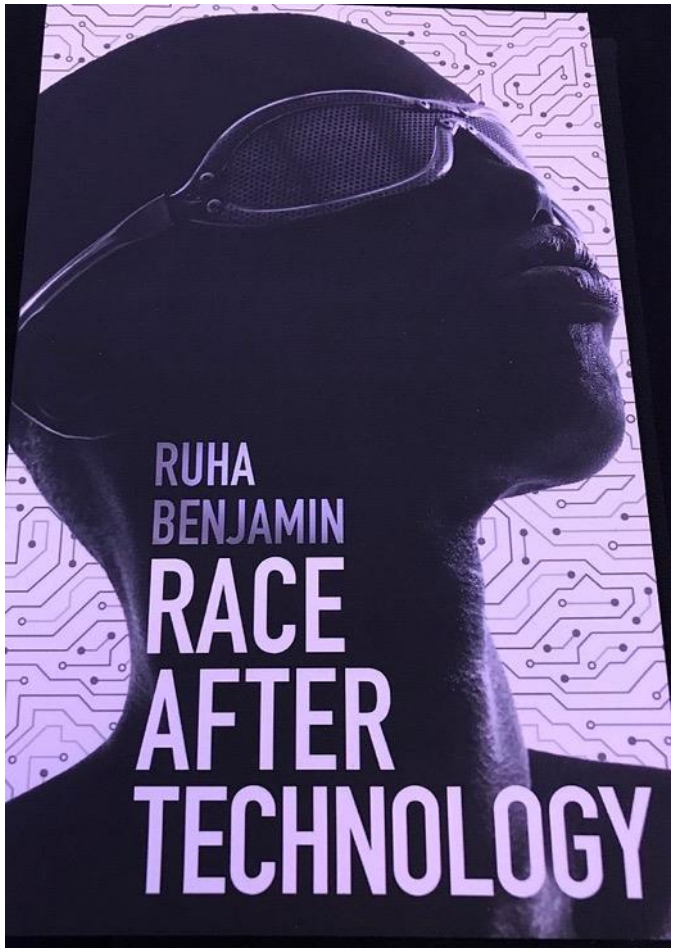
Data source: United Nations Office on Drugs and Crime

OurWorldinData.org/human-rights | CC BY

Fairness Concerns in Pre-Trial Imprisonment

- Presumption of innocence?
  - Punishment before a finding of guilt
- Discrimination on the basis of wealth
  - Bail decisions frequently made on the basis of ability to pay
    - Poor person charged with minor offense is imprisoned; rich person charged with a serious offense pays bail
- Cascading consequences
  - Job loss
  - Family separation

Technology to the Rescue?



Source: Anoushnajarian, CC-BY-4.0

Benjamin warns about the dangers of discriminatory designs that:

“explicitly work to amplify hierarches…”

“ignore and thus replicate social divisions”

“aim to fix racial bias but end up doing the opposite”

She calls this: **the New Jim Code**.

Expanding Our Imagination

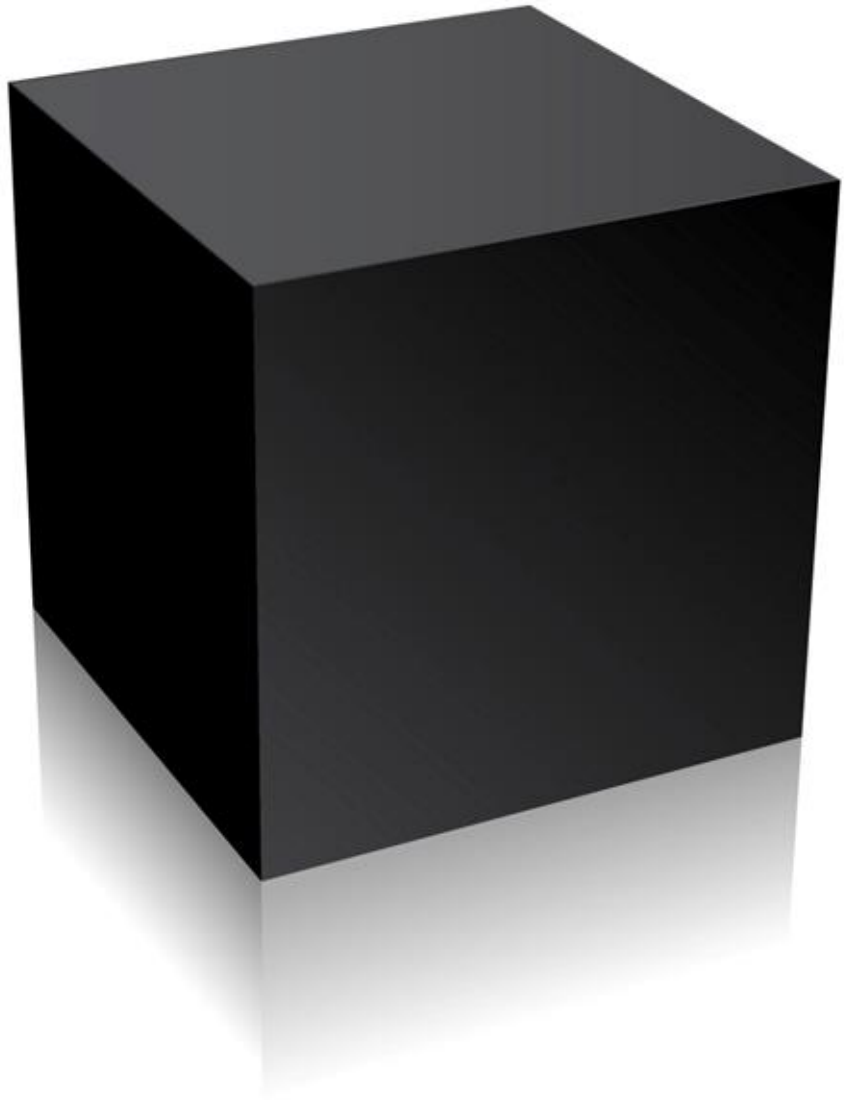
“Promise… is dangerous and insidious precisely because it is packaged as social betterment.”

“But if this company is to genuinely contribute to decarceration, it would need to shrink the carceral apparatus, not extend it and make it more encompassing.”

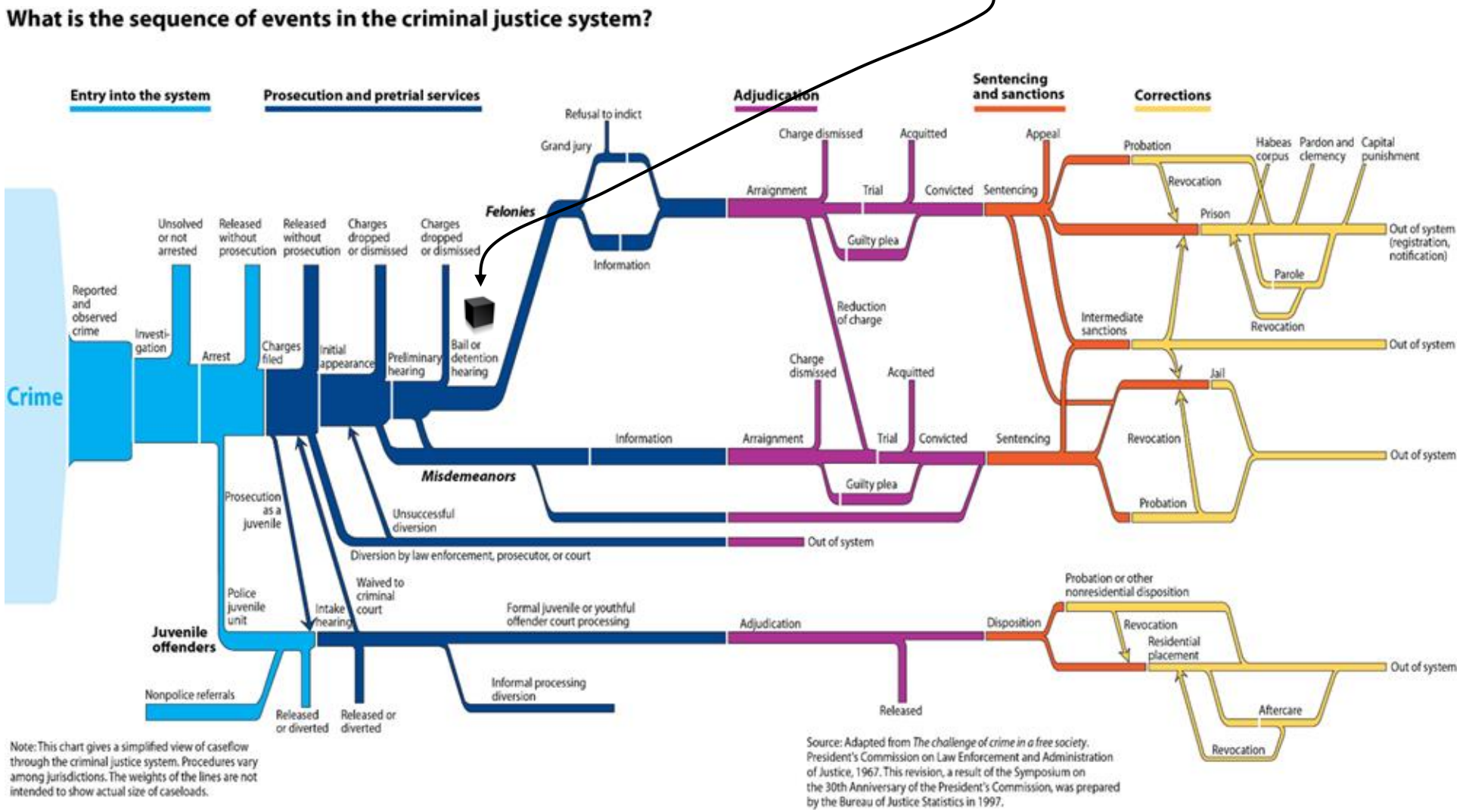
-- Ruha Benjamin, Race After Technology

We should be careful to avoid a situation in which technical fixes undermine the impetus toward more far-reaching aims.

The Concern



The Reality


What is the sequence of events in the criminal justice system?
Entry into the system
Prosecution and pretrial services
Adjudication
Sentencing and sanctions
Corrections
Crime
Reported and observed crime
Unsolved or not arrested
Investi-gation
Released without prosecution
Arrest
Released without prosecution
Charges filed
Charges dropped or dismissed
Initial appearance
Charges dropped or dismissed
Preliminary hearing
Bail or detention hearing
Refusal to indict
Grand jury
Felonies
Information
Arraignment
Charge dismissed
Acquitted
Trial
Convicted
Guilty plea
Reduction of charge
Sentencing
Appeal
Probation
Habeas corpus
Pardon and clemency
Capital punishment
Revocation
Prison
Out of system (registration, notification)
Parole
Revocation
Intermediate sanctions
Out of system
Jail
Revocation
Probation
Misdemeanors
Information
Arraignment
Charge dismissed
Acquitted
Trial
Convicted
Guilty plea
Sentencing
Out of system
Prosecution as a juvenile
Unsuccessful diversion
Diversion by law enforcement, prosecutor, or court
Out of system
Juvenile offenders
Police juvenile unit
Intake hearing
Waived to criminal court
Formal juvenile or youthful offender court processing
Adjudication
Disposition
Probation or other nonresidential disposition
Revocation
Residential placement
Out of system
Aftercare
Revocation
Nonpolice referrals
Released or diverted
Released or diverted
Informal processing diversion
Released
Note: This chart gives a simplified view of caseflow through the criminal justice system. Procedures vary among jurisdictions. The weights of the lines are not intended to show actual size of caseloads.
Source: Adapted from The challenge of crime in a free society. President's Commission on Law Enforcement and Administration of Justice, 1967. This revision, a result of the Symposium on the 30th Anniversary of the President's Commission, was prepared by the Bureau of Justice Statistics in 1997.

Today's Agenda:
Algorithmic Decision-Making: Rights and Responsibilities

1. What is Fairness? Rawls's Theory of Justice as Fairness
2. What's the point of criminal justice?
3. How to decide whether to use an algorithm
   - Simulating outcomes
   - Implementing and measuring outcomes
4. Who should decide?

For Discussion

- Imagine that Stanford (or any other selective university) proposes to admit undergraduates on the basis of an algorithmic model
  - Train the model on past applications and student admissions.
  - **What would be the benefits and risks of using automated review of applications?**

Stanford Apologizes for Limiting Jewish Admissions in the 1950s

The university also issued a report that found that some administrators had "regularly misled" people who raised concerns about anti-Jewish bias.

Source: The New York Times, October 13, 2022.

CASE STUDY

ALGORITHMIC DECISION-MAKING AND ACCOUNTABILITY

California's SB 10

STATE OF CALIFORNIA
AUTHENTICATED
ELECTRONIC LEGAL MATERIAL

Senate Bill No. 10

CHAPTER 244

An act to amend Section 27771 of the Government Code, and to add Section 1320.6 to, to add Chapter 1.5 (commencing with Section 1320.7) to Title 10 of Part 2 of, and to repeal Chapter 1 (commencing with Section 1268) of Title 10 of Part 2 of, the Penal Code, relating to pretrial release and detention.

[Approved by Governor August 28, 2018. Filed with Secretary of State August 28, 2018.]

LEGISLATIVE COUNSEL'S DIGEST

SB 10, Hertzberg. Pretrial release or detention: pretrial services.

Sources: https://campaigns.organizefor.org/petitions/bail-reform-now, California State Senate

The Choice

- Eliminate money bail
- Replace it with algorithmic “risk assessments”
- Release under pre-trial, non-monetary conditions
- 58 counties adopt (make or buy) risk assessment tool
- 2023 review for bias

How would you vote? Why? What criteria?

Criteria

1. Does it improve public safety? (Efficiency)
2. Is it fair? (Fairness)
3. Can people understand how it works? (Transparency)
4. Can people appeal its judgment? (Due Process)
5. Does it use information that an individual might reasonably expect to remain private? (Privacy)

A Victory for Fair Treatment?

- "Today, California reforms its bail system so that rich and poor alike are treated fairly…"

  - Governor Jerry Brown, Aug. 28, 2018

Source: California Governor's Office

Our Focus Today

- Last week, Mehran and I discussed issues of fairness in algorithmic decisionmaking, with a focus on COMPAS and criminal justice system
  - Four conceptualizations of fairness: anti-classification, classification parity, calibration, disparate impact
  - Kleinberg's "impossibility theorem"
- Today, we are going to focus on effectiveness/efficiency at improving public safety – **how would we know if we achieve better outcomes (for safety) by using an algorithm?**
- On Monday, we will talk about transparency and due process.

Efficiency Gains

What is the promise of using algorithmic decision-making to make bail decisions?

- It might reduce crime by more effectively/efficiently keeping dangerous criminals off of the streets

- It might reduce bias by limiting the degree to which judges take extraneous or protected characteristics into account in making their judgments

How do we know if the use of an algorithm more effectively keeps dangerous criminals off the street?

Two Questions about Efficiency

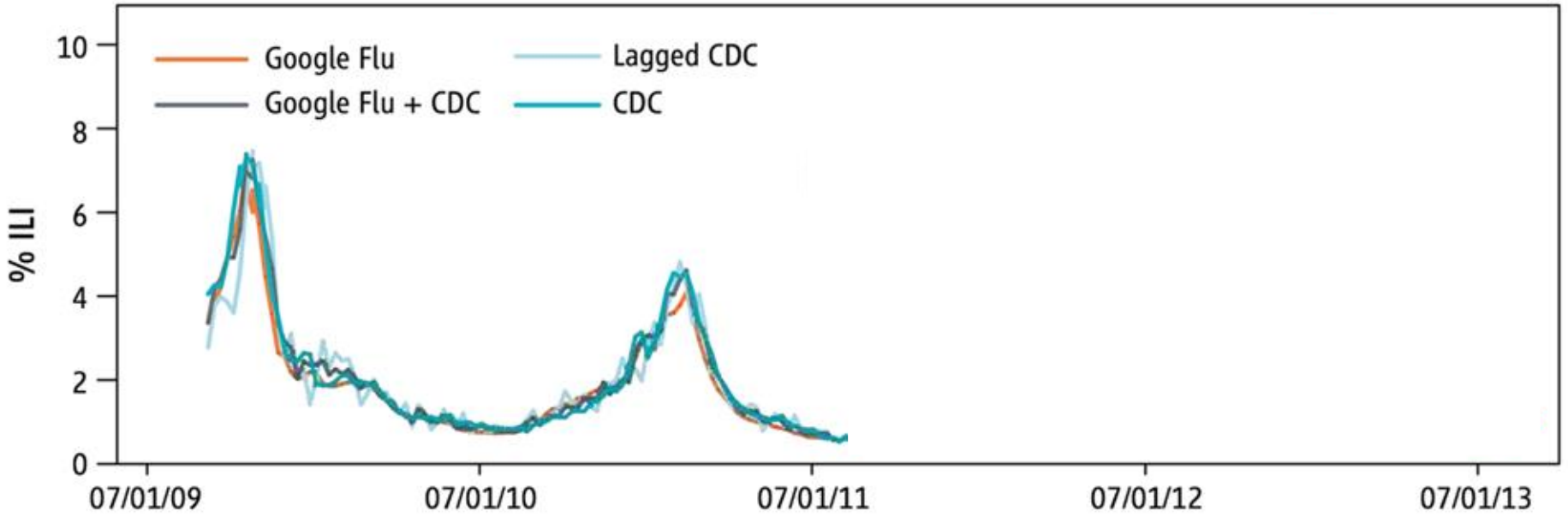
1. Does the algorithm make more accurate decisions?
   - Issue of predictive validity
   - How it compares to other decision-making approaches

2. When introduced into the judicial process, does the use of an algorithm translate into better outcomes with respect to crime and recidivism?
   - Question of real-world impact
   - Depends a great deal on implementation

Predictive Accuracy

Forbes

"Google Flu Trends identifies outbreaks more effectively than doctors can."



Source: Lazer et al., Science (2014)

Beyond Predictive Accuracy

- We typically evaluate machine learning algorithms in terms of their accuracy. We do this by making predictions “out-of-sample”: using part our data to train and part to test.
- But there is a problem in the risk assessment case: we only observe outcomes for those individuals who are released by judges. We do not know whether those who are not released would reoffend or not.
- The implication is that the algorithm may perform well for some test data, but perhaps not in the real world.
- This is especially plausible in the bail case if judges have access to information about defendants that the algorithm does not (e.g. judges see gang tattoos on young people and only release those without them, who are less likely to reoffend).

Comparing Algos & Judges

- The better question is whether the algorithm improves upon *judges'* decisions. But how can we evaluate this when we don't know whether a jailed defendant would have committed a crime if released?
- Kleinberg et al (2017) focus on defendants who were not jailed, and ask whether the algorithm could do better in those cases (by jailing additional defendants released by lenient judges) as compared to strict judges
- The results are pretty compelling:
  - The riskiest 1% of defendants, when released, fail to appear at a rate of 56% and 62% are rearrested – yet judges release nearly half of them
  - Stricter judges tend to jail defendants from throughout the risk distribution; if they jailed those with higher risk scores, they could jail half as many people and achieve the same predicted level of crime

Algorithms vs. Anybody?

TECHNOLOGY

A Popular Algorithm Is No Better at Predicting Crimes Than Random People

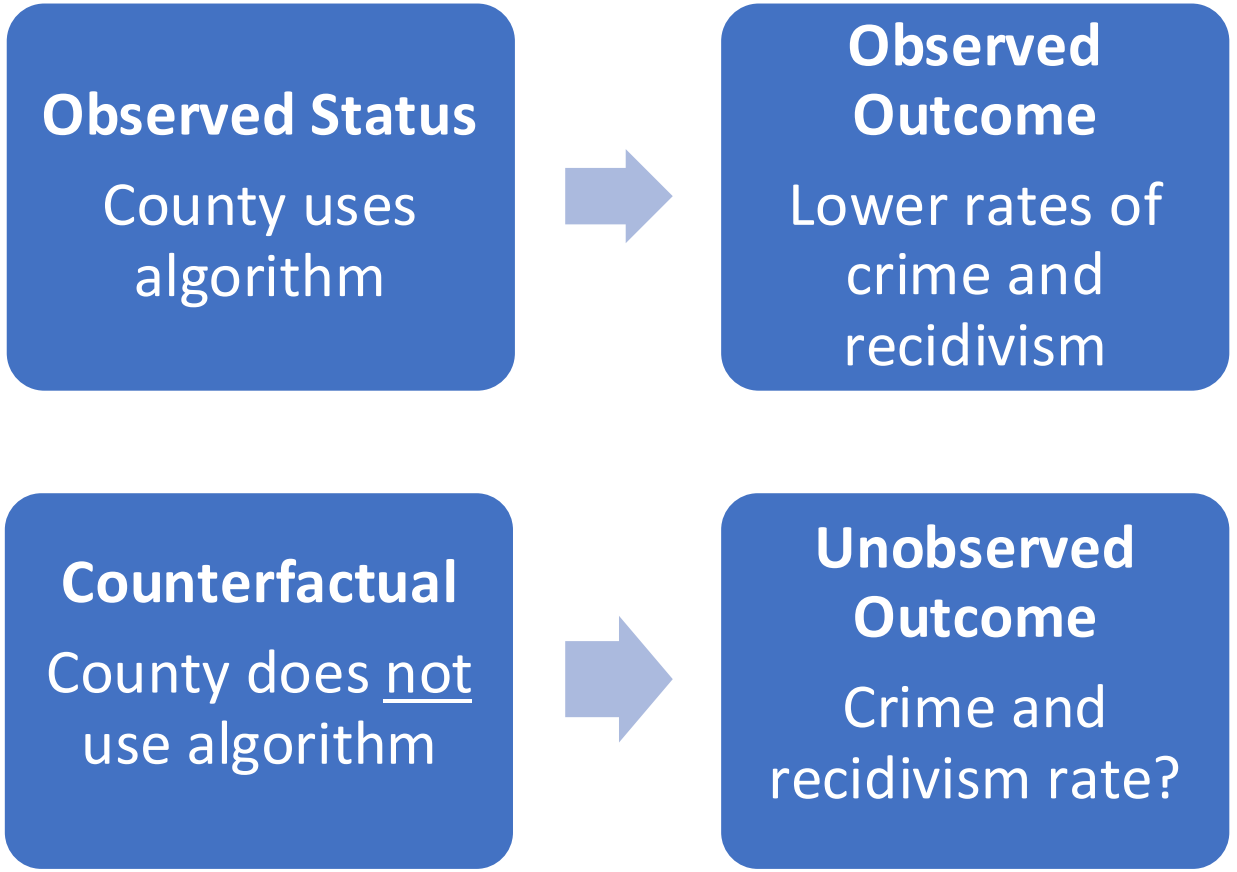
The COMPAS tool is widely used to assess a defendant's risk of committing more crimes, but a new study puts its usefulness into perspective.

The Atlantic

> "Imagine you're a judge and your court has purchased this software; the people behind it say they have big data and algorithms, and their software says the defendant is high-risk," says Farid. "Now imagine I said: Hey, I asked 20 random people online if this person will recidivate and they said yes. How would you weight those two pieces of data? I bet you'd weight them differently. But what we've shown should give the courts some pause." (A spokesperson from Equivant declined a request for an interview.)

Source: Atlantic Logo, Public Domain

Counterfactual Reasoning



A Detour on Causation

Correlation is not causation

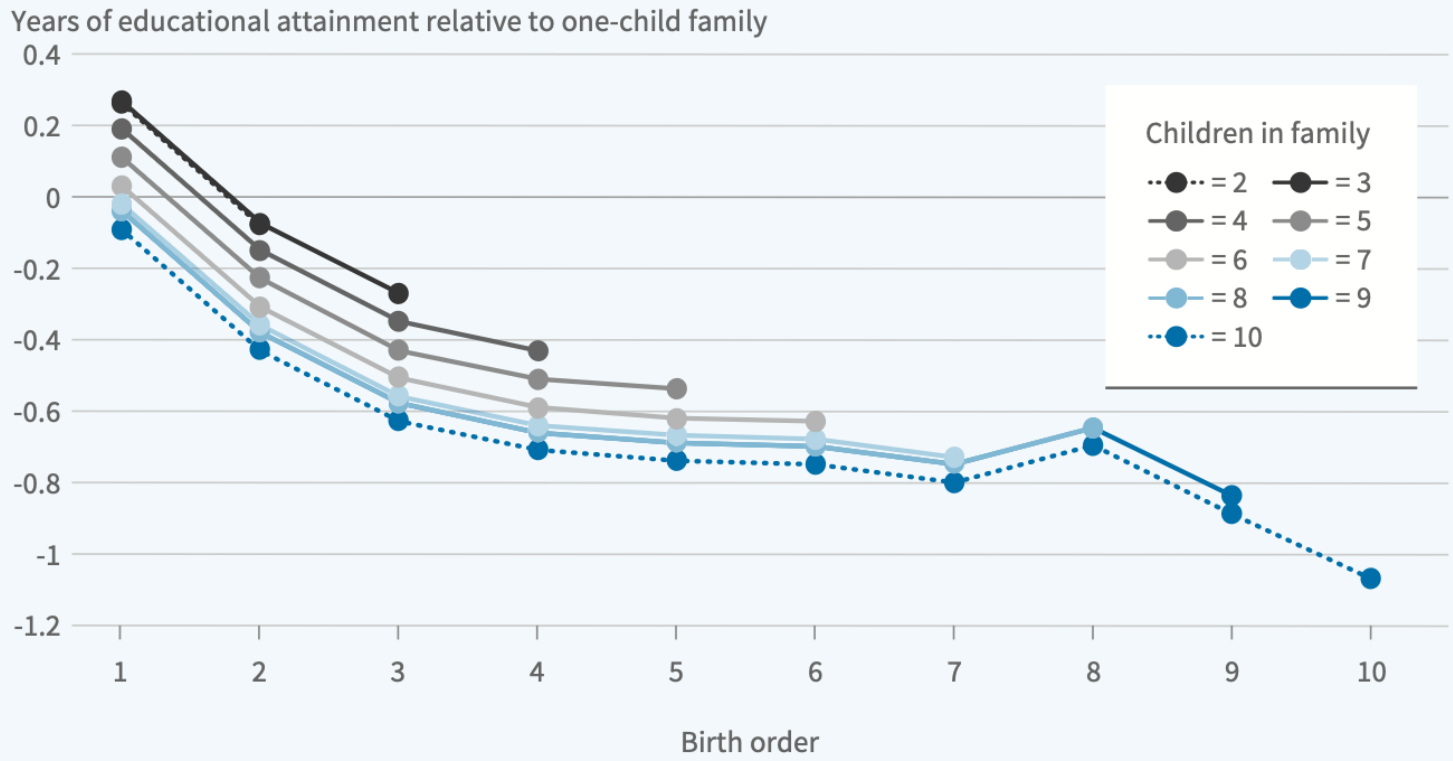
For example:

What are the predictors of academic success and educational attainment?

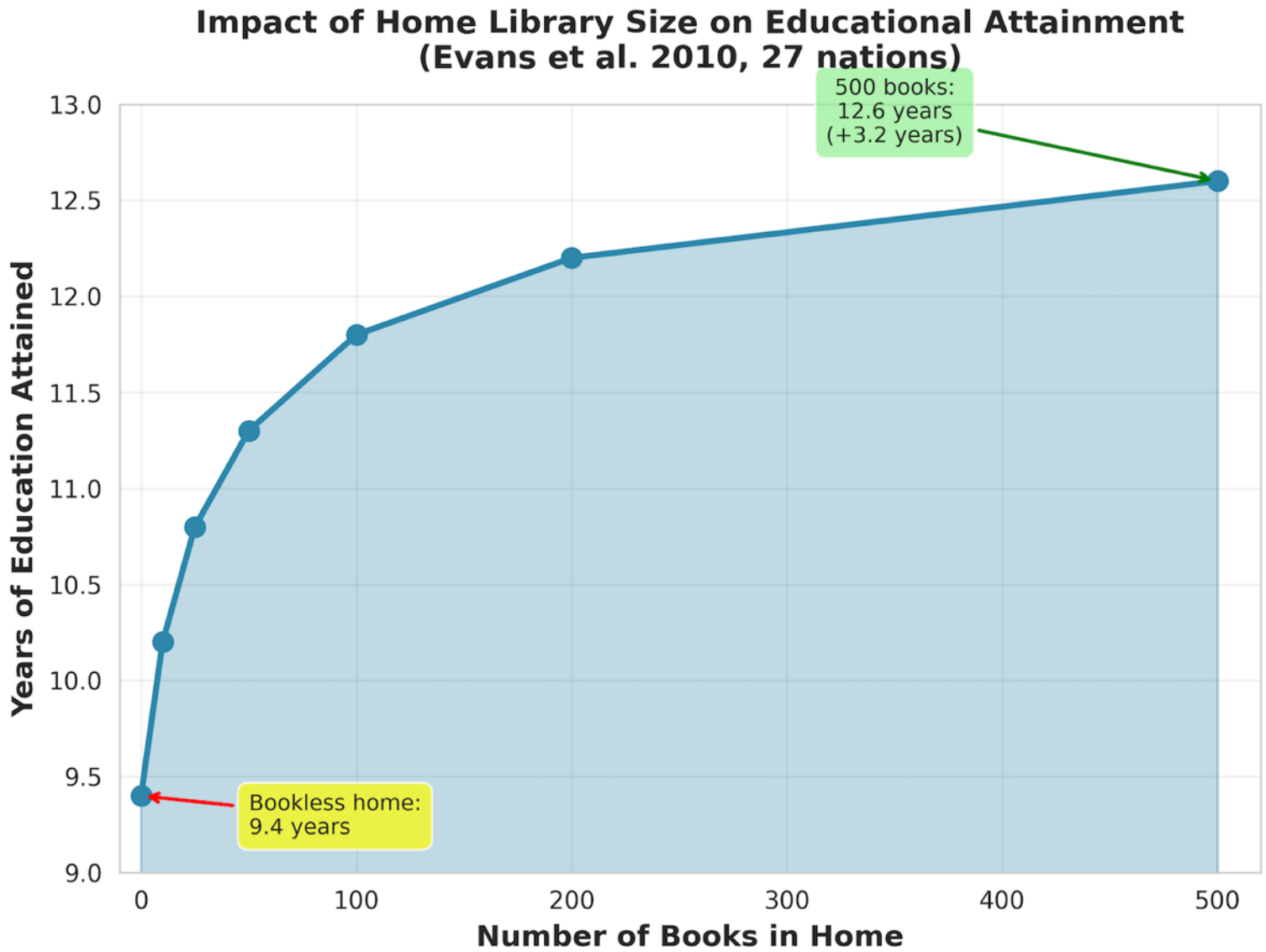
What are the causes of academic success and educational attainment?

Birth order?

Family Size, Birth Order and Educational Attainment

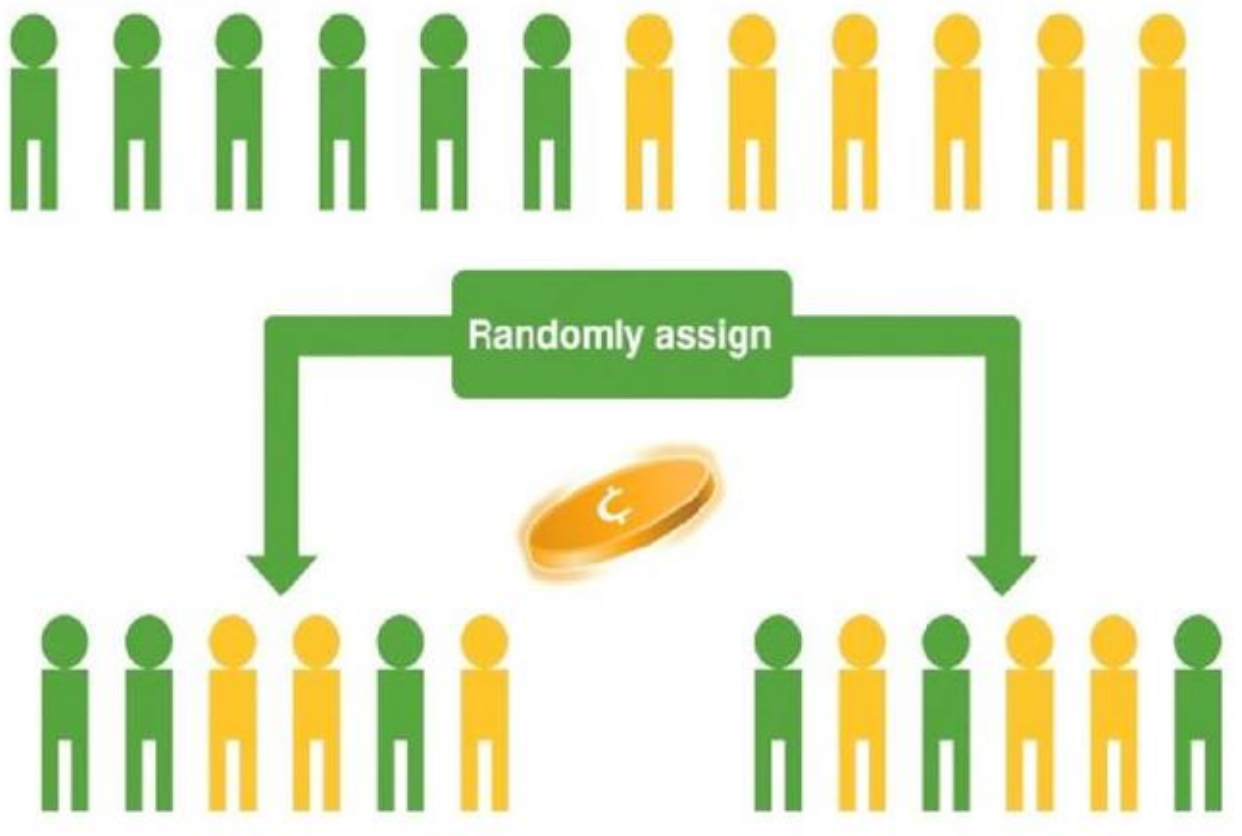


Source: S. E. Black, P. J. Devereux, and K. G. Salvanes, "*The More the Merrier? The Effect of Family Size and Birth Order on Children's Education,*" NBER Working Paper No. 10720, and The Quarterly Journal of Economics, 120(2), May 2005, pp. 669-700


Impact of Home Library Size on Educational Attainment
(Evans et al. 2010, 27 nations)
500 books:
12.6 years
(+3.2 years)
Bookless home:
9.4 years
Years of Education Attained
13.0
12.5
12.0
11.5
11.0
10.5
10.0
9.5
9.0
0
100
200
300
400
500
Number of Books in Home

Causal Inference

- Gold standard: randomized controlled trial (RCT)



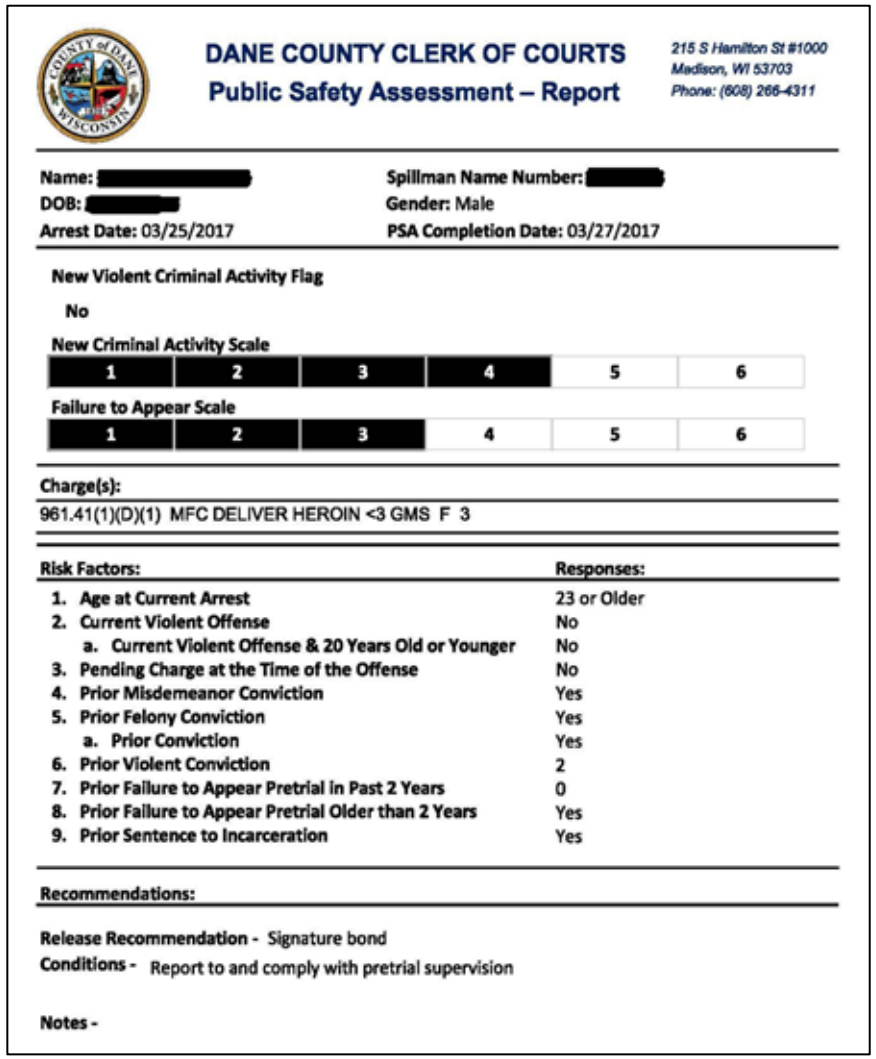
Decisions About Implementation

- Suppose you have a really awesome machine learning algorithm that predicts recidivism and you want to make it available to judges. What decisions do you need to make if you want it to be used?

1. When in the judicial process to provide the information?
2. Who to provide the information to (e.g. pre-trial services agencies, judges, etc.)?
3. How to provide the information?
4. How much the information should be weighed in making the decision?

All of these choices could have HUGE effects on the outcomes.

Dane County, WI



DANE COUNTY CLERK OF COURTS
Public Safety Assessment – Report

215 S Hamilton St #1000
Madison, WI 53703
Phone: (608) 266-4311

Name: [redacted]
DOB: [redacted]
Arrest Date: 03/25/2017

Spillman Name Number: [redacted]
Gender: Male
PSA Completion Date: 03/27/2017

New Violent Criminal Activity Flag

No

New Criminal Activity Scale

1	2	3	4	5	6

Failure to Appear Scale

1	2	3	4	5	6

Charge(s):

961.41(1)(D)(1) MFC DELIVER HEROIN <3 GMS F 3

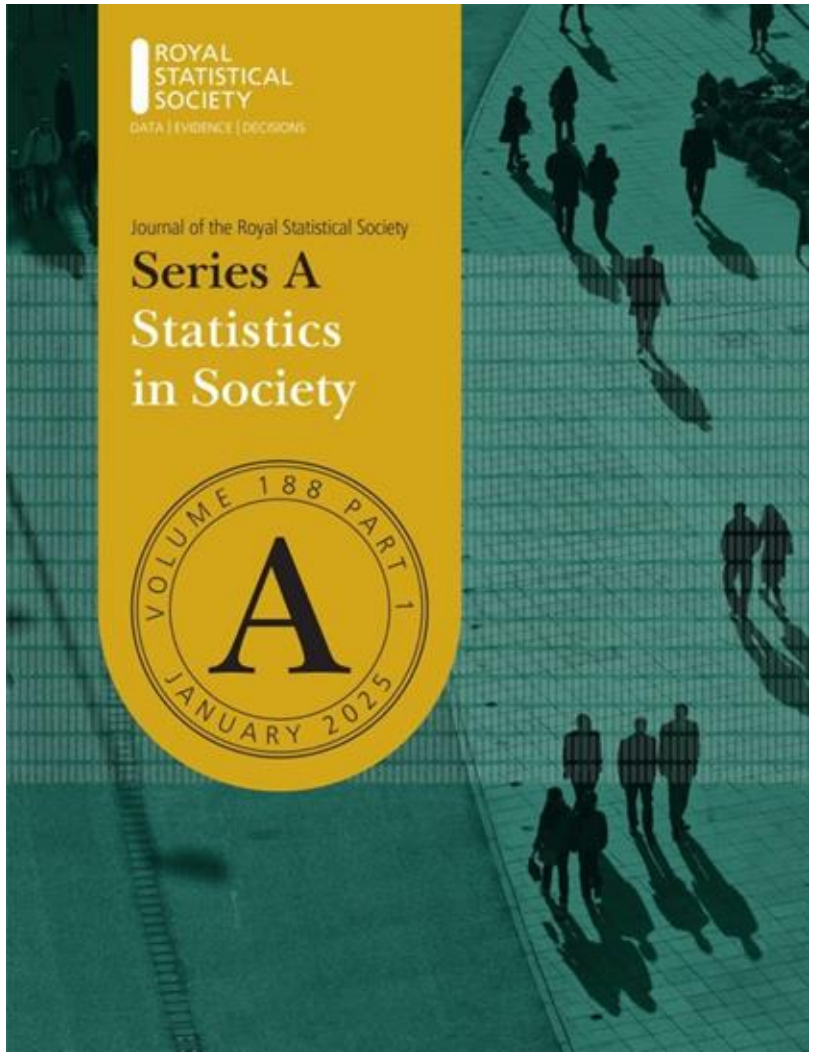
Risk Factors:	Responses:
1. Age at Current Arrest	23 or Older
2. Current Violent Offense	No
a. Current Violent Offense & 20 Years Old or Younger	No
3. Pending Charge at the Time of the Offense	No
4. Prior Misdemeanor Conviction	Yes
5. Prior Felony Conviction	Yes
a. Prior Conviction	Yes
6. Prior Violent Conviction	2
7. Prior Failure to Appear Pretrial in Past 2 Years	0
8. Prior Failure to Appear Pretrial Older than 2 Years	Yes
9. Prior Sentence to Incarceration	Yes

Recommendations:

Release Recommendation - Signature bond
Conditions - Report to and comply with pretrial supervision

Notes -

Dane County, WI



In the "first-ever randomized controlled trial that evaluates the pretrial Public Safety Assessment in the United States criminal justice system. . .
[we] show[] that providing the PSA to the judge has little overall impact on the judge's decisions and subsequent arrestee behaviour."

Source: Imai et al. (2023)

Today's Agenda:
Algorithmic Decision-Making: Rights and Responsibilities

1. What is Fairness? Rawls's Theory of Justice as Fairness
2. What's the point of criminal justice?
3. How to decide whether to use an algorithm
   - Simulating outcomes
   - Implementing and measuring outcomes
4. Who should decide?

Popular Democracy

In 2020, Ballot Proposition 25: Sustain SB 10?

In Opposition to this measure:

Total amount of reported contributions to this measure: $10,181,122*

Committee ID**	Committee Name	Total Reported Contributions
1410088	No on Prop 25 - Stop the Unfair, Unsafe and Costly Ballot Proposition, sponsored by the American Bail Association	$9,834,622

Source: California Secretary of State (https://www.sos.ca.gov/campaign-lobbying/cal-access-resources/measure-contributions/2020-ballot-measure-contribution-totals/proposition-25-referendum-overturn-2018-law-replaced-money-bail-system-system-based-public-safety-risk)

California Proposition 25, Replace Cash Bail with Risk Assessments Referendum (2020)

California Proposition 25, the **Replace Cash Bail with Risk Assessments Referendum**, was on the ballot in California as a veto referendum on November 3, 2020.[1] Proposition 25 was **defeated**, repealing the law.

A "**yes**" vote was to uphold the contested legislation, Senate Bill 10 (SB 10), which would have replaced cash bail with risk assessments for detained suspects awaiting trials.

A "**no**" vote was to repeal the contested legislation, Senate Bill 10 (SB 10), thus keeping in place the use of cash bail for detained suspects awaiting trials.

Election results

California Proposition 25		
Result	Votes	Percentage
Yes	7,232,380	43.59%
✖ **No**	**9,358,226**	**56.41%**

Results are officially certified.
Source

California Proposition 25

Election date
November 3, 2020

Topic
Civil and criminal trials

Status
✖ *Defeated*

Type
Referendum

Origin
Citizens

List of California measures

Submit

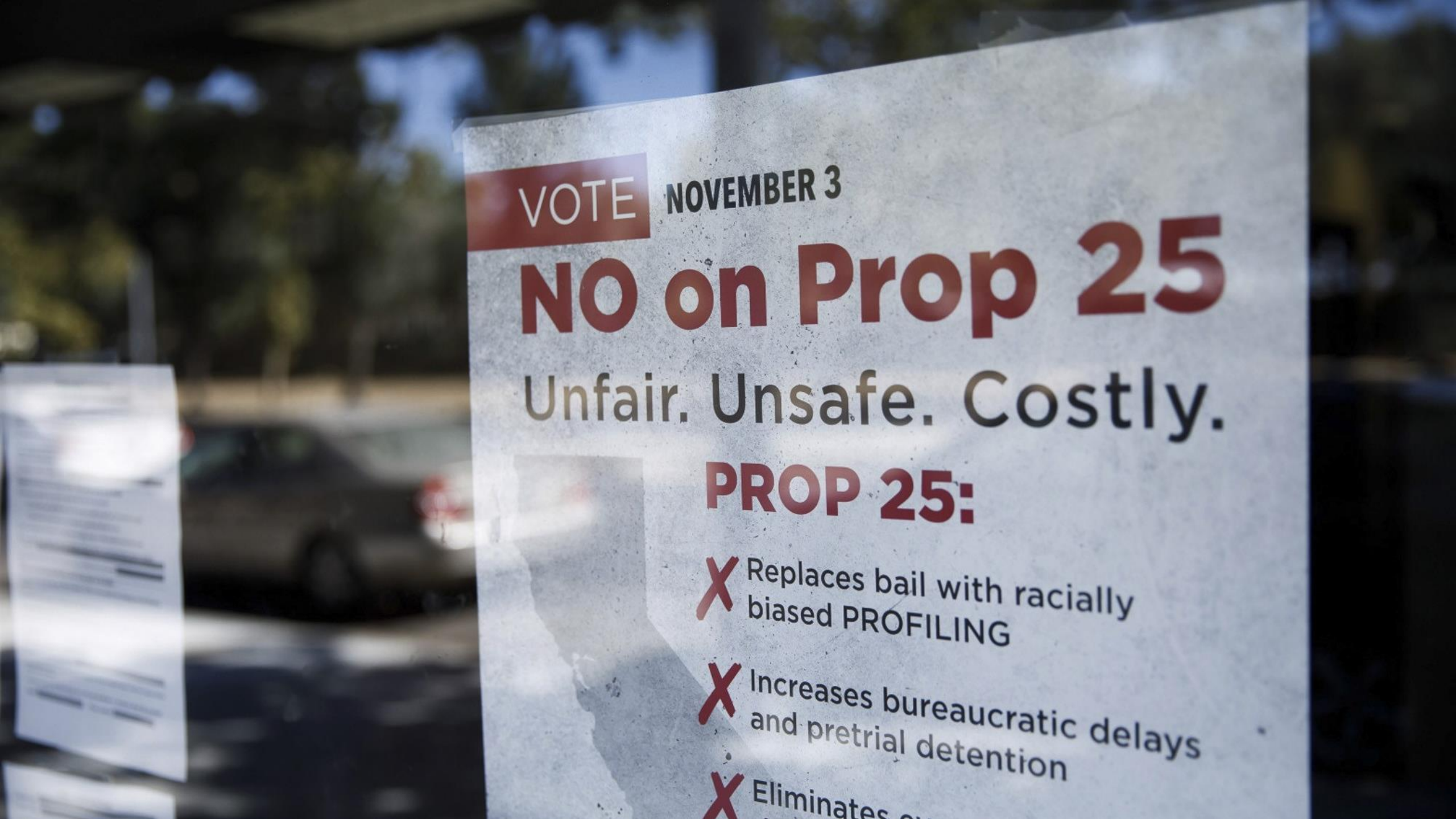
Who Supported Prop 25 and Eliminating Cash Bail?

For

- California Democratic Party, Governor Newsom
- Some criminal justice reform groups like Californians for Safety and Justice

Against

- Bail bond industry (spent millions, led the referendum)
- Law enforcement groups (California Peace Officers' Association)
- Also many progressive groups, including:
  - ACLU of Southern California
  - Human Rights Watch
  - California NAACP
  - Various criminal justice reform advocates


VOTE NOVEMBER 3
NO on Prop 25
Unfair. Unsafe. Costly.
PROP 25:
Replaces bail with racially biased PROFILING
Increases bureaucratic delays and pretrial detention