



# Ethics, Public Policy, and Technological Change

Rob Reich  
Mehran Sahami  
Head TA: Roberta Fischli

# Today's Agenda:

## Algorithmic Decision-Making: Promise and Peril

1. Introduction to Algorithmic Decision-Making
2. Algorithms and Fairness
3. What is Fairness:
  - Fair process
  - Fair outcome
4. John Rawls on Justice as Fairness

# Today's Agenda:

## Algorithmic Decision-Making: Promise and Peril

1. Introduction to Algorithmic Decision-Making
2. Algorithms and Fairness
3. What is Fairness:
  - Fair process
  - Fair outcome
4. John Rawls on Justice as Fairness



# Algorithms Shape Modern Life

## **Personalized Recommendations**

Algorithms suggest movies, products, and news to users based on their preferences, enhancing everyday experiences and convenience.

## **Automated Financial Decisions**

Credit scoring algorithms quickly assess financial trustworthiness, influencing approval for loans and credit cards in real time.

## **Matching and Screening Platforms**

Algorithms on hiring and dating platforms screen candidates and match users, shaping relationships and career opportunities.

## **Justice and Risk Assessment**

Criminal justice systems utilize algorithms for risk assessments and sentencing, impacting legal decisions and outcomes.

## **Online Search and Newsfeeds**

Algorithms curate search results and personalize newsfeeds, helping users find relevant information efficiently while shaping public opinion.

## **Medical Decision Making**

In healthcare, algorithms assist in diagnosing diseases, recommending treatments, and predicting patient outcomes, enhancing clinical decision-making.

# The Great Promise of Algorithmic Decision Making

If computer algorithms can more accurately and more efficiently outperform human at certain kinds of problem-solving, then algorithmic decision-making has enormous potential for good.

They might also be able to diminish human bias in decision-making.

**Cathy O’Neil, *Weapons of Math Destruction*:**

“A computer program could speed through thousands of résumés or loan applications in a second or two and sort them into neat lists, with the most promising candidates on top. This not only saved time but was also marketed as fair and objective. After all, it didn’t involve prejudiced humans digging through reams of paper, just machines processing cold numbers.”

# The Great Peril of Algorithmic Decision Making

- Algorithms can mask bias under a veneer of objectivity
- Algorithms can amplify bias and discrimination and exacerbate inequality
- Algorithms are often black boxes and cannot explain their decisions, thereby undermining due process
- Algorithms can automate injustice

# Today's Agenda:

## Algorithmic Decision-Making: Promise and Peril

1. Introduction to Algorithmic Decision-Making
2. Algorithms and Fairness
3. What is Fairness:
  - Fair process
  - Fair outcome
4. John Rawls on Justice as Fairness

# Algorithms and Values

*So what's the connection between algorithms and values like fairness?*

If an algorithm is a mathematical function to translate inputs into a determinative output, what do values like fairness have to do with math or logic?

Let's take an example of some immediate relevance to you...

# CS182 Grading Rubric

## Assignments and Grading Breakdown

The course includes five assignments. You will receive more information about each of the assignments well in advance of their due dates. The assignments are as follows:

- Technical assignment (algorithmic decision-making) – **due January 27 at 11:59pm PST**
- Philosophy paper – **due February 12 at 11:59pm PST** [NOTE: WIM students will have an additional revision due **March 1 at 12noon PST**]
- Technical assignment (social network) – **due February 26 at 11:59pm PST**
- Group policy assignment – **due March 12 at 11:59pm PST**
- Final reflection (take home final) – **due on March 17 at 12noon PST**

Grades will be calculated as follows:

Non-WIM Students	WIM Students
<ul style="list-style-type: none"><li>● Participation – 20%</li><li>● Technical Assignment 1 – 14%</li><li>● Philosophy Paper – 20%</li><li>● Technical Assignment 2 – 6%</li><li>● Policy Assignment – 20%</li><li>● Final Reflection – 20%</li></ul>	<ul style="list-style-type: none"><li>● Participation – 15%</li><li>● Technical Assignment 1 – 14%</li><li>● Philosophy Paper (original) – 9%</li><li>● Philosophy Paper (revision) – 21%</li><li>● Technical Assignment 2 – 6%</li><li>● Policy Assignment – 20%</li><li>● Final Reflection – 15%</li></ul>

# Possible CS182 Grading Schemes

1. Grading on curve: pre-determined number of As. Grades are determined on the basis of performance in class relative to other students in class.
2. Grading according to an external and objective standard. Everyone can earn an A, in principle.
3. Grading relative to your own past performance: that you show effort and improve over the quarter factors into your grade.

# Still More CS182 Grading Schemes

1. Everyone gets an A, conditional on full attendance and submitting work on time.
2. Grades are a function of hours worked on class assignments (as monitored by IT and course staff. More work = higher grade. Person who works the most gets the highest grade.
3. People who visit office hours get an automatic grade boost.
4. Grades are randomly assigned (eg., 50% get an A, 40% a B, 10% a C)

# What Values are at Stake in the CS182 Grading Algorithm?

1. **Fairness:** does the grading scheme treat people fairly?  
Note: multiple reasonable conceptions of fair grading
2. **Transparency:** is the grading scheme made transparent (or is it a black box?)
3. **Privacy:** are grades made public? [FERPA violation!]
4. **Due Process:** Is there a process for appeal?
5. Other values:
  - **Justice:** are grades the right or best way to evaluate students?
  - **Power:** are CAs legitimate graders?

# For Discussion

Imagine that Stanford (or another selective university) proposes to admit undergraduates on the basis of an algorithmic model.

1. Train the model on past applications and admission decisions
2. Train the model on performance of admitted students at Stanford and as alumni?

**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.

# Today's Agenda:

## Algorithmic Decision-Making: Promise and Peril

1. Introduction to Algorithmic Decision-Making
2. Algorithms and Fairness
3. What is Fairness:
  - Fair process
  - Fair outcome
4. John Rawls on Justice as Fairness

# What is Fairness?

Substantive vs. Procedural Fairness

Substantive Fairness: measured by the outcomes

Procedural Fairness: defined by a fair process

Note: Fairness can be applied to individuals or groups.

# Substantive Fairness

Fairness involves equal treatment and expectations of fair outcomes.

Three Examples:

- Orange Juice experiment
- Ultimatum game
- de Waal's monkey experiments



Orange Juice Experiment  
Inspired by economist Robert Frank

# Ultimatum Game



- Player 1 proposes some split of \$100
- Player 2 can do one of the following:
  - **Accept:** each player gets the proposed amount from the split
  - **Reject:** both players get nothing

# Competing Moral Perspectives on Fairness

- Deontological perspective
  - Principle: treat likes alike; people are owed equal treatment
  - To treat participants in significantly different ways, even if both receive a benefit, treats them unfairly and violates a duty to them
- Utilitarian perspective
  - Consider the utility gained by individuals involved
  - So long as no one is made worse off, unequal gains are beneficial

Q: In rejecting a split where someone receives less than 50%, is there some utility gained by upholding a notion of fairness?

# Is Fairness Baked Into Our Evolution?



# But is Fairness Always Equal Treatment?

No: substantive fairness might require unequal treatment!

Fairness requires **treatment as an equal**, not **equal (identical) treatment**.

Think about school funding.

Think about OAE accommodations.

# Substantive Fairness, refined

Lesson: Substantive Fairness is contextual and dependent on a social understanding of morally relevant (and irrelevant) characteristics of persons.

Aristotle:

Fairness is ‘treating likes alike, and unlikes unlike.’

So our core question becomes: In what ways are people relevantly alike and unlike?

This is not a question of math!

# Procedural Fairness

Bypass the effort to establish substantive fairness, and identify a fair process that delivers a fair outcome.



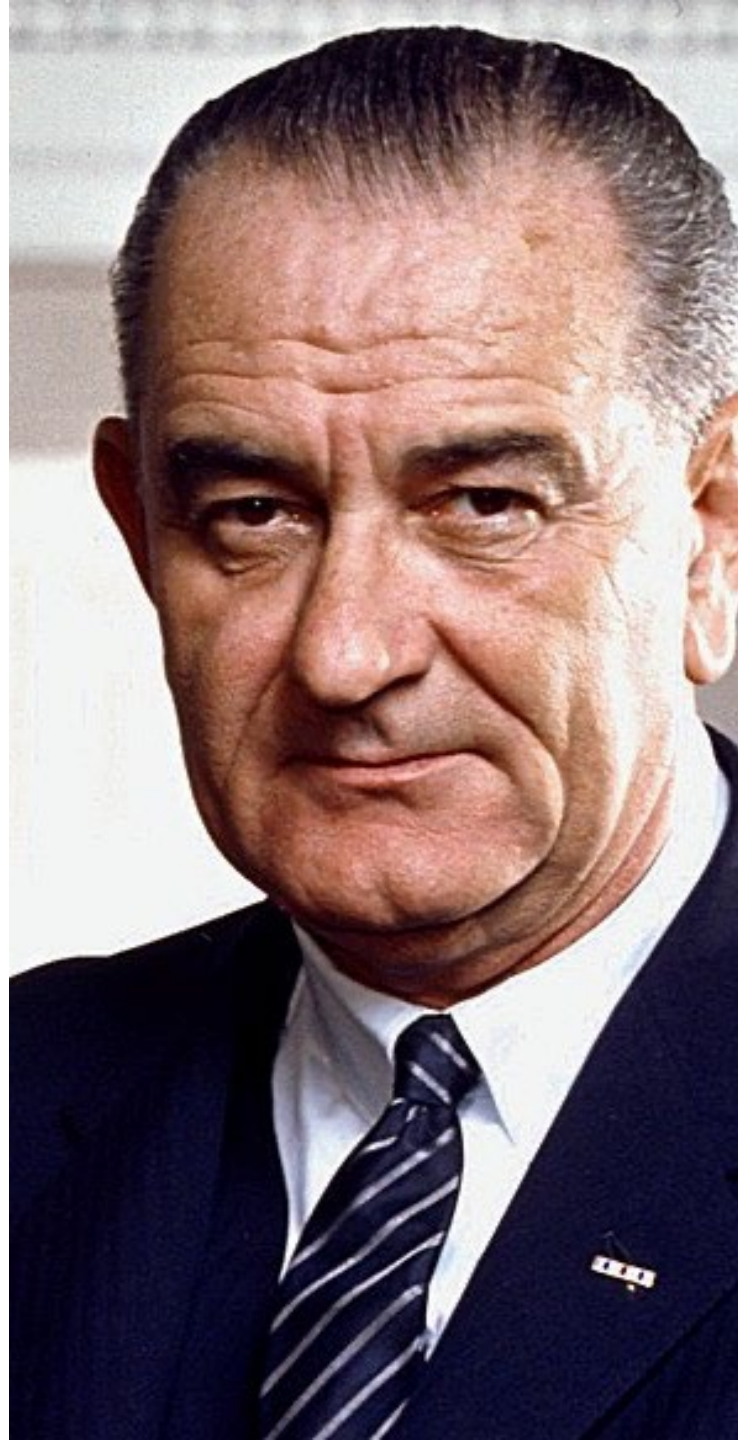
A Fair Procedure?

Perfect procedural justice



A Fair Procedure?

2. Pure procedural justice



# What about Historical Injustice?

"You do not take a person who, for years, has been hobbled by chains and liberate him, bring him up to the starting line of a race and then say, 'You are free to compete with all the others,' and still justly believe that you have been completely fair.

Thus it is not enough just to open the gates of opportunity. All our citizens must have the ability to walk through those gates."

# Procedural Fairness

Question: Can a fair process deliver a fair outcome?

We will explore these and other questions in more detail in this unit using the example of bail in the criminal justice system.

# Today's Agenda:

## Algorithmic Decision-Making: Promise and Peril

1. Introduction to Algorithmic Decision-Making
2. Algorithms and Fairness
3. What is Fairness:
  - Fair process
  - Fair outcome
4. John Rawls on Justice as Fairness  
[Will discuss next week]

# 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?

# Concluding Questions: Algorithmic Fairness

- How should we approach the idea of algorithmic fairness, and avoiding algorithmic bias? Substance or process? Individual or group?
- **What to do if there is a trade-off between predictive accuracy and fairness, transparency, privacy, explainability, due process?**
- **How does algorithmic decisionmaking reinforce or contest existing power dynamics in the world?**