Measurement as governance

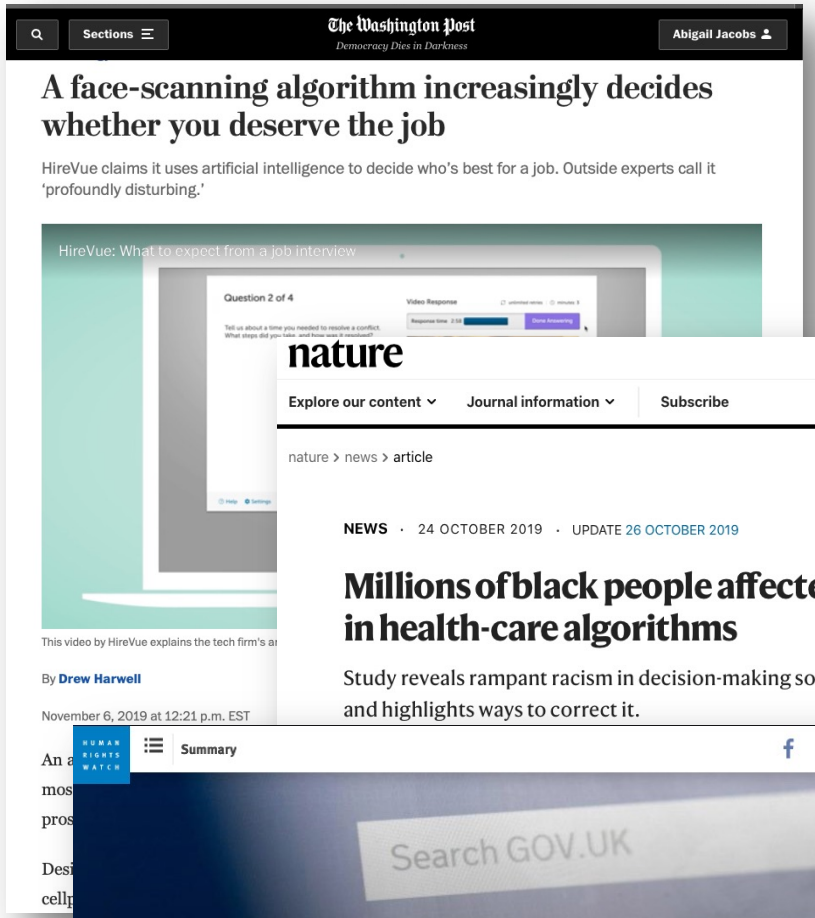
Abigail Jacobs

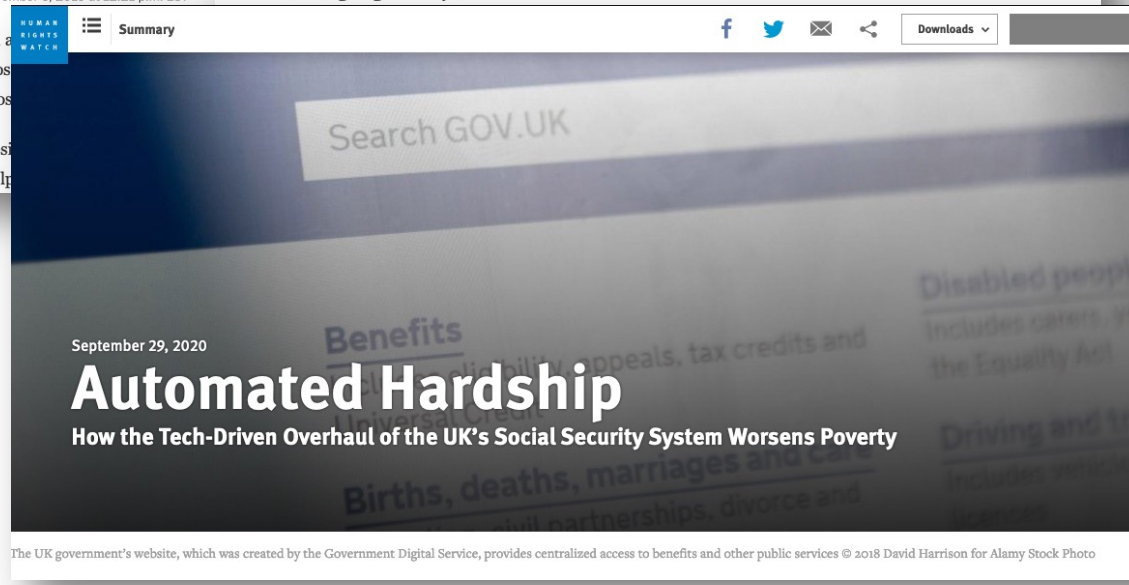
April 20, 2021

University of Michigan School of Information
Center for the Study of Complex Systems
Michigan Institute for Data Science; Center for Ethics, Society, and Computing

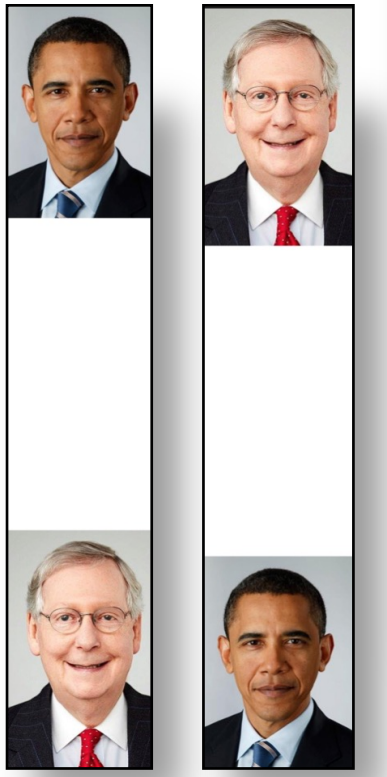
azjacobs@umich.edu | azjacobs.com | @az_jacobs

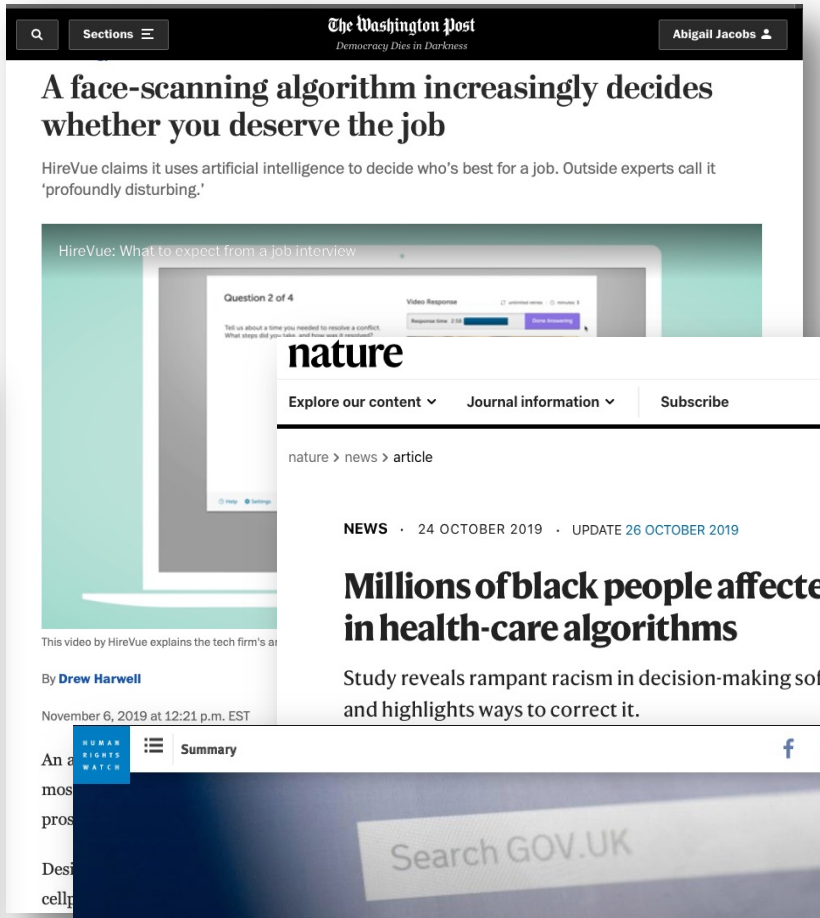
Fairness-related harms

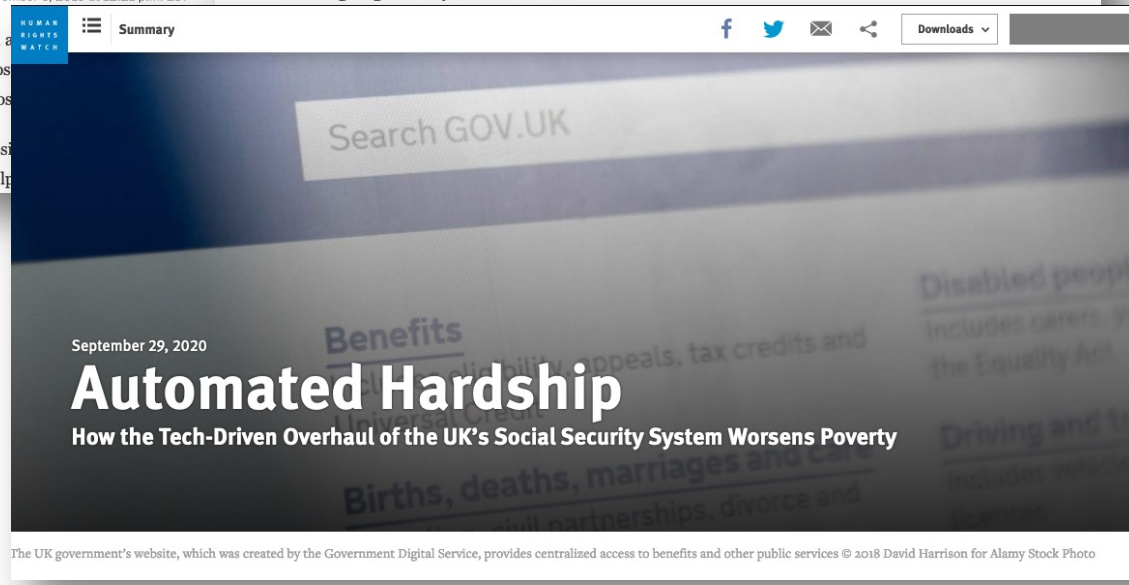




Fairness-related harms





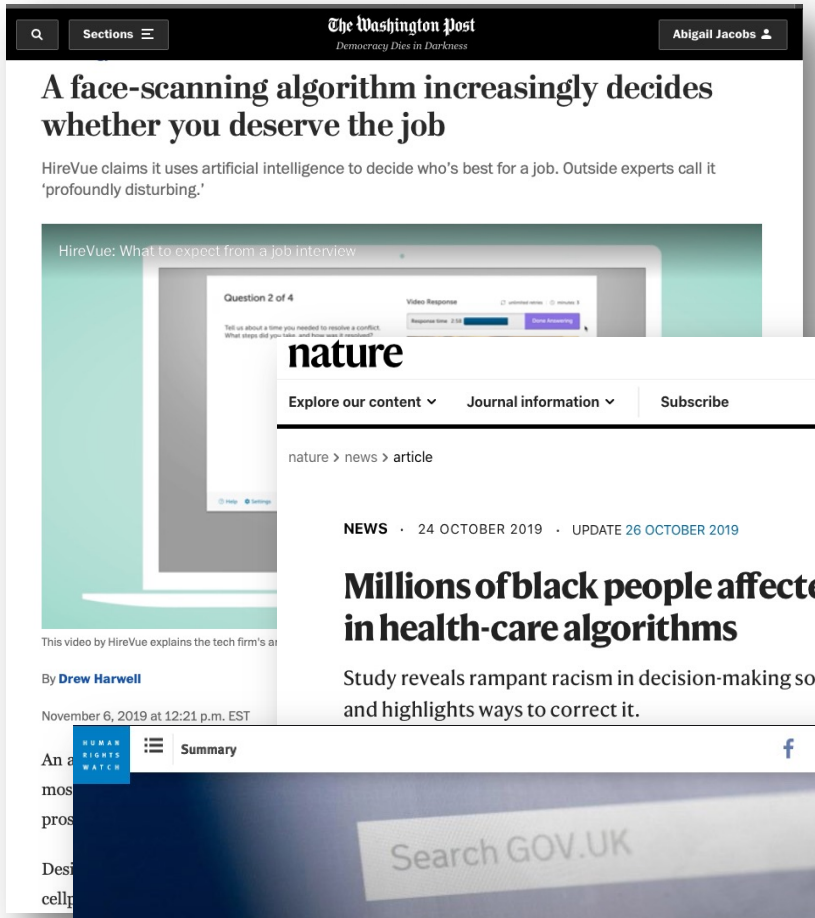


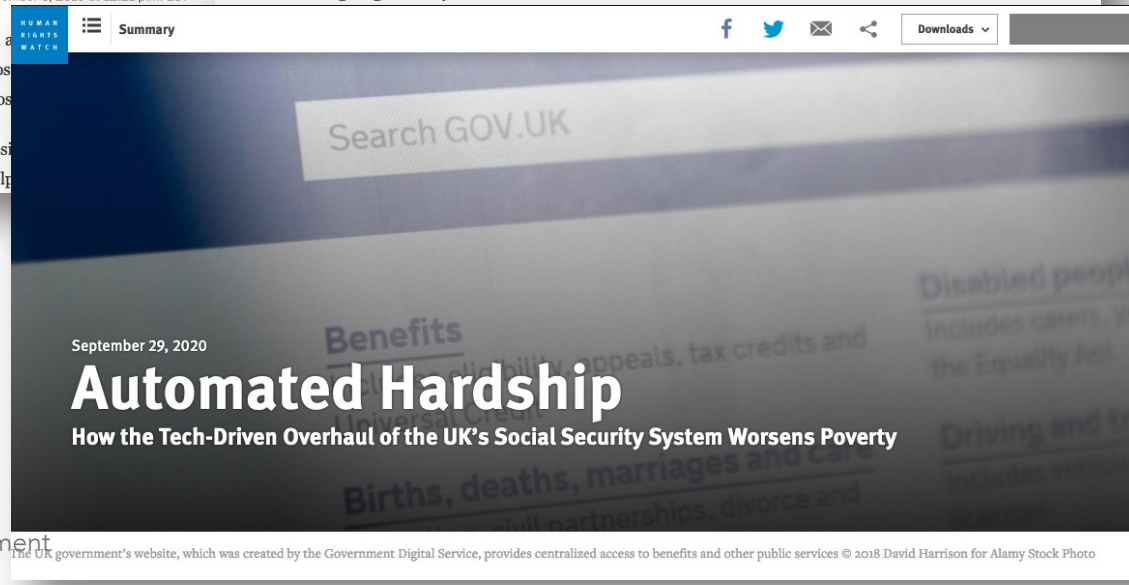
Fairness-related harms

- *Harms* (not necessarily biases)
- *Fairness-related* (not necessarily unequal)

Reframes

- Impacts
- Power
- Away from technical-only solutions





Fairness-related harms

- Why do these harms emerge?

  inequality exists

- When?

  always

  structural and individual—different levels

  accounting for specific instances useful for intervention

Assessing harms

- Audit studies
- Formal models
- Empirical evaluation, counterfactuals
- Ethnography
- Community, participatory methods

Countering with interventions

- Legal, regulatory, accountability mechanisms
- Organizational process, documentation, transparency, contestability
- Design, explainability, interpretability, reproducibility

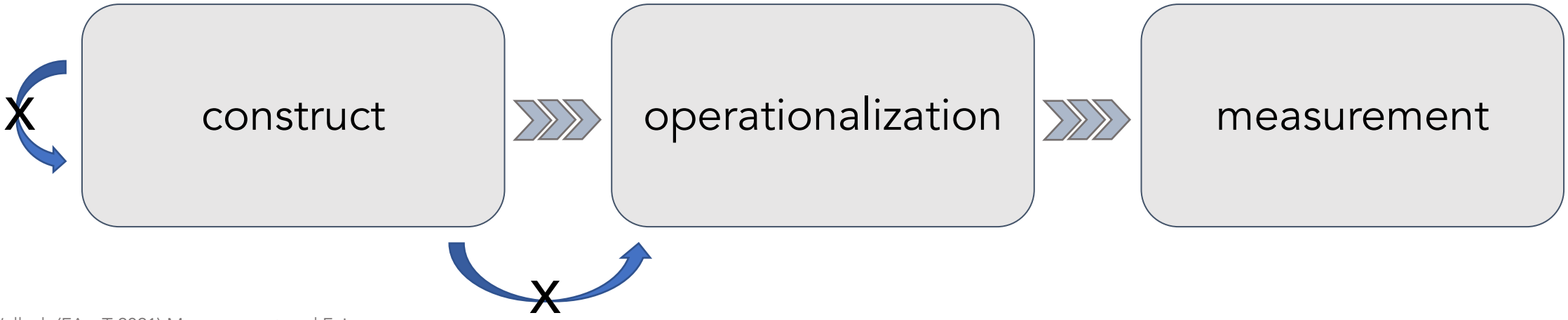
Where do fairness-related harms emerge?

*Fairness-related harms emerge when there is a **mismatch** between the thing we purport to be measuring and the thing we actually do*

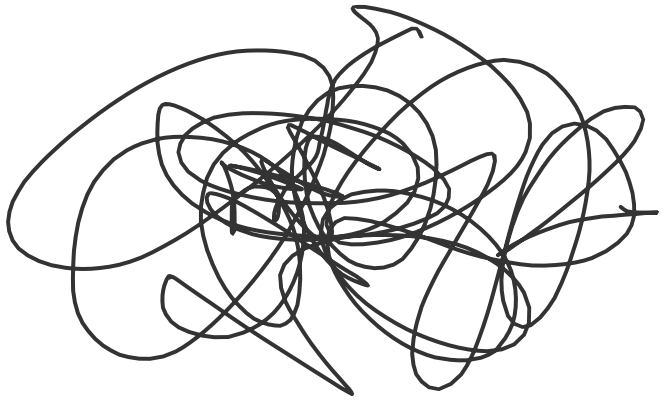
Social science world

Social science world

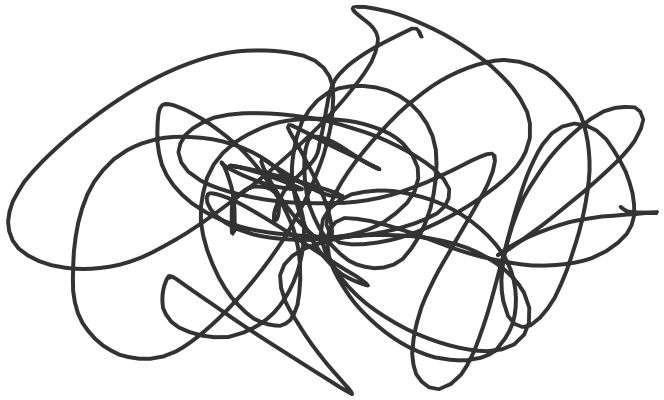
Fairness-related harms arise from mismatches in this process



ML world

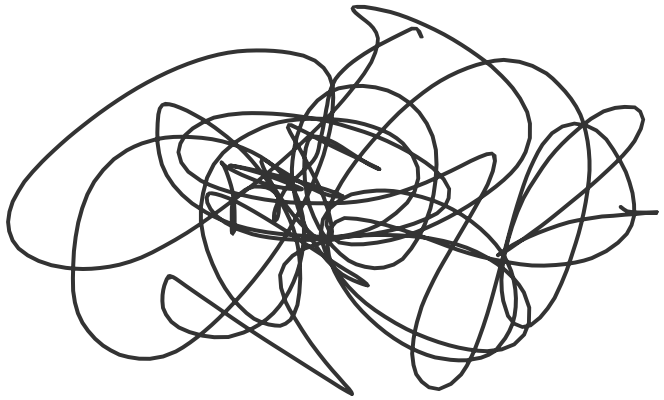


ML world



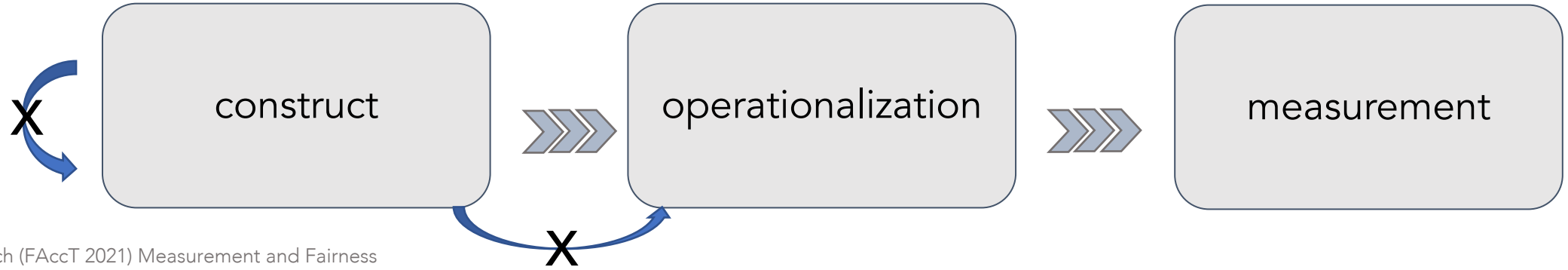
Fairness-related harms arise ??everywhere??

ML world

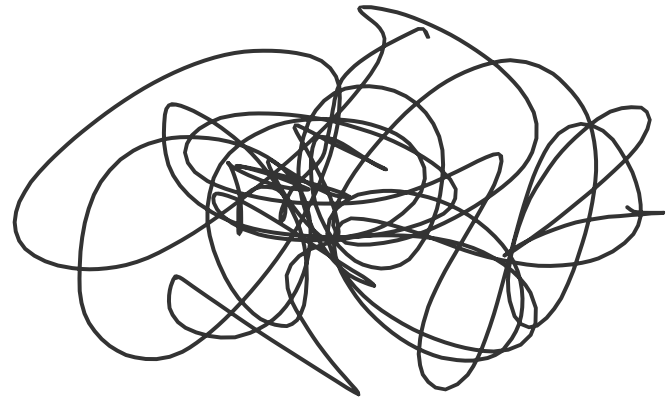


Language is power: need tools of measurement

Construct validity to diagnose/mitigate/prevent harms



Measurement & governance



Challenges to assessment & intervention

- Power is obscured
- Assumptions are hidden, often implicit
- Feedback loops
- Non-separable: technical, social, organizational problems

What we measure changes what we think we understand about the problem

[Measuring] social structure in sociotechnical systems

Organizations

Social networks

Social structure

Emergence of inequality

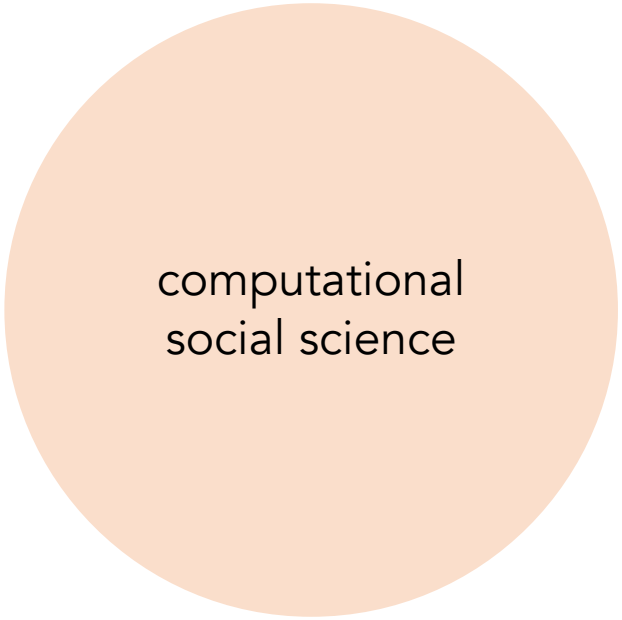
Measurement

Governance

Systems engineering

Human factors

…



Making assumptions explicit

Making assumptions

Testing assumptions

Consequences & measurement in systems

Measurement is everywhere

Creditworthiness

Teacher quality

Risk to society

Toxic language

Healthy communities

Prosocial behavior

Fairness

…

"Measurement and Fairness"
AZJ, Hanna Wallach (FAccT 2021)

Measurement is everywhere

Creditworthiness

Teacher quality

Risk to society

Toxic language

Healthy communities

Prosocial behavior

Fairness

…

unobservable theoretical **constructs**

"Measurement and Fairness"
AZJ, Hanna Wallach (FAccT 2021)

Measurement is everywhere

Creditworthiness

Teacher quality

Risk to society

Toxic language

Healthy communities

Prosocial behavior

Fairness

…

unobservable theoretical **constructs**

to *measure* the unobservable,
we necessarily **operationalize** the construct
with a measurement model

"Measurement and Fairness"
AZJ, Hanna Wallach (FAccT 2021)

Measurement is everywhere

Creditworthiness

Teacher quality

Risk to society

Toxic language

Healthy communities

Prosocial behavior

Fairness

…

unobservable theoretical **constructs**

to *measure* the unobservable,
we necessarily **operationalize** the construct
with a measurement model

and in so doing, reflect our social,
organizational, cultural, and political values

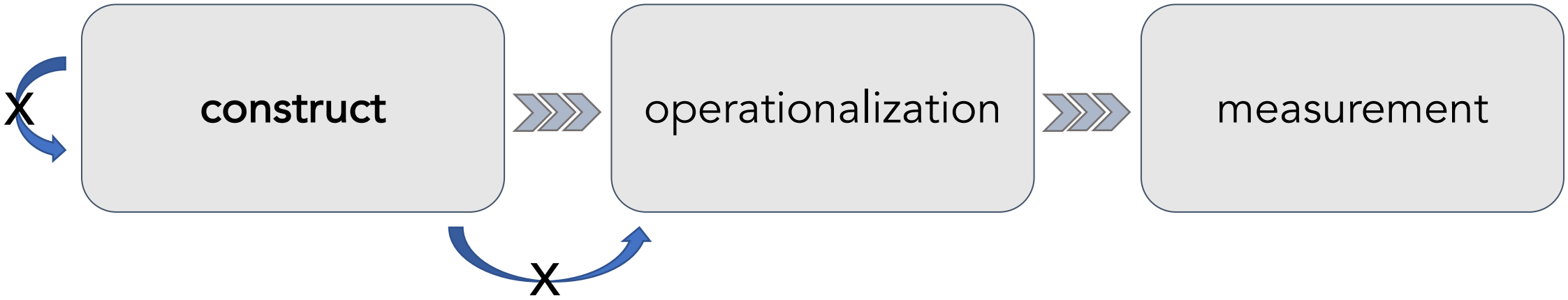
"Measurement and Fairness"
AZJ, Hanna Wallach (FAccT 2021)

The measurement process

"Measurement and Fairness"
AZJ, Hanna Wallach (FAccT 2021)

The measurement process

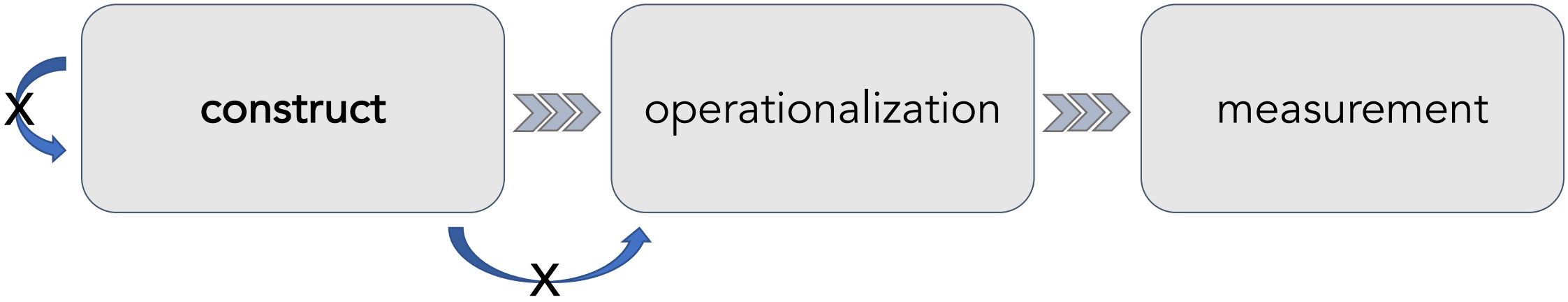
Fairness-related harms arise from **mismatches** in this process



"Measurement and Fairness"
AZJ, Hanna Wallach (FAccT 2021)

The measurement process

Fairness-related harms arise from mismatches in this process

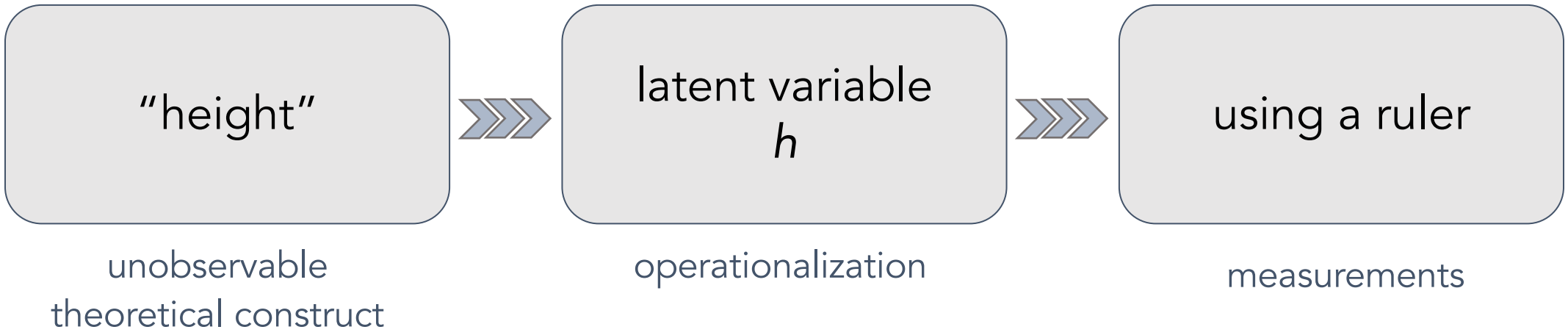


Construct reliability & validity help us interrogate this (often obscured) process

"Measurement and Fairness"
AZJ, Hanna Wallach (FAccT 2021)

Measuring height

- Assume that *h* influences observable properties of height

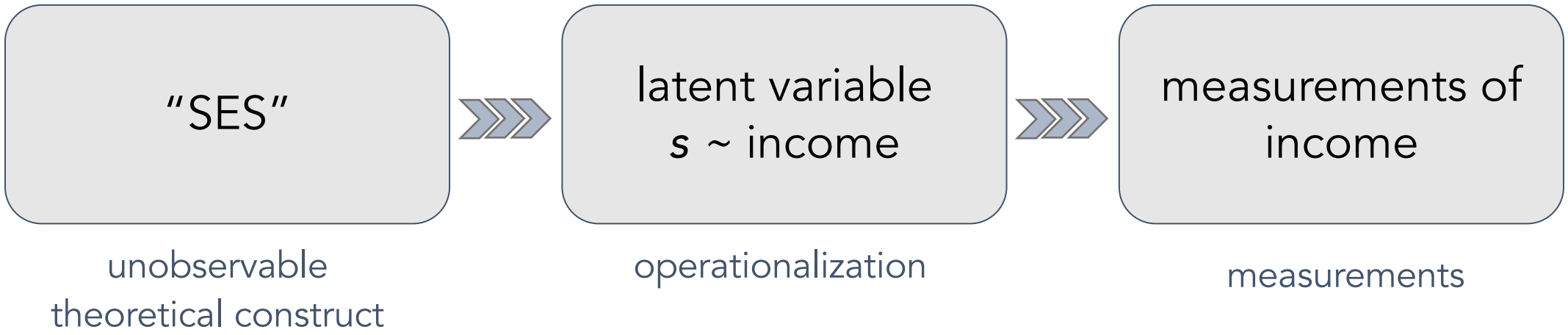


"Measurement and Fairness"
AZJ, Hanna Wallach (FAccT 2021)

Measuring socioeconomic status (SES)

- Assume that *s* influences observable properties of SES where income is the measurement model of *s*

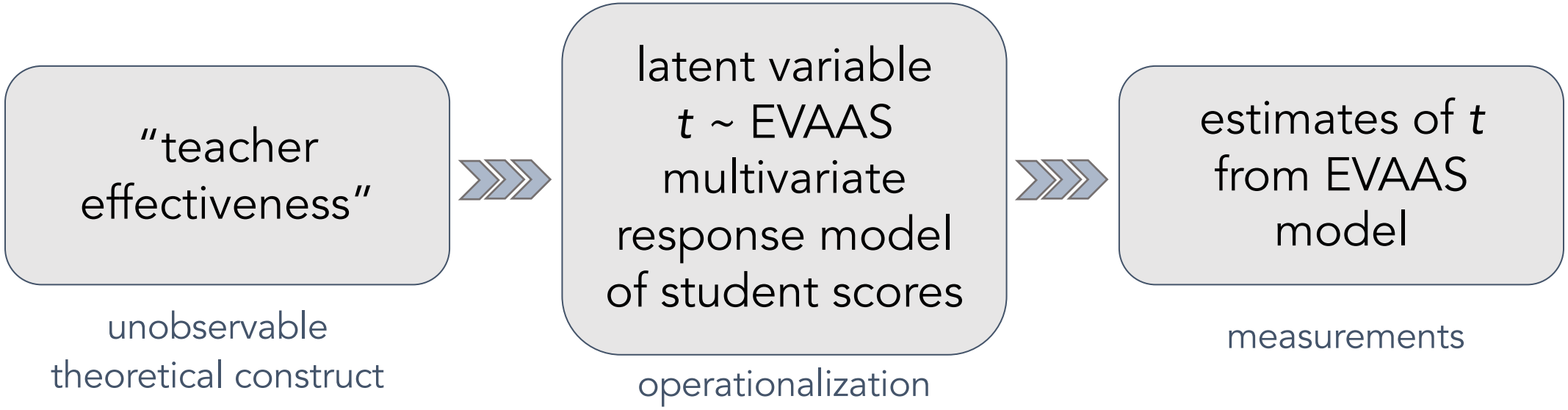
All proxies are measurement models.



"Measurement and Fairness"
AZJ, Hanna Wallach (FAccT 2021)

Measuring teacher effectiveness

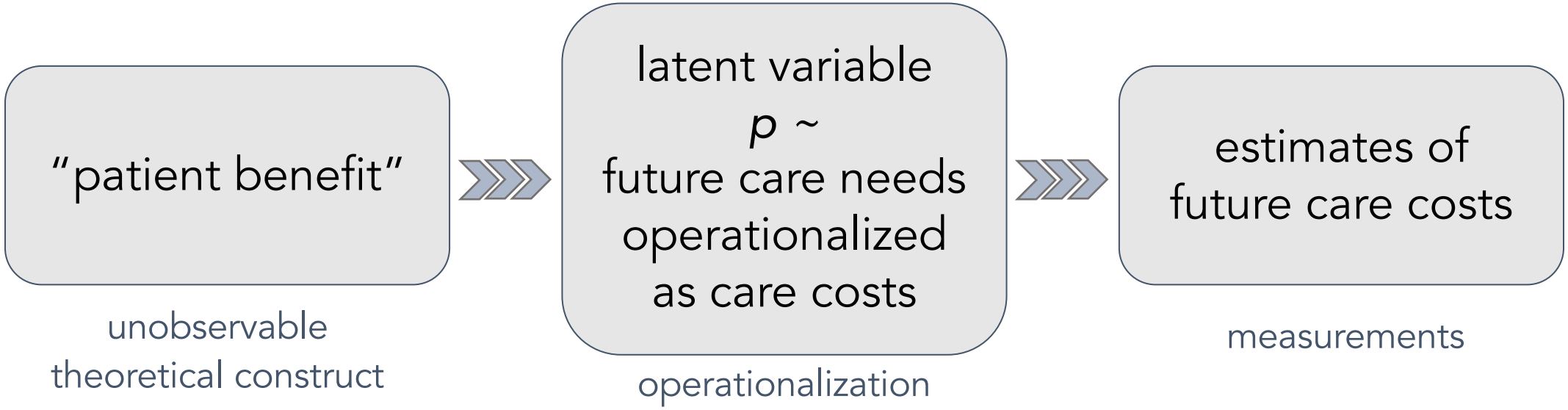
- Assume that *t* influences observable properties of teacher effectiveness using the measurement model from the Education Value-Added Assessment System (EVAAS)



"Measurement and Fairness"
AZJ, Hanna Wallach (FAccT 2021)

Measuring patient benefit

- Assume that *p* influences observable properties of patient benefit from enrollment in high-risk healthcare management programs



"Measurement and Fairness"
AZJ, Hanna Wallach (FAccT 2021)

Testing assumptions with construct reliability and validity

Making assumptions

Testing assumptions

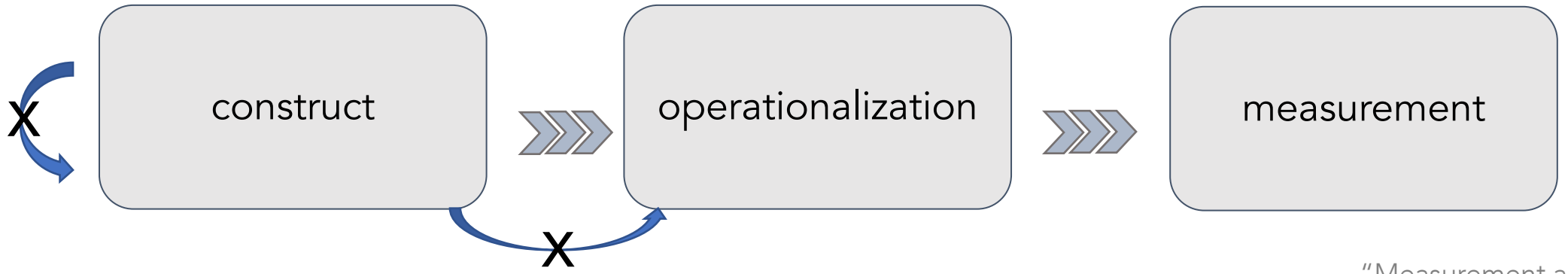
Consequences & measurement in systems

The measurement process

Fairness-related harms emerge from **mismatches** in this process

Construct reliability & validity

- identify (often hidden) assumptions
- are *generative*
- help to diagnose, mitigate, prevent fairness-related harms



"Measurement and Fairness"
AZJ, Hanna Wallach (FAccT 2021)

Construct reliability

- Roughly analogous to precision in statistics
- Can include
  - Test-retest reliability
  - Out of sample prediction

"Measurement and Fairness"
AZJ, Hanna Wallach (FAccT 2021)

Construct validity

Umbrella term

We are most influenced by

- Messick (1980s-90s)
- Loevinger (1950s-60s)
- Hand (2000s-)
- Quinn et al (2012)
- Jackman (2010s-)

We unite traditions from

- Political science
- Education testing
- Psychology
- Statistics, …

for the broader FAccT community.

"Measurement and Fairness"
AZJ, Hanna Wallach (FAccT 2021)

Umbrella of construct validity

Generative framework: Does it work (well)? Opportunity for mismatches?

Qualitative plausibility; a 'sniff test'

Capturing the construct of interest? Contestedness; substantive nature; structural assumptions

Quantitative comparison to validated measures; qualitative differences?

Confounds? Relatedness to related constructs?

Prediction of related attributes *outside* of the model

Usefulness

Measurements shape the way we understand the construct itself. Targets? Categories? Feedback loops?

- Face validity
- Content validity
- Convergent validity
- Discriminant validity
- Predictive validity
- Hypothesis validity
- Consequential validity

"Measurement and Fairness"
AZJ, Hanna Wallach (FAccT 2021)

Consequential validity

- Goodhart's Law
- Campbell's Law
- Washback
- Lucas critique
- Performativity
- …

"measurements both reflect structure in the natural world, and impose structure upon it"

Hand (2016)

Appearance of objectivity

Consequences & measurement: a systems view of fairness

Making assumptions

Testing assumptions

Consequences & measurement in systems

The substantive nature of fairness

- What about justice?
  - Content and consequential validity

"Measurement and Fairness"
AZJ, Hanna Wallach (FAccT 2021)

The substantive nature of fairness

- What about justice?
  - Content and consequential validity
- Individual vs. group fairness
  - Disagreement about operationalizations---what about values?

"Measurement and Fairness"
AZJ, Hanna Wallach (FAccT 2021)

The substantive nature of fairness

- What about justice?
  - Content and consequential validity
- Individual vs. group fairness
  - Disagreement about operationalizations---what about values?
- Individual fairness
  - Essentially contested---what about content and consequential validity?

"Measurement and Fairness"
AZJ, Hanna Wallach (FAccT 2021)

The substantive nature of fairness

- Group fairness
  - Conflicting operationalizations---predictive parity, equalized odds---reflect conflicting theoretical understandings

"Measurement and Fairness"
AZJ, Hanna Wallach (FAccT 2021)

The substantive nature of fairness

- Group fairness
  - Conflicting operationalizations---predictive parity, equalized odds---reflect conflicting theoretical understandings
- Demographic factors
  - Often essentially contested (gender, race)
  - Often implicit measurement modeling process---and risks of harms

"Measurement and Fairness"
AZJ, Hanna Wallach (FAccT 2021)

Measurement of fairness is governance

"Fairness" is an *essentially contested* construct

But to the degree to which we operationalize it anyways, need to be explicit in operationalization & of underlying values

(See also: Kasy & Abebe 2020)

Measurement as governance

Reveals:

- Types of harms
  - Debunking "de-biasing" (Blodgett et al 2020)
- Types of interventions
  - Incl organizational, technical, etc.
- Which questions get asked
- What decisions are being made, where – obscures power

"Measurement and Fairness"
AZJ, Hanna Wallach (FAccT 2021)

MC HAMMER
@MCHammer

You bore us. If science is a “commitment to truth” shall we site all the historical non-truths perpetuated by scientists ? Of course not. It’s not science vs Philosophy … It’s Science + Philosophy. Elevate your Thinking and Consciousness. When you measure include the measurer.

> **Drew** @drewgrey · Feb 22
> Replying to @MCHammer
> Philosophy is flirtation with ideas.
> Science is commitment to truth.

12:50 PM · Feb 22, 2021 · Twitter for iPhone

13.7K Retweets **5,882** Quote Tweets **78.4K** Likes

Consequential validity … zooming out

Durability of social structures – adaptation, reinforcement, obscured technical governance decisions

- Race as and of technology (Benjamin; Roberts)
- Myth and tool (Haraway)
- Reproduction of status, value (incl. Jasanoff, Perry)

Socially constituted nature of things being measured means this consequential validity

Language is power

Theory of measurement provides language & framework for unpacking systems

Measurement is everywhere

Harms emerge from mismatches in the measurement process

Construct reliability & validity help us interrogate this often-obscured process to prevent & mitigate harms

"Measurement and Fairness"
AZJ, Hanna Wallach (FAccT 2021)

“Language is power, life and the instrument of culture, the instrument of domination and liberation”

Angela Carter

Consequential validity helps us bring in systems-level feedback

Reconciles understandings of risk, systems, social structure

Some paths forward / learning at the interface

Unintended behaviors generate harms

AI "accidents" typically treated like '**component accidents**' (Perrow 1984)

common

probably could have been fixed if you tried harder

a band-aid exists

example: word embeddings

Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings (Bolukbasi et al 2016)

"Unsafe at any AUC: Uncovering sociotechnical control for responsible AI"
AZJ, Joshua Kroll, Andrew Smart, Elana Zeide

AI “accidents”

Possibly (most likely!) doing what ‘it’s supposed to do’

Glitches reveal **underlying social structure**—incl. algorithmic oppression, New Jim Code (Noble, Benjamin)

- Harms emerge from a disconnect between intended and operationalized constructs
- Emerge from training data, from problem formulation, from design, from designers, from formalization, interaction, amplification, etc.
- Reliability != safety (Leveson)

“Unsafe at any AUC: Uncovering sociotechnical control for responsible AI”
AZJ, Joshua Kroll, Andrew Smart, Elana Zeide

Unpacking mechanisms:
The need for structural explanations

- Problem isn't a component failure, but *structural*
  - Requires structural explanations (Haslanger 2016; Leveson)
  - Understanding *mechanism* (Hedstrom, Elster, Bearman)

- Need to understand the *resilience* of social structure
  - Durability of inequality
  - Level of site of justice (Wark/Binns)

- Lessons from other high-risk technologies
  - Major accidents emerge from organizational-cultural-technical contexts
  - Lessons? Assessment and audits; regulation; rules vs. standards; adopting a systems perspective *beyond* reliability
  - The issue is "not risk, but power" (Perrow 1994/1999; cf. Hopkins 1999)

Case study

Case study

- Decentralization
- Brokerage
- Inequality

Internet-human infrastructures
Lessons from Havana's StreetNet

Abigail Jacobs & Michaelanne Dye
University of Michigan

Empirical opportunities & challenges

- Spillovers
- Depth
- Seamfulness

- Ethics
- Integrating multiple modes of data and observations
- Empirical/theoretical gap between observed, deep individual contexts & system-level structures

Empirical opportunities & challenges

Exist across contexts and scales

Reveal *new questions and methods* to study diverse systems using mixed-methods

Make the physical, social, organizational, and political actors *visible*— paving the way to understand the *evolution and governance* of the Web, Internet, and computational systems more broadly

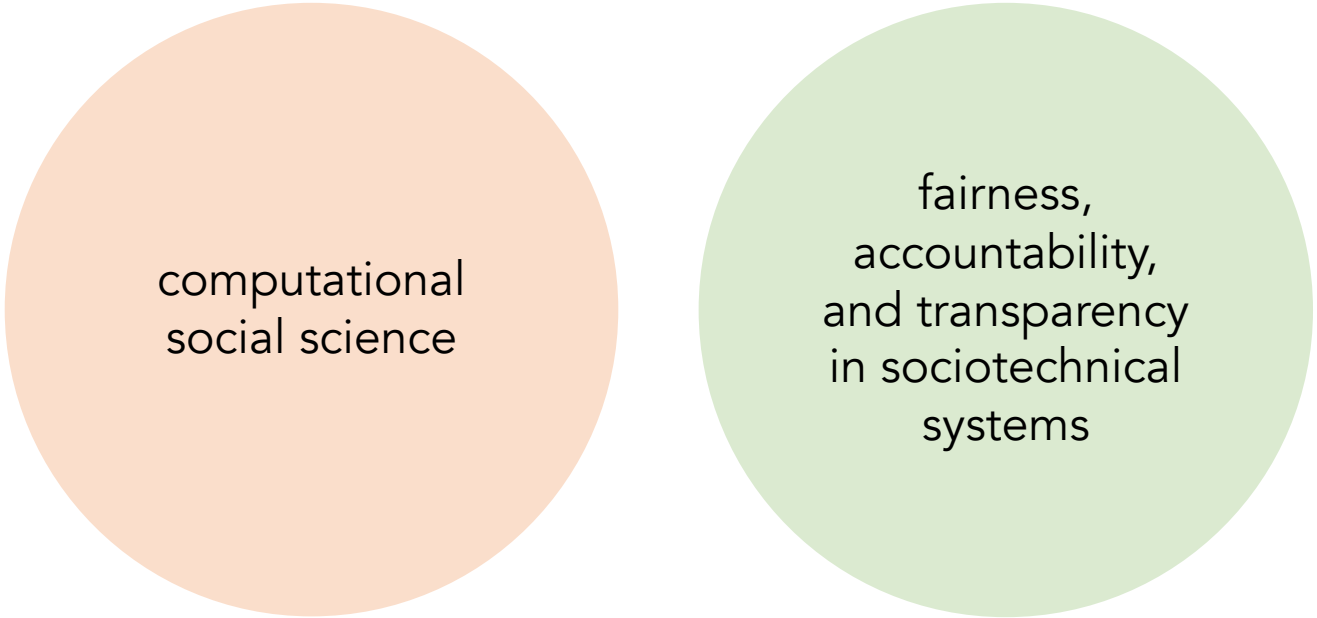
Measurement is governance

Measurement is everywhere, often implicit

Harms emerge from mismatches in the measurement process

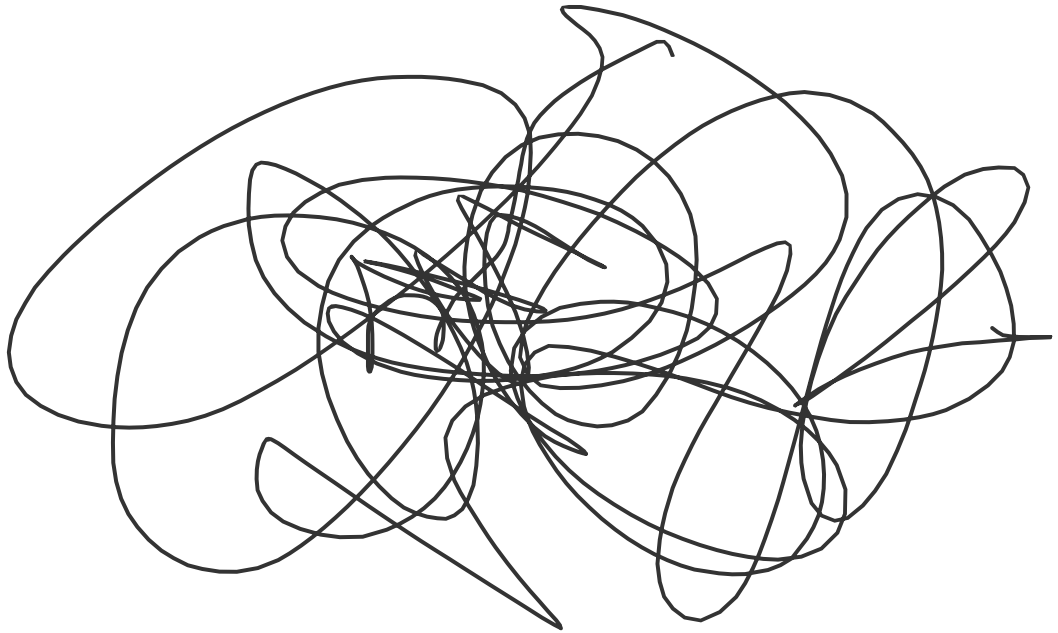
Construct reliability & validity help us interrogate this often-obscured process to prevent & mitigate harms

Reconciles understandings of risk, systems, social structure



[Measuring] social structure in sociotechnical systems

Thank you!

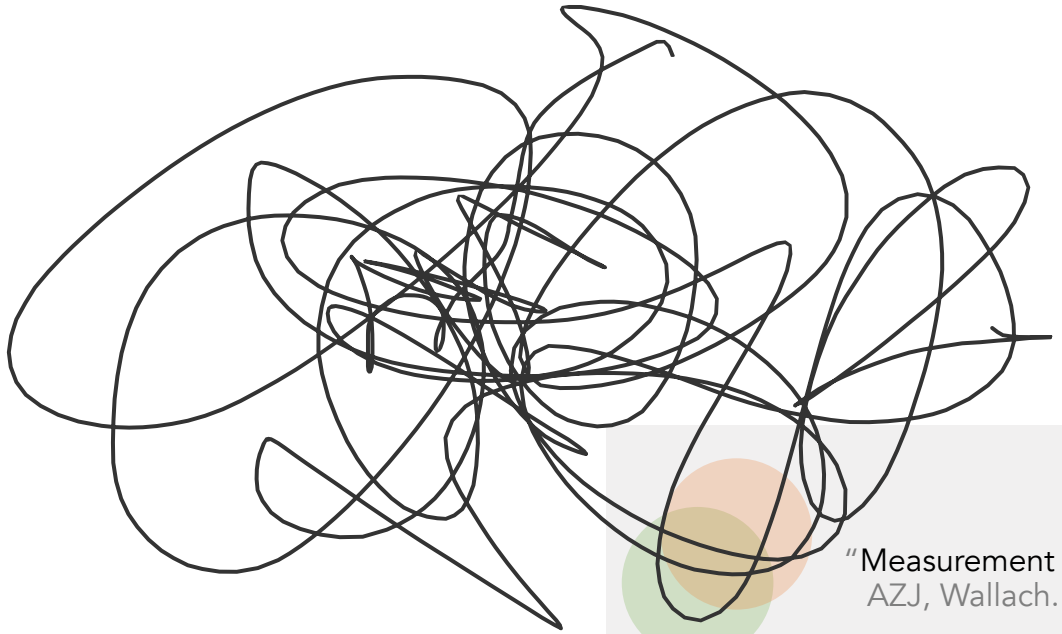


Abigail Jacobs

azjacobs@umich.edu

azjacobs.com | @az_jacobs

Thank you!



Abigail Jacobs

azjacobs@umich.edu

azjacobs.com | @az_jacobs

"Measurement and Fairness"
AZJ, Wallach. FAccT 2021

"Translation tutorial: The meaning and measurement of 'bias': Lessons from natural language processing"
AZJ, Blodgett, Barocas, Daumé, Wallach. FAccT 2020

"Unsafe at any AUC: Uncovering sociotechnical control for responsible AI"
in prep—AZJ, Kroll, Smart, Zeide

"Internet-human infrastructures: Lessons from Havana's StreetNet "
AZJ, Dye. Working paper