Discriminatory & Liberatory Algorithms: Restructuring Algorithmic "Fairness"

Manuel Sabin
UC Berkeley

Algorithmic Fairness

- COMPAS: Algorithm predicting recidivism
- Recidivism: Whether a person awaiting trial for a suspected crime will commit “further” crimes if not detained

Bernard Parker, left, was rated high risk; Dylan Fugett was rated low risk. (Josh Ritchie for ProPublica)

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

For those that didn't recidivate, blacks were twice as likely than whites to be given a high risk score

Algorithmic Fairness

- Creating more “Fair” recidivism predictors is commonly used as motivation for new work in the field
- ProPublica was a “watchdog” effort
- “Fair” recidivism predictors are still from the POV of the jailer

Hard to not talk about COMPAS when talking about Algorithmic Fairness because it’s always talked about in Algorithmic Fairness

Algorithmic “Fairness”

- Creating more “Fair” recidivism predictors is commonly used as motivation for new work in the field
- ProPublica was a “watchdog” effort
- “Fair” recidivism predictors are still from the POV of the jailer

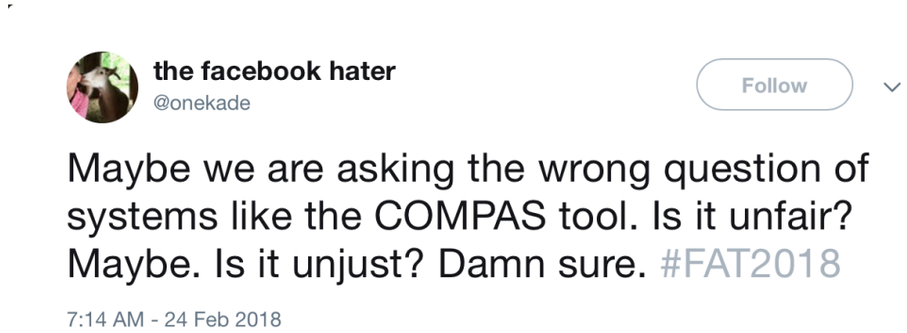
Algorithmic “Fairness”

- Normativity is the phenomenon in human societies of designating some actions or outcomes as good or desirable or permissible and others as bad or undesirable or impermissible
- Value judgement
- Culturally dependent and shifting
- “Fair” is a normative term

No math
We’ll use normative a lot but will try to not have too much jargon

Fair normative -> Do you believe Prison Industrial Complex and punitive justice? Or rehabilitative or restorative justice? These are normative. If you believe PIC is inherently “unfair,” can you have a “fair” recidivism predictor?

Algorithmic “Fairness”



Algorithmic “Fairness”

- What do we do if those studying carcerality say creating “Fair” recidivism predictors isn’t helpful?

- Algorithms, even ones attempting “Fairness,” codify and legitimize the systems using them, and thus the main question of this nascent field is this:

    What should our *role* be as Fairness researchers?

I’ll try to make progress towards answering this question by the end of the talk with a framework for a restructuring of the field.

But first we’ll further motivate a need for change with 5 traps the field seems prone to. After, we’ll introduce the framework and argue its merit by its lack of susceptibility to these traps

Fairness and Abstraction in Sociotechnical Systems

[Selbst-boyd-Friedler-Venkatasubramanian-Vertesi19]

“Technical systems are subsystems. Fairness and justice are properties of social and legal systems like employment and criminal justice, not properties of the technical tools within. **To treat fairness and justice as terms that have meaningful application to technology separate from a social context is therefore to make a category error**, or as we posit here, an abstraction error.”

This paper says that the notions of abstraction and portability that we love in computer science are not compatible with sociotechnical systems

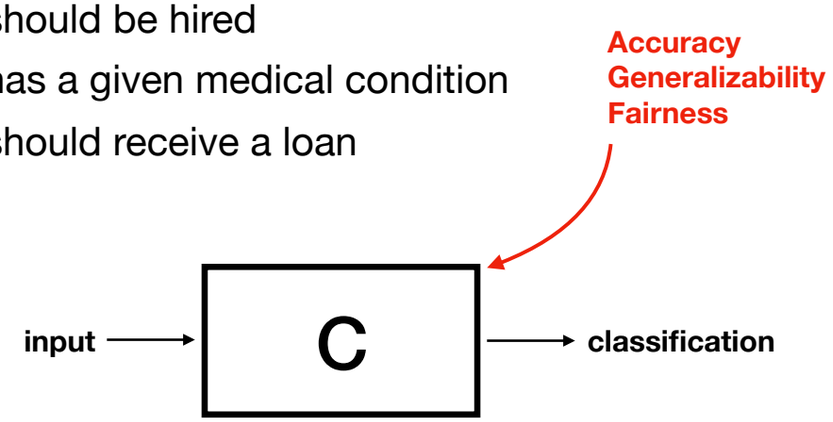
Five Traps

- Framing Trap
- Portability Trap
- Formalism Trap
- Ripple Effect Trap
- Solutionism Trap

Framing Trap

Failure to model the entire system over which a social criterion, such as fairness, will be enforced

- Say I want to classify,
  - Who will recidivate
  - Who should be hired
  - Who has a given medical condition
  - Who should receive a loan
  - etc.



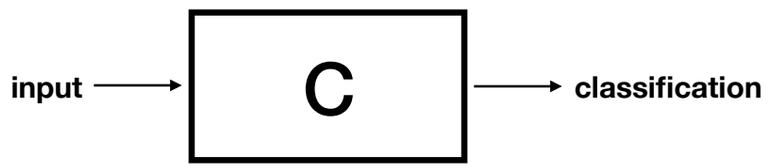
In many Fairness talks you see something like this

Notice the POV we take here

Framing Trap

Failure to model the entire system over which a social criterion, such as fairness, will be enforced

Algorithmic Frame: Representation of input and labeling of outputs are assumed and now we want to build a classifier



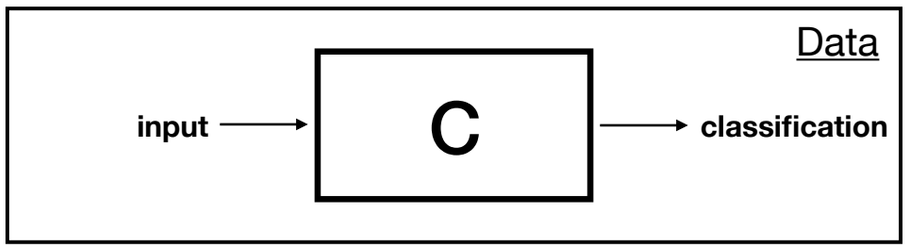
- End-to-end guarantees are its accuracy and ability to generalize to new inputs from the same distribution
- Impossible to define “Fairness” at this frame

We don’t even think about representation of input as having categories of groups of races or the sort yet

Framing Trap

Failure to model the entire system over which a social criterion, such as fairness, will be enforced

Data Frame: Representation of input and labeling of outputs are interrogated and are part of the engineering of the technology



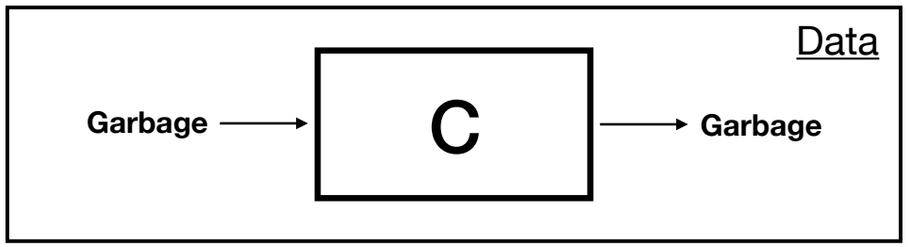
- End-to-end guarantees are the same as Algorithmic frame *plus* statistical parities on categories of the data
- But are statistical parities a good proxy for “Fairness?”

Break data into groups of protected classes

Framing Trap

Failure to model the entire system over which a social criterion, such as fairness, will be enforced

Data Frame: Representation of input and labeling of outputs are interrogated and are part of the engineering of the technology

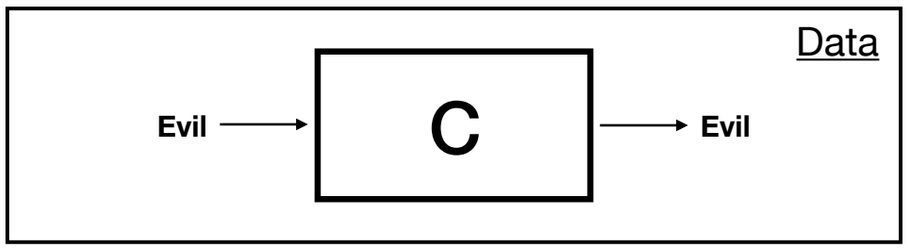


- End-to-end guarantees are the same as Algorithmic frame *plus* statistical parities on categories of the data
- But are statistical parities a good proxy for “Fairness?”

Framing Trap

Failure to model the entire system over which a social criterion, such as fairness, will be enforced

Data Frame: Representation of input and labeling of outputs are interrogated and are part of the engineering of the technology



- End-to-end guarantees are the same as Algorithmic frame *plus* statistical parities on categories of the data
- But are statistical parities a good proxy for “Fairness?”

If data is historically racist, then C learns to be racist….except faster

This isn’t just a matter of making C tamper-resilient or some similar crypto-like notion. Let’s see why

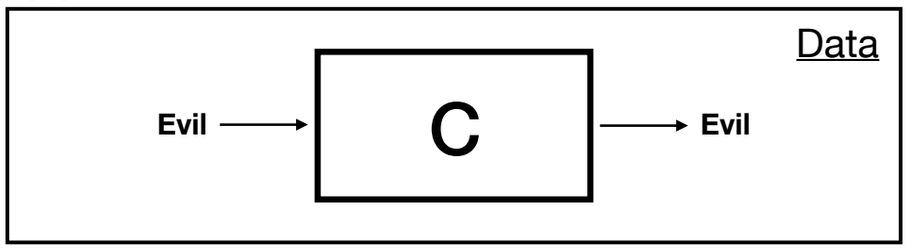
Framing Trap

Failure to model the entire system over which a social criterion, such as fairness, will be enforced

- Say I want to classify,
  - Who will recidivate
  - Who should be hired
  - Who has a given medical condition
  - Who should receive a loan
  - etc.

Who collects data and how?

Who chooses relevant features?

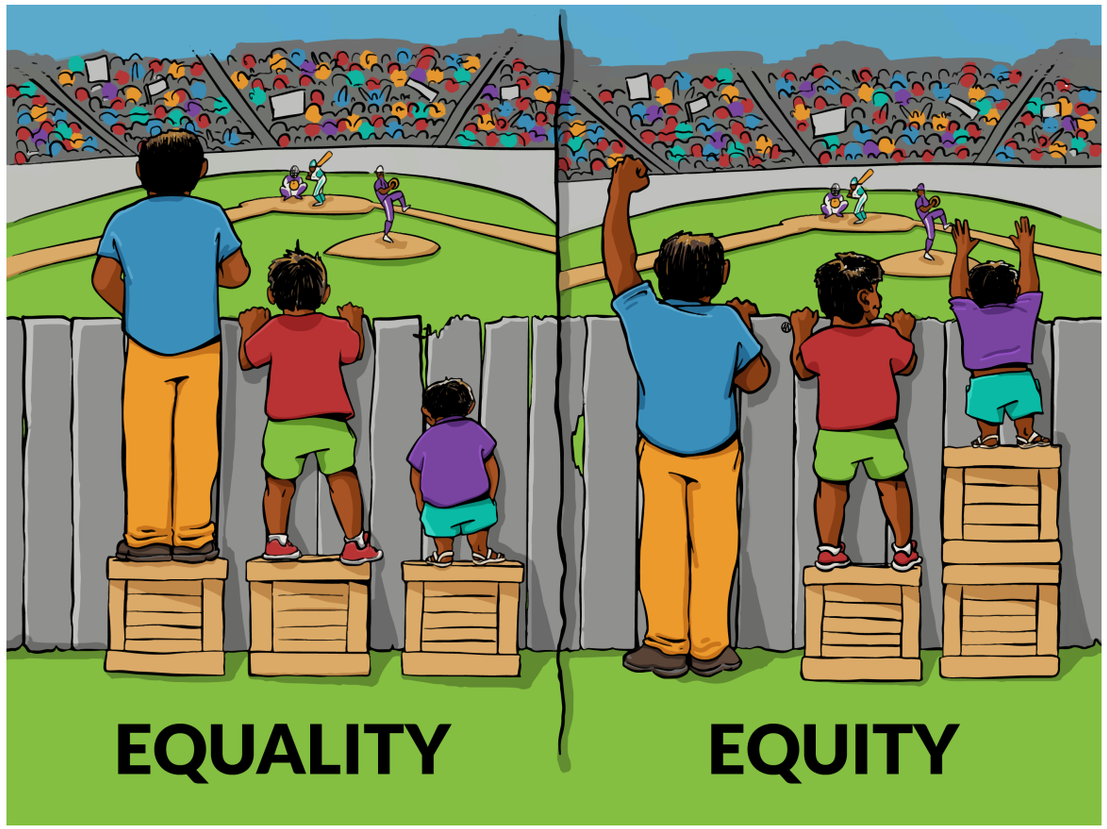


If choosing who to admit to college, are we using SAT scores as a feature? Does that privilege affluence?

Recidivism isn't committing a crime. It's committing a crime and getting arrested for it. If a black area is overpoliced, then the 'data collection' of recidivism statistics will be skewed.

At the end of the day, are these algorithms "Fair" in the social justice way we've been using the word?

Statistical parity across categories may not capture that



From http://interactioninstitute.org/illustrating-equality-vs-equity/

At the end of the day we want a notion of equity or justice. We may not even want things to be “Fair.”

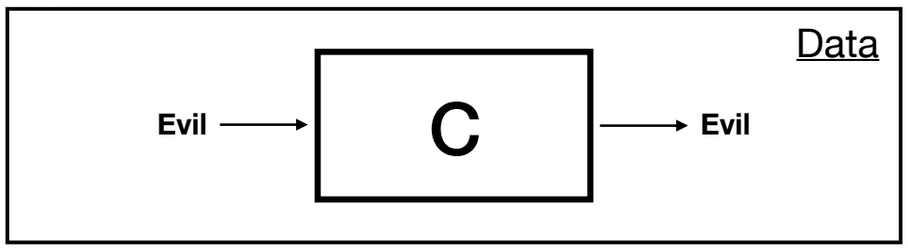
Framing Trap

Failure to model the entire system over which a social criterion, such as fairness, will be enforced

- Say I want to classify,
  - Who will recidivate
  - Who should be hired
  - Who has a given medical condition
  - Who should receive a loan
  - etc.

Who collects data and how?

Who chooses relevant features?



These questions can be argued to be in the data frame but they're often not. And they really start getting at the frame we need to be in…

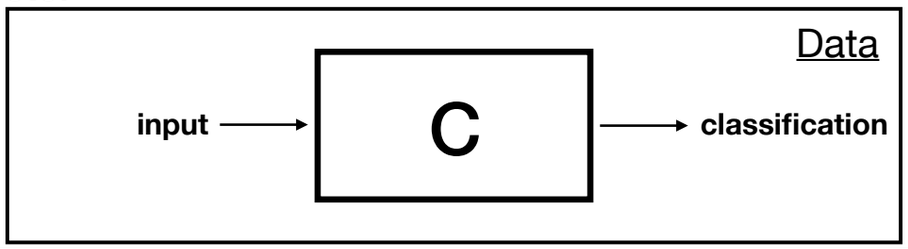
Framing Trap

Failure to model the entire system over which a social criterion, such as fairness, will be enforced

- Say I want to classify,
  - Who will recidivate
  - Who should be hired
  - Who has a given medical condition
  - Who should receive a loan
  - etc.

Who collects data and how?

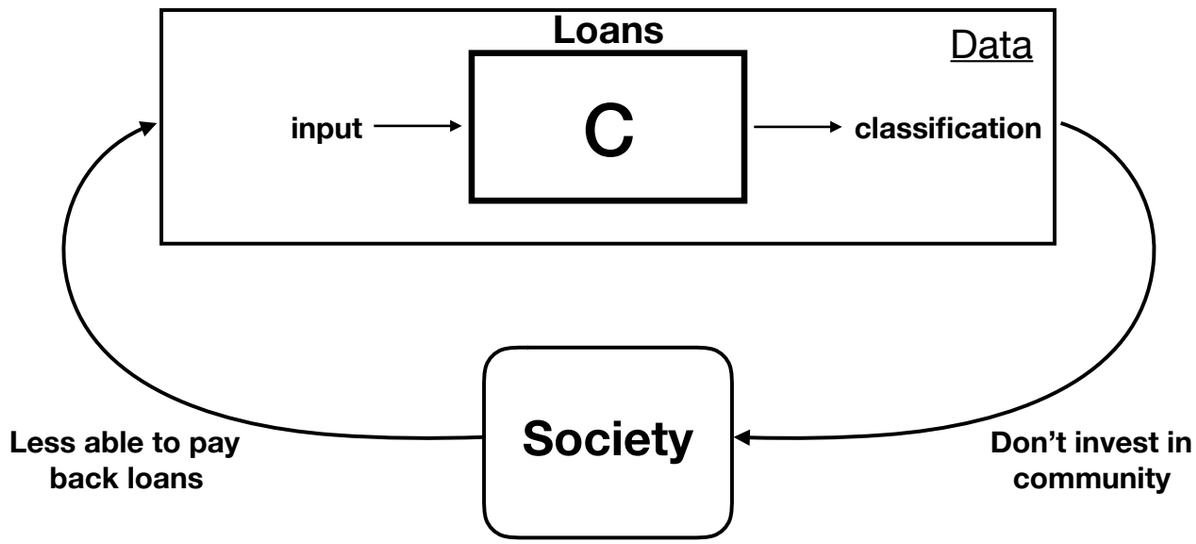
Who chooses relevant features?



These questions can be argued to be in the data frame but they're often not. And they really start getting at the frame we need to be in…

Framing Trap

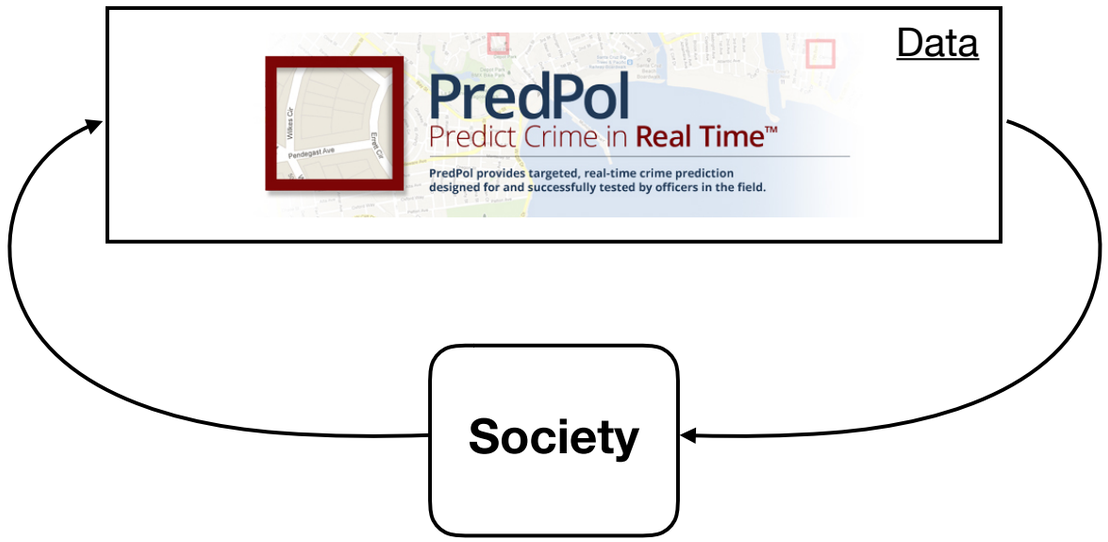
Failure to model the entire system over which a social criterion, such as fairness, will be enforced



Sociotechnical frame

Framing Trap

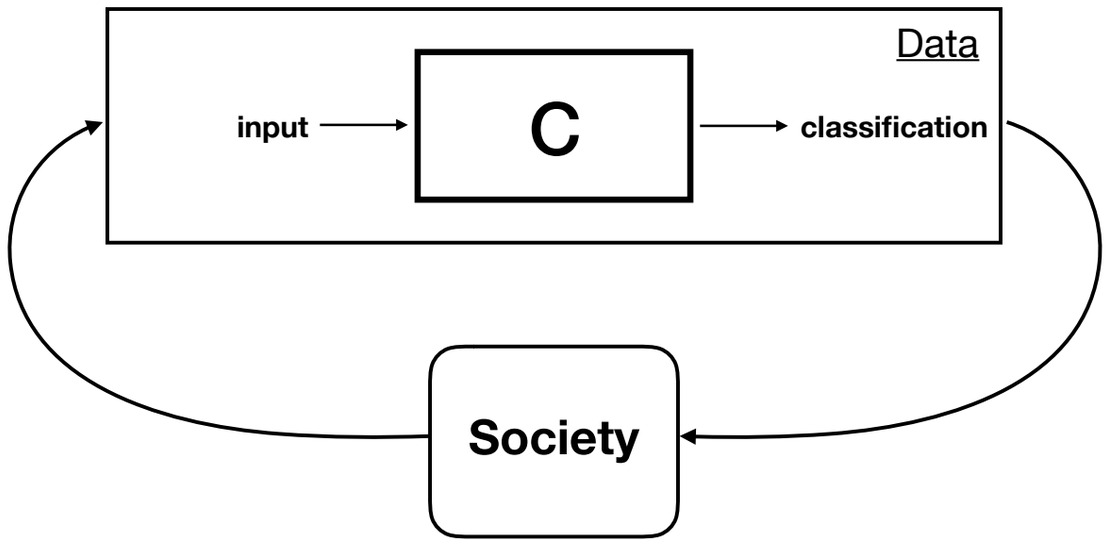
Failure to model the entire system over which a social criterion, such as fairness, will be enforced



PredPol doesn’t predict crimes, it predicts arrests. It’s a self-fulfilling prophecy

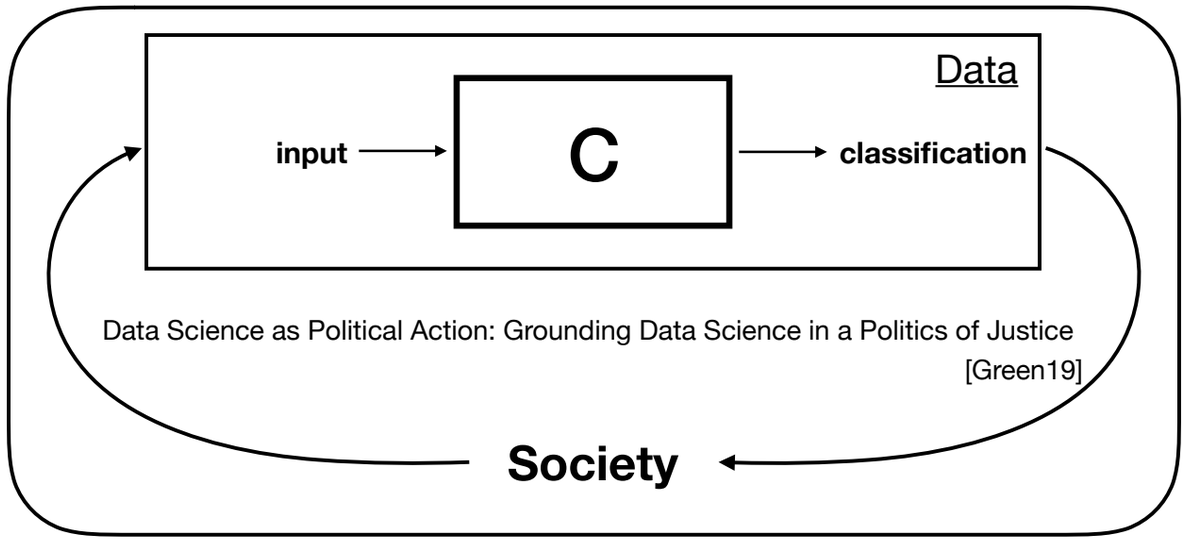
Framing Trap

Failure to model the entire system over which a social criterion, such as fairness, will be enforced



Framing Trap

Failure to model the entire system over which a social criterion, such as fairness, will be enforced



This whole frame is set up within society as well, not outside of it

Should we do actuarial risk assessment for loans?
What does is mean to call someone “competent” for a job?

Regardless of what your answers are for those questions, we each have worldview and politics and those are imbued into the algorithms and what we consider “Fair”

Even if it’s apolitical. Neutrality is a stance, and it’s one that’s often not in favor of the marginalized.

Five Traps

- Framing Trap
- Portability Trap
- Formalism Trap
- Ripple Effect Trap
- Solutionism Trap

Five Traps

- Framing Trap
- Portability Trap
- Formalism Trap
- Ripple Effect Trap
- Solutionism Trap

Portability Trap

Failure to understand how repurposing algorithmic solutions designed for one social context may be misleading, inaccurate, or otherwise do harm when applied to a different context

- Ex: False positives not so bad in resume screening, but very bad for predicting recidivism
- “Legal scholar and philosopher Deborah Hellman has argued that what we mean by "discrimination" is actually wrongful discrimination: we make distinctions all the time, but only cultural context can determine when the basis for discrimination is morally wrong”

Race is the product of racism;
Racism is not the product of race.
- Dorothy Roberts

Race should only ever be considered to combat Racism.

Portability Trap

Failure to understand how repurposing algorithmic solutions designed for one social context may be misleading, inaccurate, or otherwise do harm when applied to a different context

- Ex: False positives not so bad in resume screening, but very bad for predicting recidivism
- “Legal scholar and philosopher Deborah Hellman has argued that what we mean by "discrimination" is actually wrongful discrimination: we make distinctions all the time, but only cultural context can determine when the basis for discrimination is morally wrong”

The tall guy was discriminated against in the Equality v Equity picture. He had no box

We’ll come back to this wrongful discrimination idea

Five Traps

- Framing Trap
- Portability Trap
- Formalism Trap
- Ripple Effect Trap
- Solutionism Trap

Five Traps

- Framing Trap
- Portability Trap
- Formalism Trap
- Ripple Effect Trap
- Solutionism Trap

Formalism Trap

Failure to account for the full meaning of social concepts such as fairness, which can be procedural, contextual, and contestable, and cannot be resolved through mathematical formalisms

Formalism Trap

Failure to account for the full meaning of social concepts such as fairness, which can be procedural, contextual, and contestable, and cannot be resolved through mathematical formalisms

- Inherent Trade-Offs in the Fair Determination of Risk Scores [Kleinberg-Mullainathan-Raghavan17]

Formalism Trap

Failure to account for the full meaning of social concepts such as fairness, which can be ***procedural, contextual, and contestable****, and cannot be resolved through mathematical formalisms*

- Inherent Trade-Offs in the Fair Determination of Risk Scores [Kleinberg-Mullainathan-Raghavan17]

Formalism Trap

Failure to account for the full meaning of social concepts such as fairness, which can be ***procedural, contextual, and contestable****, and cannot be resolved through mathematical formalisms*

- Inherent Trade-Offs in the Fair Determination of Risk Scores [Kleinberg-Mullainathan-Raghavan17]
- Disparate Impact: 80% rule codified in Algorithmic Fairness, but, in law, this kicks off a procedural review of Disparate Impact

whether the plaintiff can show an equally effective but less discriminatory alternative that the defendant refused to use

Formalism Trap

Failure to account for the full meaning of social concepts such as fairness, which can be ***procedural, contextual, and contestable****, and cannot be resolved through mathematical formalisms*

- Inherent Trade-Offs in the Fair Determination of Risk Scores [Kleinberg-Mullainathan-Raghavan17]
- Disparate Impact: 80% rule codified in Algorithmic Fairness, but, in law, this kicks off a procedural review of Disparate Impact
- Fairness is cultural and will change over time. “To set them in stone—or in code—is to pick sides, and to do so without transparent process violates democratic ideals.”

Five Traps

- Framing Trap
- Portability Trap
- Formalism Trap
- Ripple Effect Trap
- Solutionism Trap

Five Traps

- Framing Trap
- Portability Trap
- Formalism Trap
- Ripple Effect Trap
- Solutionism Trap

Ripple-Effect Trap

Failure to understand how the insertion of technology into an existing social system changes the behaviors and embedded values of the pre-existing system

- Sociotechnical feedback loop: loans
- “Incapacitation is only one of at least seven existing rationales for punishment: prevention, incapacitation, rehabilitation, deterrence, education, retribution, and restoration.”
- Recidivism prediction for sentencing focuses on incapacitation
- Risk assessment may privilege the social value that is quantified

It’s like if you drop you keeps in the dark on the street and only look under the streetlamp because that’s where the light is

Five Traps

- Framing Trap
- Portability Trap
- Formalism Trap
- Ripple Effect Trap
- Solutionism Trap

Five Traps

- Framing Trap
- Portability Trap
- Formalism Trap
- Ripple Effect Trap
- Solutionism Trap

Solutionism Trap

Failure to recognize the possibility that the best solution to a problem may not involve technology

- “Because fair-ML is rooted in computer science, there is no concept of the system without a technical intervention” (If you have a hammer…)

[Toyoma13]

- Objective: Keep track of sales and stock of agricultural goods at local shops in an African country so that the non-profit could target specific shopkeepers for training and inventory support
- Idea: Shop owners send in sales data via SMS texts
- Alternative Suggestion: Human operator call the shops
- “Problem:” Voice calls require little design. No need for HCI

Solutionism Trap

Failure to recognize the possibility that the best solution to a problem may not involve technology

- “Because fair-ML is rooted in computer science, there is no concept of the system without a technical intervention” (If you have a hammer…)

[Toyoma13]

- “Sometimes the core tenets of HCI lead us away from HCI… Sometimes the right thing is to walk away. Not all problems are HCI problems, and we don’t want to become the proverbial hammer that sees everything as a nail.”

Solutionism Trap

Failure to recognize the possibility that the best solution to a problem may not involve technology

- “Because fair-ML is rooted in computer science, there is no concept of the system without a technical intervention” (If you have a hammer…)

[Toyoma13]

- “Sometimes the core tenets of HCI lead us away from HCI… Sometimes the right thing is to walk away. Not all problems are HCI problems, and we don’t want to become the proverbial hammer that sees everything as a nail.”
- Don’t just do something, stand there!

Maybe we don’t need a “Fair” facial recognition algorithm that tries to classify people’s genders
Still instantiates a binary, static notion of gender

Solutionism Trap

- Algorithmic Fairness realizes that technical solutions to social problems are often harmful
- …But is Algorithmic Fairness a technical solution to *that* social problem?

Change

- How do we start? We can't all be experts in every intricately sociotechnical domain
- But people were already doing "Algorithmic Fairness"…it was just problem-specific
  - Calling out *unfair* algs in carcerality, hiring, loans, etc.
  - Using algorithms to increase Equity
- TCS has brought abstraction to this effort
- …And this is being said to not be possible for "Fairness"

That's a rough crash-course of sociotechnical type of thinking. You shouldn't be experts on it now if you weren't already, but it should cause enough concern to realize that some change is needed

Crypto wants *zero* info leakage. That's formalizable. Fairness isn't

Change

- Because “Fairness” is normative, change needed is mostly cultural
- Science & Technology Studies (STS) is “an anthropological study of the act of science”
  - Research is a social activity and has cultural norms
  - Conference vs Journals
  - “First Do No Harm” vs “Move Fast, Break Things” vs No Principle
  - Multidisciplinary vs Interdisciplinary
  - Community input required?

The Myth in Methodology: Towards a Recontextualization of Fairness in Machine Learning [Green-Hu18]

Multi where people work in parallel but may talk past each other

We need to think in these terms

That these are cultural choices we get to make at the start of this field

Change

- Community input required?

“I have not yet personally read a published HCI-for-development paper in which at least one author did not spend time interacting with intended users or beneficiaries. Meanwhile, as a paper reviewer for HCI publications, whenever I see studies with no time in the field, the papers are rejected—typically by unanimous reviewer decision. I can’t say the same thing for our non-HCI computer science or engineering cousins.”

[Toyoma13]

Change

- Problem: Cultural change is hard.
- 5 Traps paper argues for “heterogeneous engineers”
  - Are we *able* to write from perspective of the marginalized
  - Will TCS start doing IRBs?
- Structural change that embeds cultural change
  - 5 Traps paper recommends including a section in all papers, detailing how the traps are avoided

Discriminatory & Liberatory Algorithms

- No meaningful fixed definition of “Fairness”
- Is all Fairness work void then? **No.**

¬ Algorithmic “Fairness” Test:

If all you care about is this specific statistical parity, and you forget all history and context, then your alg is “Fair”

A large takeaway from this talk should be a salvaging of Algorithmic Fairness in light of the the critiques against the field, including the 5 traps paper

Discriminatory & Liberatory Algorithms

- No meaningful fixed definition of “Fairness”
- Is all Fairness work void then? **No.**

¬ Algorithmic “Fairness” Test:

Even if all you care about is this specific statistical parity, and *even if* you forget all history and context, then your alg is *still* “Discriminatory”

Discriminatory & Liberatory Algorithms

- No meaningful fixed definition of “Fairness”
- Is all Fairness were void then? **No.**

¬ Algorithmic “Fairness” Test: **Increasingly sophisticated suite of statistical tests**

Even if all you care about is this specific statistical parity, and *even if* you forget all history and context, then your alg is *still* warranting contextual/procedural review

Do away with the word Fairness

This sort of thing actually happens in law, like in Disparate Impact

We have this sort of thing in crypto vs cryptanalysis. Except crypto is formalizable and the directions partition each other.
Fairness doesn’t have utility in one direction. In the other direction it does, so long as it’s phrased as kicking off contextual/procedural review

Discriminatory & Liberatory Algorithms

Discriminatory Algorithms

- **Goal: Expose algorithmic injustices**

Discriminatory & Liberatory Algorithms

Watchdog

- **Goal: Expose algorithmic injustices**

Discriminatory & Liberatory Algorithms

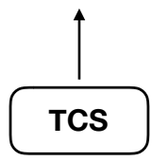
Watchdog

- **Goal: Expose algorithmic injustices**
- **Frame: Sociotechnical, interdisciplinary with activists and community**

Discriminatory & Liberatory Algorithms

Watchdog

- **Goal: Expose algorithmic injustices**
- **Frame: Sociotechnical, interdisciplinary with activists and community**



- **Goal: Develop toolkit of statistical tests to be used in watchdog efforts**
- **Frame: Data frame, possibly abstract**

No normative language

Do away with the word Fairness. No more Algorithmic “Fairness”

Discriminatory & Liberatory Algorithms

Watchdog

Problem-specific and normative
Merit argued interlacing quantitative and qualitative reasoning

- Goal: Expose algorithmic injustices
- Frame: Sociotechnical, interdisciplinary with activists and community
- Ultimate Goal: Gather enough evidence and develop enough language to enact policy change or guide activism with activists and community

TCS

- Goal: Develop toolkit of statistical tests to be used in watchdog efforts
- Frame: Data frame, possibly abstract

No normative language

This negates most of the traps by definition. This makes sense, this was the problem-centric activist-type work that preceded Algorithmic Fairness.

Fairness papers can be rewritten with the negation of Fairness definitions in mind.

Discriminatory & Liberatory Algorithms

Watchdog

- Goal: Expose algorithmic injustices
- Frame: Sociotechnical, interdisciplinary with activists and community
- Ultimate Goal: Gather enough evidence and develop enough language to enact policy change or guide activism with activists and community

TCS

- Goal: Develop toolkit of statistical tests to be used in watchdog efforts
- Frame: Data frame, possibly abstract

No normative language

Problem-specific and normative
Merit argued interlacing quantitative and qualitative reasoning

Liberatory Algorithms

- Goal: Use algorithms to expose or correct injustices
- Frame: Sociotechnical, interdisciplinary with activists and community
- Ultimate Goal: Algorithms to put power and autonomy into the hands of the marginalized

Specific population, context, point in time

Algorithm predicting police officers at risk of profiling or being violent, ML used on police body cams to find that officers speak much less respectfully to black people

Hesitant to draw arrow from TCS to Liberatory
TCS born out of Discriminatory Algs, but then tried to be Liberatory
Toolkit bad analogy, TCS makes raw material

Ruha Benjamin writes a lot about using algorithms for liberatory purposes

Might be making an app for community organizing or problem-specific data collection

May be more of an HCI problem than anything.
Then what's the unifying theme?

Discriminatory & Liberatory Algorithms

Watchdog

- Goal: Expose algorithmic injustices
- Frame: Sociotechnical, interdisciplinary with activists and community
- Ultimate Goal: Gather enough evidence and develop enough language to enact policy change or guide activism with activists and community

TCS

- Goal: Develop toolkit of statistical tests to be used in watchdog efforts
- Frame: Data frame, possibly abstract

No normative language

Problem-specific and normative
Merit argued interlacing quantitative and qualitative reasoning

Liberatory Algorithms

- Goal: Use algorithms to expose or correct injustices
- Frame: Sociotechnical, interdisciplinary with activists and community
- Ultimate Goal: Algorithms to put power and autonomy into the hands of the marginalized

Specific population, context, point in time

Overall Goal: A more Just world with respect to algorithm usage

"Restructuring" is a bit of a lie. This just susses out where the entrance of TCS fits into these efforts that were implicitly *ongoing*

Race is the product of racism;
Racism is not the product of race.
- Dorothy Roberts

Race should only ever be considered to combat Racism.

Race should only ever be considered
to expose or correct against Racism

Discriminatory & Liberatory Algorithms

Watchdog

- Goal: Expose algorithmic injustices
- Frame: Sociotechnical, interdisciplinary with activists and community
- Ultimate Goal: Gather enough evidence and develop enough language to enact policy change or guide activism with activists and community

TCS

- Goal: Develop toolkit of statistical tests to be used in watchdog efforts
- Frame: Data frame, possibly abstract

No normative language

Problem-specific and normative
Merit argued interlacing quantitative and qualitative reasoning

Liberatory Algorithms

- Goal: Use algorithms to expose or correct injustices
- Frame: Sociotechnical, interdisciplinary with activists and community
- Ultimate Goal: Algorithms to put power and autonomy into the hands of the marginalized

Specific population, context, point in time

Overall Goal: A more Just world with respect to algorithm usage

What is TCS's *role* in this?

Discriminatory & Liberatory Algorithms

- Not just a renaming of different objectives of the field but attaches frame, methodologies, and *domain* boundaries to the categories
- Abstraction is isolated to developing a suite of statistical tools. All else is at the sociotechnical frame
- This framework is a social construct. But so is the current organically formed one
- This one is just intentional. And it's less susceptible to the traps

Five Traps

- Framing Trap
- Portability Trap
- Formalism Trap
- Ripple Effect Trap
- Solutionism Trap

Framework would get people to avoid the traps whether or not they’re trained to know how to or even spot them
(like the peer review suggestions gets people thinking sociotechnically whether or not they’re trained how)

We argue that the DLA framework aims at the same goal of producing a more just world with respect to algorithm usage while being much less prone to these traps.

Discriminatory & Liberatory Algorithms

Watchdog

- Goal: Expose algorithmic injustices
- Frame: Sociotechnical, interdisciplinary with activists and community
- Ultimate Goal: Gather enough evidence and develop enough language to enact policy change or guide activism with activists and community

TCS

- Goal: Develop toolkit of statistical tests to be used in watchdog efforts
- Frame: Data frame, possibly abstract

No normative language

Problem-specific and normative
Merit argued interlacing quantitative and qualitative reasoning

Liberatory Algorithms

- Goal: Use algorithms to expose or correct injustices
- Frame: Sociotechnical, interdisciplinary with activists and community
- Ultimate Goal: Algorithms to put power and autonomy into the hands of the marginalized

Specific population, context, point in time

Overall Goal: A more Just world with respect to algorithm usage

What is TCS's *role* in this?

Five Traps

- Framing Trap
- Portability Trap
- Formalism Trap
- Ripple Effect Trap
- Solutionism Trap

Discriminatory & Liberatory Algorithms

Watchdog

- Goal: Expose algorithmic injustices
- Frame: Sociotechnical, interdisciplinary with activists and community
- Ultimate Goal: Gather enough evidence and develop enough language to enact policy change or guide activism with activists and community

TCS

- Goal: Develop toolkit of statistical tests to be used in watchdog efforts
- Frame: Data frame, possibly abstract

No normative language

Problem-specific and normative
Merit argued interlacing quantitative and qualitative reasoning

Liberatory Algorithms

- Goal: Use algorithms to expose or correct injustices
- Frame: Sociotechnical, interdisciplinary with activists and community
- Ultimate Goal: Algorithms to put power and autonomy into the hands of the marginalized

Specific population, context, point in time

Overall Goal: A more Just world with respect to algorithm usage

What is TCS's *role* in this?

Discriminatory & Liberatory & Other Algorithms?

Recidivism Prediction

- Not a watchdog effort
- Not Liberatory: Written for use by jailer and not the affected population

Where does this effort exist in framework?

By *default* it doesn't!

Discriminatory & Liberatory & Other Algorithms?

But…

- Prison Industrial Complex still exists
- Recidivism predictors still exist

Don't we want less discriminatory ones as a 'harm reduction' tactic in the meantime?

We welcome such sociotechnical arguments!

Or maybe we think tech is going to be better than human error

Make that argument! Are there studies? We can't just presume it.

Besides, the tech is *used* by humans. It's still a judge taking predictor values into account, deciding when to use it, getting numb to it, not wanting tech oversight, etc. We need to take sociotechnical frame of how the tech will actually be used to call it fair

Discriminatory & Liberatory & Other Algorithms?

The key point is that its existence is not presumed.

Efforts that introduce *technology* into *society* must have their merit *justified* through a *sociotechnical* argument

Discriminatory & Liberatory Algorithms

Watchdog

- Goal: Expose algorithmic injustices
- Frame: Sociotechnical, interdisciplinary with activists and community
- Ultimate Goal: Gather enough evidence and develop enough language to enact policy change or guide activism with activists and community

TCS

- Goal: Develop toolkit of statistical tests to be used in watchdog efforts
- Frame: Data frame, possibly abstract

No normative language

Problem-specific and normative
Merit argued interlacing quantitative and qualitative reasoning

Liberatory Algorithms

- Goal: Use algorithms to expose or correct injustices
- Frame: Sociotechnical, interdisciplinary with activists and community
- Ultimate Goal: Algorithms to put power and autonomy into the hands of the marginalized

Specific population, context, point in time

Overall Goal: A more Just world with respect to algorithm usage

What is TCS's *role* in this?

These efforts we say *should* exist, and everything else needs to be argued for

Part 2

- Framework separated from instantiations of it with terminology & methodology
  - Framework draws domain boundaries and “Frame abstractions” for them
  - Methodology/terminology may be normative and evolving
- *Possible* methodology
  - Medical community: Community-Based Participatory Research
  - Feminist theory: Center the affected and most marginalized populations
  - Sociology: Heterogeneous engineering
  - HCI: *Bridging AI and HCI: Incorporating Human Values into the Development of AI Technologies* - Haiyi Zhu
  - Community Organizing: Power analysis

I’ll be co-running a workshop at FAT* on introducing community organizing principles and methodologies with Ezra Goss, a student in Georgia Tech’s Human-Centered Computing group, Lily Hu, a math and philosophy of tech student at Harvard, and Stephanie Teeple, an MD/PhD at UPenn using quantitative and qualitative tools to understand ML’s impact on marginalized communities in medicine

This work came out of and is continuing to form out of conversations with them

Conclusion

- DLA has been ongoing. This just names it and makes its domains explicit. And places the entrance of TCS amongst the ongoing efforts
- This framework is a social construct
- But so is the organic thing going on now. This one is just intentional. And it's less susceptible to the traps
- Hopefully fields can talk past each other less
- Can spend less time on critiques and more time doing constructive work under these principles

Social construct: Words matter. We might be used to swapping out dummy variables for things, but the words we use have real-world effects. AI Winter happened because normative terms got conflated with technical terms. But while the stakes there was the funding of researchers, the human cost of misuse of normative terms may be much higher here

I think the computer has from the beginning been a fundamentally conservative force. It has made possible the saving of institutions pretty much as they were, which otherwise might have had to be changed… banks were faced with the fact that the population was growing at a very rapid rate…
if the computer had not been invented, what would the banks have had to do? They might have had to decentralize, or they might have had to regionalize in some way. In other words, it might have been necessary to introduce a social invention, as opposed to the technical invention.

What the coming of the computer did, "just in time," was to make it unnecessary to create social inventions, to change the system in any way. So in that sense, the computer has acted as fundamentally a conservative force, a force which kept power or even solidified power where it already existed.

-Joseph Weizenbaum

Q: Did you have these concerns when you were designing the banking system?
A: Not in the slightest. It was a very technical job…there were a number of very, very difficult problems…It was a whale of a lot of fun attacking those hard problems, and it never occurred to me at the time that I was cooperating in a technological venture which had certain social side effects which I might come to regret. That never occured to me; I was totally wrapped up in my identity as a professional, and besides, it was just too much fun.

This field is in its formative stages and I'm not saying we're at this point yet or that we've been adopted enough to be here, but we should look to this as a cautionary tale when we're dealing with sociotechnical systems.