Algorithmic Economics

Sendhil Mullainathan

joint work with

Jens Ludwig, Ashesh Rambachan, Amanda Agan

and Diag Dvaneport

Why AI Needs Behavioral Economics

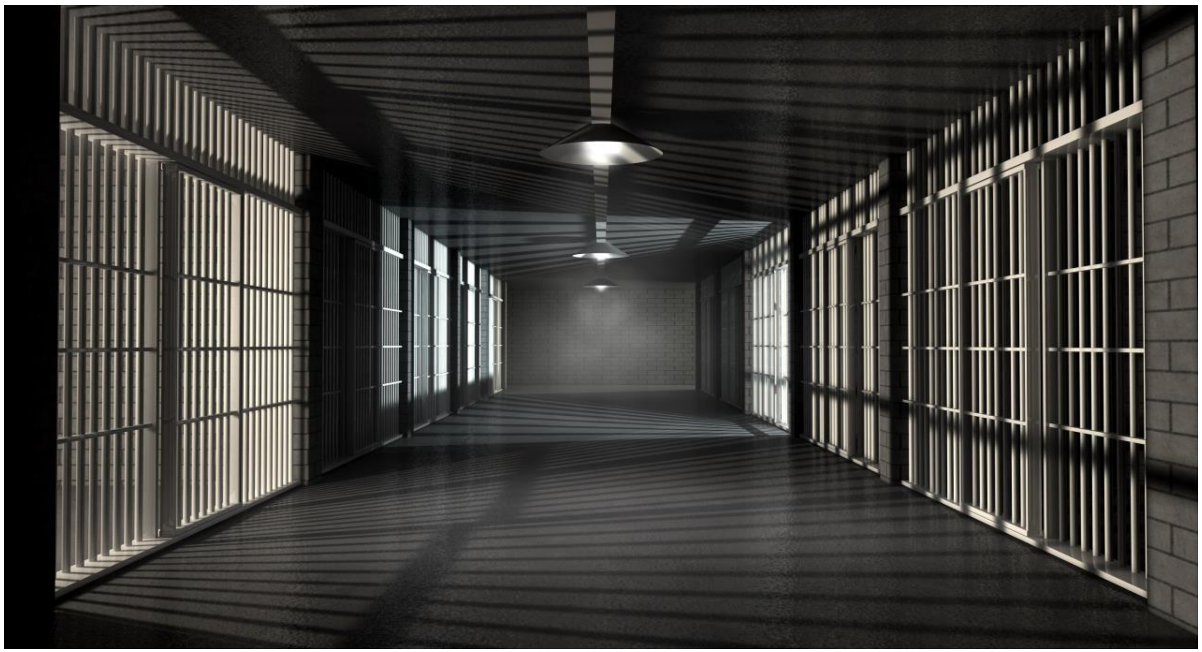
Why Behavioral Economics Needs AI

Why AI Needs Behavioral Economics

Why Behavioral Economics Needs AI

12 million arrests

Where to wait for trial?



12 million arrests

Where to wait for trial?

Implications for both society and arrestee

Jail stay: 2-3 months (as high as 9-12)

Law dictates jailing decision

Flight risk:
Will defendant appear at trial?

Public safety risk
Will defendant commit crime?

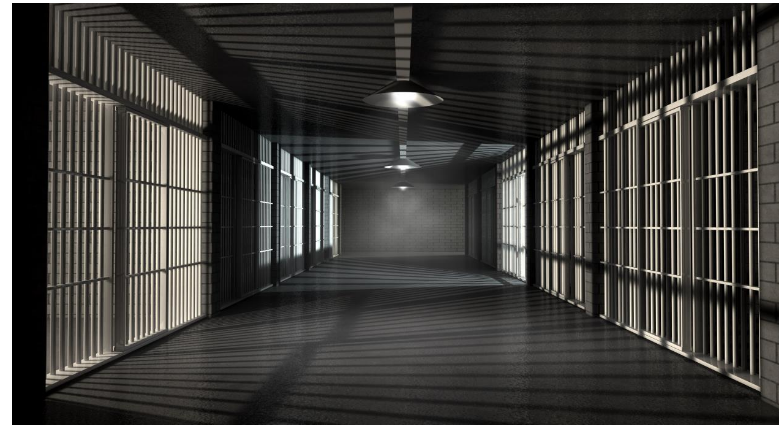
Judge must *predict risk*
And jail those with high risk

Judge is doing what <u>supervised learning</u> algorithms do

Machine Learning

Given X (image, soundwave,…)
Predict Y (face?, object,…)

Prediction Policy



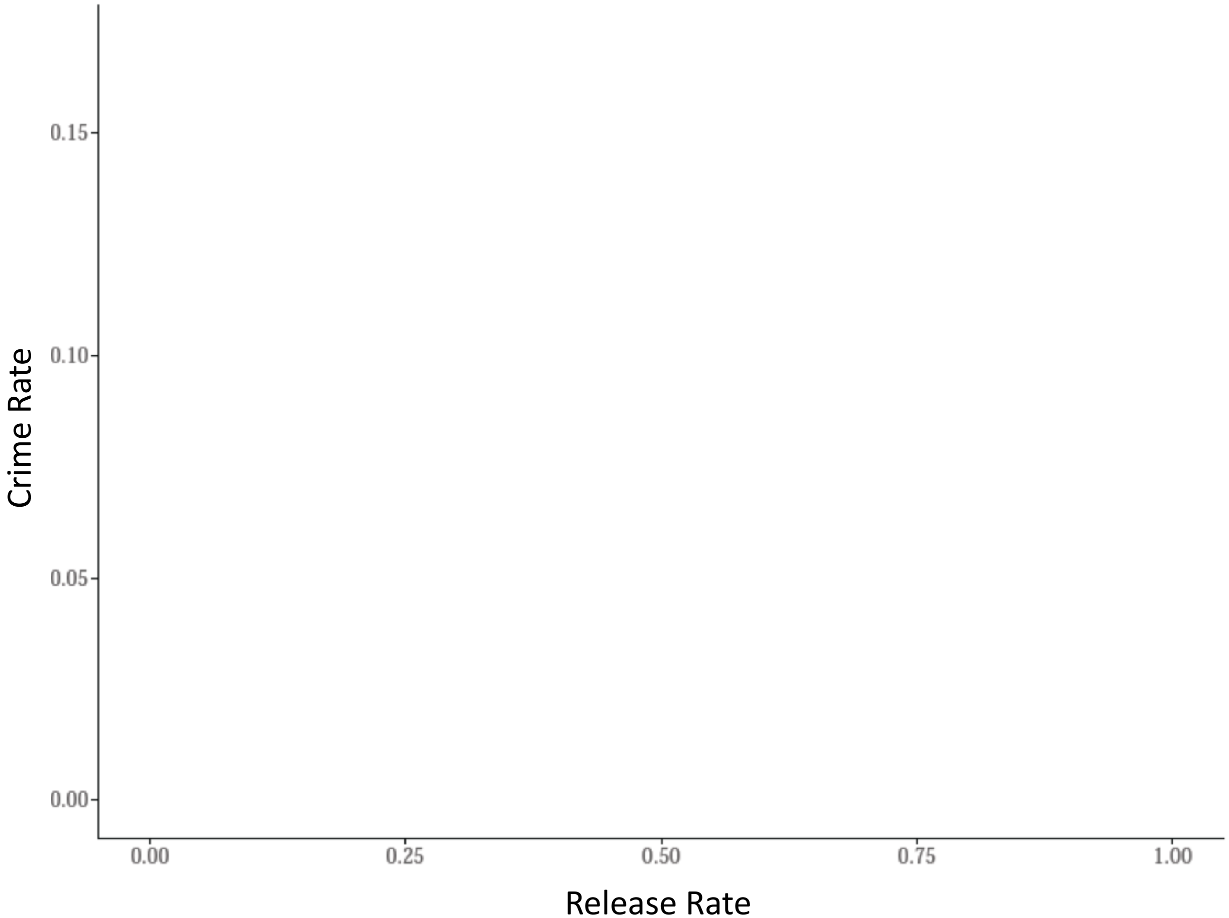
Given X (defendant characteristics)
Predict Y (flight risk)

Build
Defendant Risk Predictor

Evaluate
Compare to Judge

Data from New York State

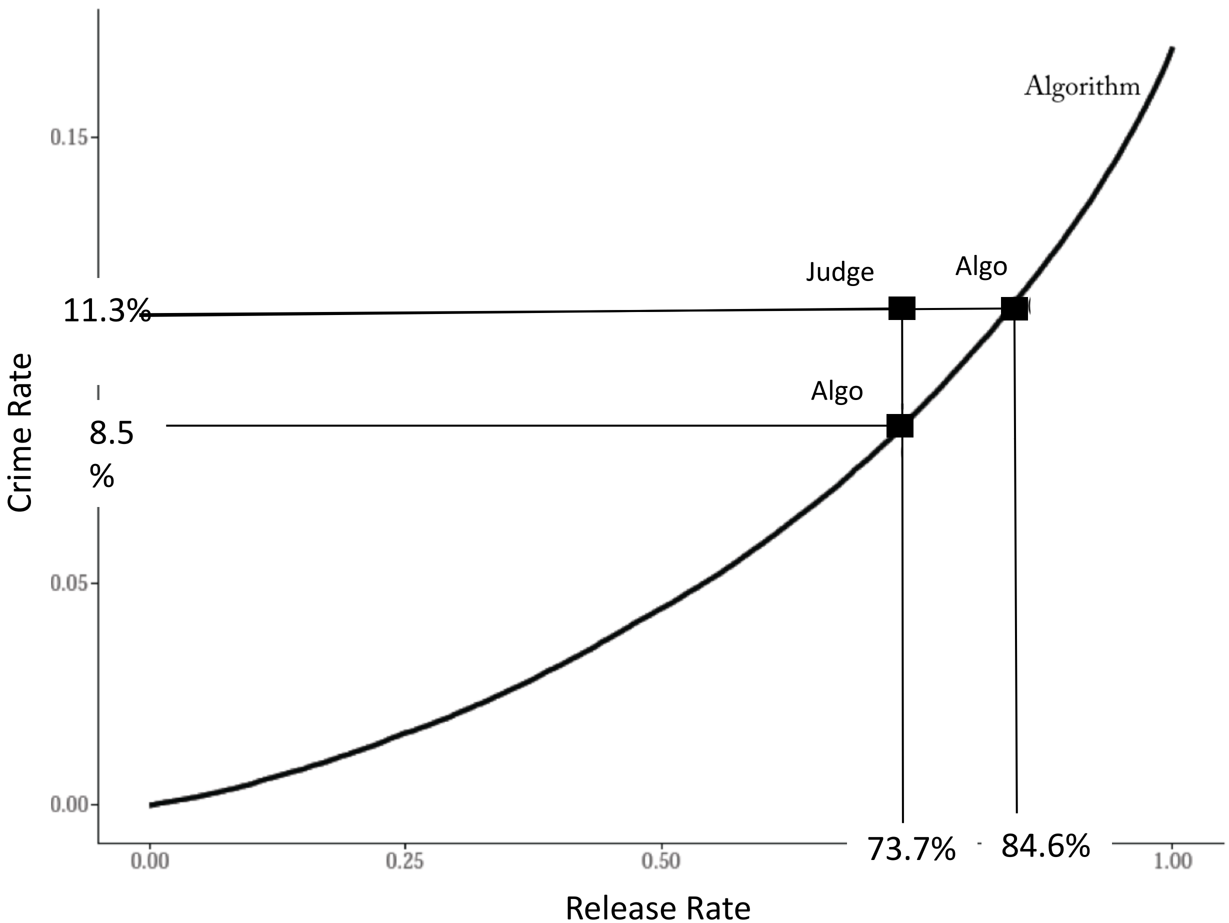
- 2011-13
- 758,027 cases
- Many judges
- Input variables: RAP sheet, current charge
- Label: Flight (failure to appear) or Arrested for crime



Rank all defendants by predicted risk

Calculate crime rate for releasing x% of them

Need a benchmark!



Keep jail population fixed
Reduce crime by 24%

Keep crime fixed
Reduce jail population by 40%

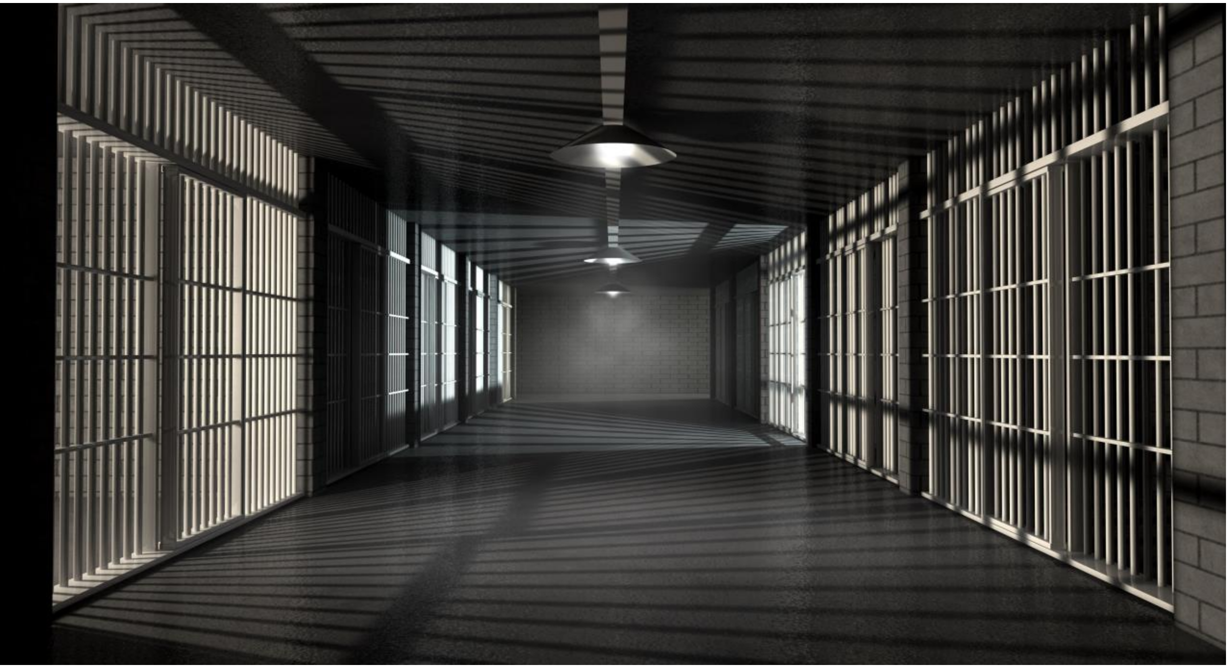
Close Riker's in July

Some subtle issues here
(see paper)

Selective labels
Omitted payoff bias
Racial bias
Equity

Rambachan (2021)

Can algorithms improve judicial decision making?



Keep crime fixed
Reduce jail population by 40%

Keep jail population fixed
Reduce crime by 24%

Can **reduce** racial disparities

The key is to build the algorithm correctly

Kleinberg et. al. (2018) "Human Decisions and Machine Predictions," QJE

Can algorithms improve judicial decision making?

Can algorithms improve decision making?

Key feature: Decision depends on some prediction

Domestic violence

Can algorithms improve decision making?

Key feature: Decision depends on some prediction

Domestic violence

Financial advice

<u>Can</u> algorithms improve decision making?

Key feature: Decision depends on some prediction

Domestic violence

Financial advice

Job search advice

Can algorithms improve decision making?

Key feature: Decision depends on some prediction

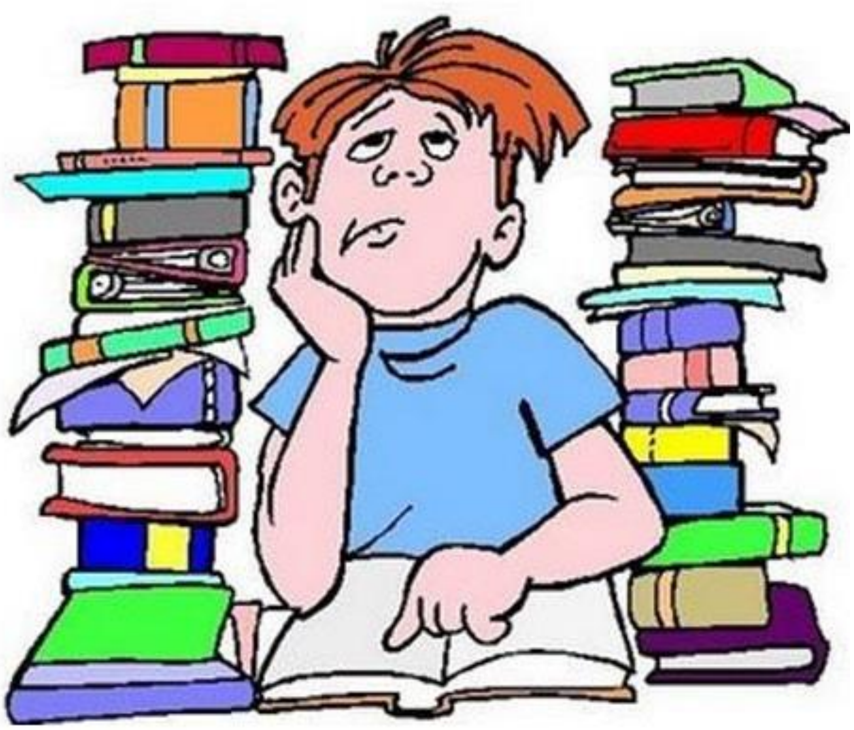
Domestic violence

Financial advice

Job search advice

Who will make a good teacher?

Can algorithms improve decision making?



Key feature: Decision depends on some prediction

Domestic violence

Financial advice

Job search advice

Who will make a good teacher?

Student advising

How much can algorithms improve decision making?

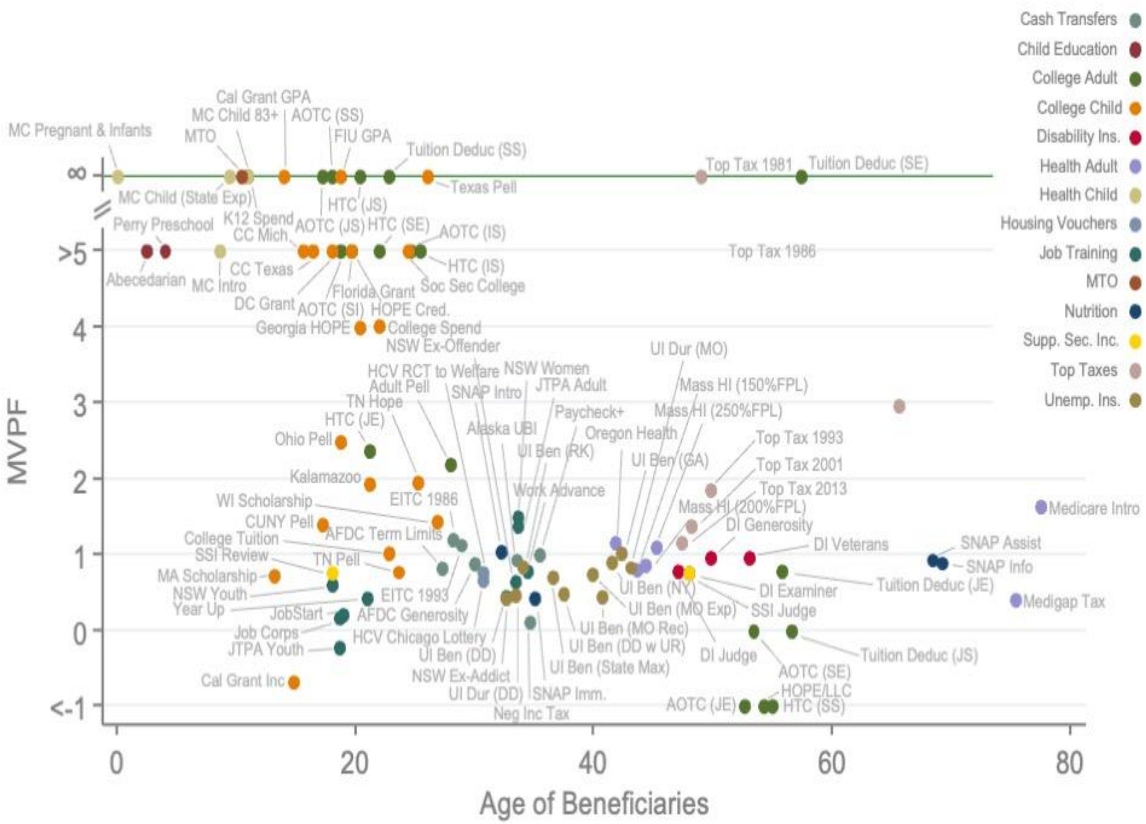
The marginal value of public funds (MVPF) of a policy is

$$MVPF = \frac{Benefit\ to\ Society}{Net\ Cost\ to\ Government}$$

We have a way of answering this question

We have calculated this for a whole suite of programs

How much can algorithms improve decision making?



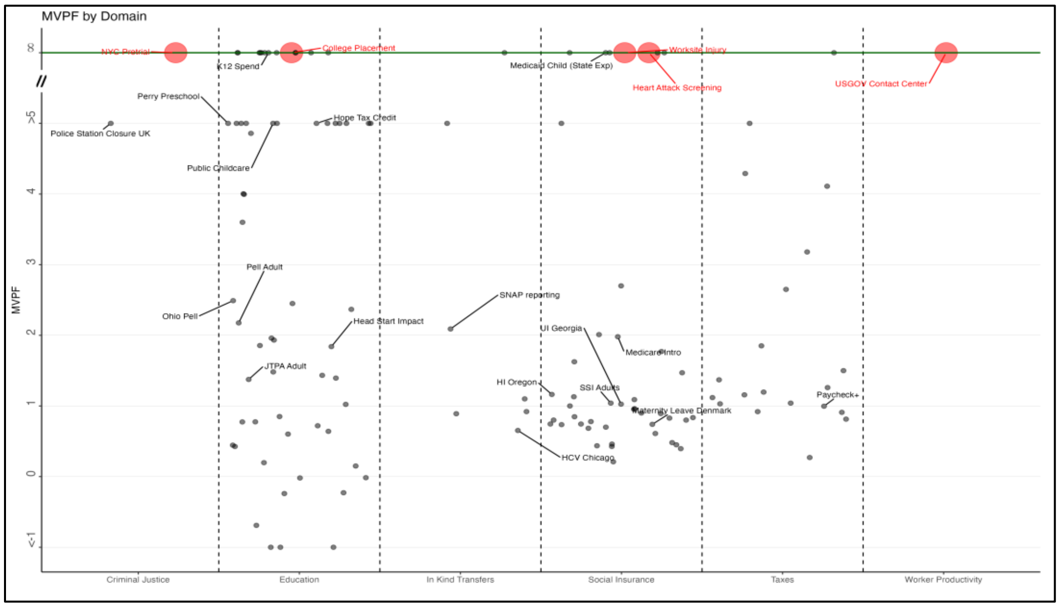
We have a way of answering this question

We have calculated this for a whole suite of programs

Hendren & Sprung-Keyser (2020): unified analysis of 133 historical policy changes.

Note: a very few programs have *infinite* return

<u>How much</u> can algorithms improve decision making?



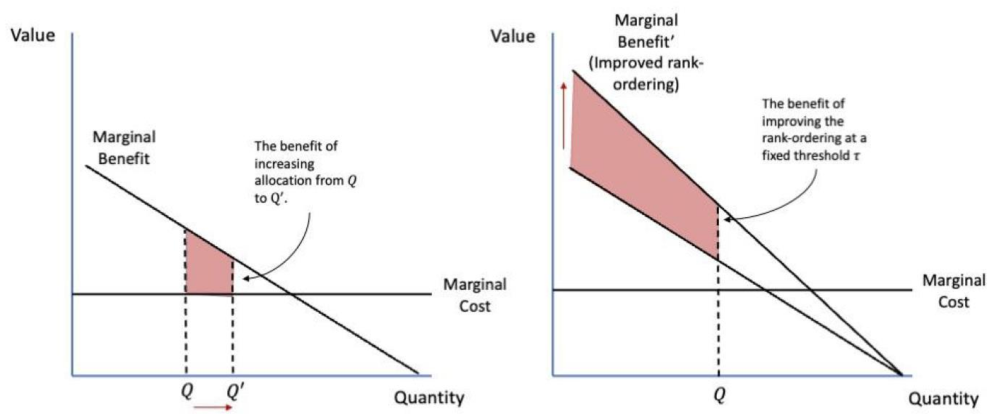
Let’s calculate it for algorithms

Every algorithm has infinite returns

Only 19 out of 133 historical policies studied by Hendren & Sprung-Keyser (2020) have MVPF = Infinity.

Why can algorithms improve decision making?

Figure 1: Stylized illustration of the social welfare gains from algorithmic re-ranking of who is prioritized for services



This argues we should invest more <u>research</u> in algorithms

Why might these benefits be so consistently big?

Reason 1: Improved ranking has first order social gains

Why can algorithms improve decision making?

This argues we should invest more research in algorithms

Why might these benefits be so consistently big?

Reason 1: Improved ranking has first order social gains

Reason 2: Algorithms have low marginal cost

Reason 3: Behavioral economics


LORAX
Breaking Bad
how I met your mother
HICK
30 ROCK
AZIZ ANSARI
SASSY PANTS

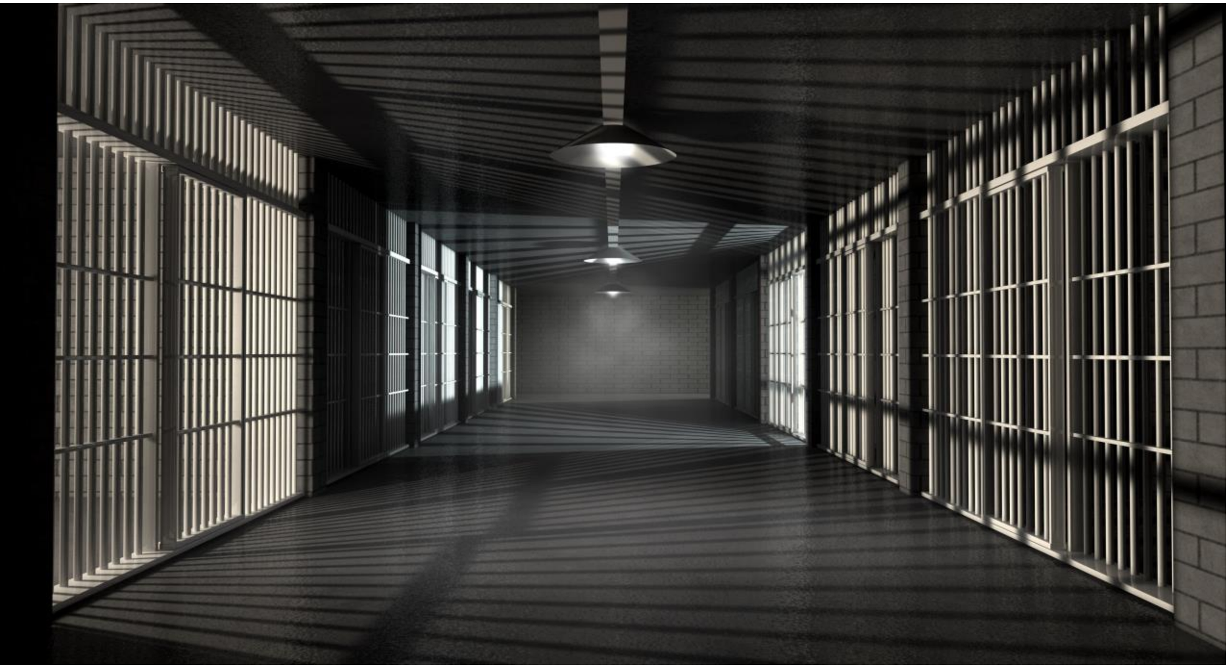
Why AI Needs Behavioral Economics

Why Behavioral Economics Needs AI

Algorithms can help us tackle social problems
(Choice architecture 2.0)

Many social problems are ultimately behavioral economics problems

Why can algorithms improve judicial decision making?



Why are judges doing so badly?

How would we go about answering this question?

Generate a hypothesis

Then test it

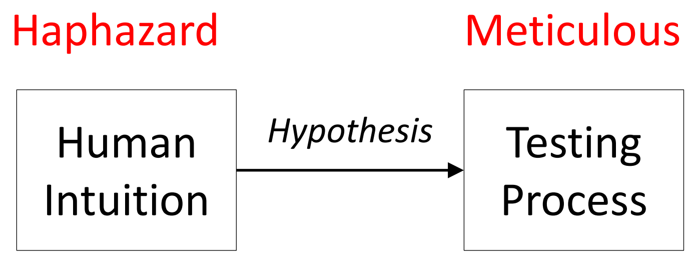
Hypothesis → Testing Process

But what about this bit?

Where do hypotheses come from?

Vast array of methods

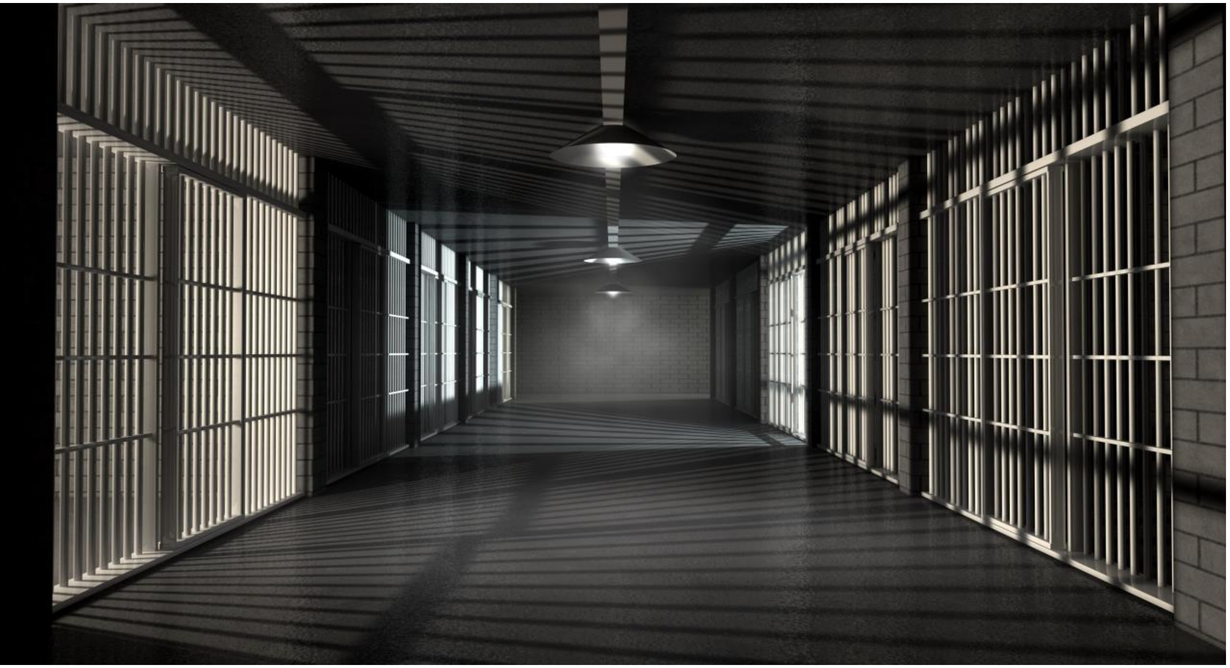
- Lab experiments
- Field experiments
- Structural estimation
- Observational causal inference



Something inconsistent here: if judges are bad at pattern detection….

Why can’t SL algorithms help us generate overlooked hypotheses?

<u>Why</u> can algorithms improve judicial decision making?



Let's build a predictive model of the <u>judge</u>

Ludwig & Mullainathan

Which factors predict judge decision?

1. Mugshot
2. Current charge: Violent crime
3. Current charge: Property crime
4. Current charge: Felony crime
5. Current charge: Drug crime

Not facial features we coded but the pixels themselves….

Form a new predictor using only the face

Large fraction of the predictable variation comes from the face

Key Steps

1. Check that there is *new* signal in face
   – Race, skin color, age all in face
   – So are factors identified by psychologists
   – Can even collect human predictions based on face

   – All these are rediscovered by algorithm
   – But collectively explain little of what algorithm has found

Key Steps

1. Check that there is *new* signal in face

2. Build a way to communicate with algorithm

3. Give that to subjects, *not just us looking at it*

Mugshot GAN

(GAN = generative adversarial network)

Build a mugshot-specific GAN
Pre-trained GANs not enough

Generates very realistic synthetic mugshots

Two real mugshots
Two GAN generated

Mugshot GAN

(GAN = generative adversarial network)

Build a mugshot-specific GAN
Pre-trained GANs not enough

Generates very realistic
synthetic mugshots

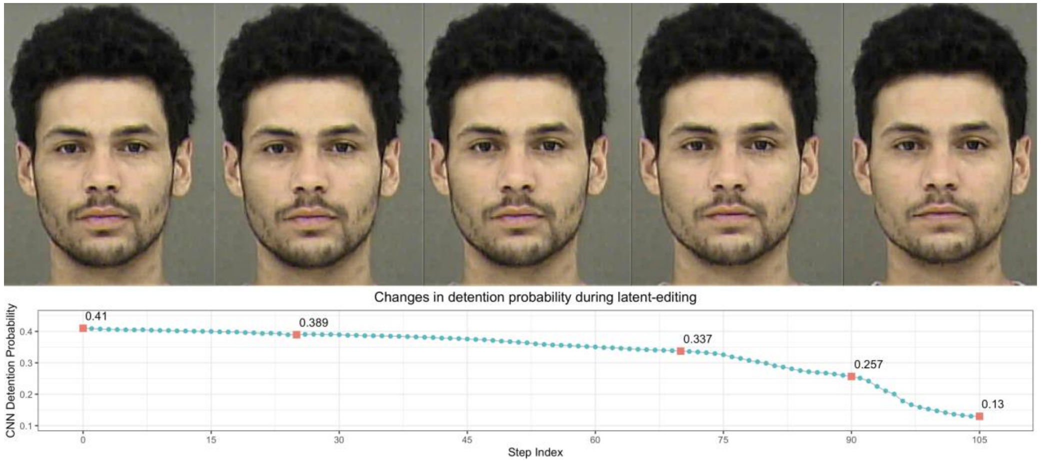
Two real mugshots
Two GAN generated

Start with a synthetic mugshot

<u>Morph</u> to increase detention risk

Mugshot GAN

(GAN = generative adversarial network)



Build a mugshot-specific GAN
Pre-trained GANs not enough

Generates very realistic
synthetic mugshots

Two real mugshots
Two GAN generated

Start with a synthetic mugshot

<u>Morph</u> to increase detention risk

Key Steps

1. Check that there is *new* signal in face

2. Build a way to communicate with algorithm
   – Not off the shelf

3. Give that to subjects, *not just us looking at it*

Applying Pipeline to Judge Prediction

Judge Predictor

Data Manifold (GAN)

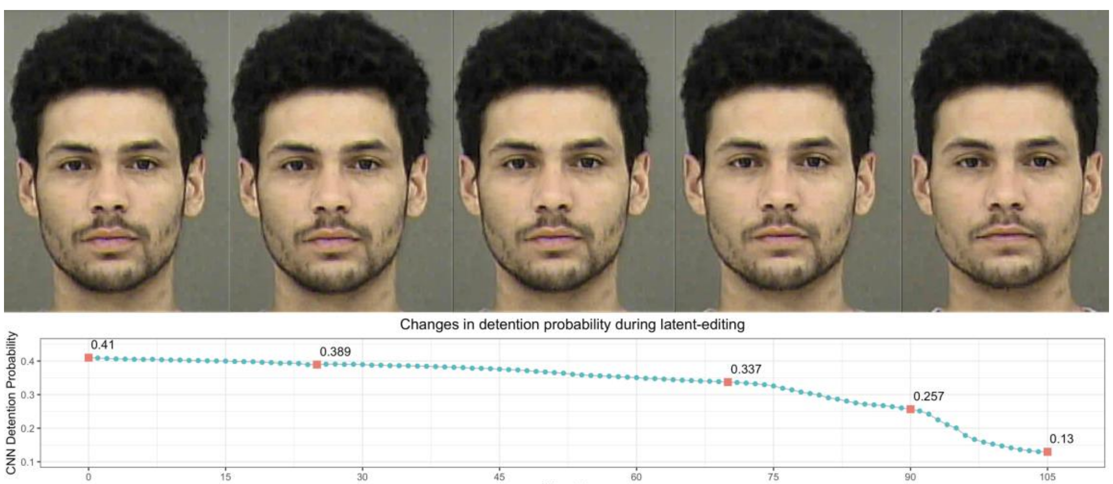
→ Morphing Tool → Image Pairs → Subjects Inspect → Hypothesis

Applying Pipeline to Judge Prediction

Judge Predictor

Data Manifold (GAN)

→ Morphing Tool → Image Pairs → Subjects Inspect → Hypothesis



All morphs: bottom to top decile

.13 to **.41** prediction detention

Applying Pipeline to Judge Prediction

Judge Predictor

Data Manifold (GAN)

→ Morphing Tool → **Image Pairs** → Subjects Inspect → Hypothesis

Applying Pipeline to Judge Prediction

Judge Predictor

Data Manifold (GAN)

Morphing Tool → Image Pairs → Subjects Inspect → Hypothesis

MTurk Task – guess which is higher predicted release

Subjects given feedback – chance to learn

Subjects given incentives

Then asked at end to name features

Can users guess better than chance?

Can users articulate something meaningful (and do they agree)?

Does that articulated feature actually correlate with algorithm prediction?

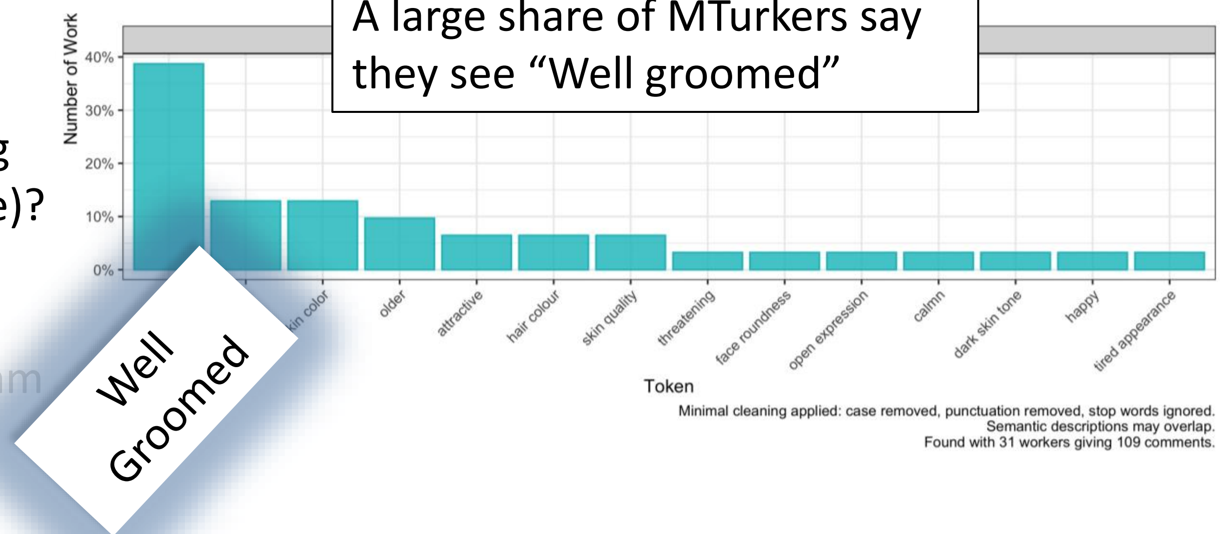
More where this comes from
Orthogonalize and iterate

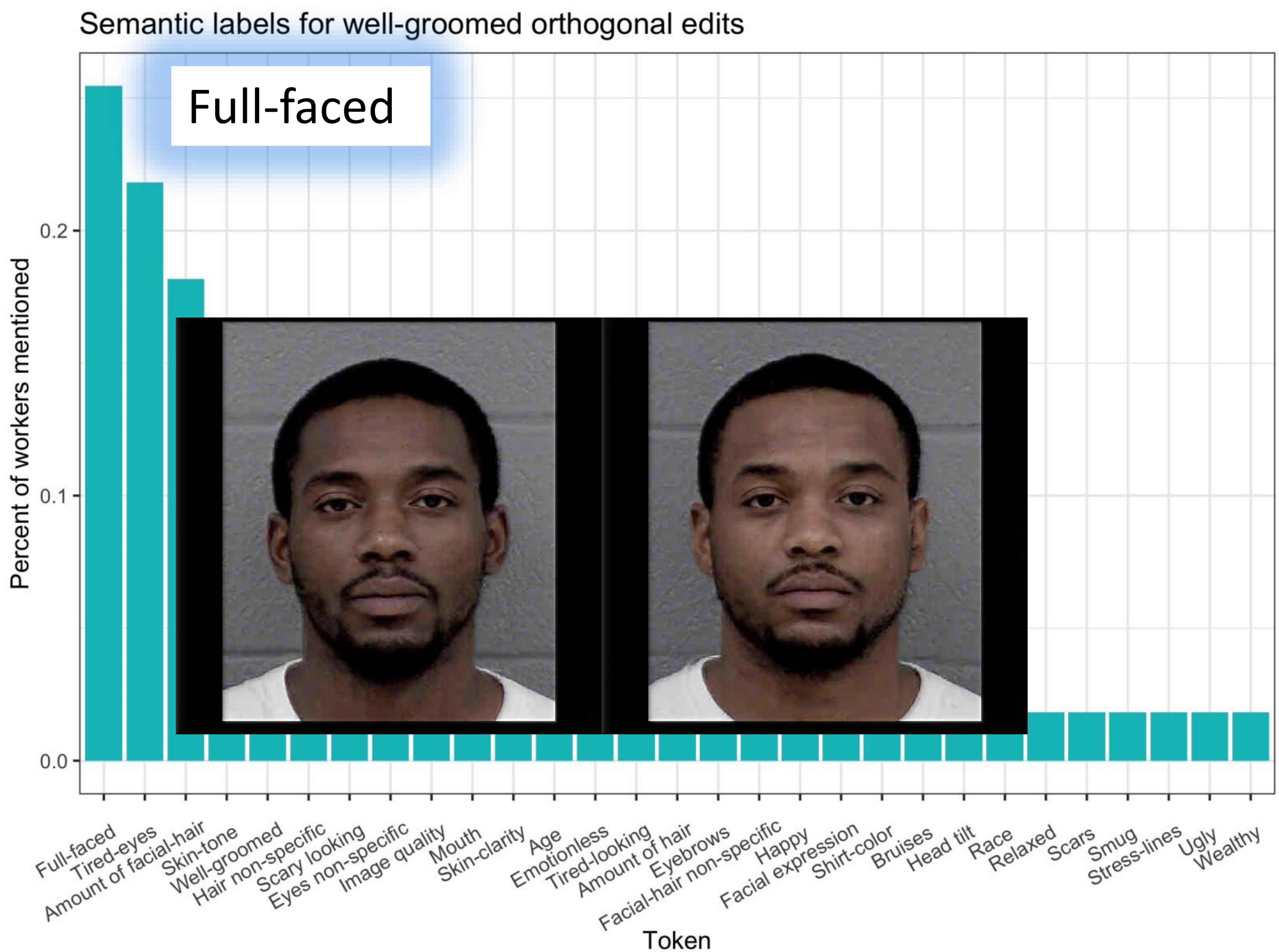
Can users guess better than chance?

Can users articulate something meaningful (and do they agree)?

Does that articulated feature actually correlate with algorithm prediction?

A large share of MTurkers say they see “Well groomed”

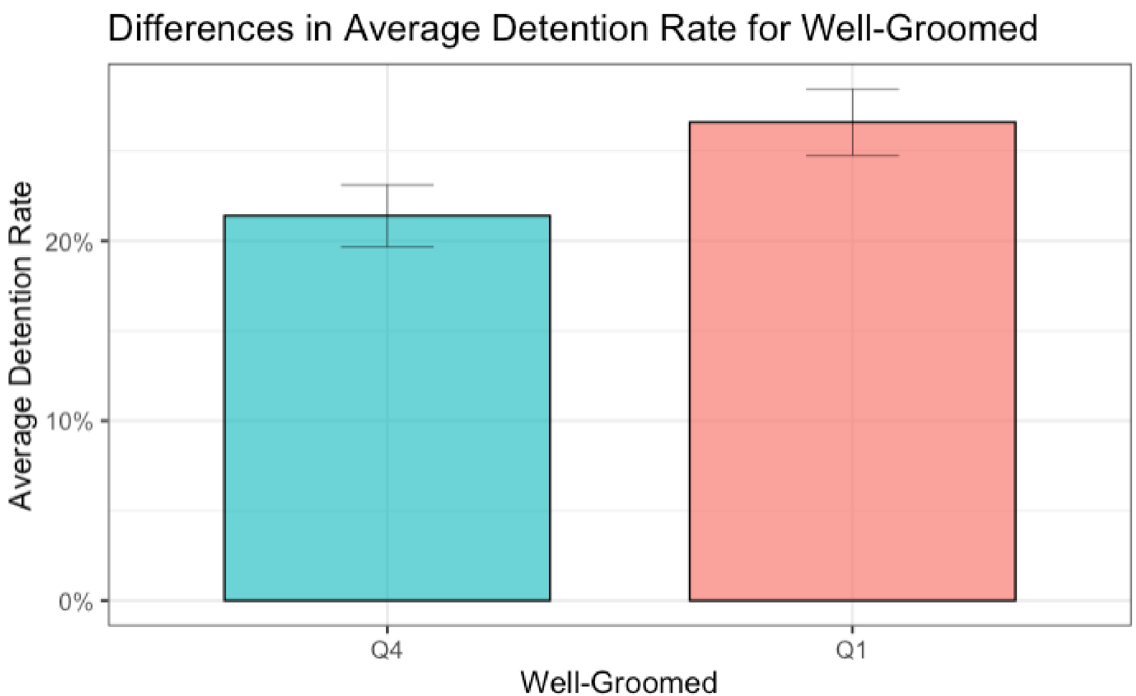



Semantic labels for well-groomed orthogonal edits
Full-faced
Percent of workers mentioned
0.2
0.1
0.0
Full-faced
Tired-eyes
Amount of facial-hair
Skin-tone
Well-groomed
Hair non-specific
Scary looking
Eyes non-specific
Image quality
Mouth
Skin-clarity
Age
Emotionless
Tired-looking
Amount of hair
Eyebrows
Facial-hair non-specific
Happy
Facial expression
Shirt-color
Bruises
Head tilt
Race
Relaxed
Scars
Smug
Stress-lines
Ugly
Wealthy
Token

Proof of Concept

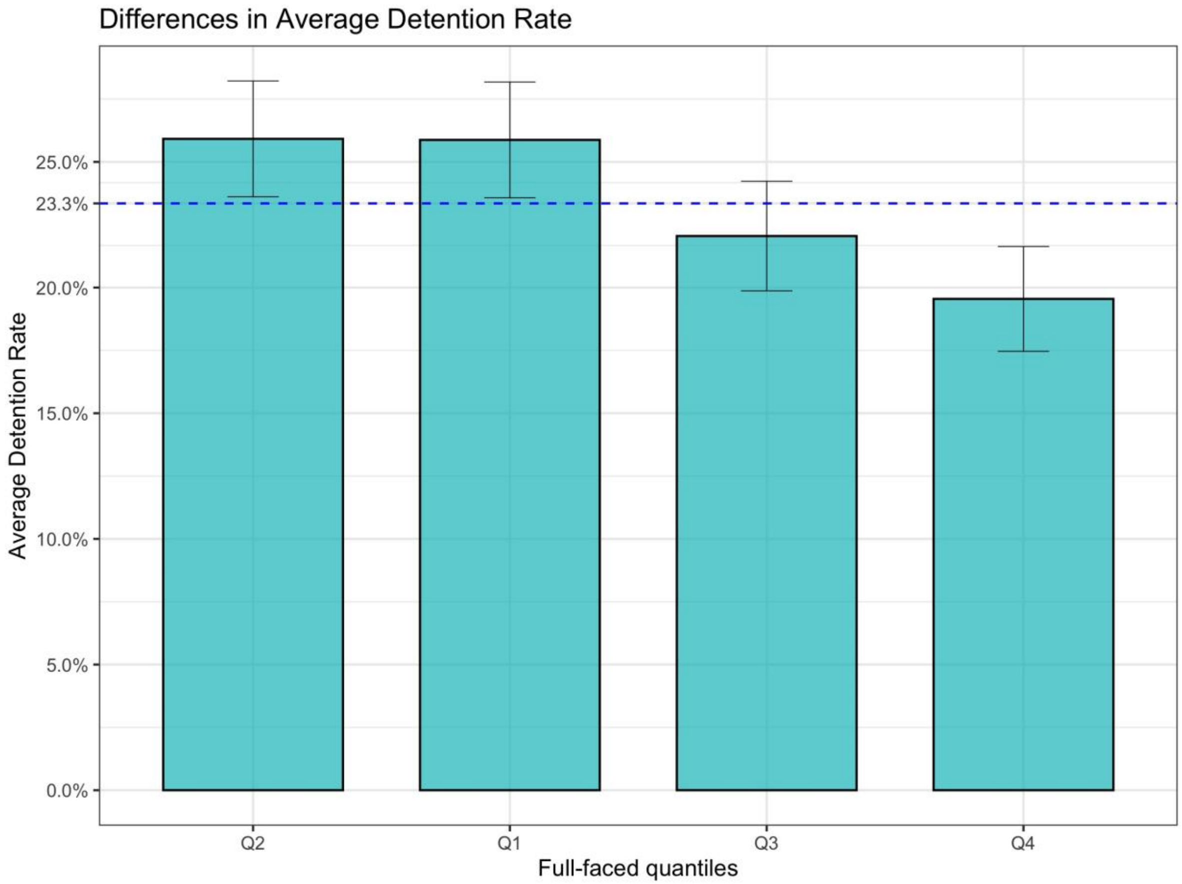
- Do these actually predict judge behavior?
  - Yes and at large magnitudes

Well Groomed is Important



Those not well-groomed are 28.7% more likely to be detained

Full Faced is Important



Fuller-faced less likely to be detained

Top-bottom quartile difference is ~7 ppt

That's 30% of overall detention rate

Proof of Concept

- Do these actually predict judge behavior?
  - Yes and at large magnitudes
- Are these “new”?
  - Experiments with public defenders
- They even seem to work “causally” in lab experiments
  - Control for predicted risk
  - Lab experiments that manipulate full-faced and well-groomed
- But…

Proof of Concept

- This is not a paper about well groomed or full faced
- It is not a paper even about faces
- It is not a paper about judges
- It is about a new way to study people
- Transcends hypothesis generation – can help improve **formal theories**

Theories of Choice Under Uncertainty

Expected Utility Theory

Theories of Choice Under Uncertainty

Expected Utility Theory

Allais Paradox

Lottery A:
For sure: $1 million

Lottery B:
89% chance: $1 million
10% chance: $5 million
1%   chance: <u>nothing</u>

Most people pick this

Why risk losing the million?

Theories of Choice Under Uncertainty

Expected Utility Theory

Allais Paradox

Lottery A:
For sure: $1 million

Lottery B:
89% chance: $1 million
10% chance: $5 million
1% chance: nothing

Lottery C:
89% chance: 0
11% chance: $1 million

Lottery D:
90% chance: $0
10% chance: $5 million

Most people pick this

Anyway it's a long shot,
go for the big bucks

Theories of Choice Under Uncertainty

Expected Utility Theory

Allais Paradox

Lottery A:
For sure: $1 million

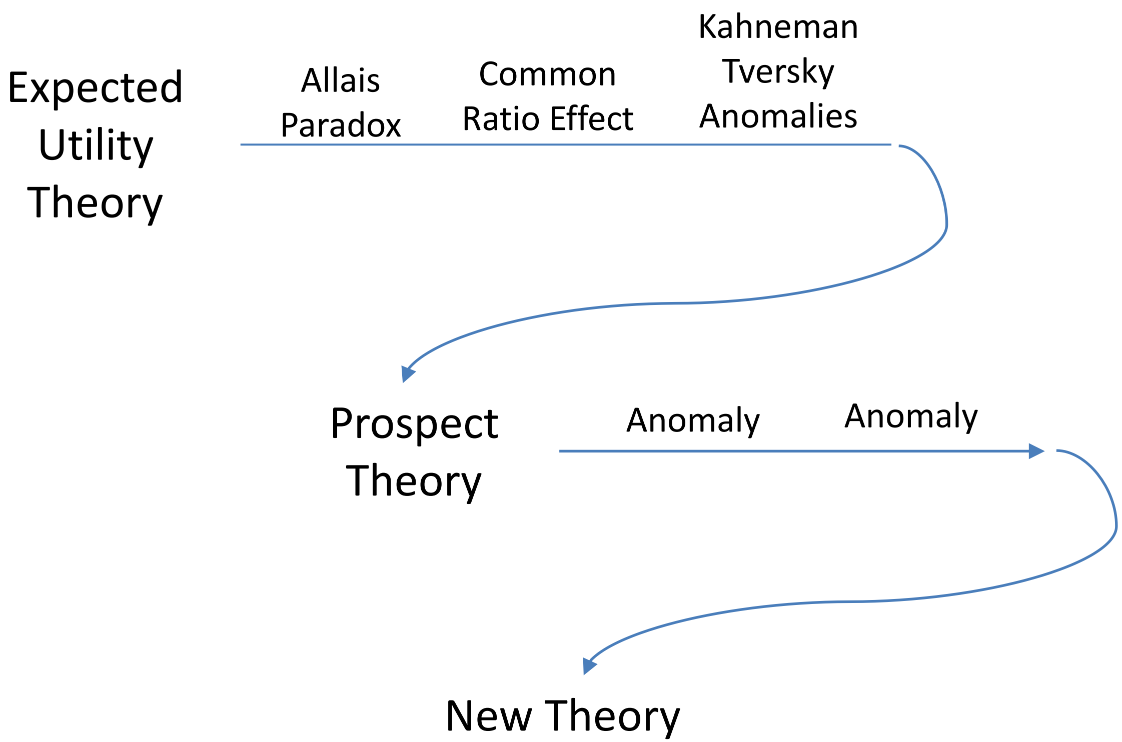
Lottery B:
89% chance: $1 million
10% chance: $5 million
1% chance: nothing

Lottery C:
89% chance: 0
11% chance: $1 million

Lottery D:
90% chance: $0
10% chance: $5 million

Allais (1953): These choices violate independence axiom
Inconsistent with expected utility theory

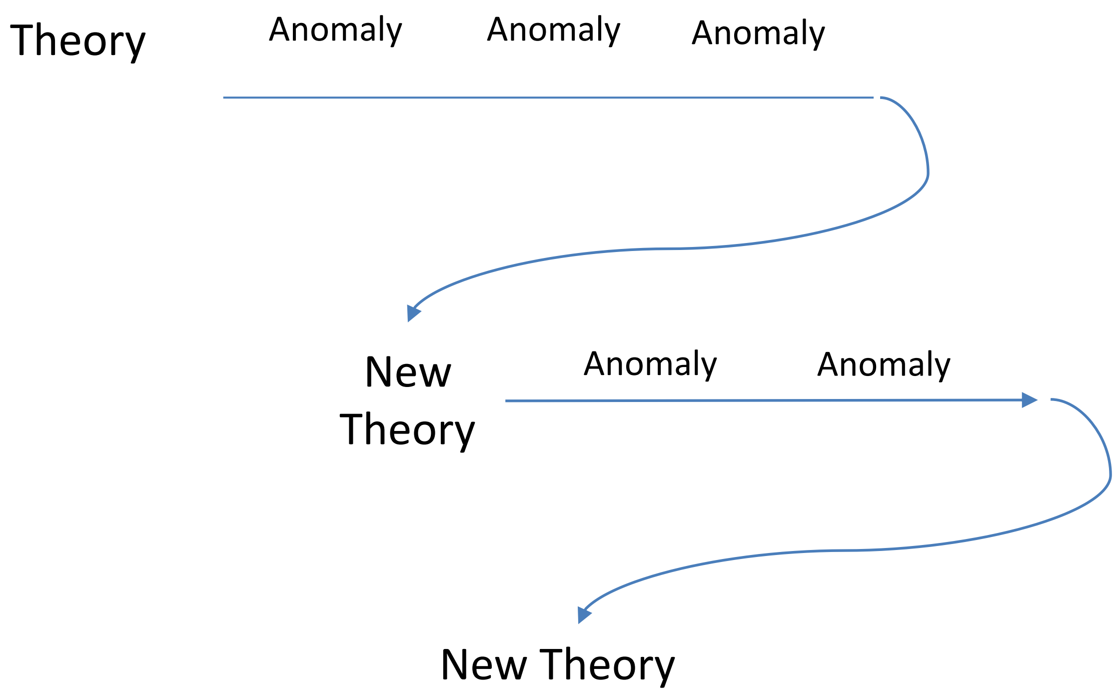
Theories of Choice Under Uncertainty



Salience theory – Bordalo et al. (2012)
Simplicity preferences – Oprea (2022), Puri (2022).
Cognitive uncertainty – Enke & Graeber (2023)
And many more…

Harless & Camerer (1994) sits above this process: compare how well theories can explain these anomalies?

Anomalies Improve Theories



Nash
Subgame Perfect
Bayesian Subgame Perfect
.
.
.
.

Anomalies Improve Theories

Theory

Anomaly Anomaly Anomaly

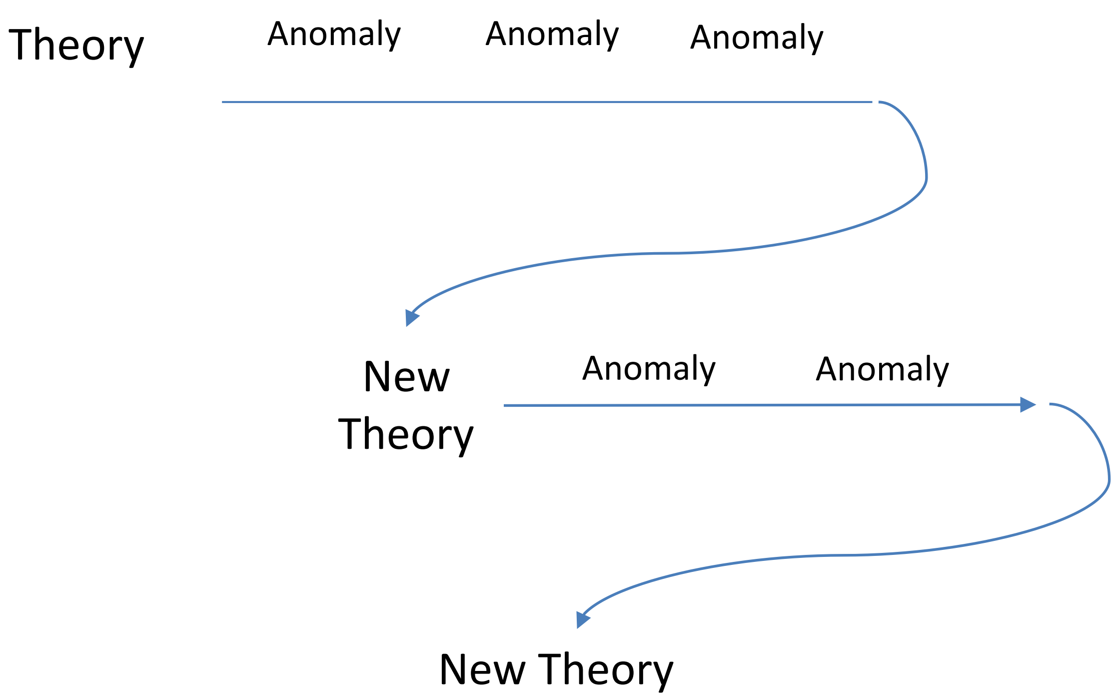
New Theory

Anomaly Anomaly

New Theory

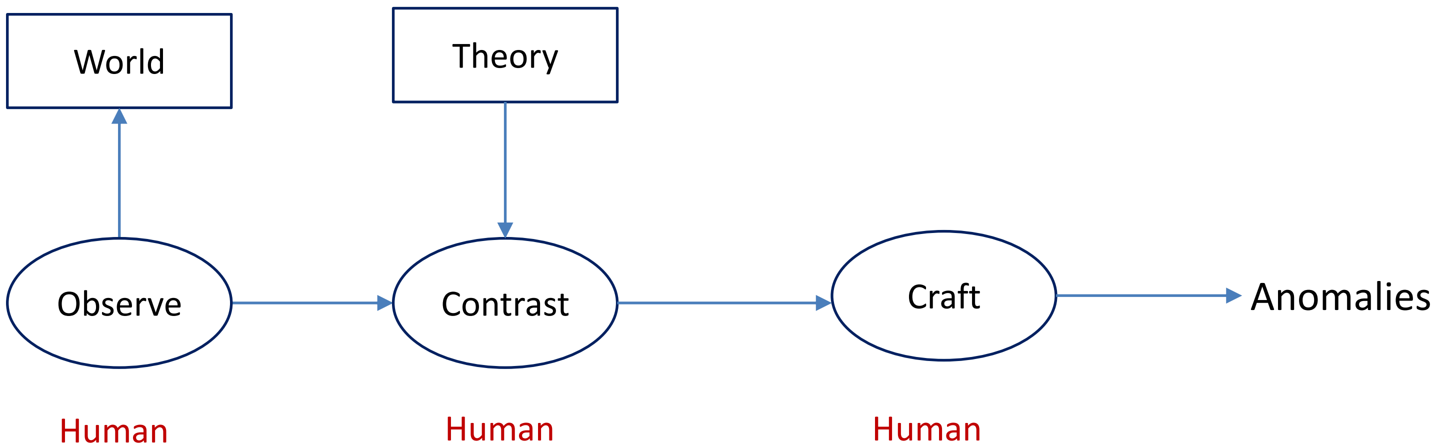
CAPM
CCAPM
3 Factor Models
.
.
.
.

Anomalies Central to Evolution of Theories

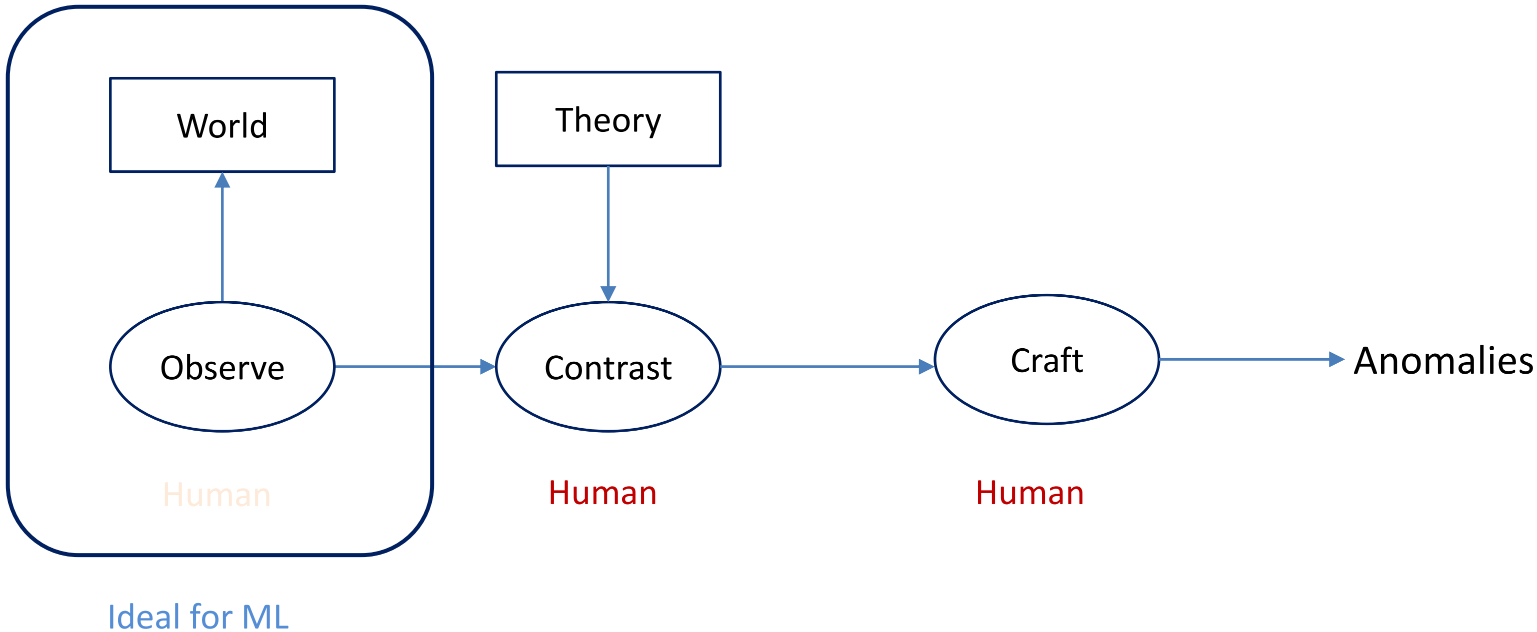


How do we usually find anomalies?

Current Anomaly Generation

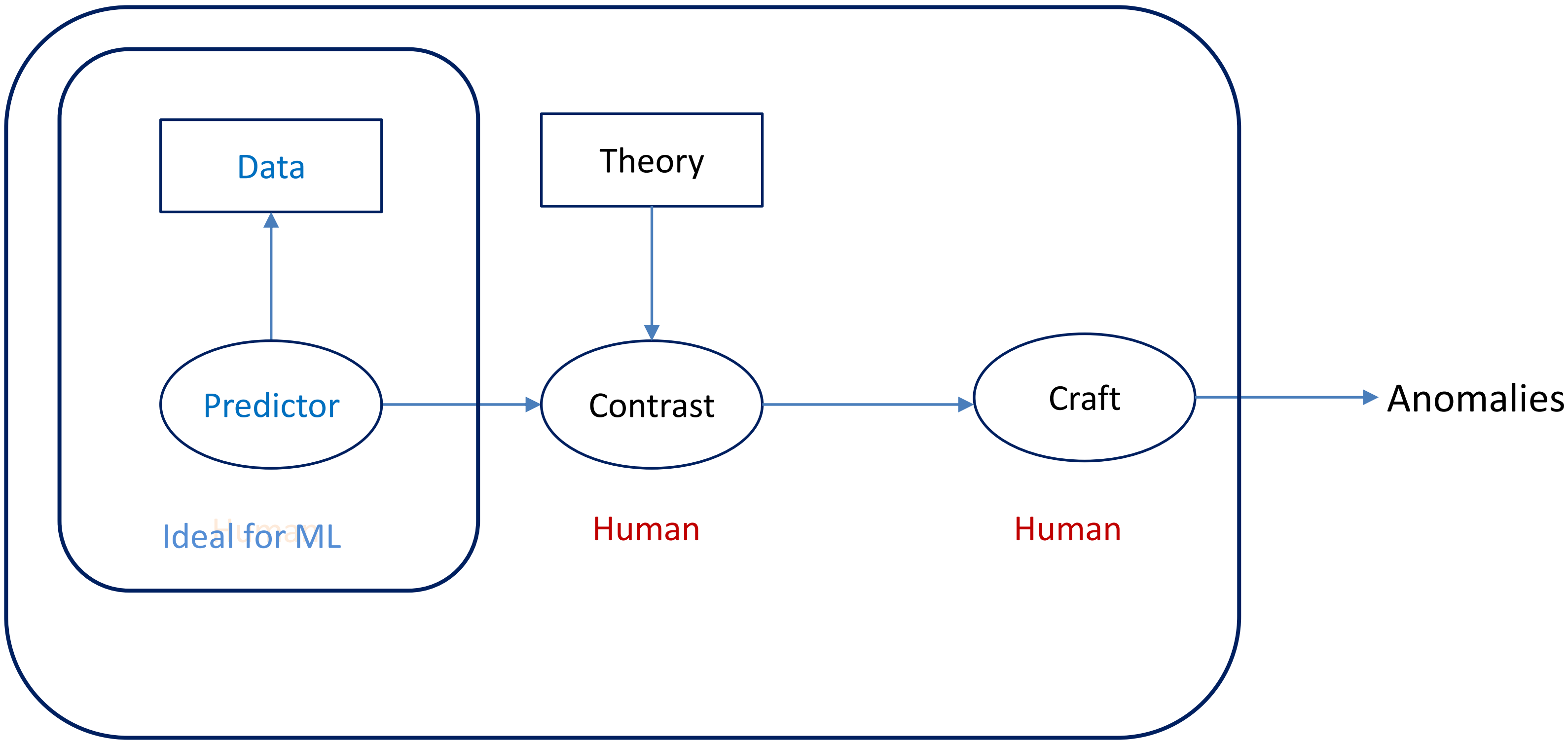


Current Anomaly Generation



Algorithms can see things in data that we might not

Algorithmic Anomaly Generation



Build Tool to Fully Automate Anomaly Generation

(1) Given a dataset
(2) Given a theory
(3) Produce any anomalies

Mullainathan and Ramba
(2023)

Why AI Needs Behavioral Economics

Why Behavioral Economics Needs AI

Algorithms can help us tackle social problems
(Choice architecture 2.0)

Many social problems are ultimately behavioral economics problems

Algorithms can help us *understand people*

Newsfeed

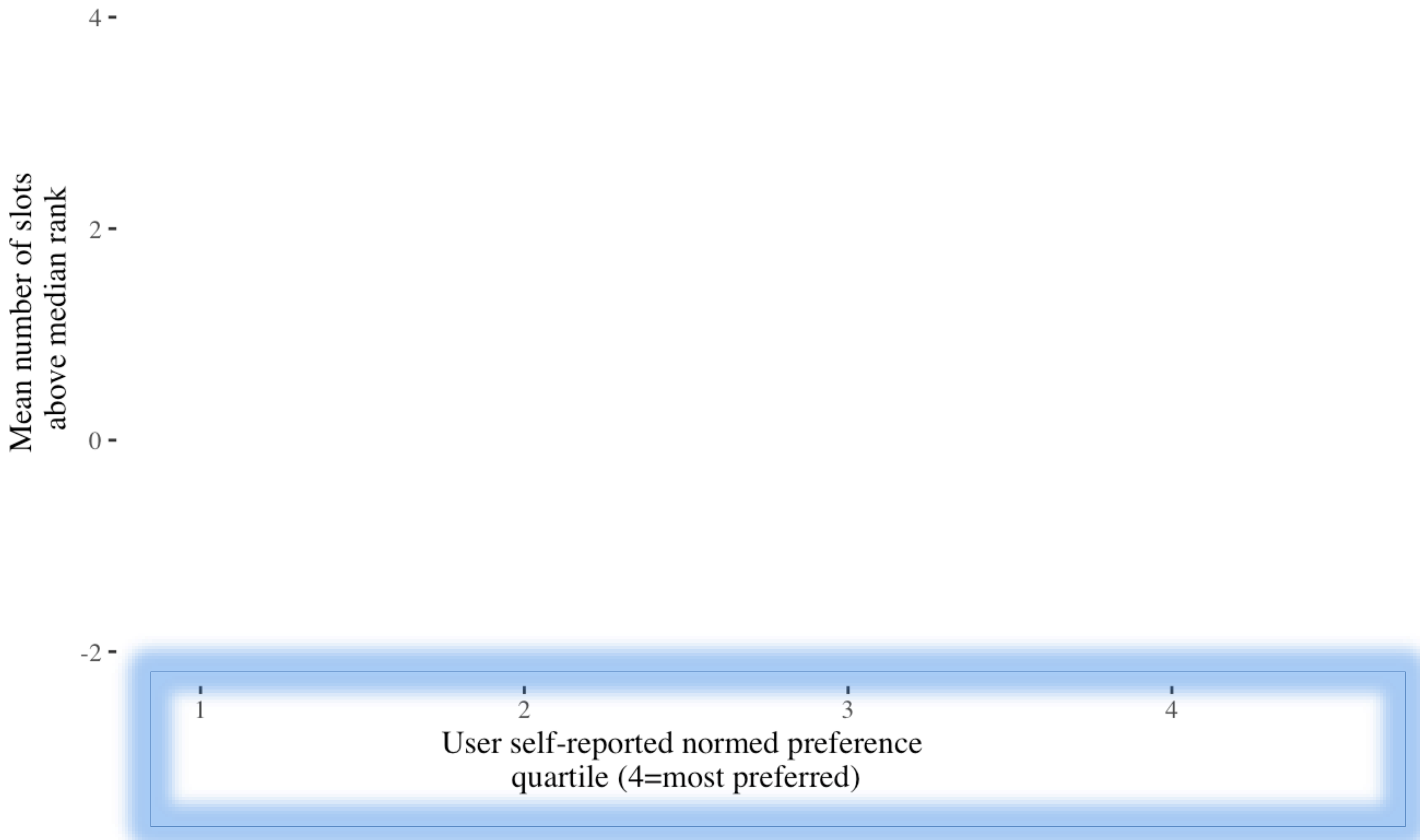
Subjects screenshare while scrolling through FB

Answer questions about how much they like posts

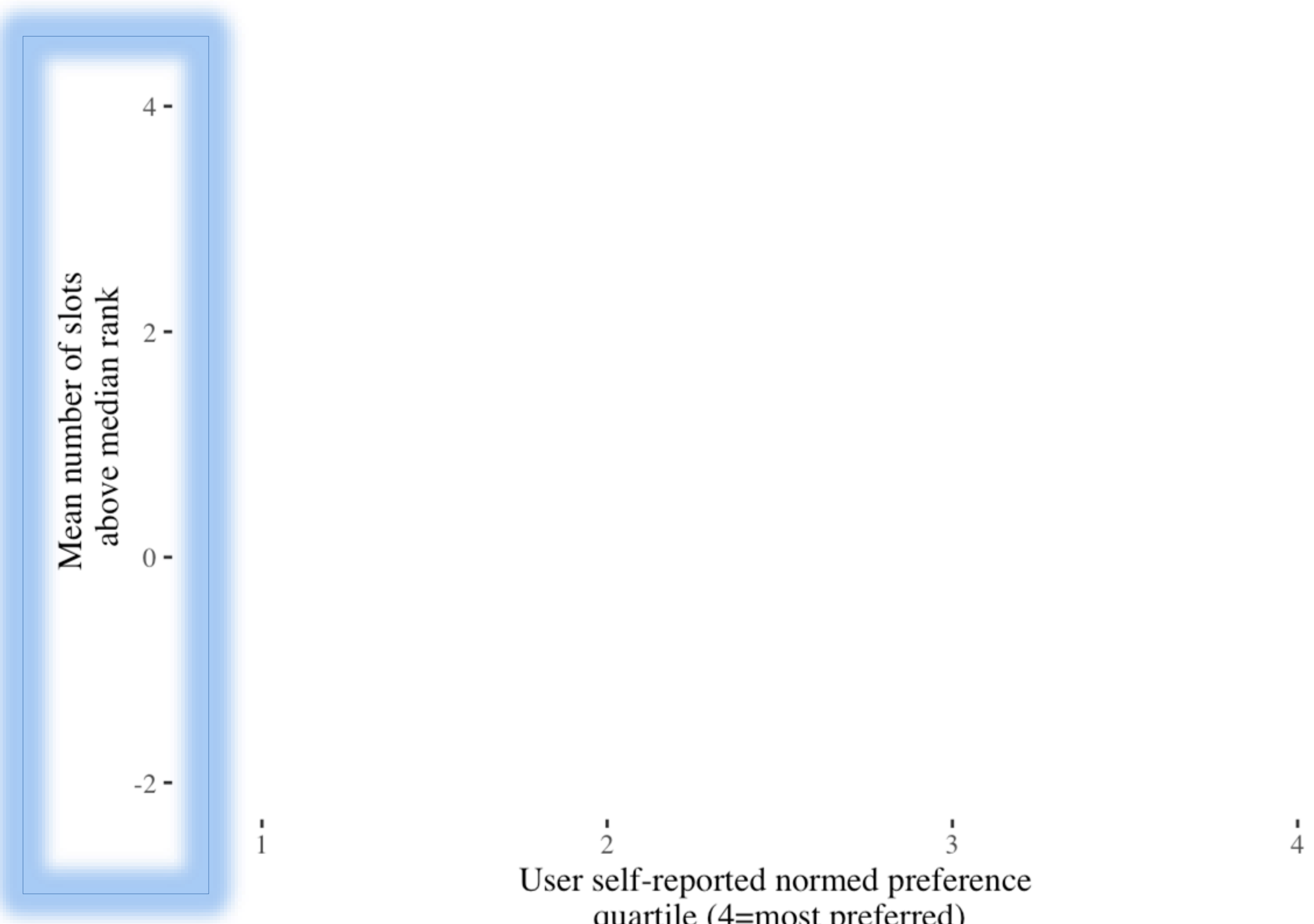
We track details of post

Agan, Davenport, Ludwig & Mullainathan

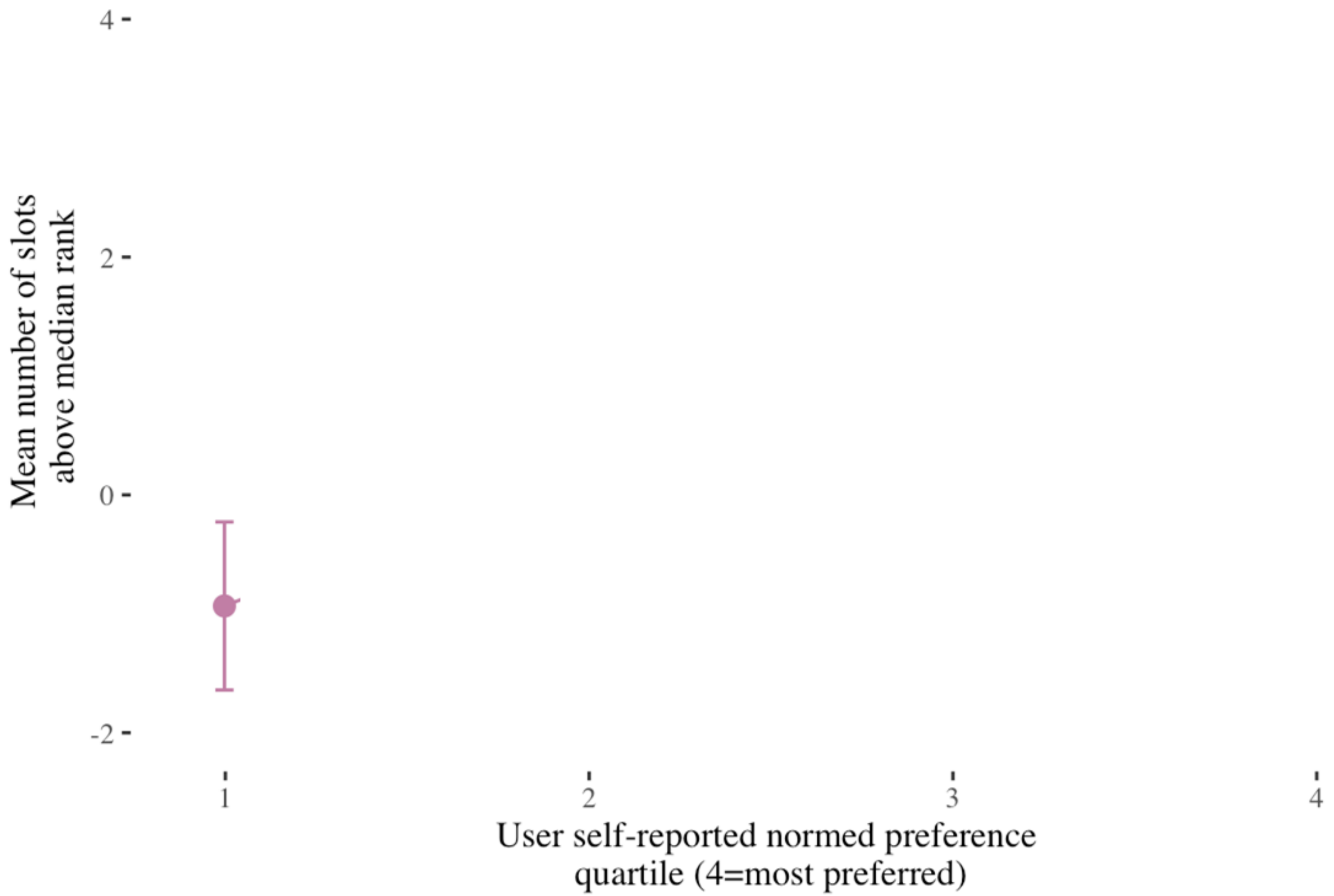
Newsfeed Ranking of Posts



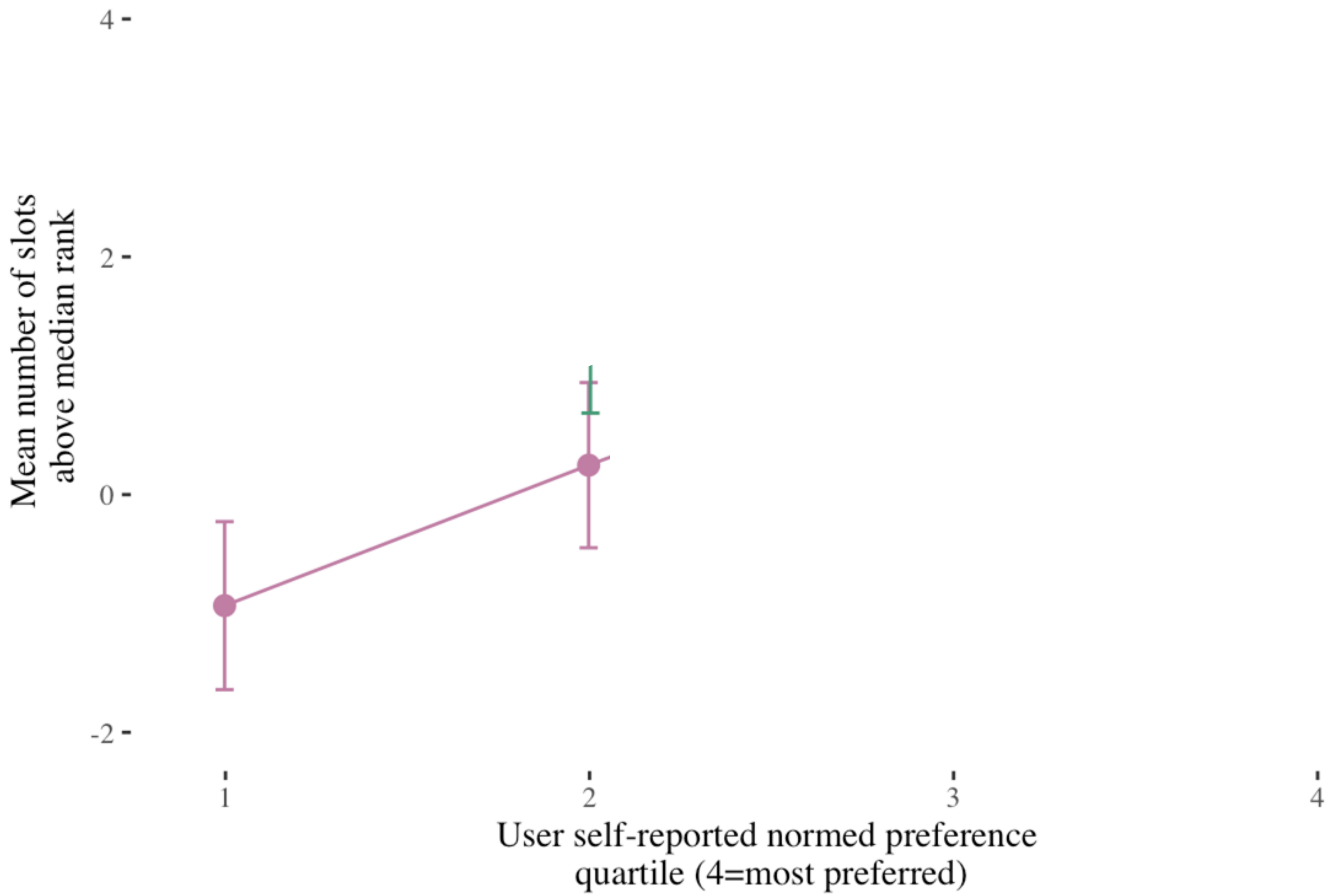
Newsfeed Ranking of Posts



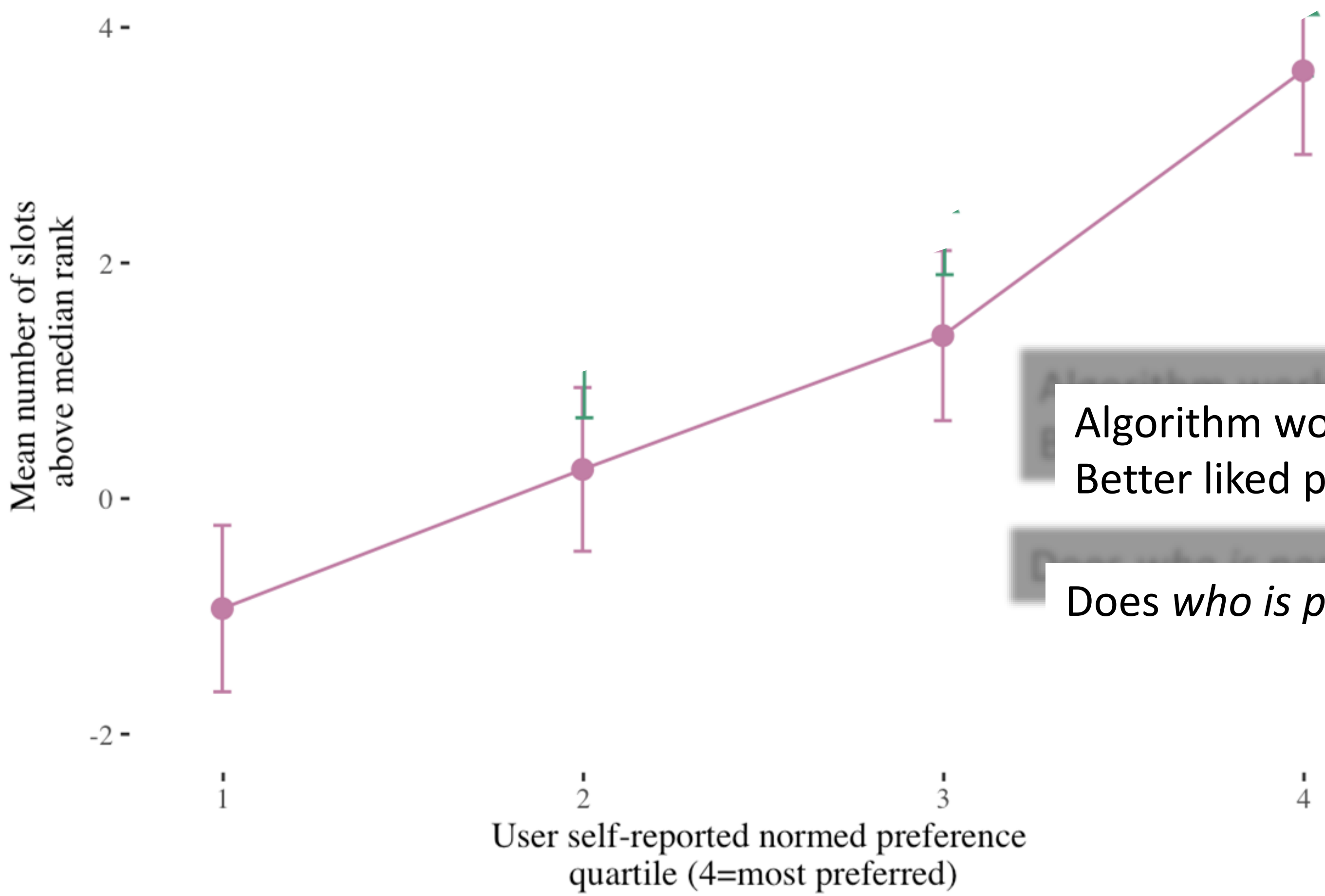
Newsfeed Ranking of Posts



Newsfeed Ranking of Posts



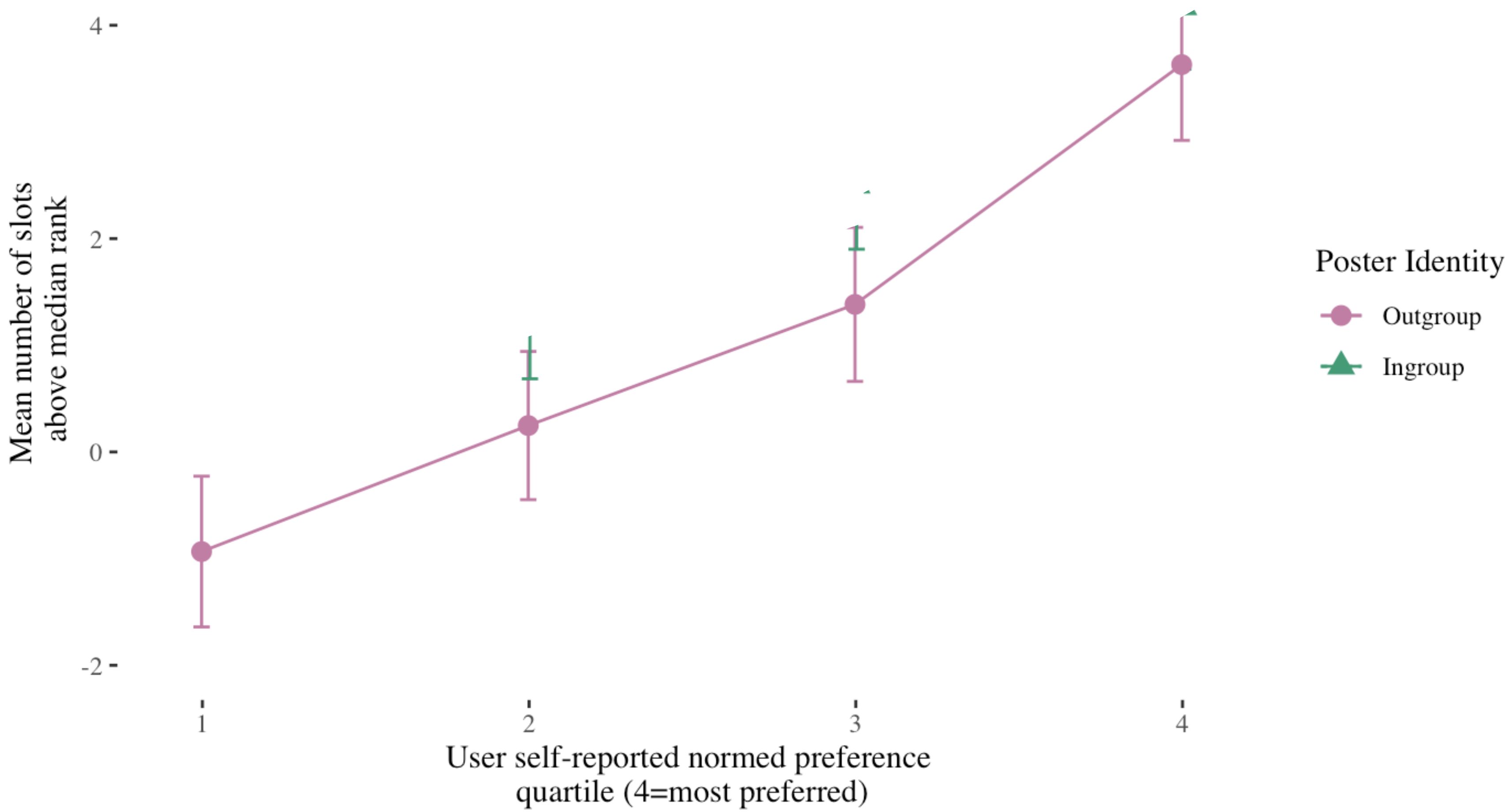
Newsfeed Ranking of Posts



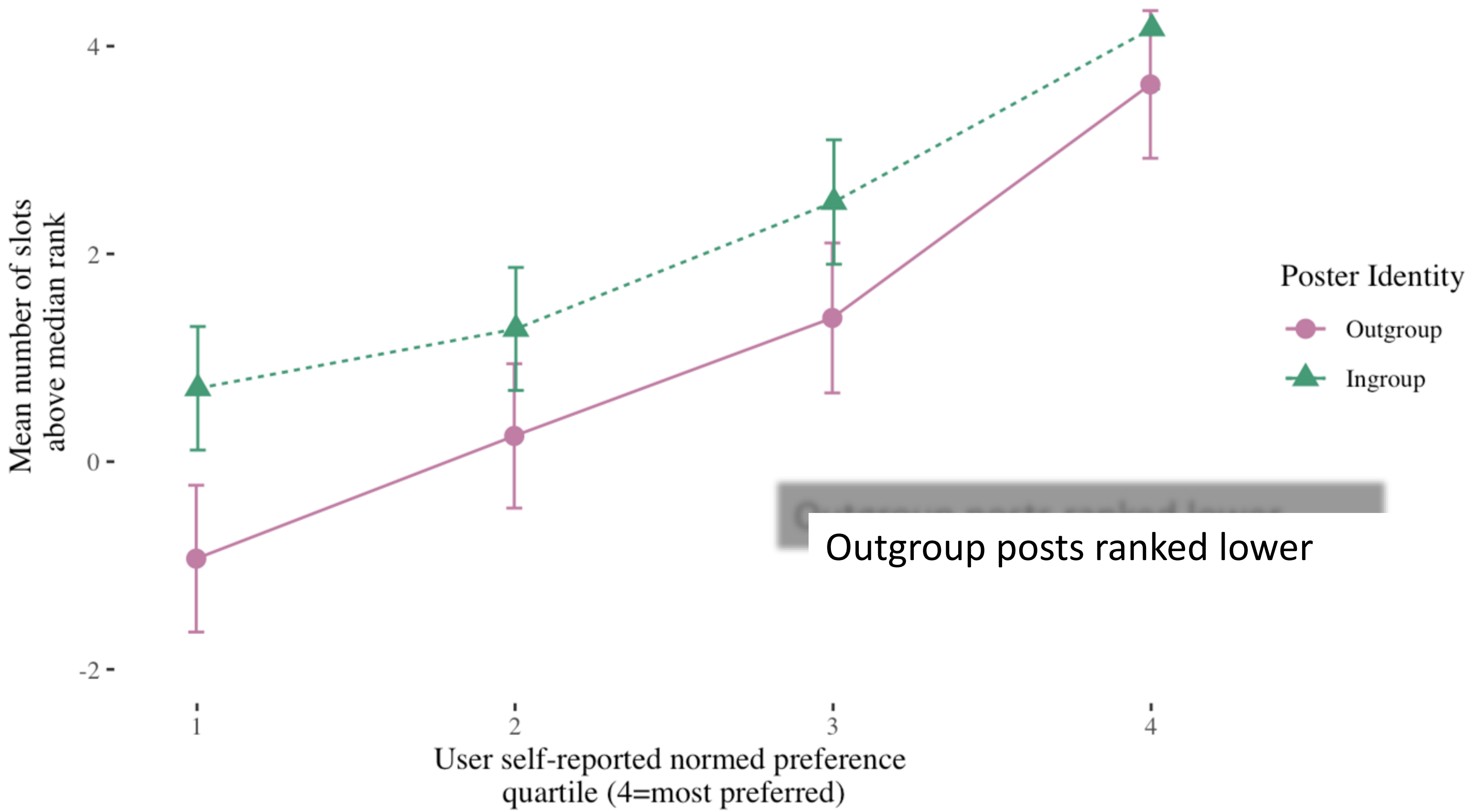
Algorithm works:
Better liked posts ranked higher

Does *who is posting* matter?

Newsfeed Ranking of Posts



Newsfeed Ranking of Posts



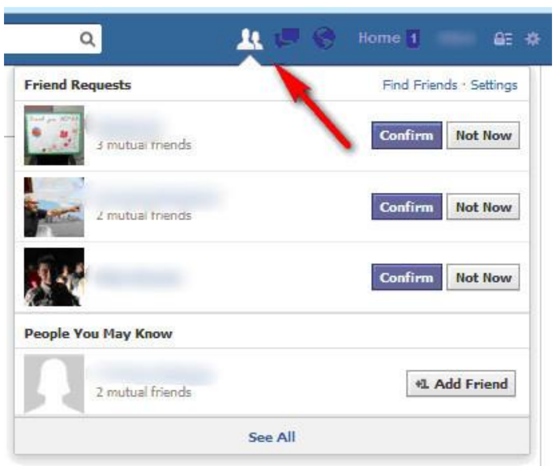
Newsfeed Ranking of Posts

Outgroup posts ranked lower <u>conditional on preference</u>

Audited second algorithm

Newsfeed

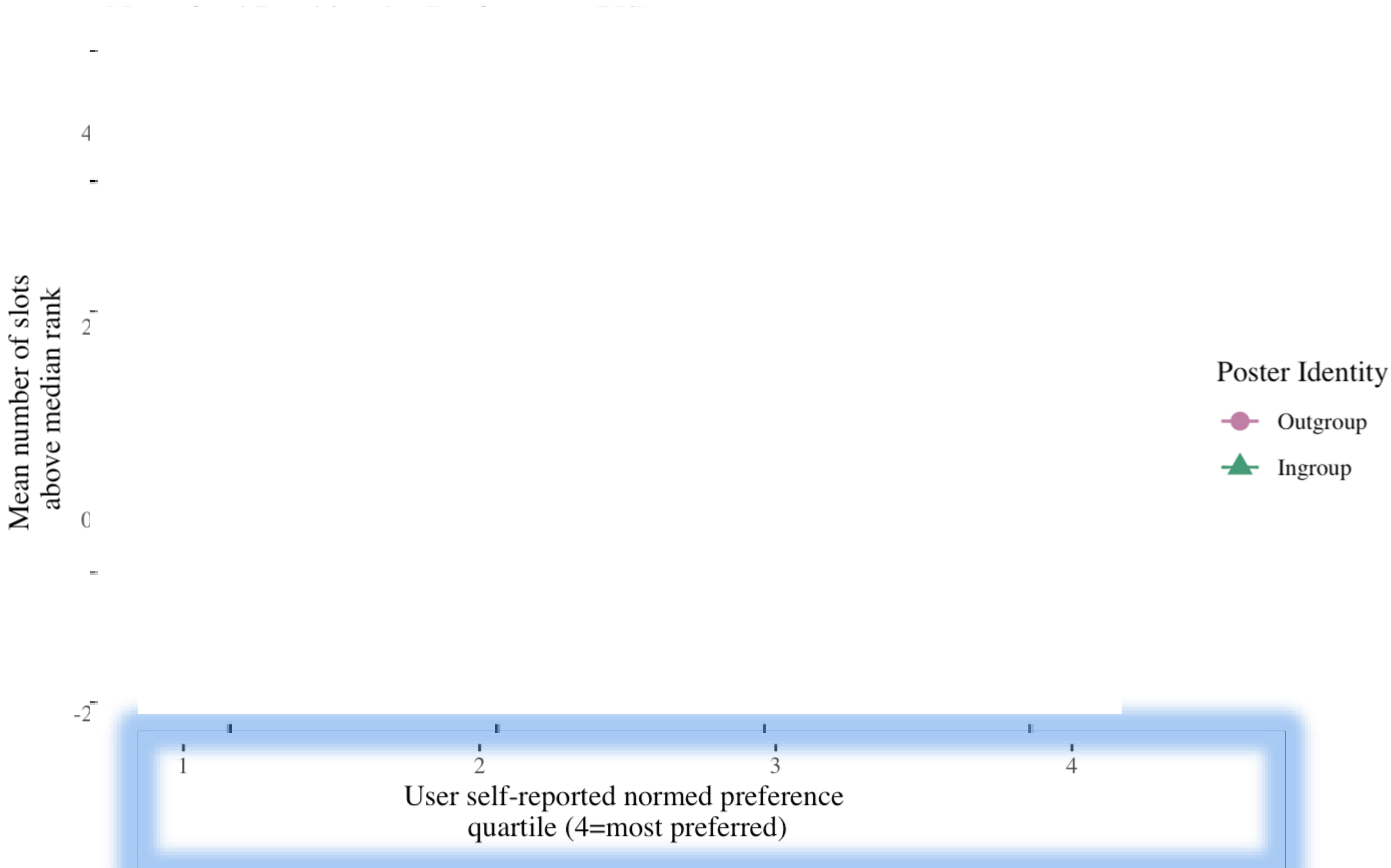
Interacting with friends



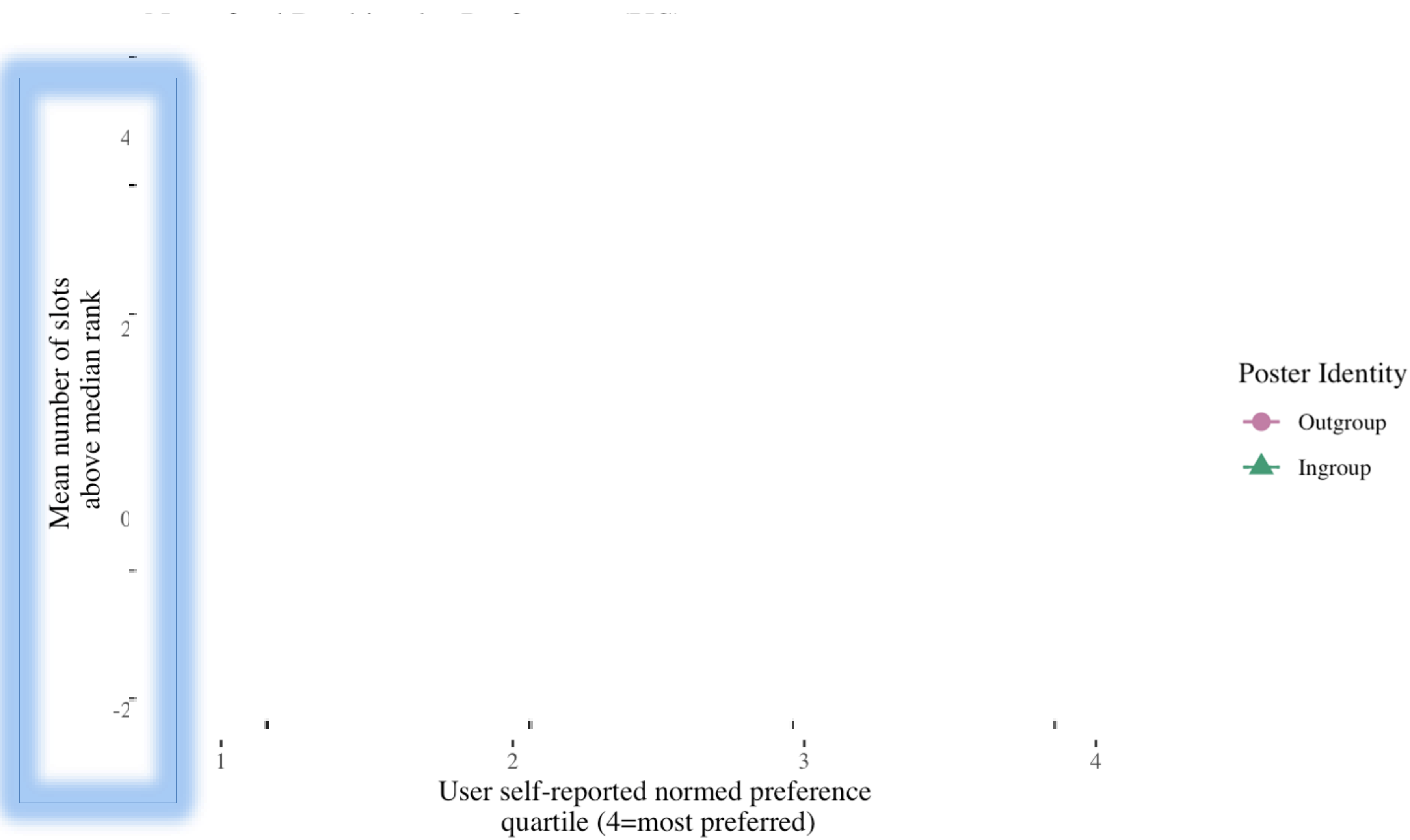
People You May Know

Making new friends

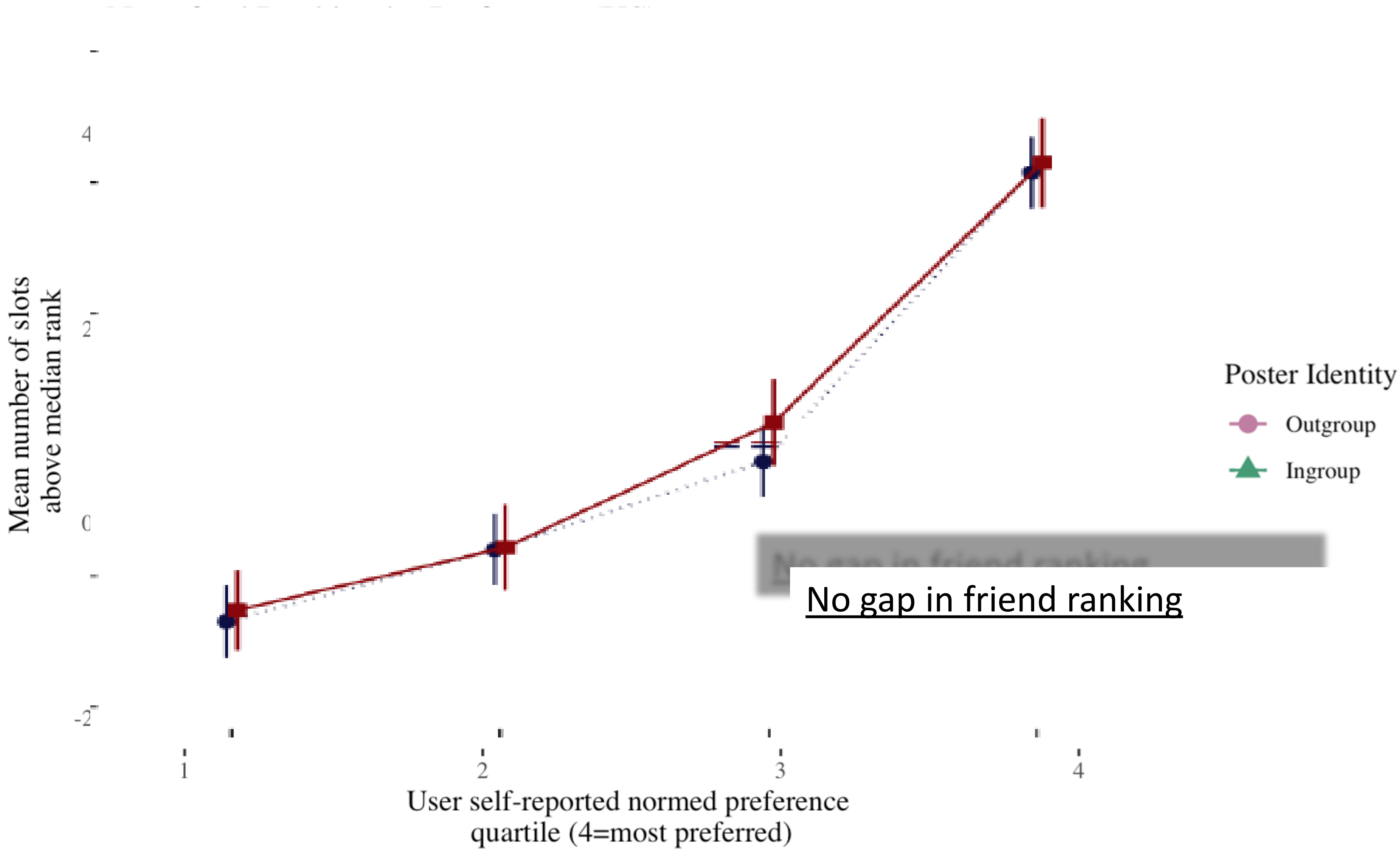
Audit of PYMK



Audit of PYMK



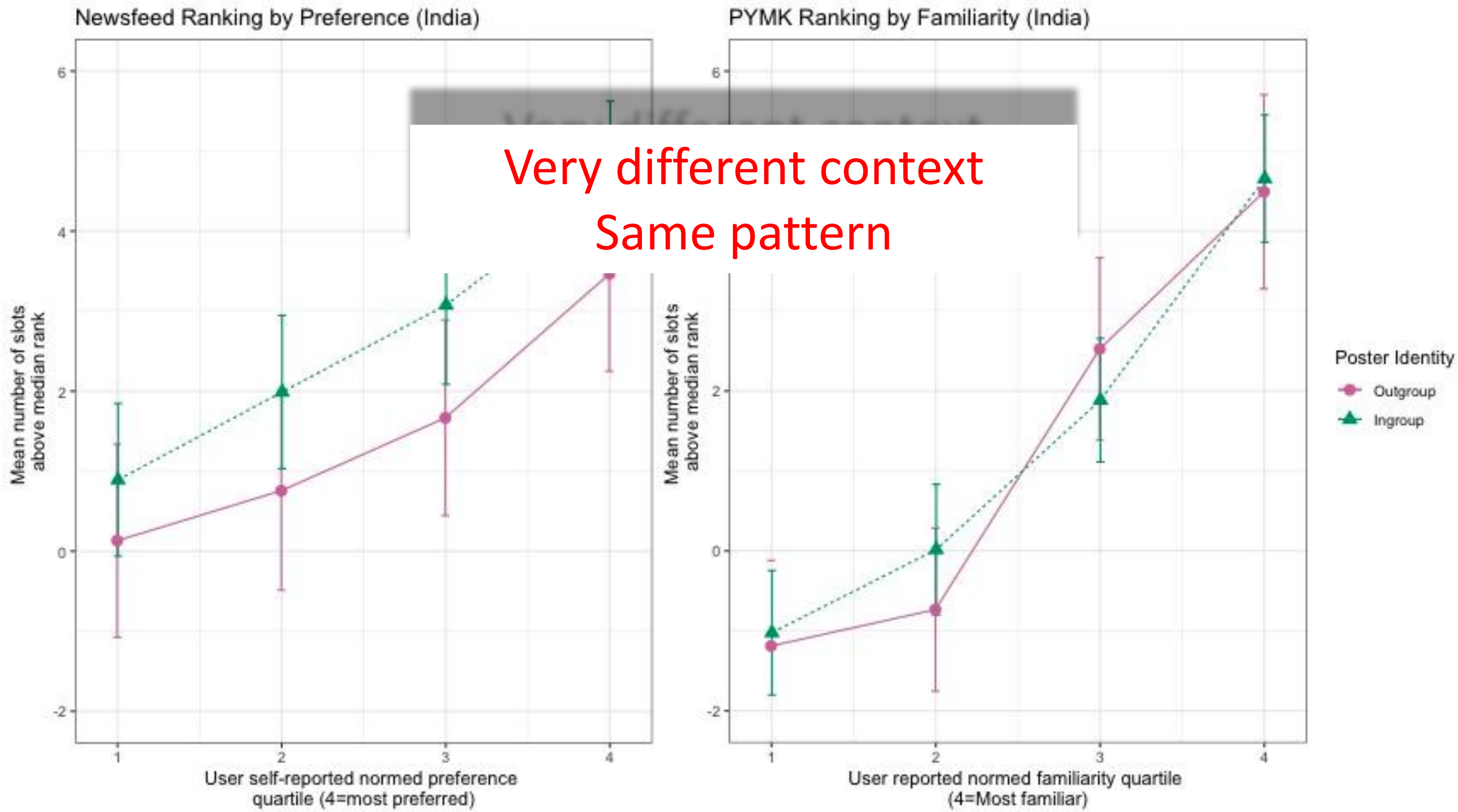
Audit of PYMK



Redid both audits in India
Muslim/Hindu

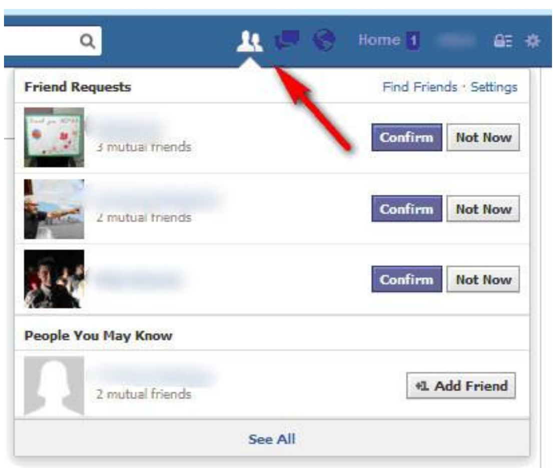
Newsfeed

People You May Know



Very different context
Same pattern

Newsfeed



People You May Know

Why the algorithmic bias?

Why in NewsFeed but not in PYMK?

Existing explanations of bias predict no difference or opposite sign

Cannot give a definitive answer without working with Facebook

But here's one **hypothesis**

What went wrong?

My host was exceptionally considerate

But I'm never going back there

What went wrong?

I went wrong

I clearly have a problem with Doritos

What went wrong?

Host went wrong?

Heart was in right place:
Give me what I <u>want</u>

What went wrong?

Host went wrong?

Heart was in right place:
Give me what I <u>want</u>

Head was not:
Gave me what I <u>ate</u>

Her mistake:

Choices = Preferences

Choices ≠ Preferences

Well understood behavioral science fact

Here – self control

e.g. want-should conflict

But can happen for **many** reasons

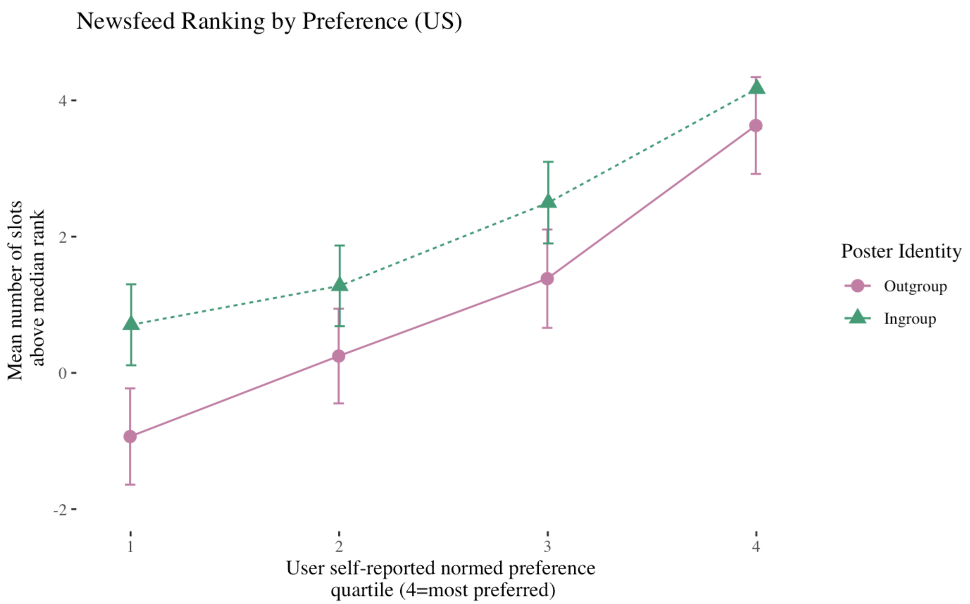
Choices ≠ Preferences

~~Well understood~~ behavioral science fact

Widely ignored behavioral science fact

Especially when building algorithms

Choices ≠ Preferences



Newsfeed algorithm trained on behavior – how I interact with posts

But clicks may
not reflect preferences

Choices ≠ Preferences

"There can be no peace until they renounce their Rabbit God and accept our Duck God."

Widely documented –
Ingroup bias

Favor own group <u>more</u> in behavior than in preference

So an algorithm trained on my behaviors produces more ingroup favoritism than **I myself** want….

But how does this explain PYMK vs Newsfeed

Ingroup Bias

"There can be no peace until they renounce their Rabbit God and accept our Duck God."

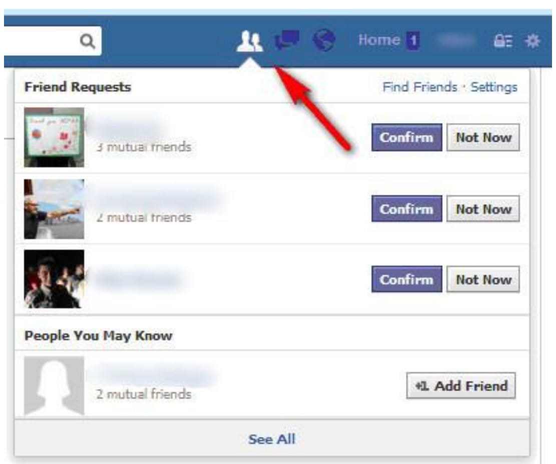
Bias has structure:

Larger when behaving <u>automatically</u>…

- Low deliberation
- Quick choices
- Low consequences

Ingroup Bias

Newsfeed



People You May Know

Bias has structure:

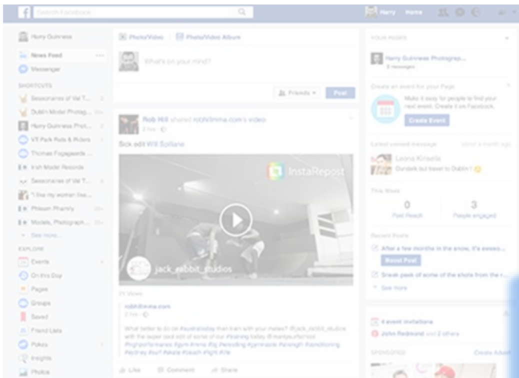
Larger when behaving <u>automatically</u>…

- Low deliberation
- Quick choices
- Low consequences

How do you interact with Facebook posts?

How do you decide who to friend?

Ingroup Bias

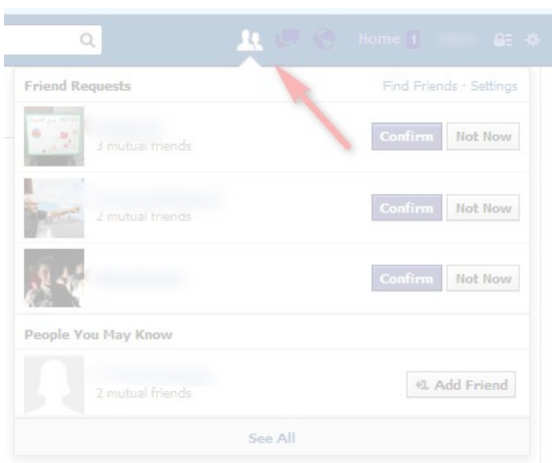


Newsfeed

Behavior less thoughtful
More ingroup favoritism

Algorithm biased

How do we test?



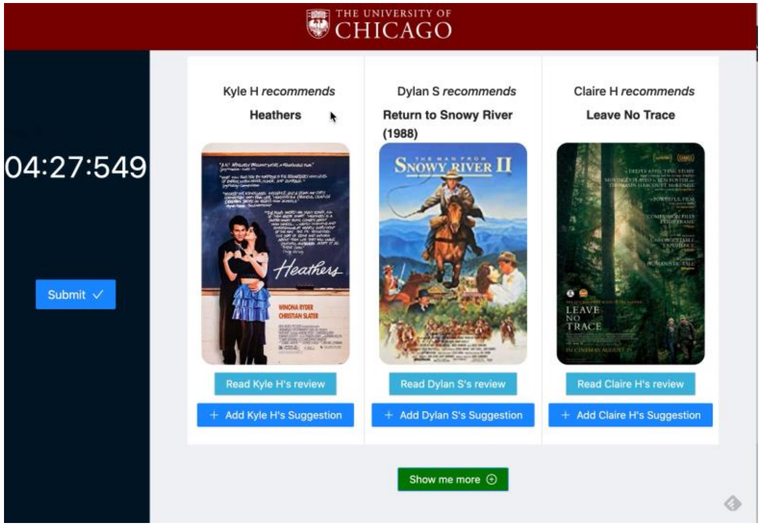
People You May Know

Behavior more thoughtful
Less ingroup favoritism

Algorithm less biased

Lab study that manipulates automaticity

Create two worlds



Simulate scrolling

Manipulate automaticity

Preferences the same
Behavior different

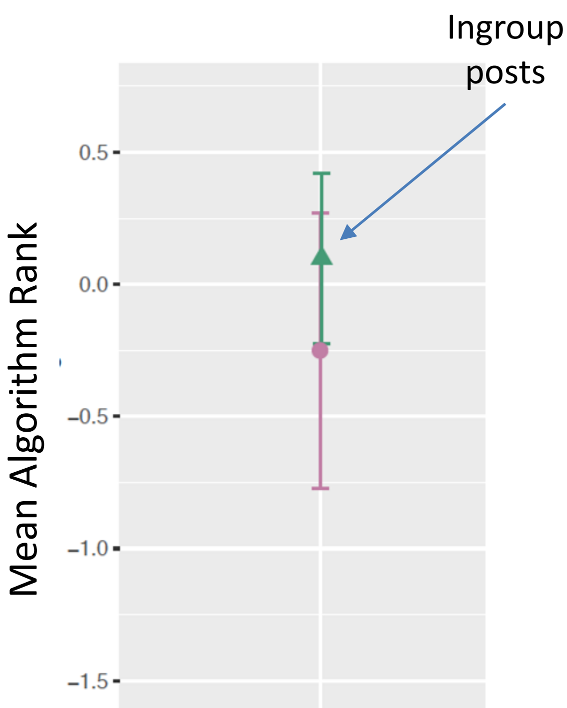
Train an algorithm on each of these conditions

deliberate behavior — automatic behavior

"PYMK" — "Newsfeed"

Lab study that manipulates automaticity

Algorithms trained on two worlds



Trained on deliberate behavior | Trained on automatic behavior

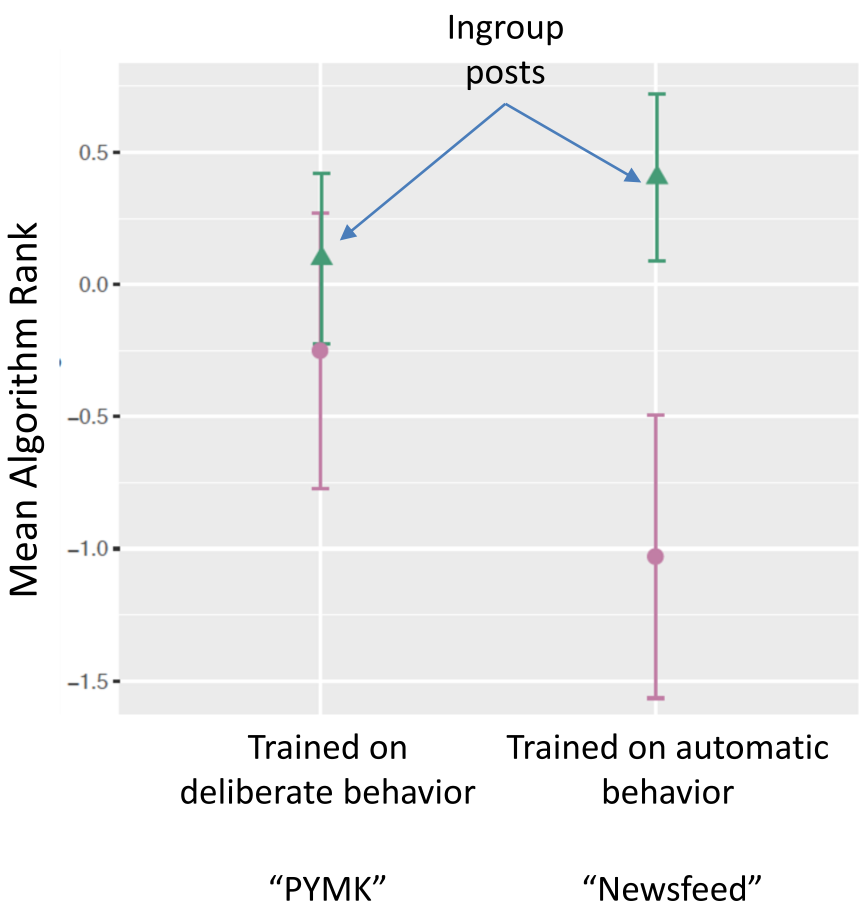
“PYMK” | “Newsfeed”

How does algorithm rank posts?

Does it create a bias?

Lab study that manipulates automaticity

Algorithms trained on two worlds



How does algorithm rank posts?

Does it create a bias?

It does when behavior is automatic

Broader lessons

Algorithms that Misunderstand us

ML approach

- Predict easy to measure behavior
- Use predictions to drive “decisions”

Algorithms that Misunderstand us

ML approach

Naïve presumption – measured behavior
revealed our preferences, objectives, feelings..

Problem transcends social media

Recommender systems

Hiring algorithms

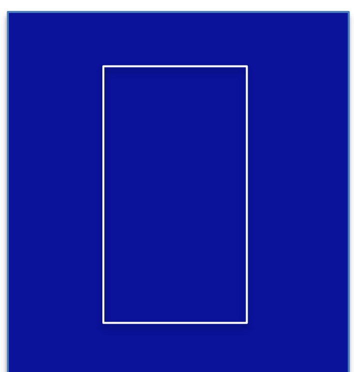
Pricing algorithms

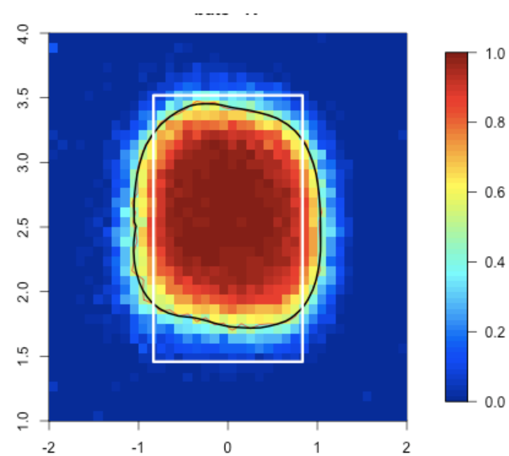
Algorithms that misunderstand us

ML approach

Naïve presumption – measured behavior
revealed our preferences, objectives, feelings..

Problem transcends automaticity or want-should conflict



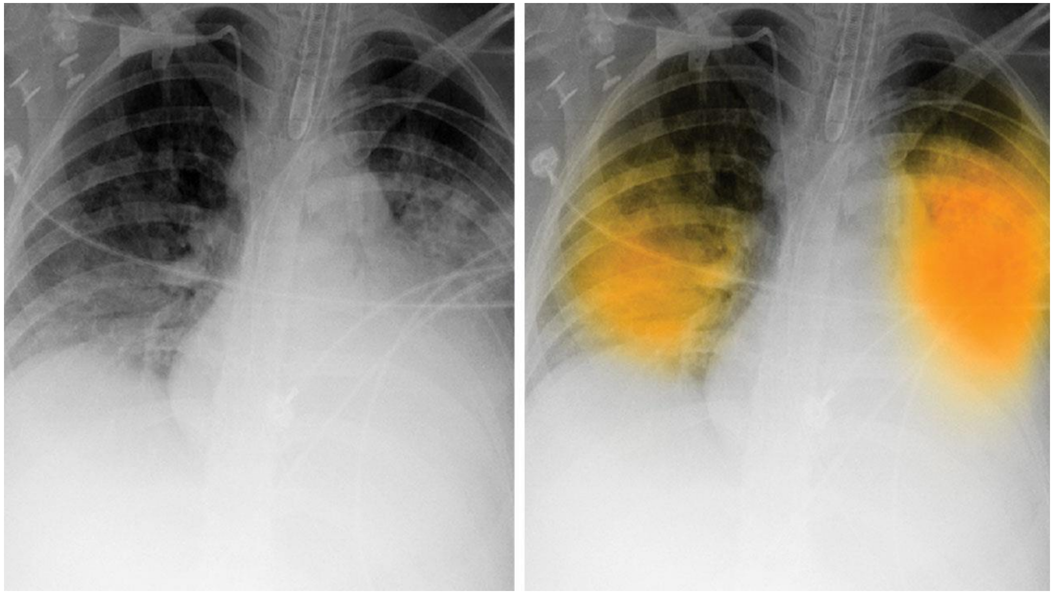


Algorithms that misunderstand us

ML approach

Naïve presumption – measured behavior
revealed our preferences, objectives, feelings..

Problem transcends automaticity or want-should conflict



Algorithms that misunderstand us

ML approach

Naïve presumption – measured behavior
revealed our preferences, objectives, feelings..

How do we fix this?

Use known psychology to model
behavior relates to user's mind

Incorporate behavioral insights into how we build
machine learning models

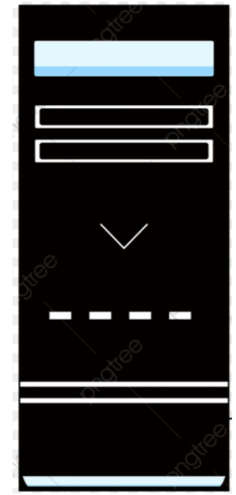
ML needs structural modeling

Why AI Needs Behavioral Economics

Why Behavioral Economics Needs AI

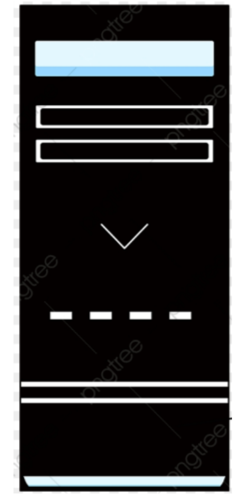
Algorithms have (often implicit) models of people
Those models are naive and likely wrong.
Those errors can have massive consequences

Humans



Machine “intelligence”

\>



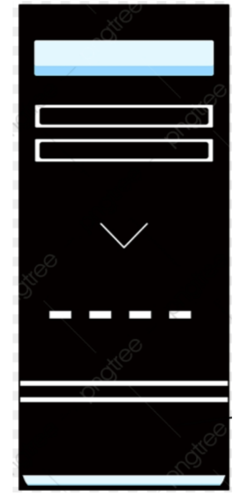
Humans

Machine "intelligence"

Understand hidden "meaning" behind data

Trapped inside data

<



Humans

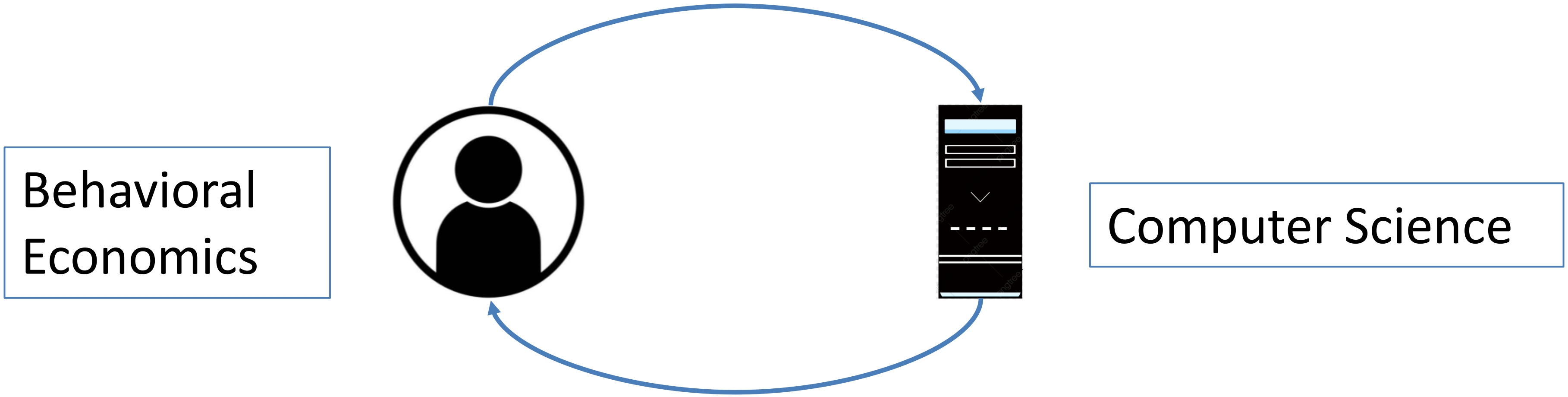
Understand hidden "meaning" behind data

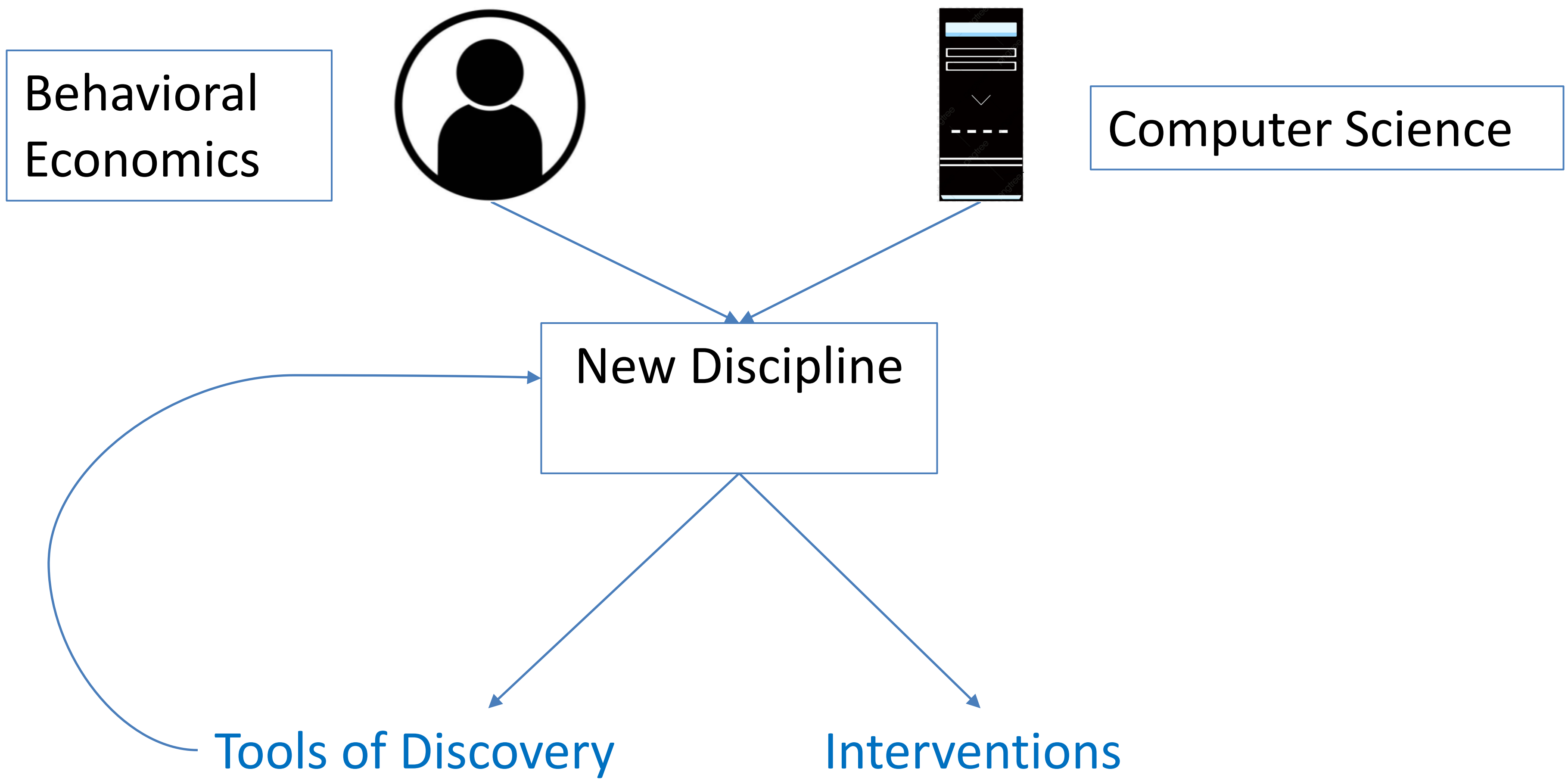
We are limited in what we see in the data

Machine "intelligence"

Trapped inside data

Can see things in data we cannot


Behavioral Economics
Computer Science


Behavioral Economics
Computer Science
New Discipline
Tools of Discovery
Interventions