CSC2552
Topics in Computational Social Science: AI, Data, and Society

Lecture 1: Introduction to Computational Social Science

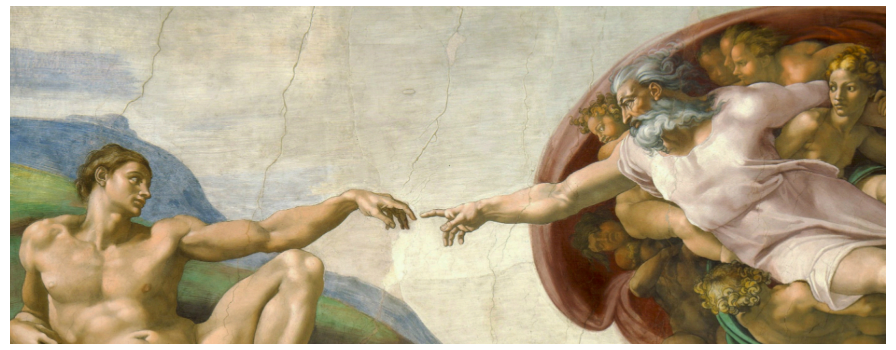
Prof. Ashton Anderson, Fall 2023

Computational social science in 7 easy pieces

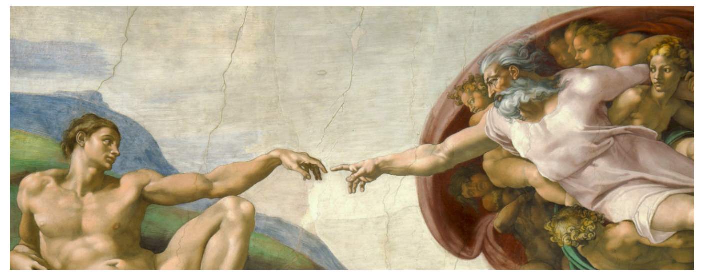
Week	Date	Topic	Reviews Due	Textbook Readings
1	9/7	Introduction to computational social science		Ch. 1
2	9/14	Introduction to computational social science cont'd		Ch. 1
3	9/21	Observational studies 1	9/20 9:00pm	Ch. 2
4	9/28	Observational studies 2	9/27 9:00pm	Ch. 2
5	10/5	Experiments 1	10/4 9:00pm	Ch. 4
6	10/12	Project proposals		
7	10/19	Experiments 2	10/18 9:00pm	Ch. 4
8	10/26	Asking questions	10/25 9:00pm	Ch. 3
9	11/2	Deep learning	11/1 9:00pm	
10	11/16	Ethics in computational social science	11/15 9:00pm	Ch. 6
11	11/23	Project presentations (Part 1)		
12	11/30	Project presentations (Part 2)		

Readymades

Custommades



Ways of doing computational social science



Observational analyses

Human computation

Natural experiments

Surveys

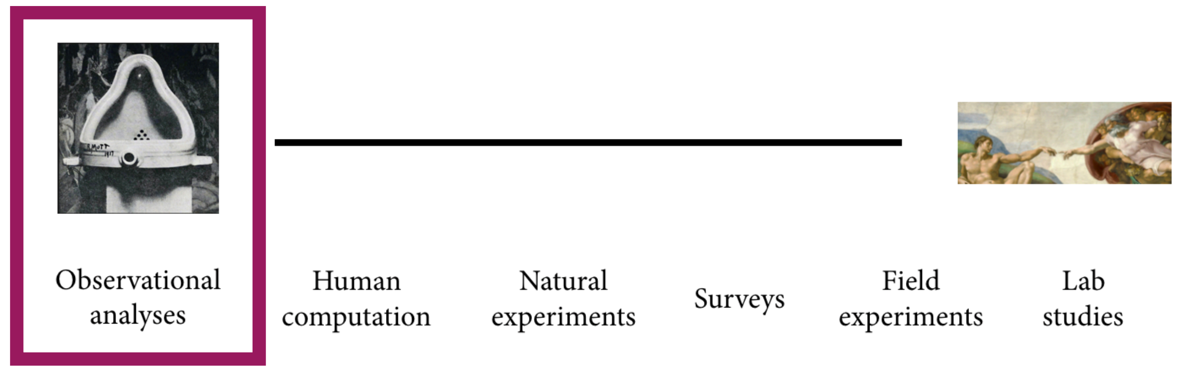
Field experiments

Lab studies

Observational analyses of existing data

- Massive datasets of all kinds of human behaviour are now available for study
  - Wikipedia, GPS traces, health databases, Facebook, Twitter, Reddit, reviews, purchases, dating, invitations, exercise apps, etc., etc…
- Key part of the "socioscope": huge traces of things that we couldn't see before
- Lack of detail/fidelity in individual records is hopefully made up for by large numbers of records (hope: small noisy errors cancel out, big patterns are signal)

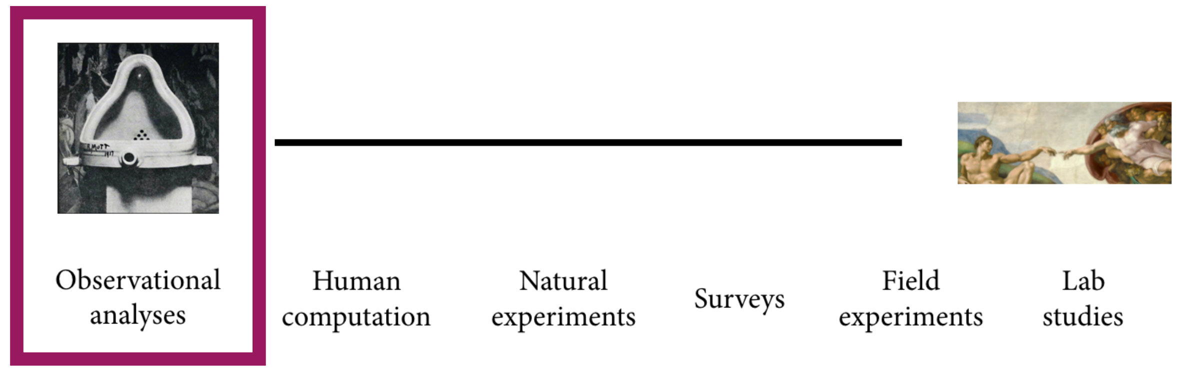
"Big data" / "Found data"



Ten common characteristics of big data

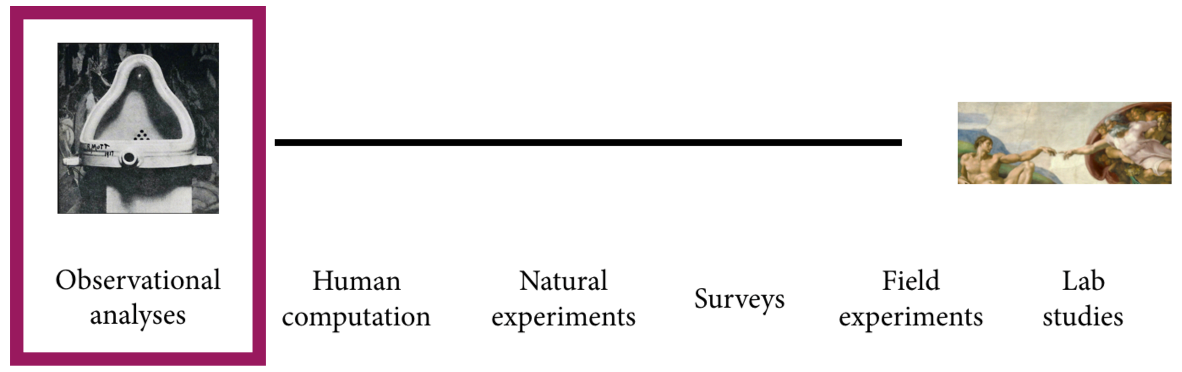
- **Big**: statistical power, rare events, fine resolution
- **Always-on**: unexpected events, real-time measurement
- **Nonreactive**: measurement probably won't change behaviour

- **Incomplete**: probably won't have the ideal information you want
- **Inaccessible**: difficult to access (gov't, companies)
- **Nonrepresentative**: bad out-of-sample generalization (good in-sample)
- **Drifting**: Population drift, usage drift, system drift
- **Algorithmically confounded**: want to study behaviour, not an algorithm
- **Dirty**: Junk, spam
- **Sensitive**: Private, hard to tell what's sensitive



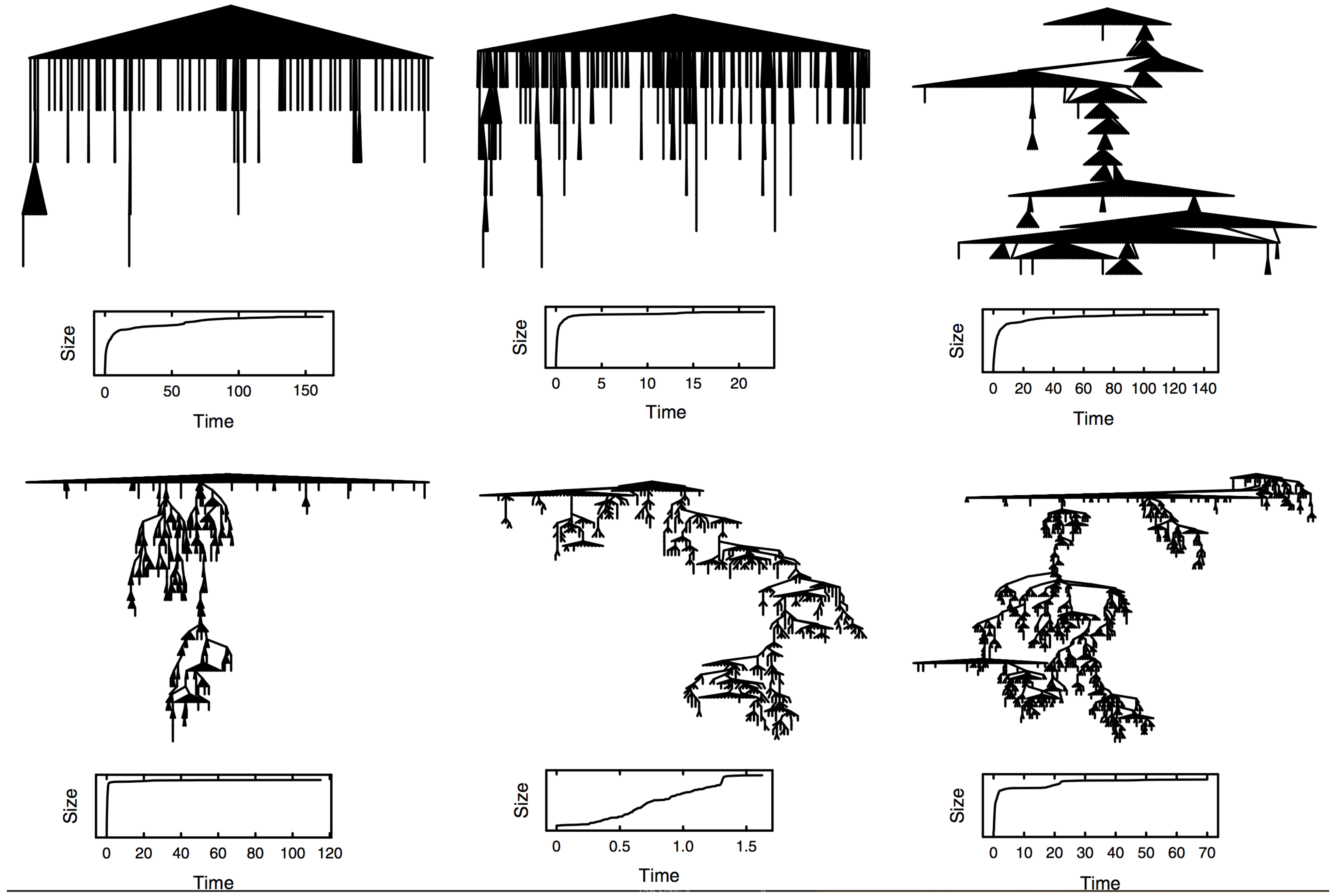
Observing Behaviour: Three research strategies

1. Counting things
2. Forecasting/nowcasting
3. Approximating experiments

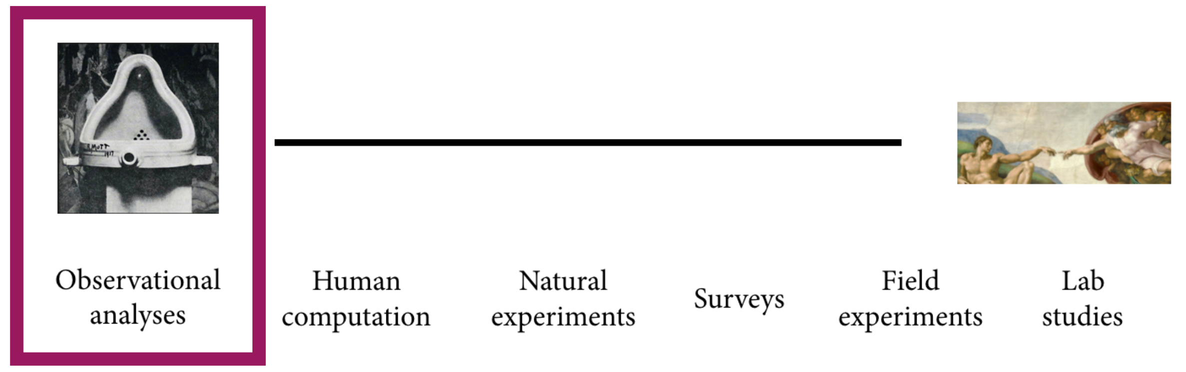


Observing Behaviour: 1. Counting Things

Example: Measuring viral vs. broadcast diffusion on Twitter



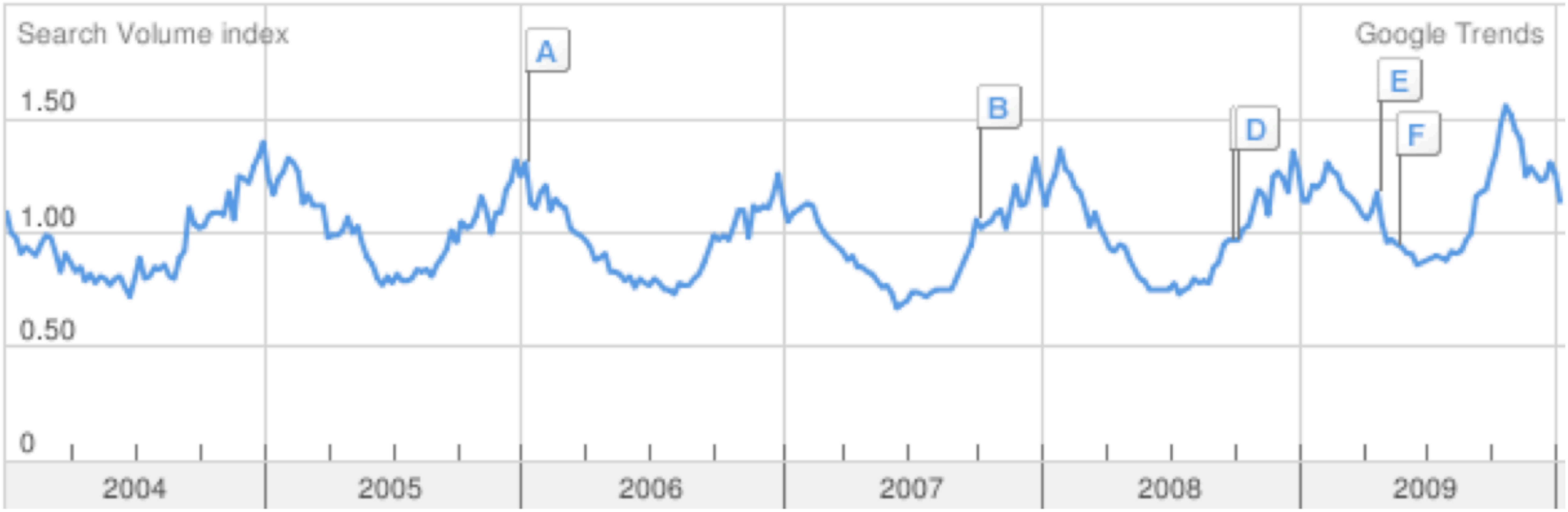
With newfound datasets and computational resources, many valuable initial contributions are measurements of quantities we couldn't measure before → counting at scale



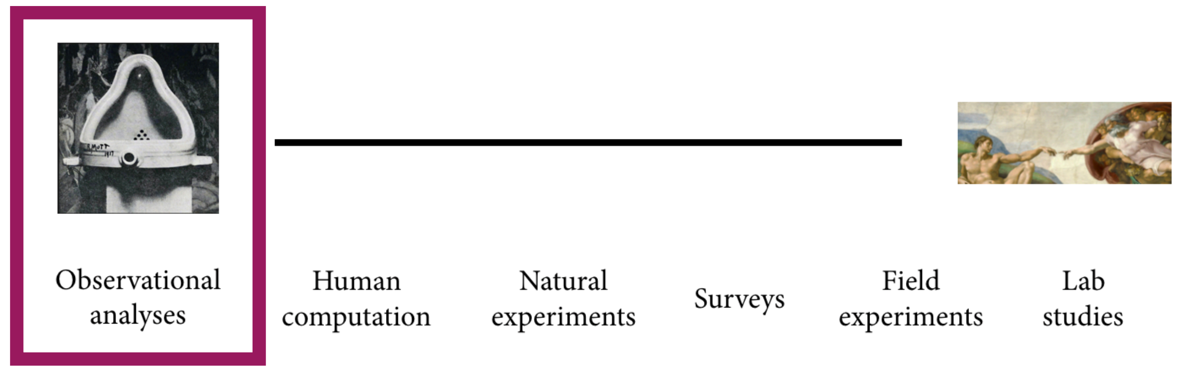
Observing Behaviour: 2. Nowcasting

Google Flu Trends

Idea: find 50 most correlated search query volume trends with flu data

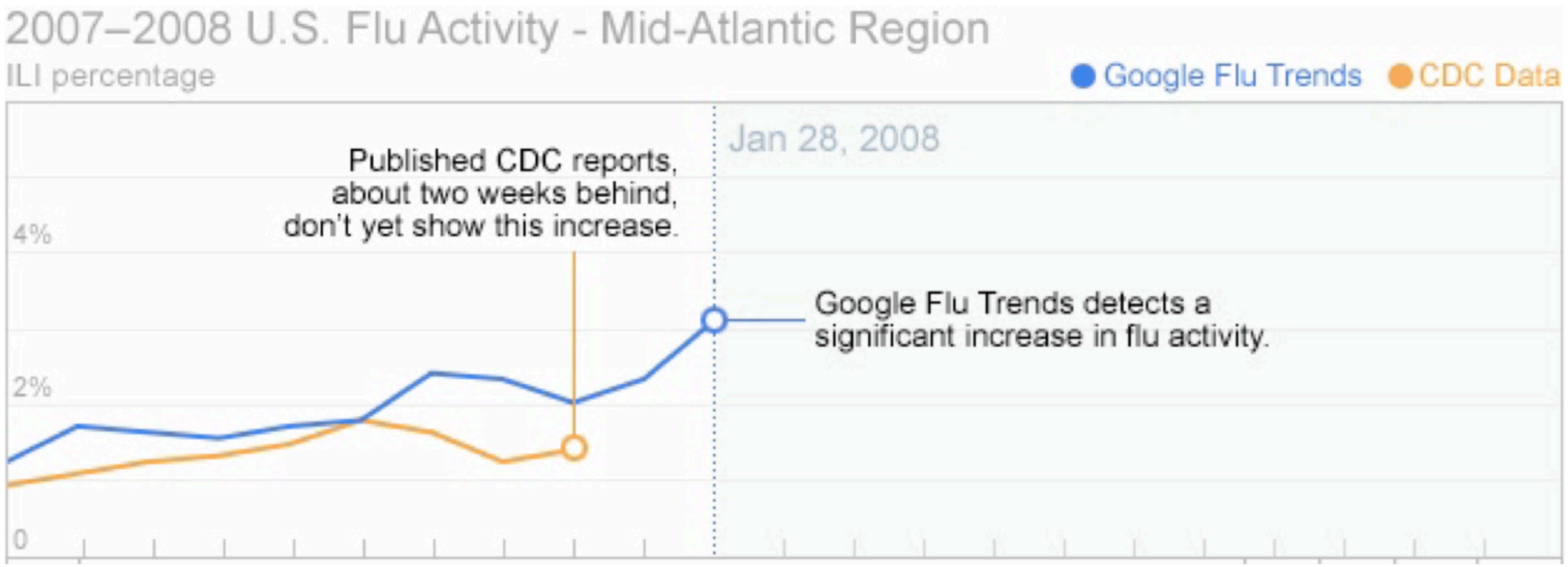


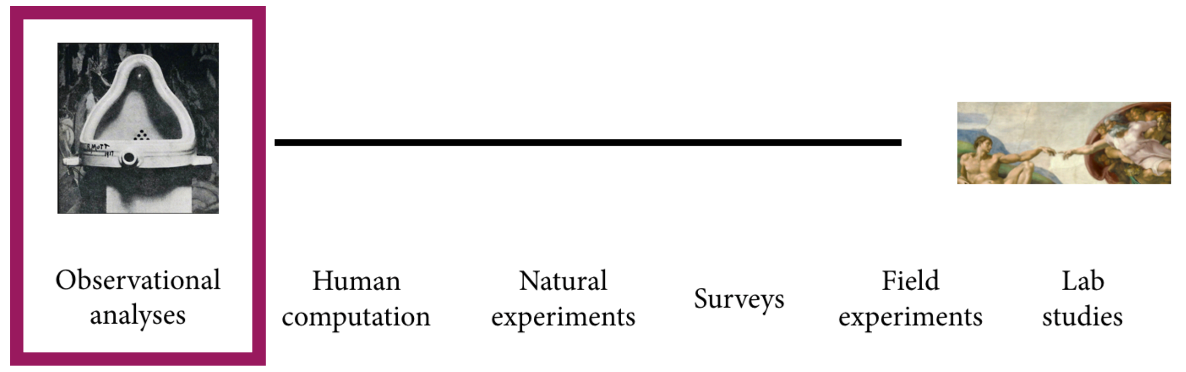
Search volume for the term "cough"



Observing Behaviour: 2. Nowcasting

The flu has a 1-2 week lag from when cases are reported to when the CDC releases official stats





Observing Behaviour: 2. Nowcasting

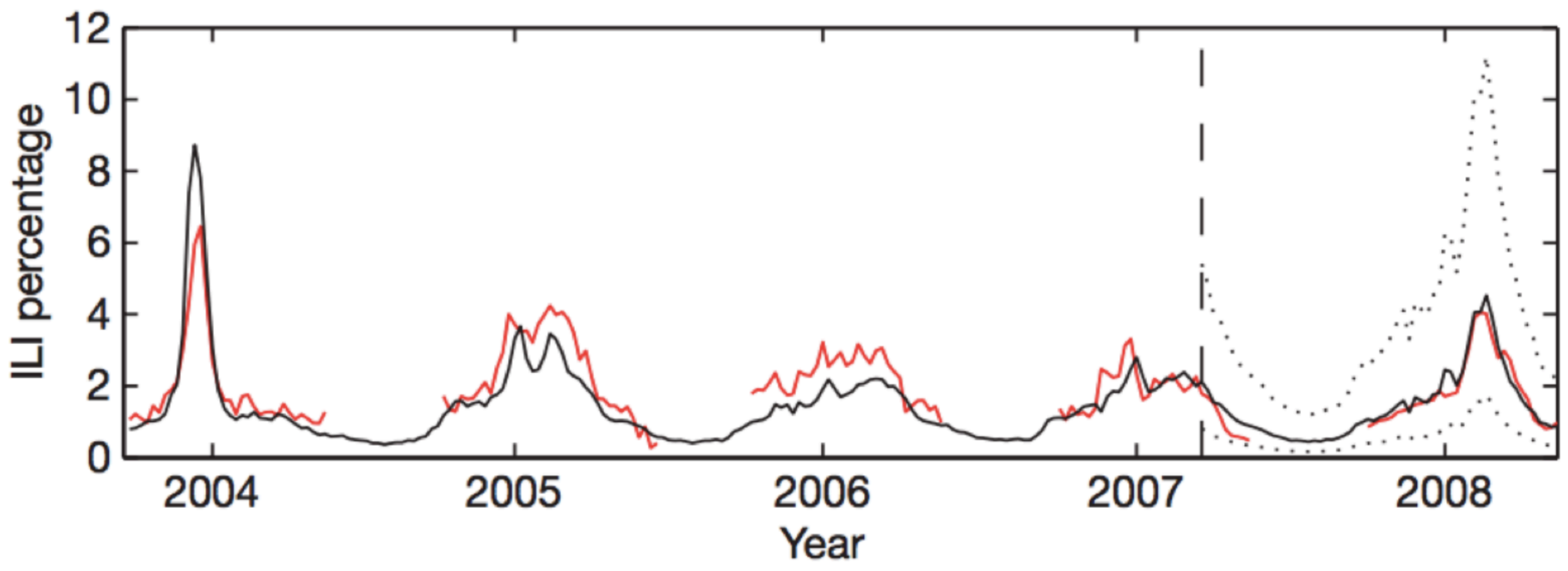
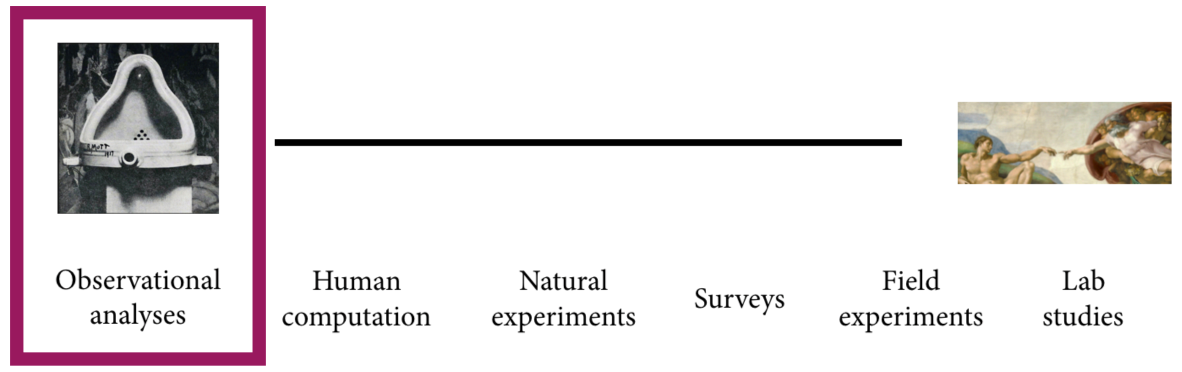
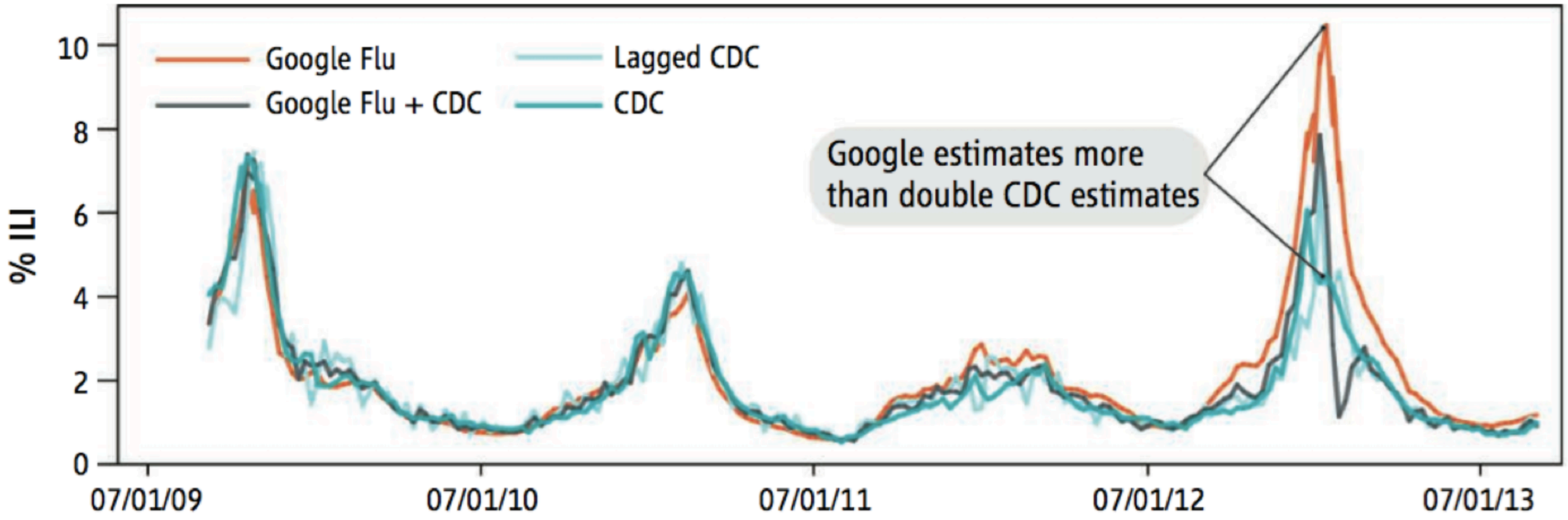


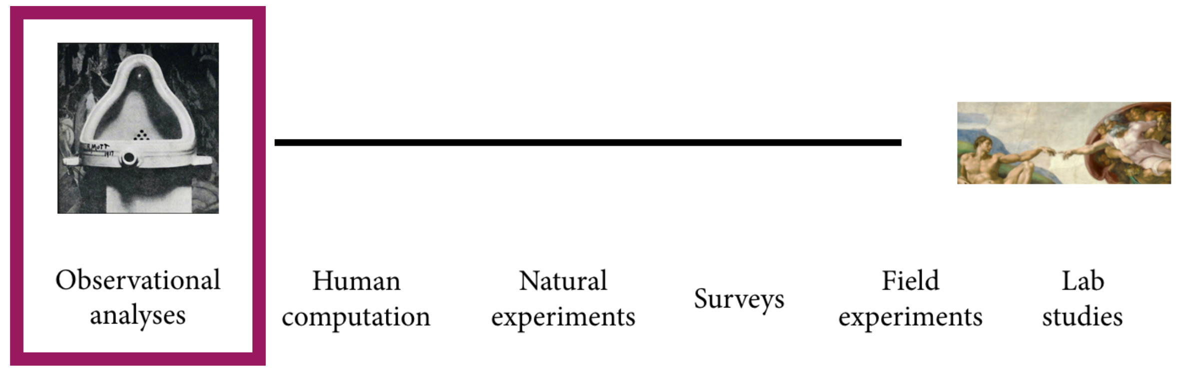
Figure 2 | A comparison of model estimates for the mid-Atlantic region (black) against CDC-reported ILI percentages (red), including points over which the model was fit and validated. A correlation of 0.85 was obtained over 128 points from this region to which the model was fit, whereas a correlation of 0.96 was obtained over 42 validation points. Dotted lines indicate 95% prediction intervals. The region comprises New York, New Jersey and Pennsylvania.



Observing Behaviour: 2. Nowcasting



Soon after Google Flu Trends launched, it was drastically off



Observing Behaviour: 2. Nowcasting

Media attention

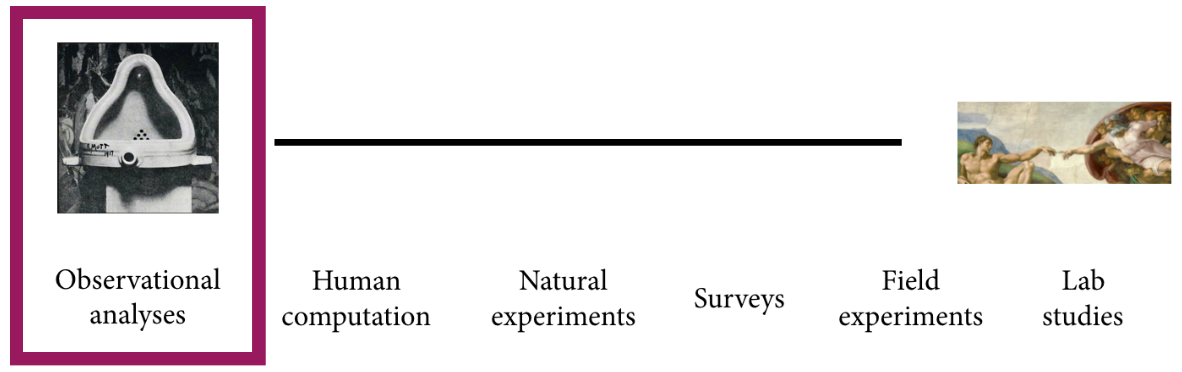
- "Bird flu", "swine flu"

Algorithm changes

- Starting suggesting search terms

"Social hacking" (Trolling)

- Hey look we can screw up Google's flu predictions

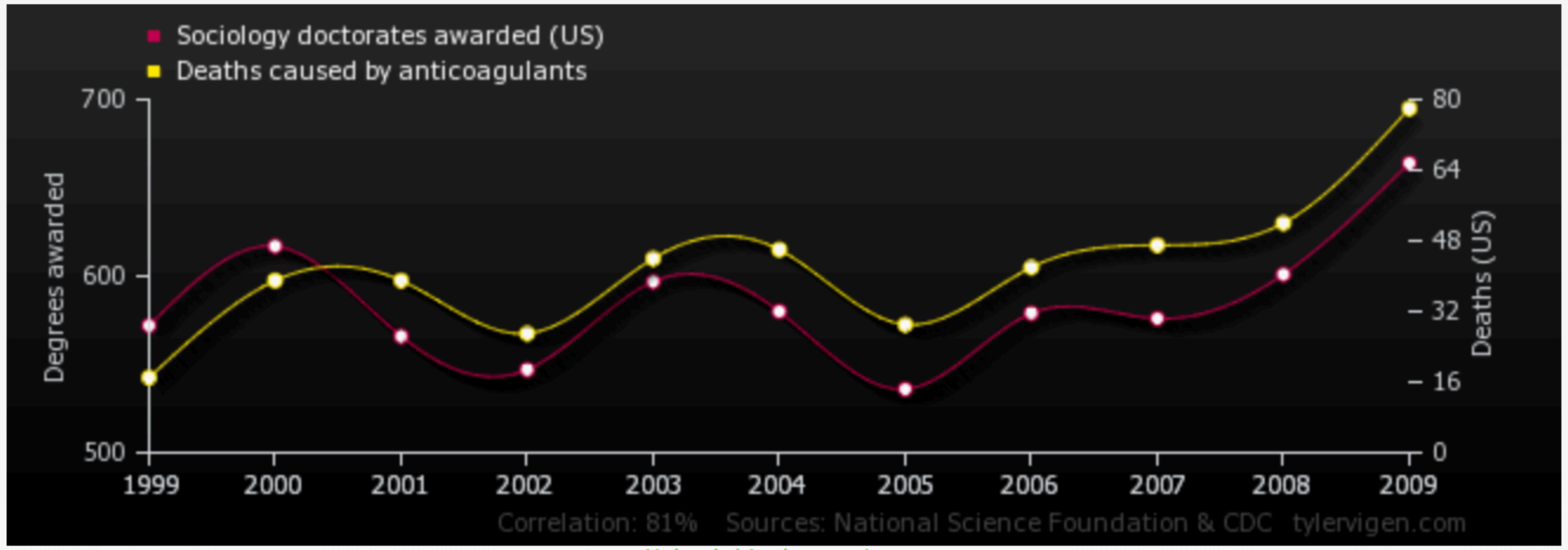


Correlation and causation

Sociology doctorates awarded (US)

correlates with

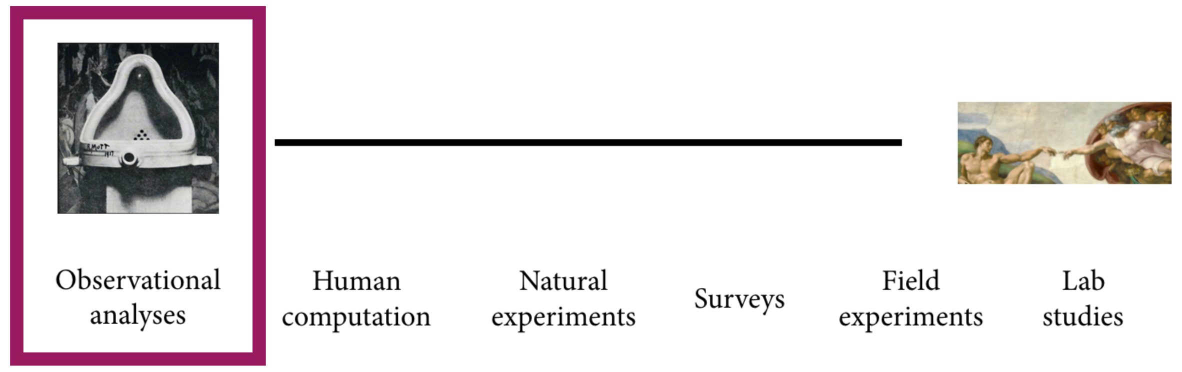
Deaths caused by anticoagulants



Upload this chart to imgur

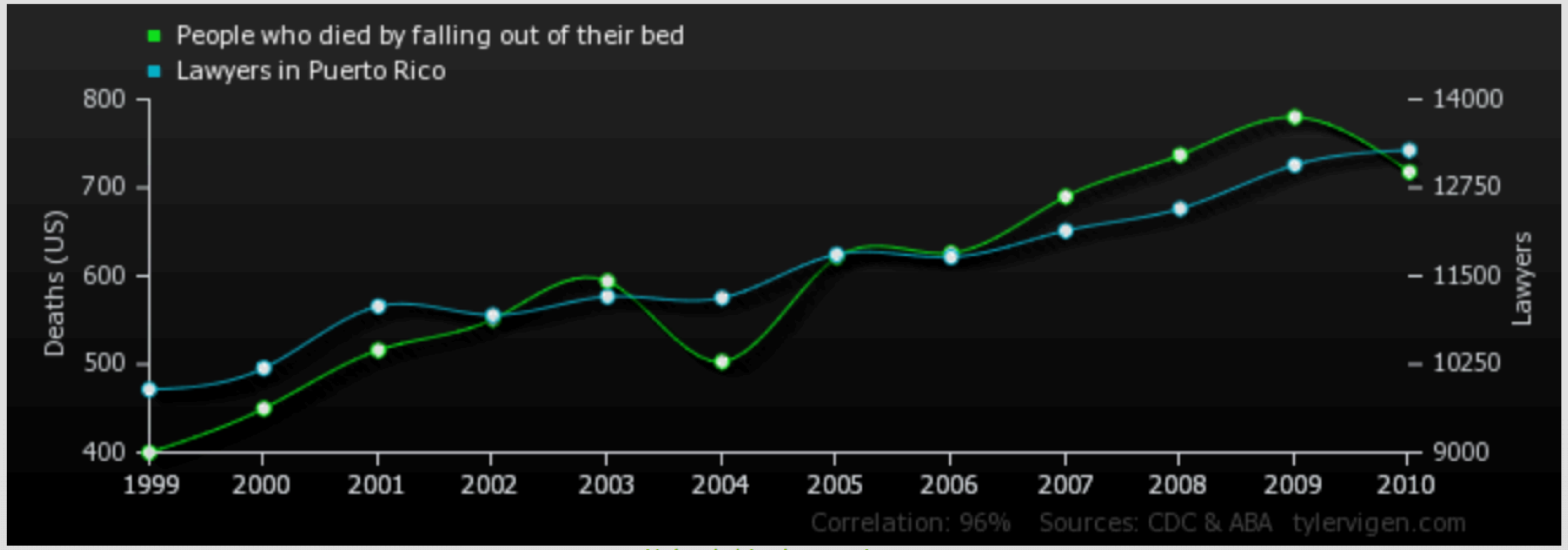
	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
Sociology doctorates awarded (US) *Degrees awarded (National Science Foundation)*	572	617	566	547	597	580	536	579	576	601	664
Deaths caused by anticoagulants *Deaths (US) (CDC)*	17	39	39	27	44	46	29	42	47	52	78

Correlation: 0.811086



Correlation and causation

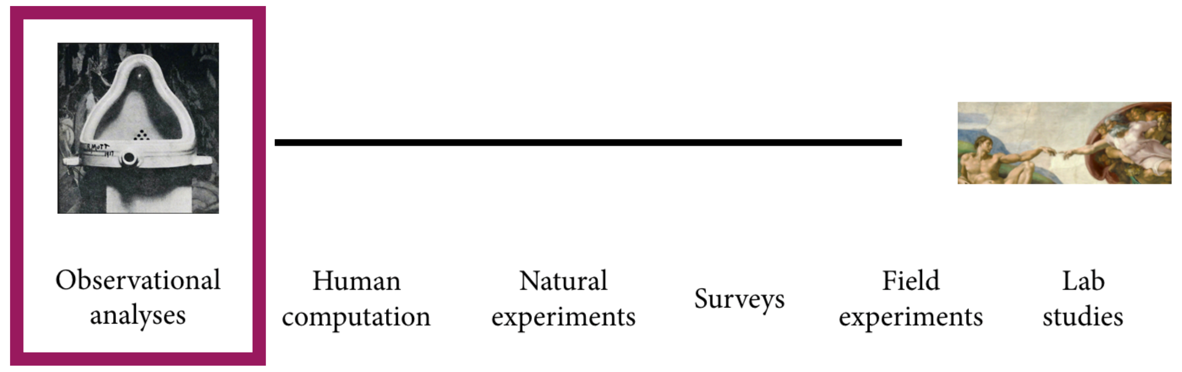
People who died by falling out of their bed
correlates with
Lawyers in Puerto Rico



Upload this chart to imgur

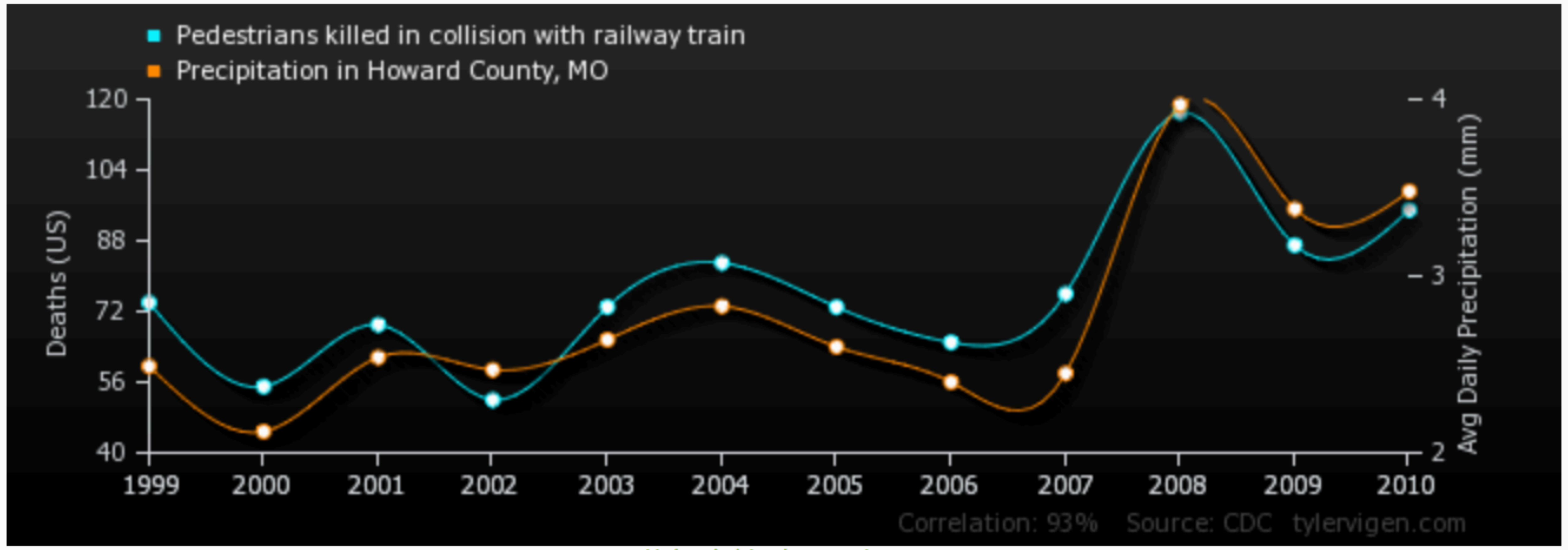
	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
People who died by falling out of their bed *Deaths (US) (CDC)*	400	450	516	551	594	503	621	626	690	737	780	718
Lawyers in Puerto Rico *Lawyers (ABA)*	9,892	10,195	11,071	10,947	11,209	11,191	11,805	11,767	12,142	12,454	13,071	13,282

Correlation: 0.957087



Correlation and causation

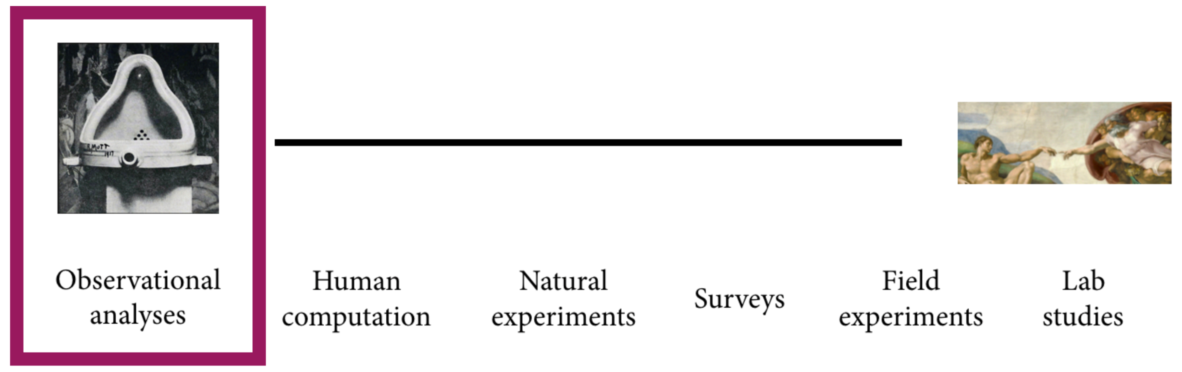
Pedestrians killed in collision with railway train
correlates with
Precipitation in Howard County, MO



Upload this chart to imgur

	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Pedestrians killed in collision with railway train *Deaths (US) (CDC)*	74	55	69	52	73	83	73	65	76	117	87	95
Precipitation in Howard County, MO *Avg Daily Precipitation (mm) (CDC)*	2.49	2.12	2.54	2.47	2.64	2.83	2.6	2.4	2.45	3.97	3.38	3.48

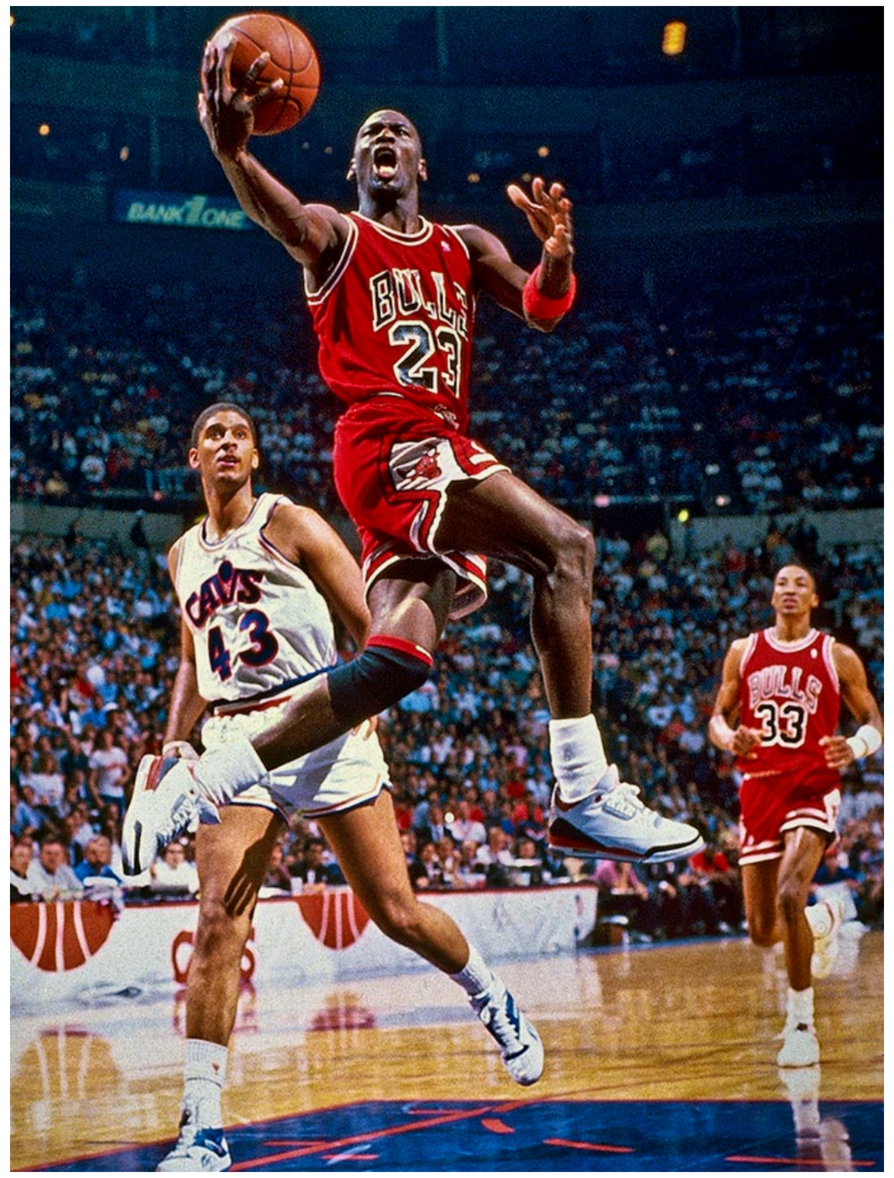
Correlation: 0.92783

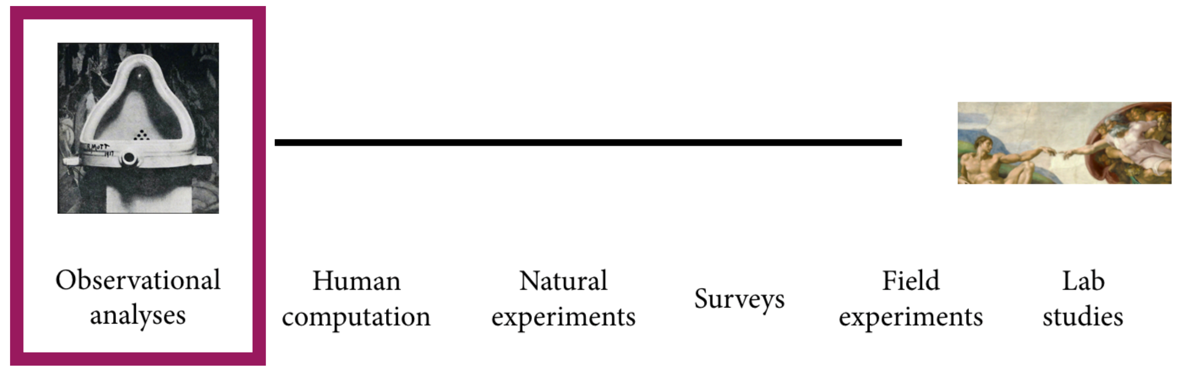


Perils of big data

“When you have large amounts of data, your appetite for hypotheses tends to get even larger. And if it’s growing faster than the statistical strength of the data, then many of your inferences are likely to be false. They are likely to be white noise.”

— Michael Jordan

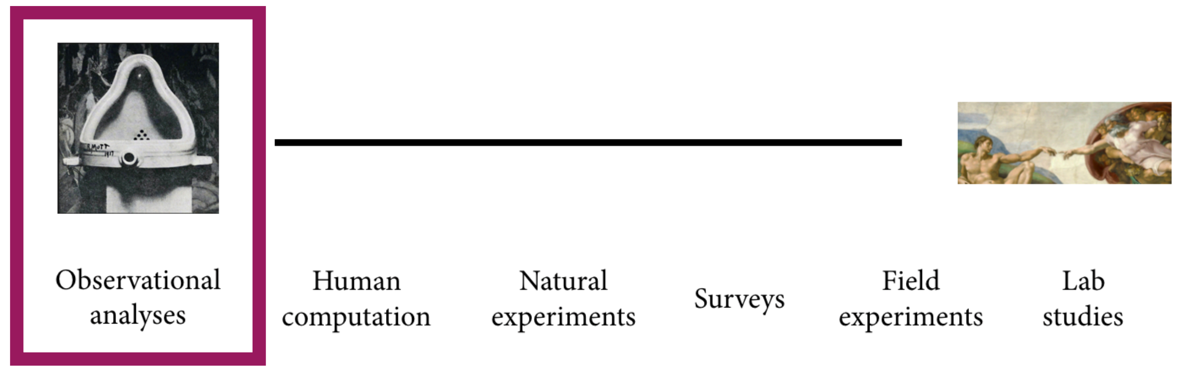




Perils of big data

"When you have large amounts of data, your appetite for hypotheses tends to get even larger. And if it's growing faster than the statistical strength of the data, then many of your inferences are likely to be false. They are likely to be white noise."

— Michael Jordan



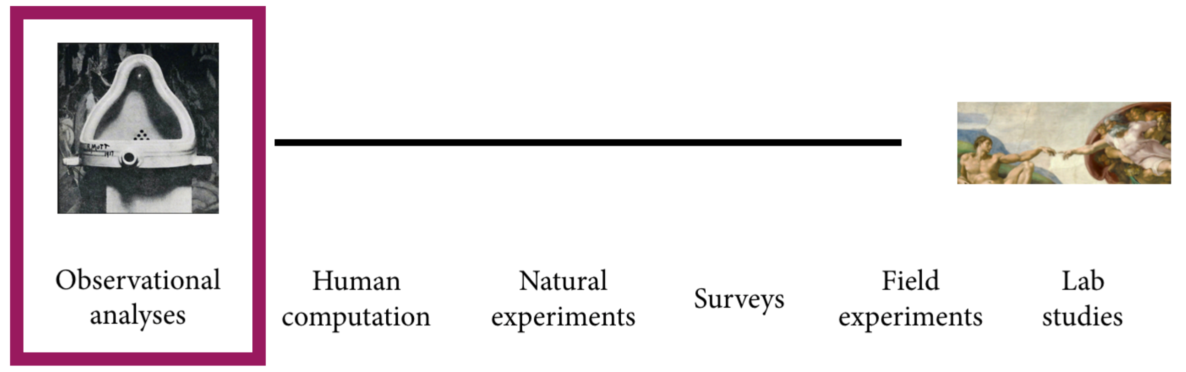
Observing Behaviour: 3. Approximating Experiments

Some clever strategies allow us to do "causal inference": make causal claims from observational data (i.e. arrive at experiment-like conclusions without actually running an experiment)

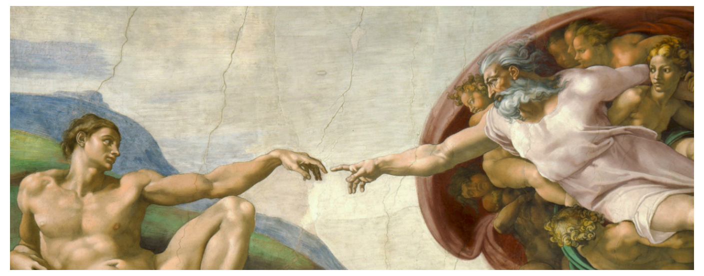
One well-known technique is instrumental variables: exploit natural variation in something to make a causal claim

Rain → Exercise

Friends exercising → You exercise?



Ways of doing computational social science



Observational analyses

Human computation

Natural experiments

Surveys

Field experiments

Experiments

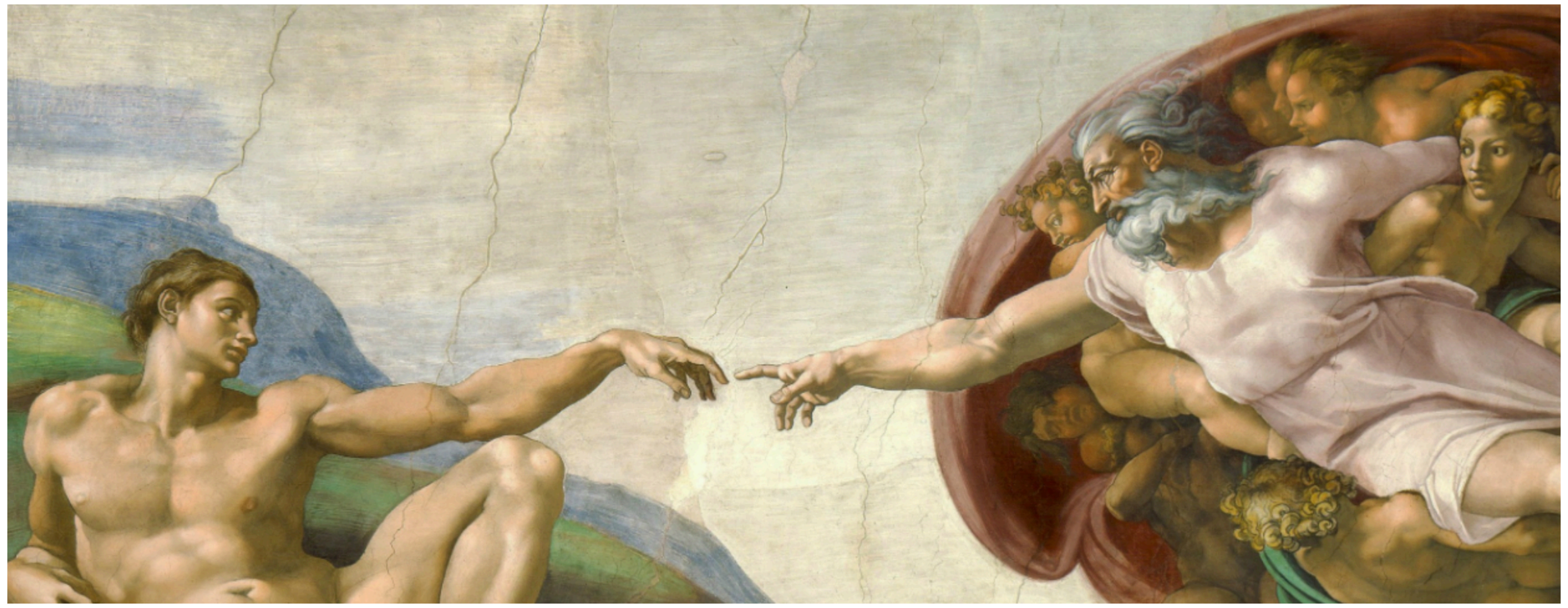
Experiments

On the other end of the spectrum is experimentation

The goal is to learn about causal relationships (cause-and-effect questions)

The strategy is to directly manipulate the environment and observe the consequences

Design the ideal scenario that will create just the data you need to answer your question



Experiments

Here, researchers intervene in the world to isolate and study a specific question

Nomenclature:

“Experiment”: perturb and observe

“Randomized controlled experiment”: Intervene for one group, don’t for another (randomly)

Correlation is not causation

Observational data often riddled by unknown or hard-to-control *confounding variables*

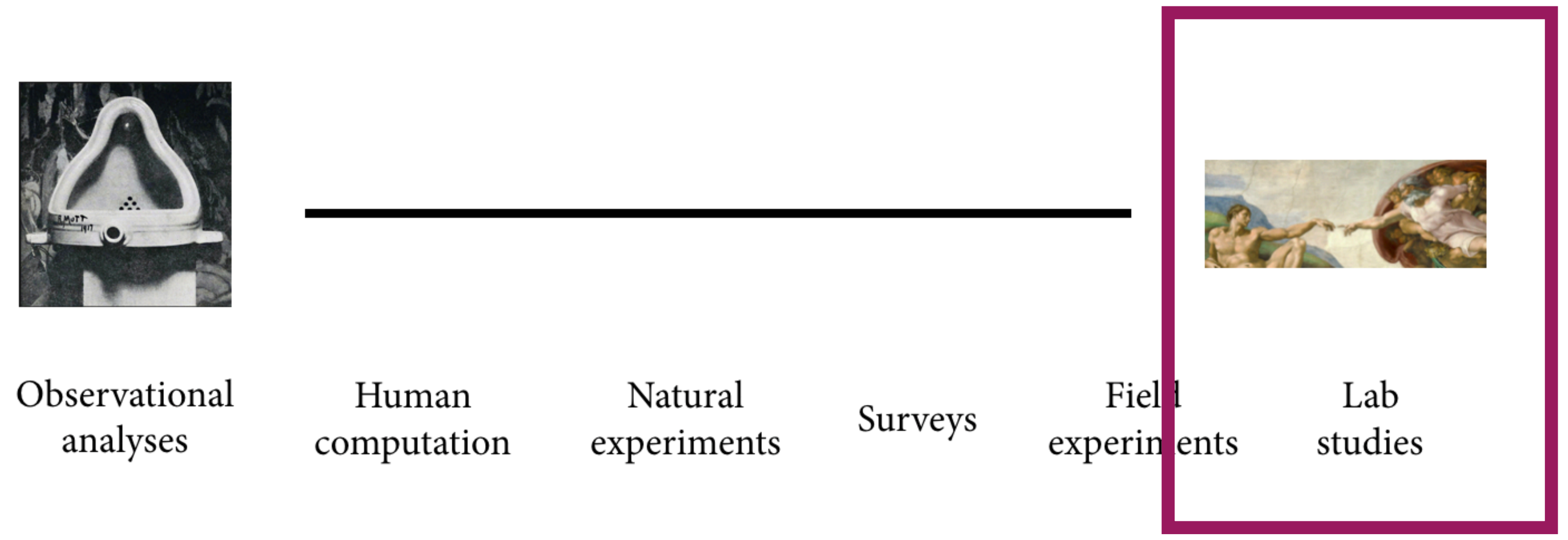
E.g. Do students learn more in schools that offer high teacher salaries?

What’s an observational way to study this question?

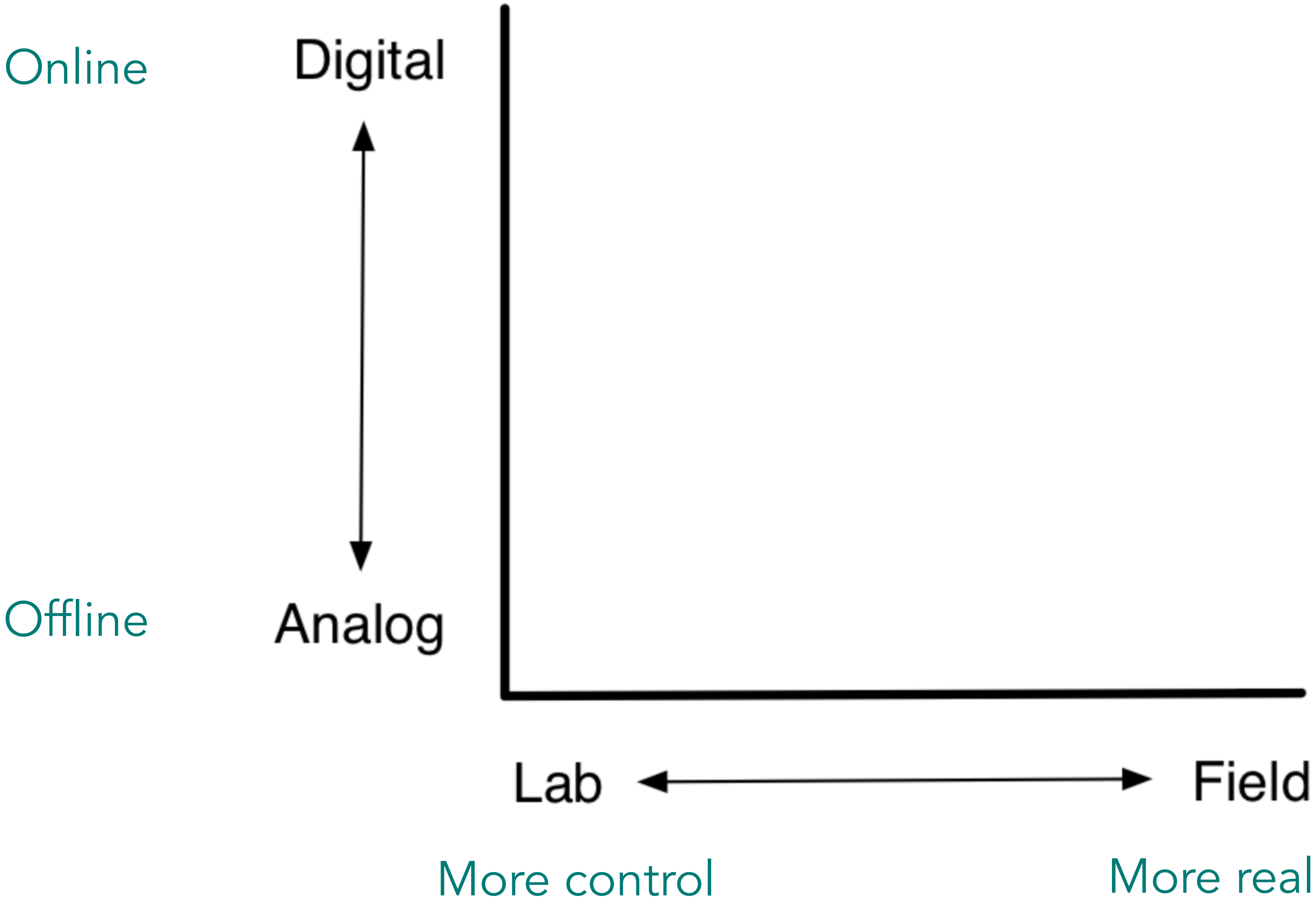
What’s wrong with it?

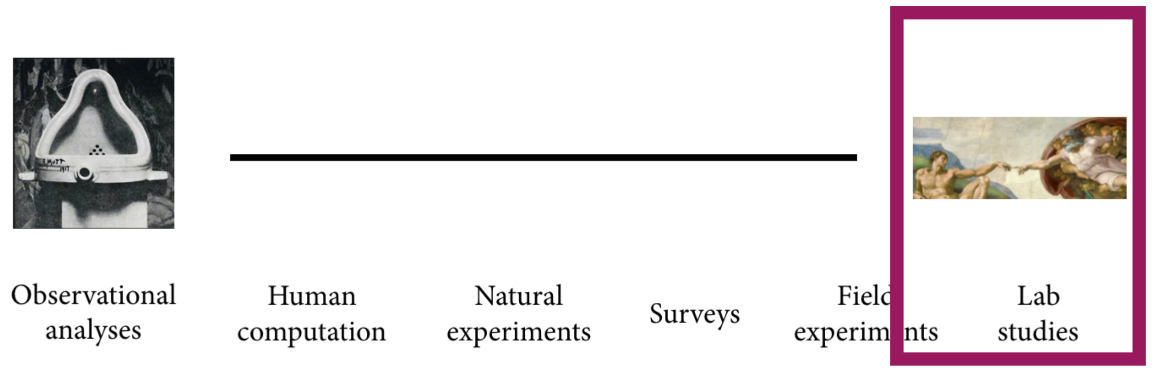
What’s an experimental way to study this question?

What’s wrong with it?

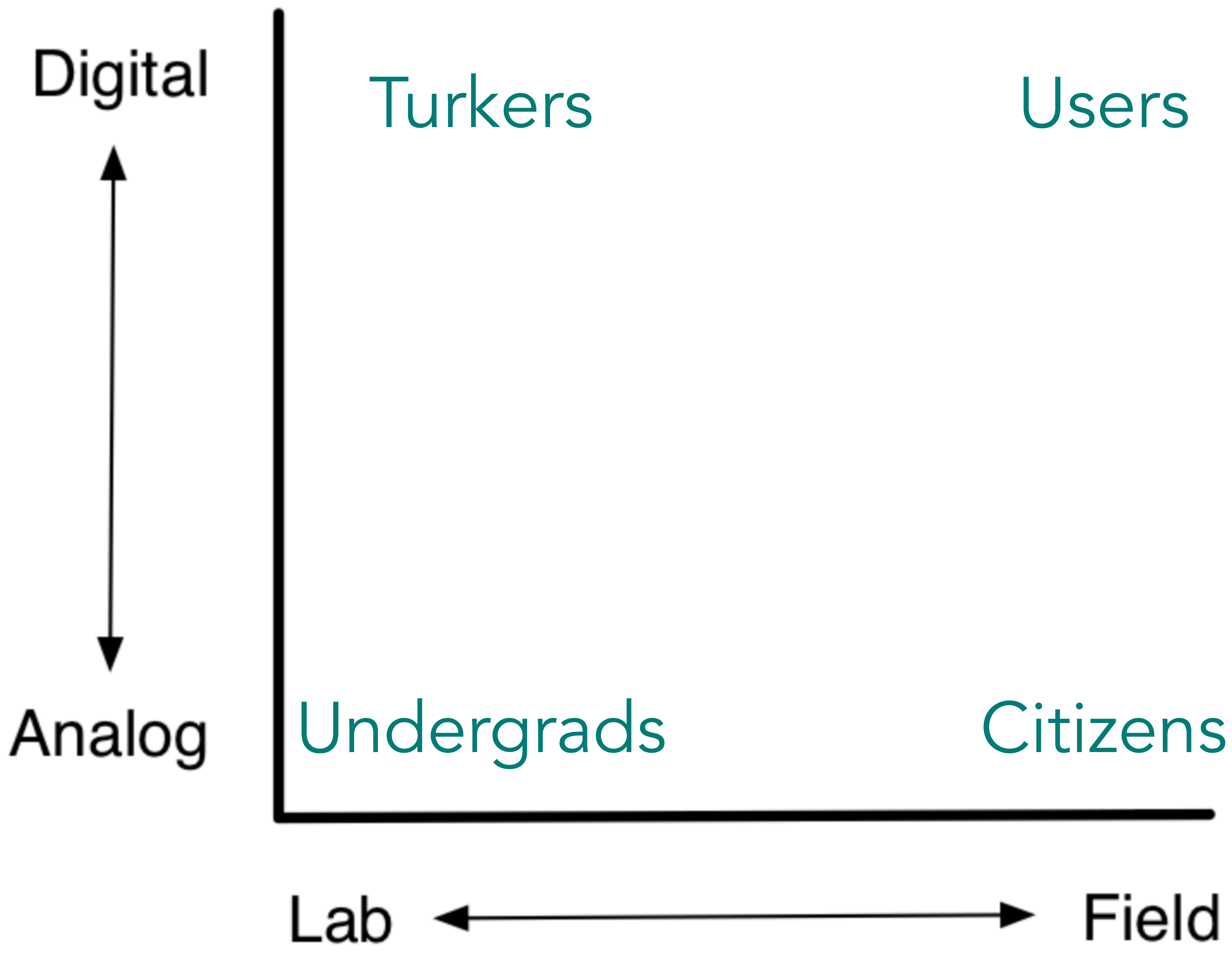


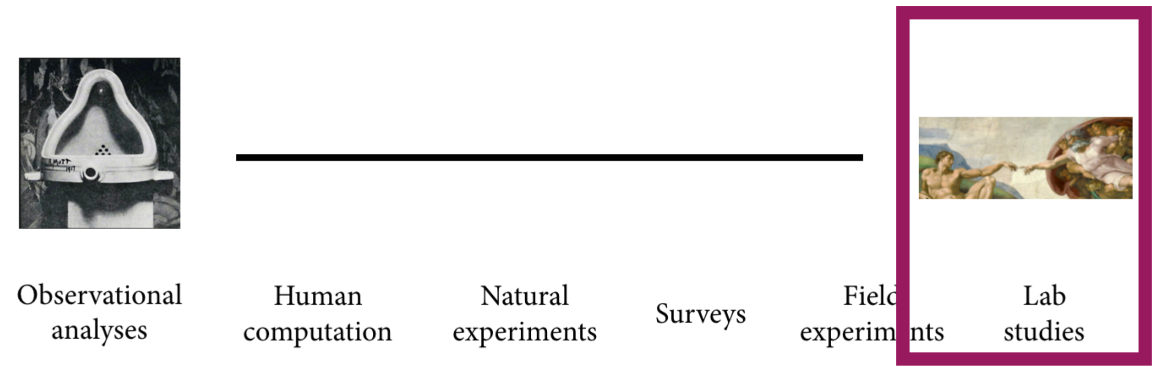
Experiments





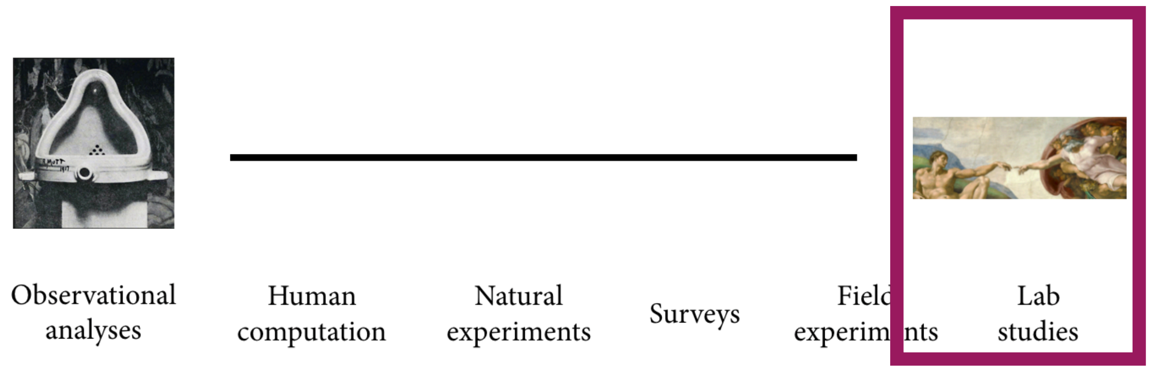
Experiments





Three major components of rich experiments

1. Validity
2. Heterogeneity
3. Mechanisms

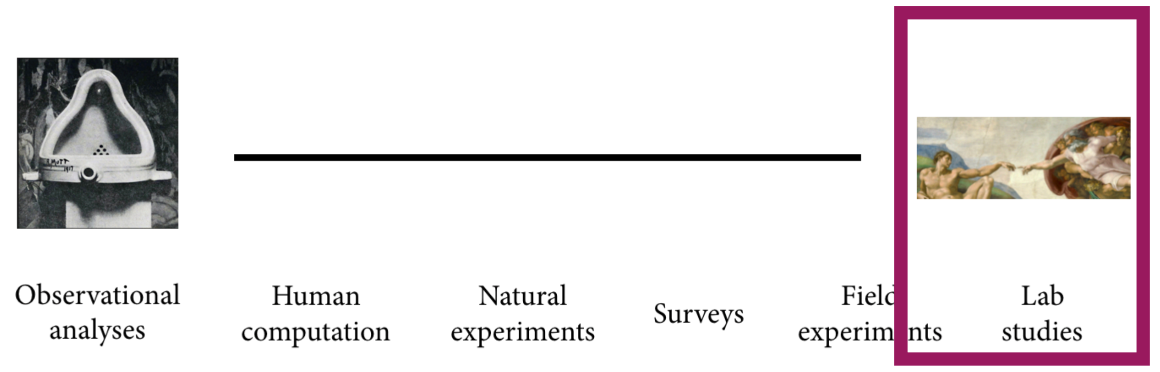


Three major components of rich experiments: 1. Validity

Validity: How general are the results?

Types of validity:

1. Statistical conclusion validity: were the stats done right?
2. Internal validity: was the experiment done right?
3. Construct validity: are we measuring the right thing?
4. External validity: is this applicable in other settings?

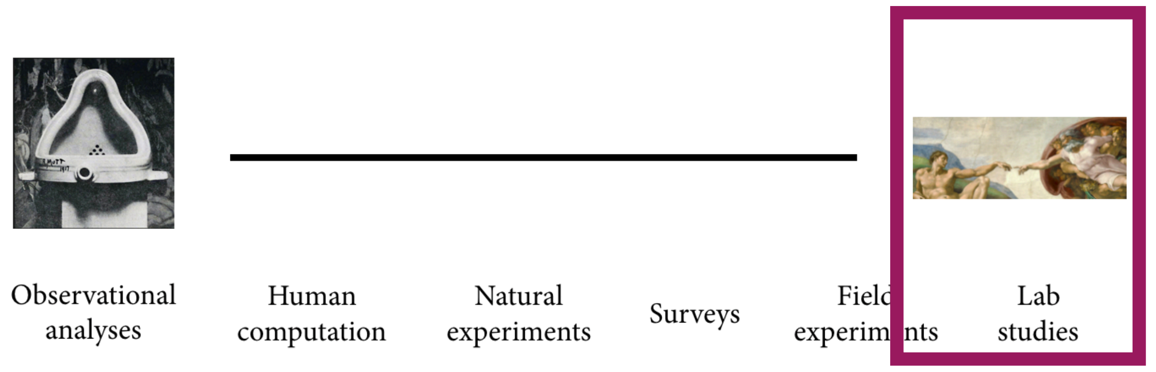


Three major components of rich experiments: 2. Heterogeneity

Barebones experiment: measure the average treatment effect (ATE)

But in social research, people almost always vary.

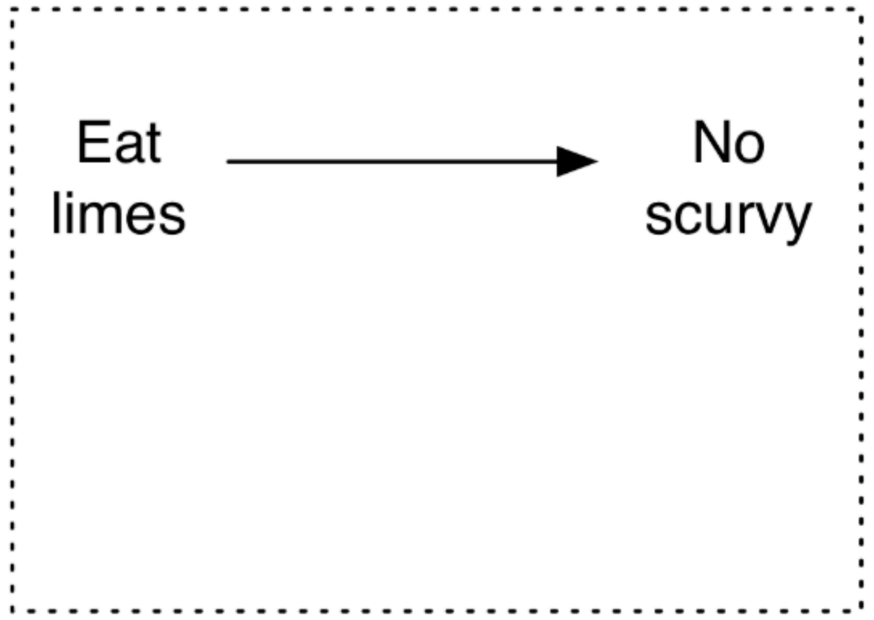
Digital research presents many more opportunities to measure how causes affect people differently



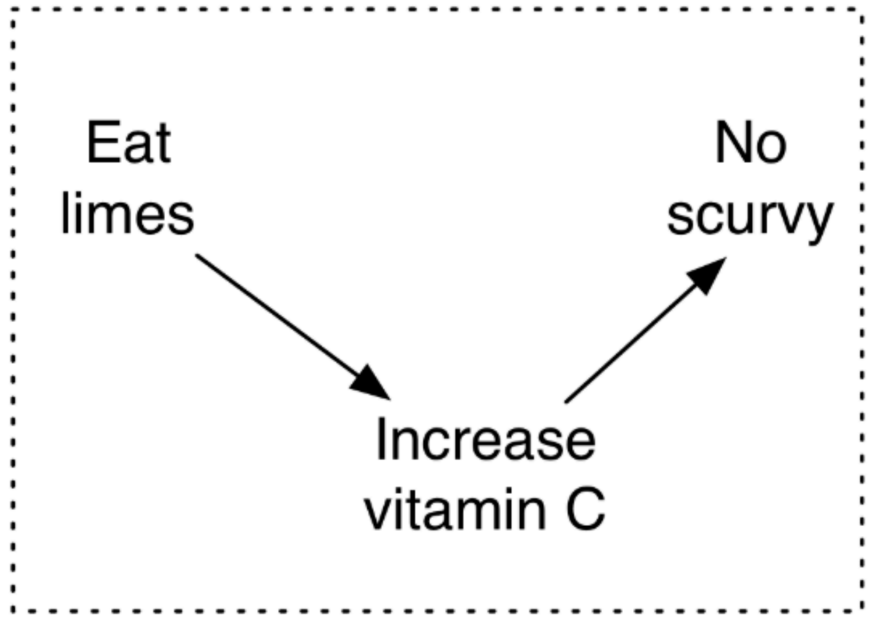
Three major components of rich experiments: 3. Mechanisms

Barebones experiment: measure what happened.

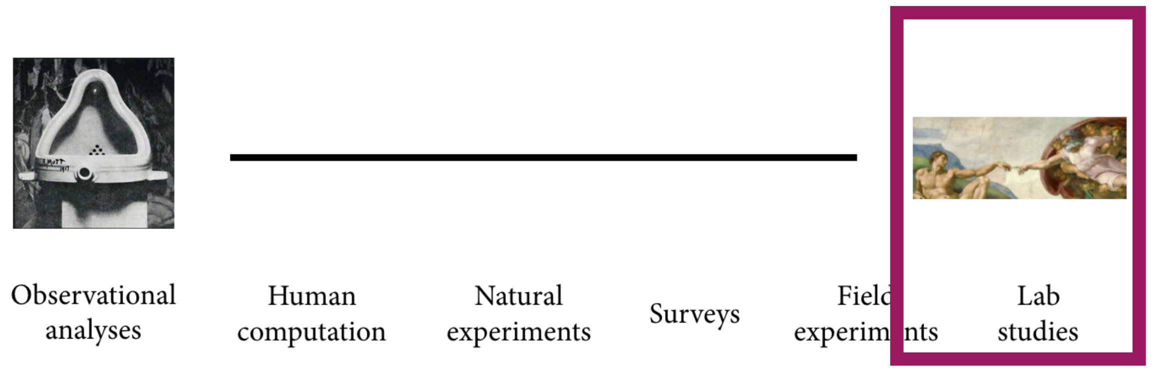
Mechanisms: why and how did it happen?



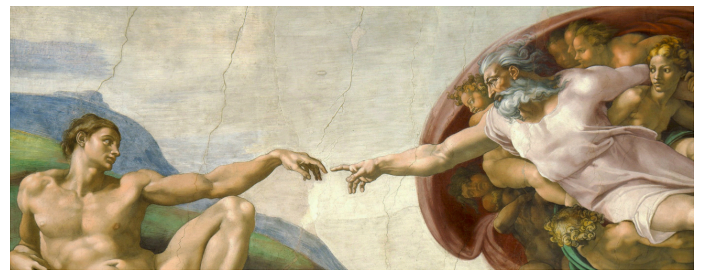
Causal effect
without mechanism



Causal effect
with mechanism



Ways of doing computational social science



Observational analyses

Human computation

Natural experiments

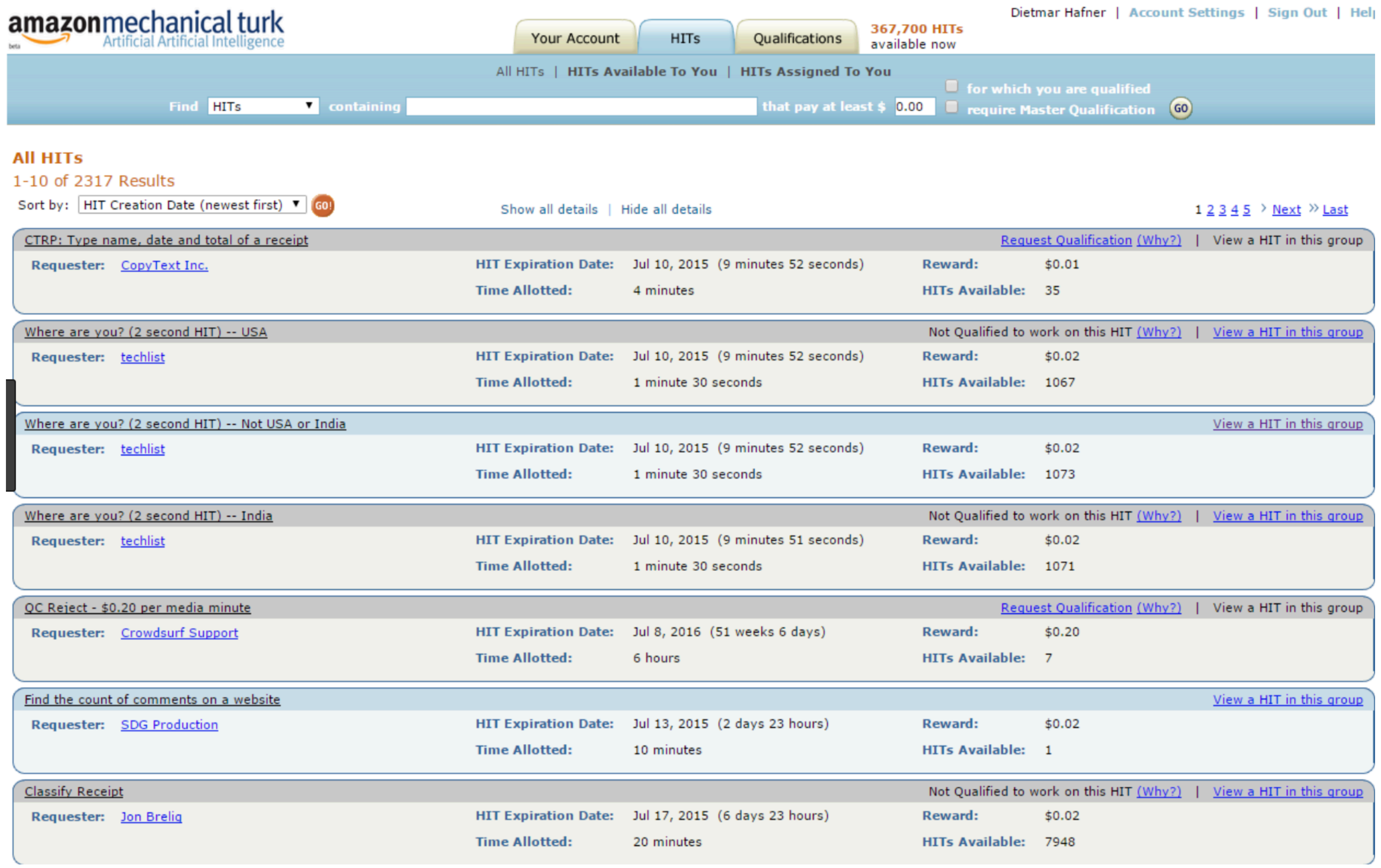
Surveys

Field experiments

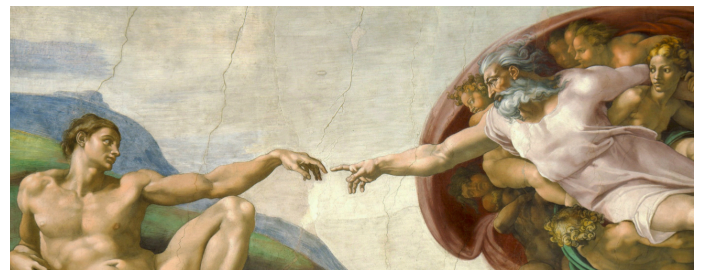
Experiments

Human computation

- Online crowdsourcing platforms allow dividing work into microtasks
- Human-in-the-loop computing, modern-day lab studies, mass collaboration to build big resources (Wikipedia etc.)



Ways of doing computational social science



Observational analyses

Human computation

Natural experiments

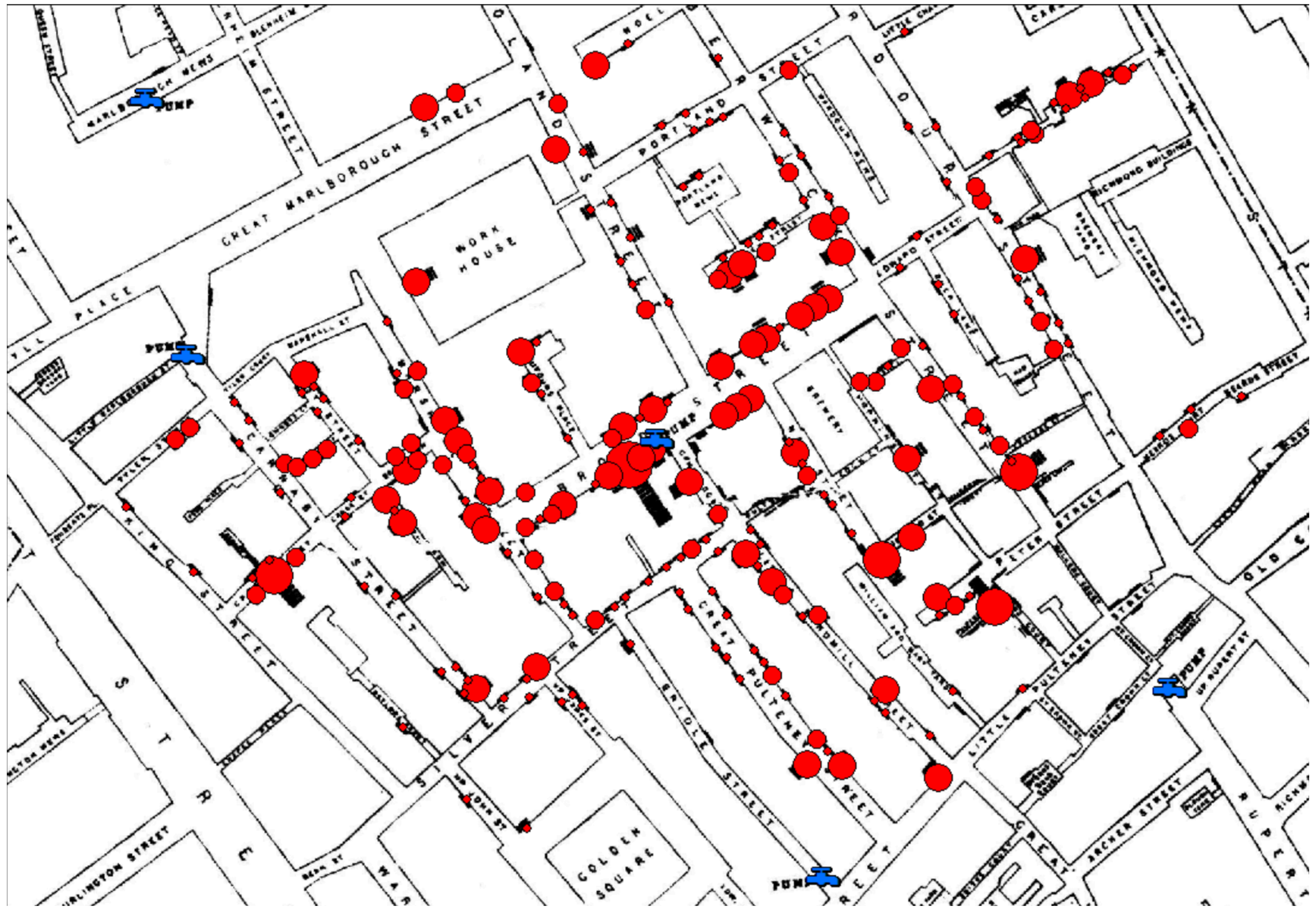
Surveys

Field experiments

Experiments

Natural experiments

Sometimes observational data has some random component you can exploit, and analyze as a “natural” experiment

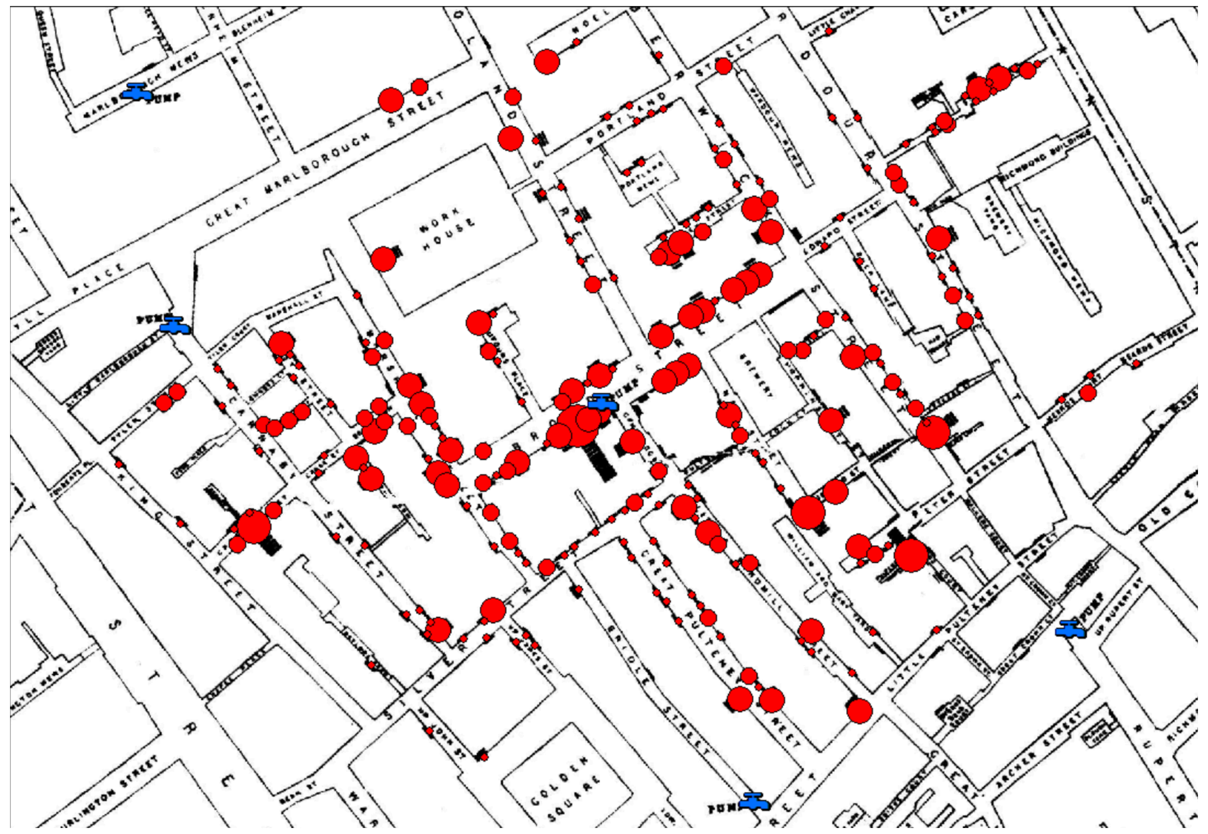


Cholera outbreak in London in 1850s

Natural experiments

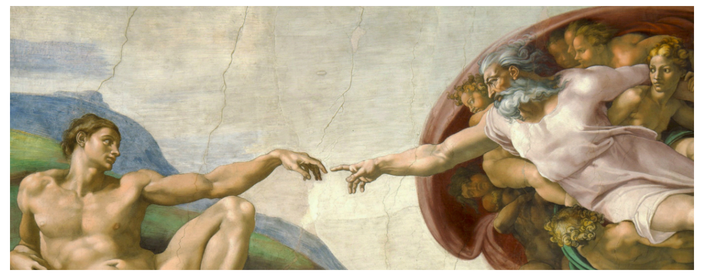
- Physician John Snow produced a map suggesting particular water was the culprit
- Two main water suppliers: one from downstream Thames where raw sewage was dumped in the water (high attack rates), and one from upstream (low attack rates)
- Which supplier you had was arbitrary (varied even within same house, same neighbourhood, etc.)
- Exposure to polluted water was as-if random

Now: in large datasets, more opportunities to identify and argue for as-if random assignment



Cholera outbreak in London in 1850s

Ways of doing computational social science



Observational analyses

Human computation

Natural experiments

Surveys

Field experiments

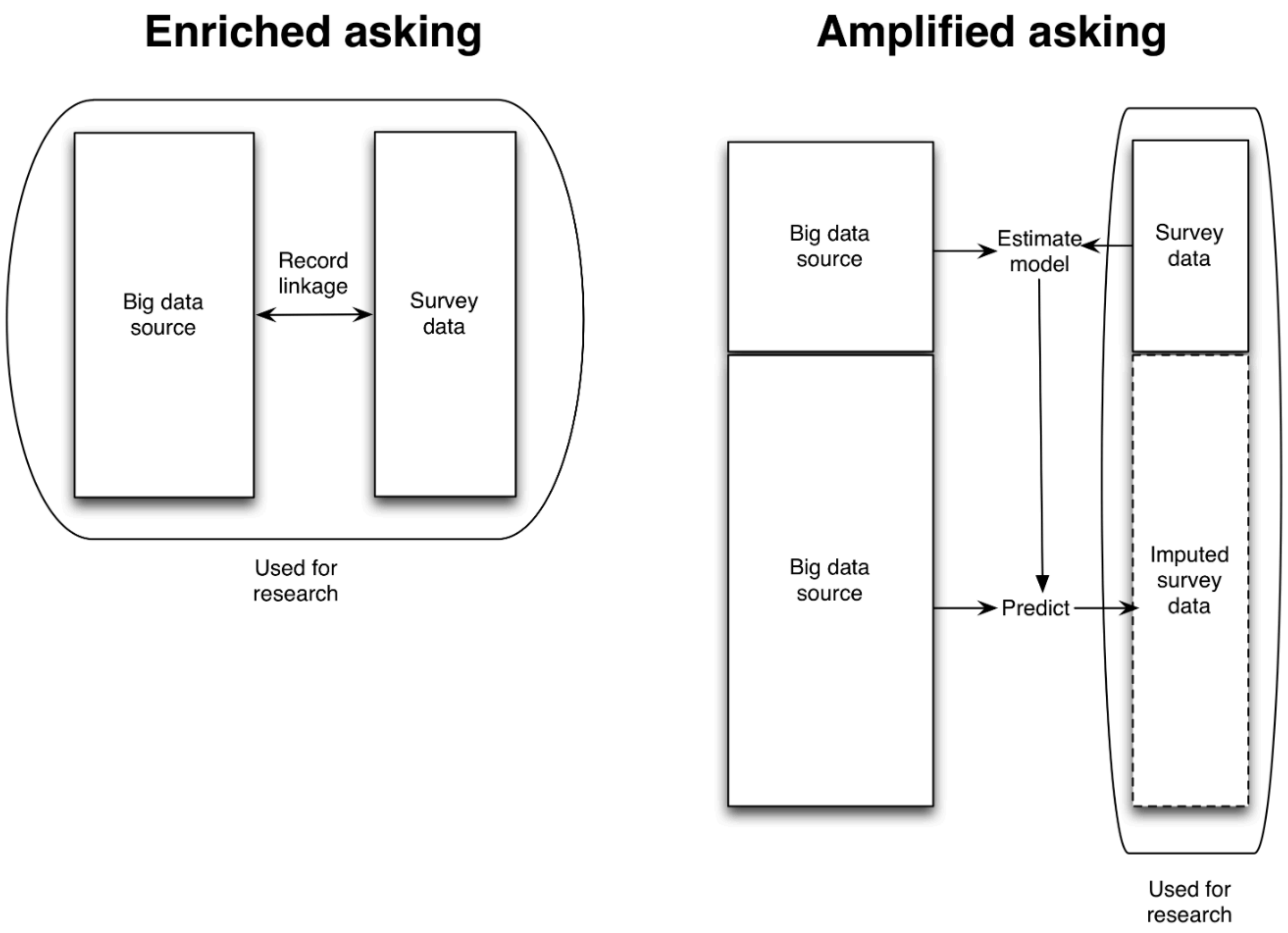
Experiments

Surveys: asking questions

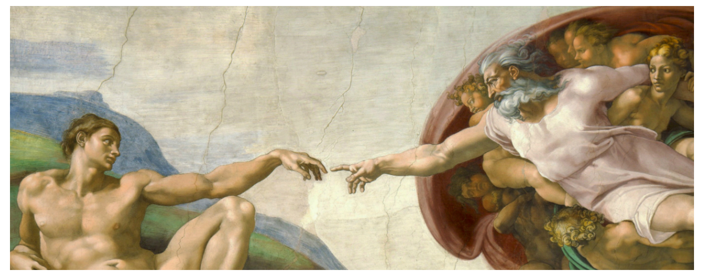
Social research has a unique advantage: we can ask our subjects what they're thinking!

Still the best way to learn the answer to many questions

In the digital era, there are new ways of asking questions



Ways of doing computational social science



Observational analyses

Human computation

Natural experiments

Surveys

Field experiments

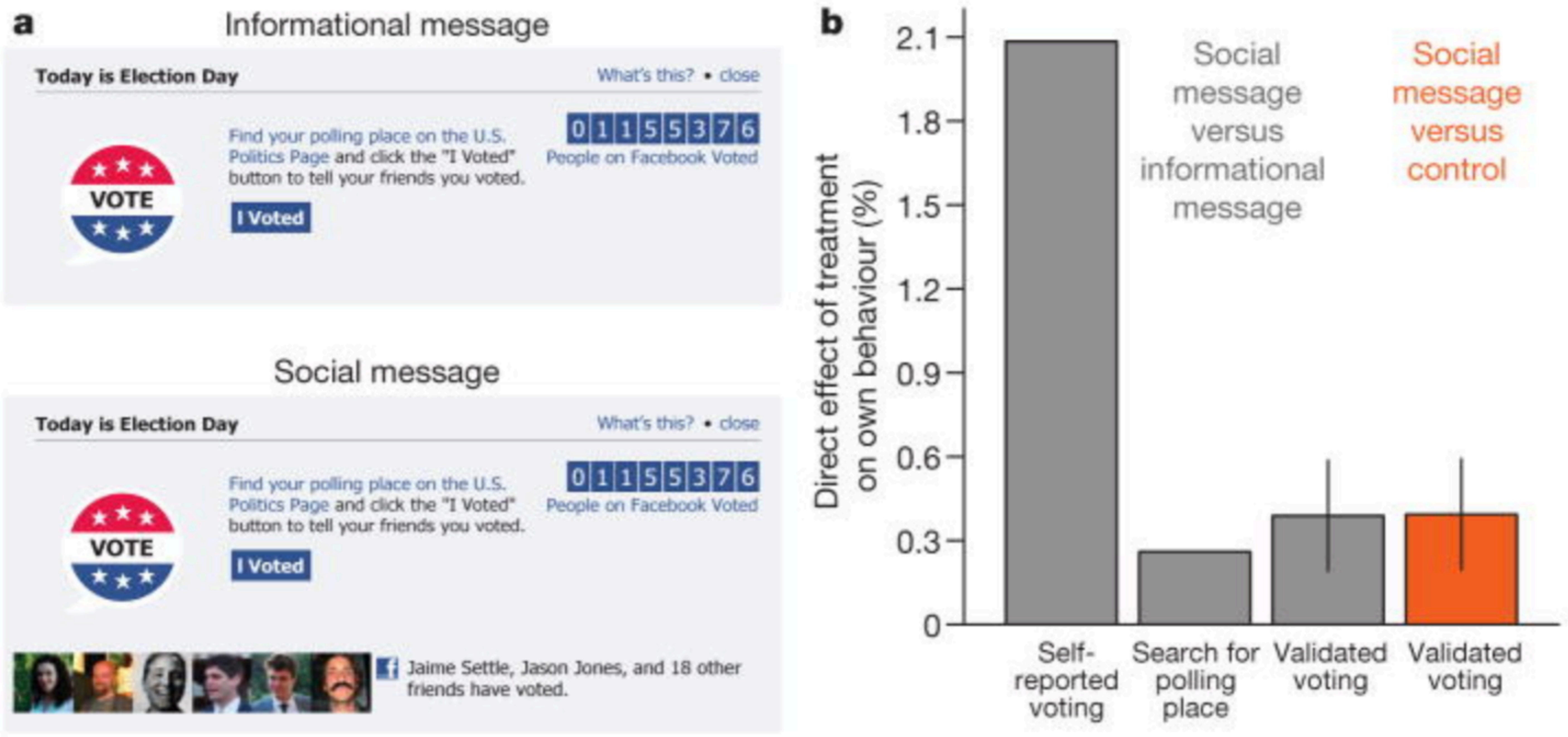
Experiments

Field experiments

- Introducing a treatment into a real system
- Much more possible now with algorithmic systems

Voting experiment on Facebook

Figure 1



The experiment and direct effects

a, **b**, Examples of the informational message and social message Facebook treatments (**a**) and their direct effect on voting behaviour (**b**). Vertical lines indicate s.e.m. (they are too small to be seen for the first two bars).

~300,000 more validated votes

AI & Society: Algorithmic decision-making

St. George's Hospital in the UK developed an algorithm to sort medical school applicants. Algorithm trained to mimic past admissions decisions made by humans.

But past decisions were biased against women and minorities. It codified discrimination.

Web search ads for “Kristen Haring”

Ads by Google

We Found:Kristen Haring

1) Contact **Kristen Haring** - Free Info! 2) Current Phone, Address & More.

www.peoplesmart.com/**Kristen**

- Search by Phone
- Background Checks
- Public Records
- Search by Email
- Search by Address
- Criminal Records

Kristen Haring

Public Records Found For: **Kristen Haring**. Search Now.

www.publicrecords.com/

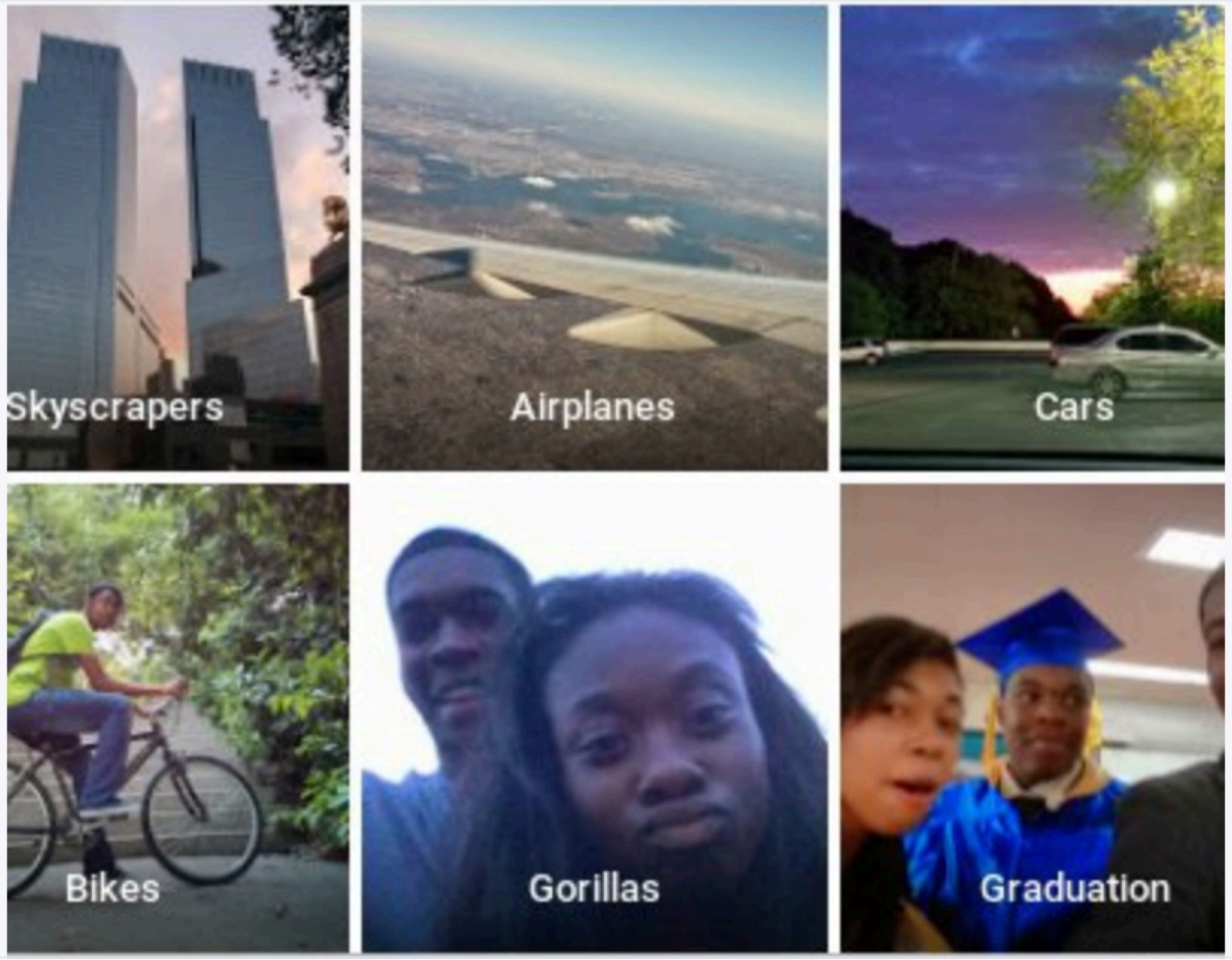
Web search ads for “Latanya Farrell”

Ads related to **latanya farrell**

[Latanya Farrell, Arrested?]
www.instantcheckmate.com/
1) Enter Name and State. 2) Access Full Background Checks Instantly.

[Latanya Farrell]
www.publicrecords.com/
Public Records Found For: **Latanya Farrell**. View Now.

Image labeling gone wrong



Jacky Alciné
@jackyalcine

Google Photos, y'all fucked up. My friend's not a gorilla.

8:22 PM - Jun 28, 2015

226 3,214 2,067

Image searching for "CEO"

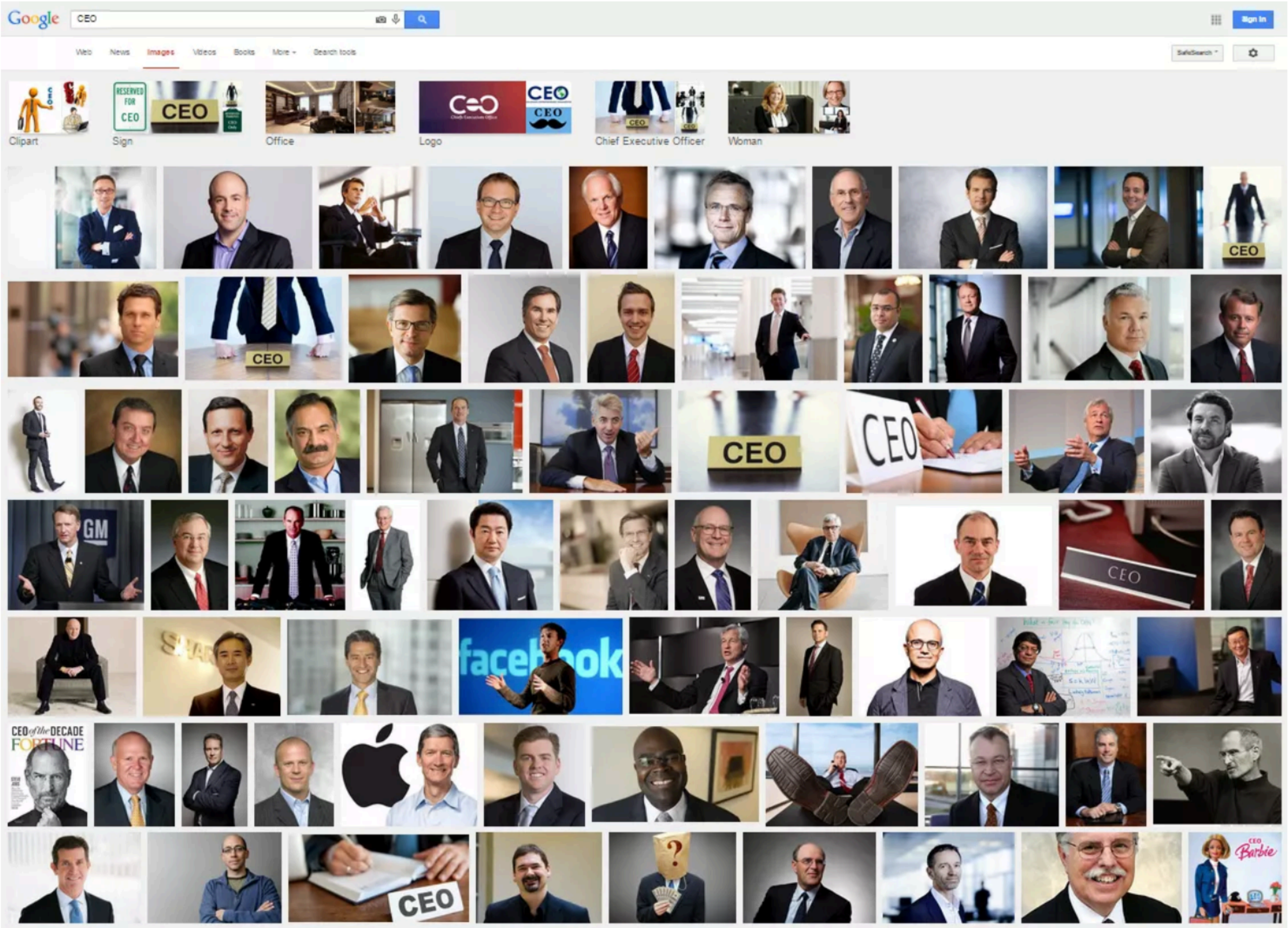


Image searching for "CEO"

Last nail in the coffin: this picture is from an Onion article.

Ethics and privacy

Experimental evidence of massive-scale emotional contagion through social networks

Adam D. I. Kramer[a,1], Jamie E. Guillory[b,2], and Jeffrey T. Hancock[b,c]

Facebook's Users Outraged Over Emotion Experiment

Facebook reveals news feed experiment to control emotions

Facebook emotion experiment sparks criticism

Facebook Tinkers With Users' Emotions in News Feed Experiment, Stirring Outcry

Facebook conducted secret psychology experiment on users' emotions

Everything We Know About Facebook's Secret Mood Manipulation Experiment

Computational social science

Game-changing opportunity to improve our understanding of human behaviour and have positive societal impact.

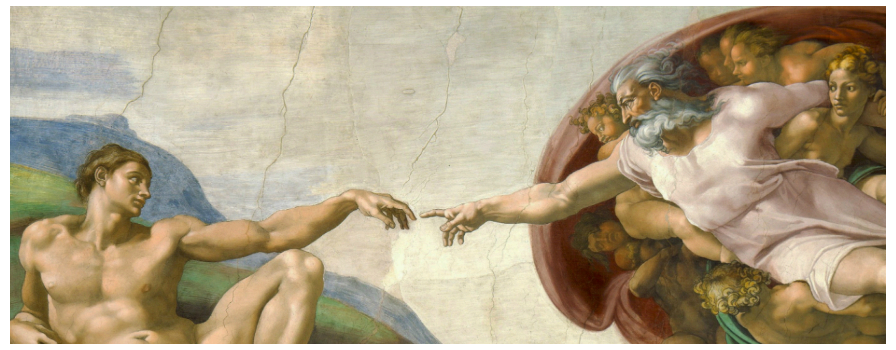
Doing so requires addressing serious technical, scientific, and ethical challenges.

Computational social science in 7 easy pieces

Week	Date	Topic	Reviews Due	Textbook Readings
1	9/7	Introduction to computational social science		Ch. 1
2	9/14	Introduction to computational social science cont'd		Ch. 1
★ 3	9/21	Observational studies 1	9/20 9:00pm	Ch. 2
★ 4	9/28	Observational studies 2	9/27 9:00pm	Ch. 2
★ 5	10/5	Experiments 1	10/4 9:00pm	Ch. 4
6	10/12	Project proposals		
★ 7	10/19	Experiments 2	10/18 9:00pm	Ch. 4
★ 8	10/26	Asking questions	10/25 9:00pm	Ch. 3
★ 9	11/2	Deep learning	11/1 9:00pm	
★ 10	11/16	Ethics in computational social science	11/15 9:00pm	Ch. 6
11	11/23	Project presentations (Part 1)		
12	11/30	Project presentations (Part 2)		

Readymades

Custommades



Observational studies 1

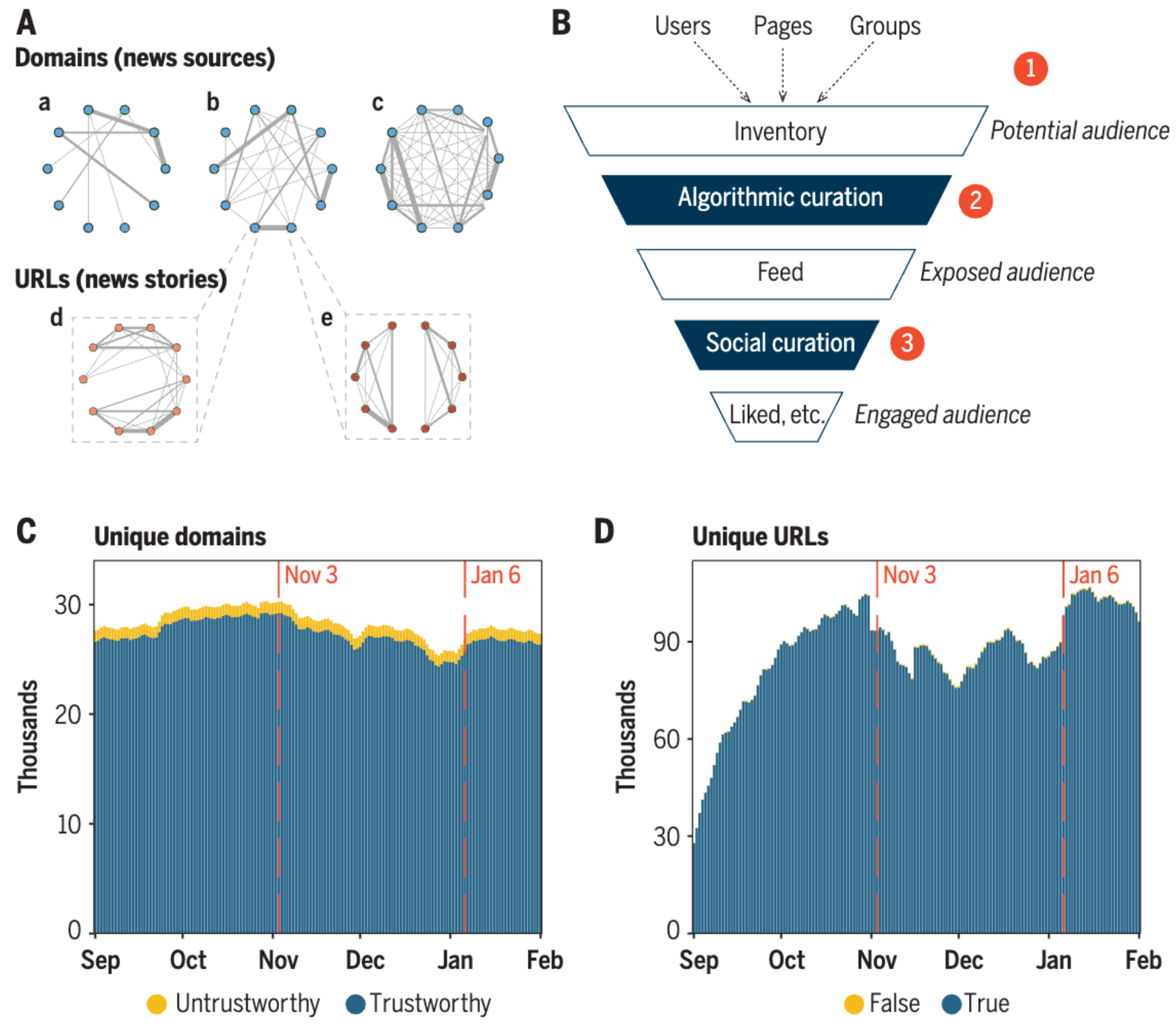
SOCIAL MEDIA

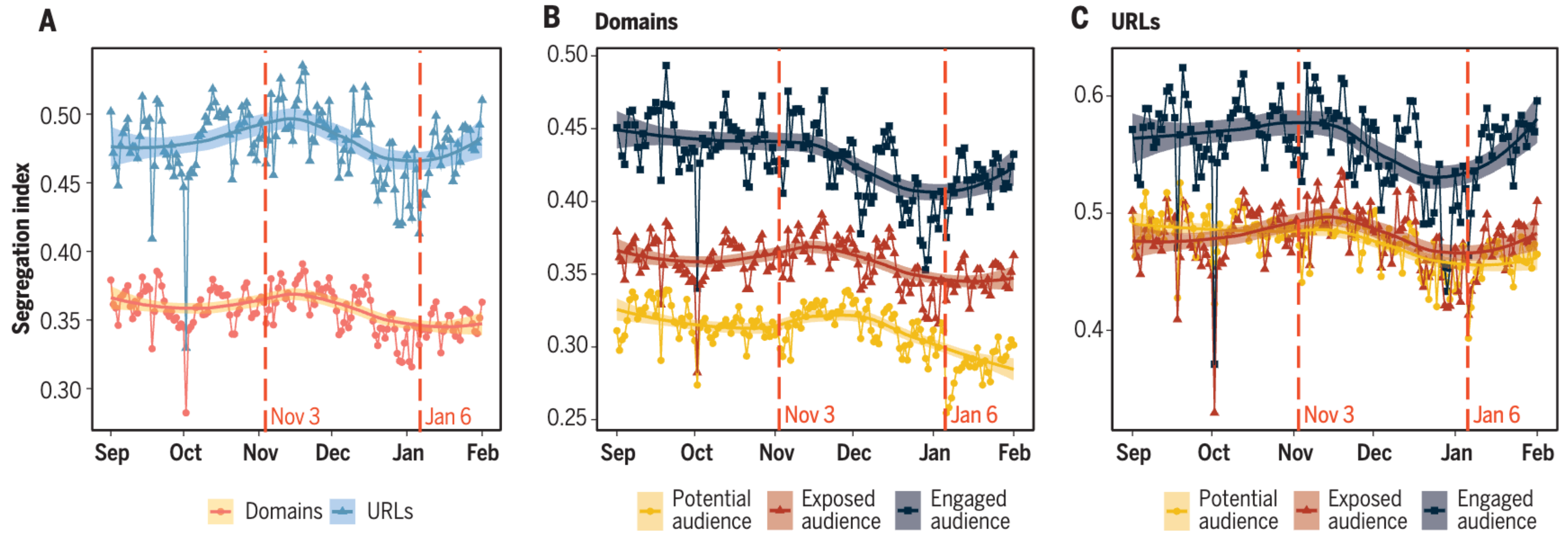
Asymmetric ideological segregation in exposure to political news on Facebook

Sandra González-Bailón[1]*, David Lazer[2], Pablo Barberá[3], Meiqing Zhang[3], Hunt Allcott[4], Taylor Brown[3], Adriana Crespo-Tenorio[3], Deen Freelon[1], Matthew Gentzkow[5], Andrew M. Guess[6], Shanto Iyengar[7], Young Mie Kim[8], Neil Malhotra[9], Devra Moehler[3], Brendan Nyhan[10], Jennifer Pan[11], Carlos Velasco Rivera[3], Jaime Settle[12], Emily Thorson[13], Rebekah Tromble[14], Arjun Wilkins[3], Magdalena Wojcieszak[15,16], Chad Kiewiet de Jonge[3], Annie Franco[3], Winter Mason[3], Natalie Jomini Stroud[17,18], Joshua A. Tucker[19,20]

Science, 2023

Analysis of news exposure during US 2020 election for 208M US Facebook users





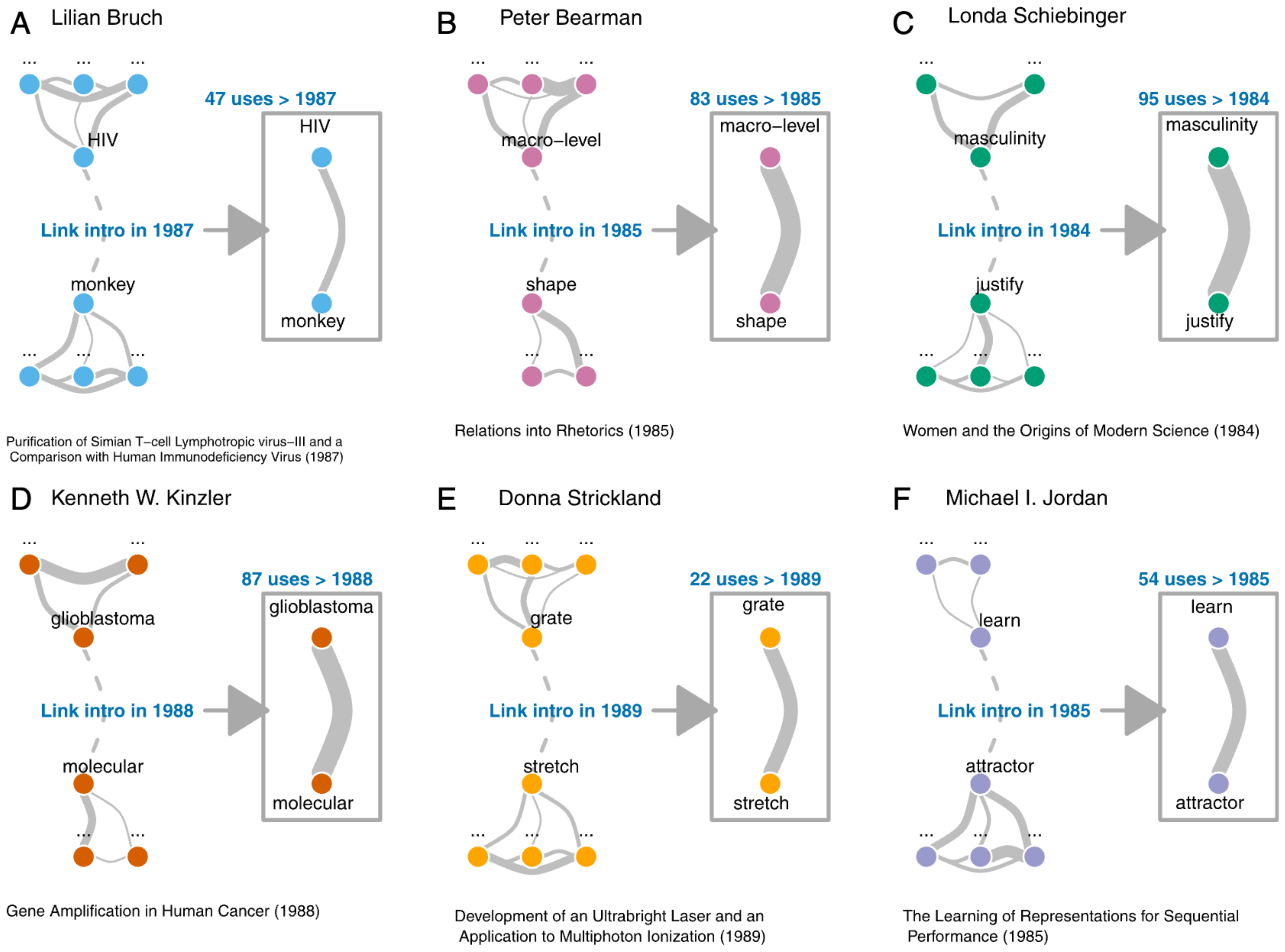
Observational studies 1

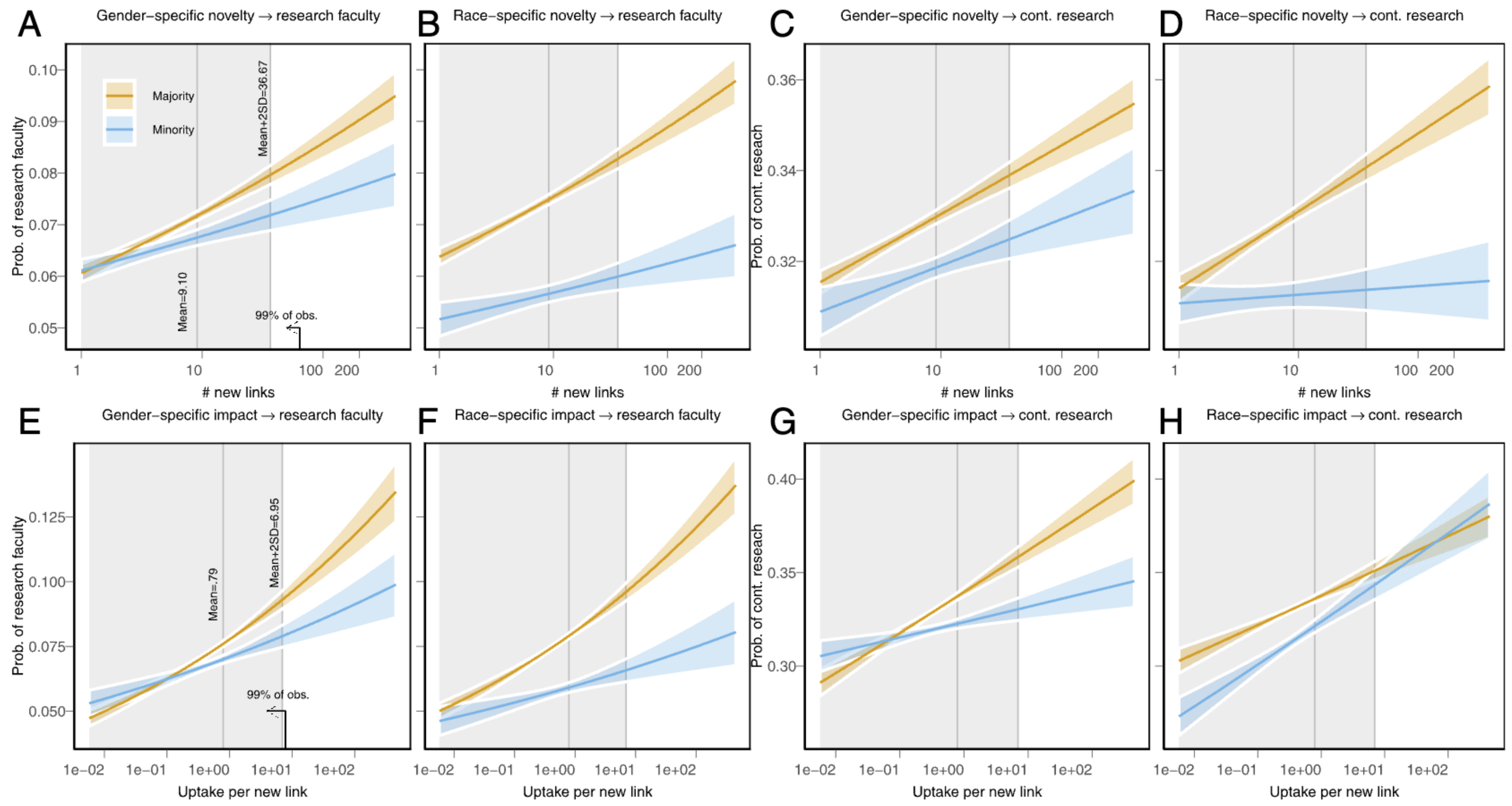
The Diversity–Innovation Paradox in Science

Bas Hofstra[a,1], Vivek V. Kulkarni[b], Sebastian Munoz-Najar Galvez[a], Bryan He[b], Dan Jurafsky[b,c], and Daniel A. McFarland[a,1]

PNAS, 2020

Analysis of all 1.2M US PhD students 1977–2015 on the diversity paradox in science: diversity breeds innovation, yet under-represented groups are less successful





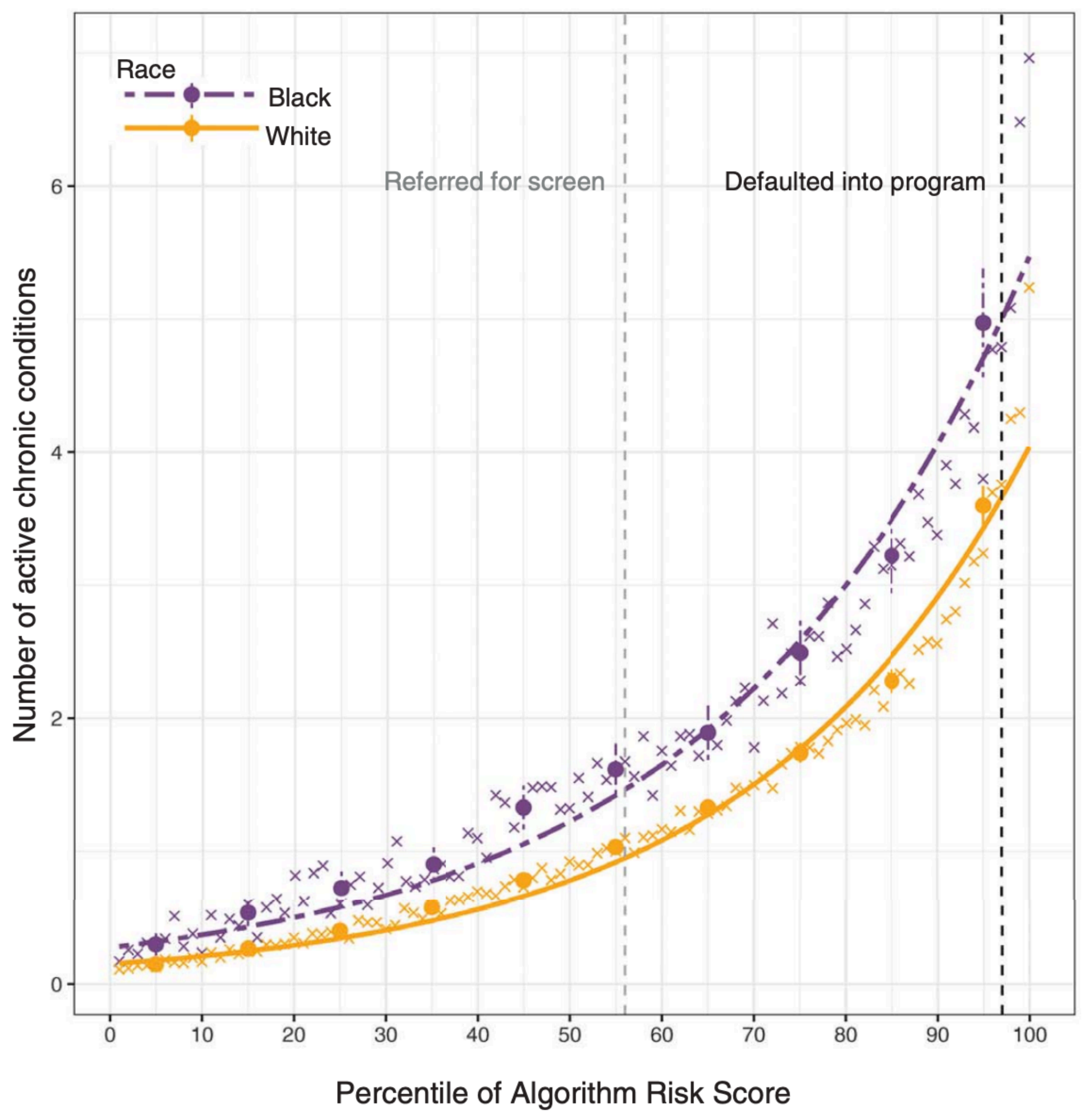
Observational studies 2

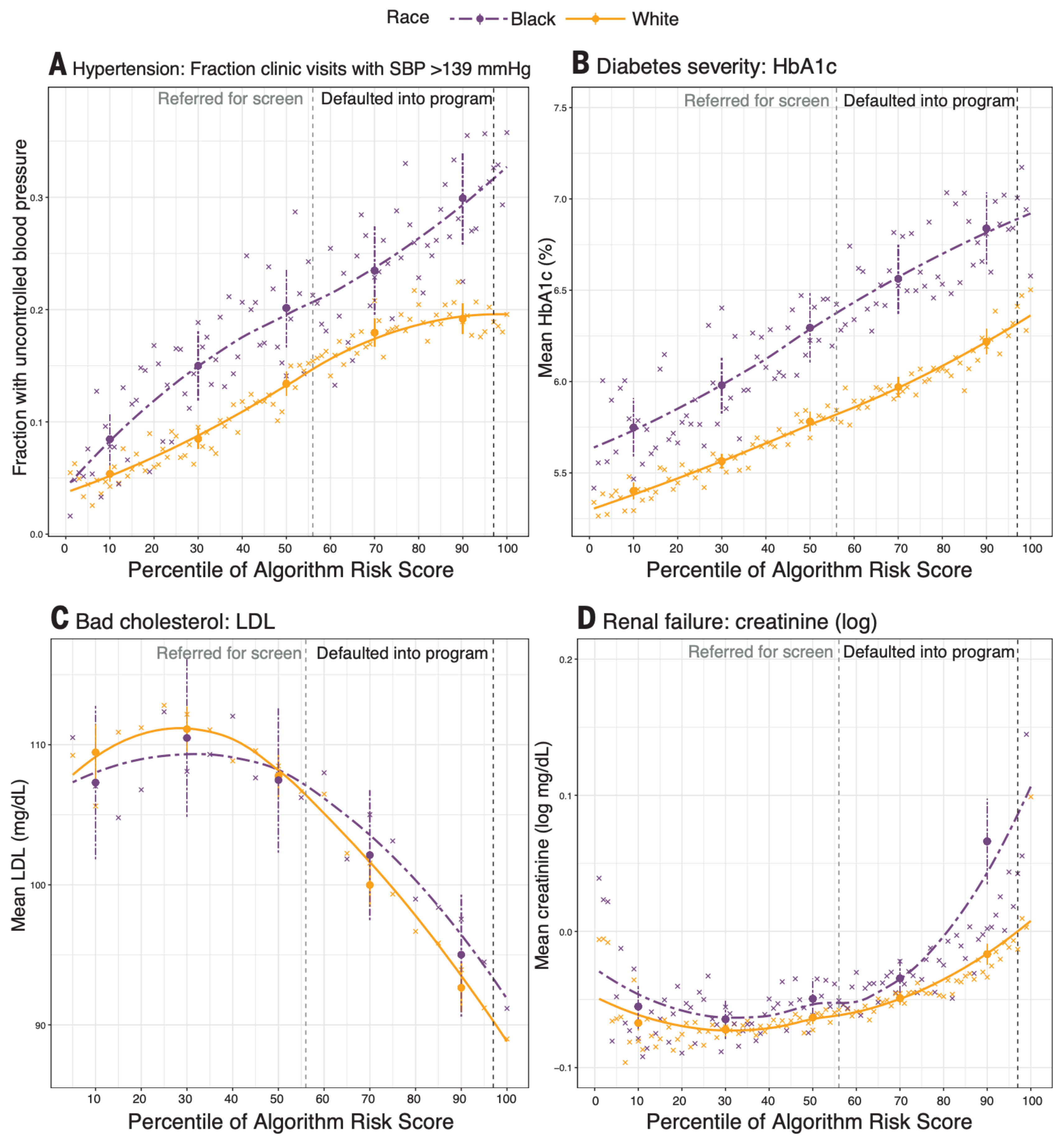
Dissecting racial bias in an algorithm used to manage the health of populations

Ziad Obermeyer[1,2]*, Brian Powers[3], Christine Vogeli[4], Sendhil Mullainathan[5]*†

Science, 2021

Measuring algorithmic bias in a high-stakes health setting





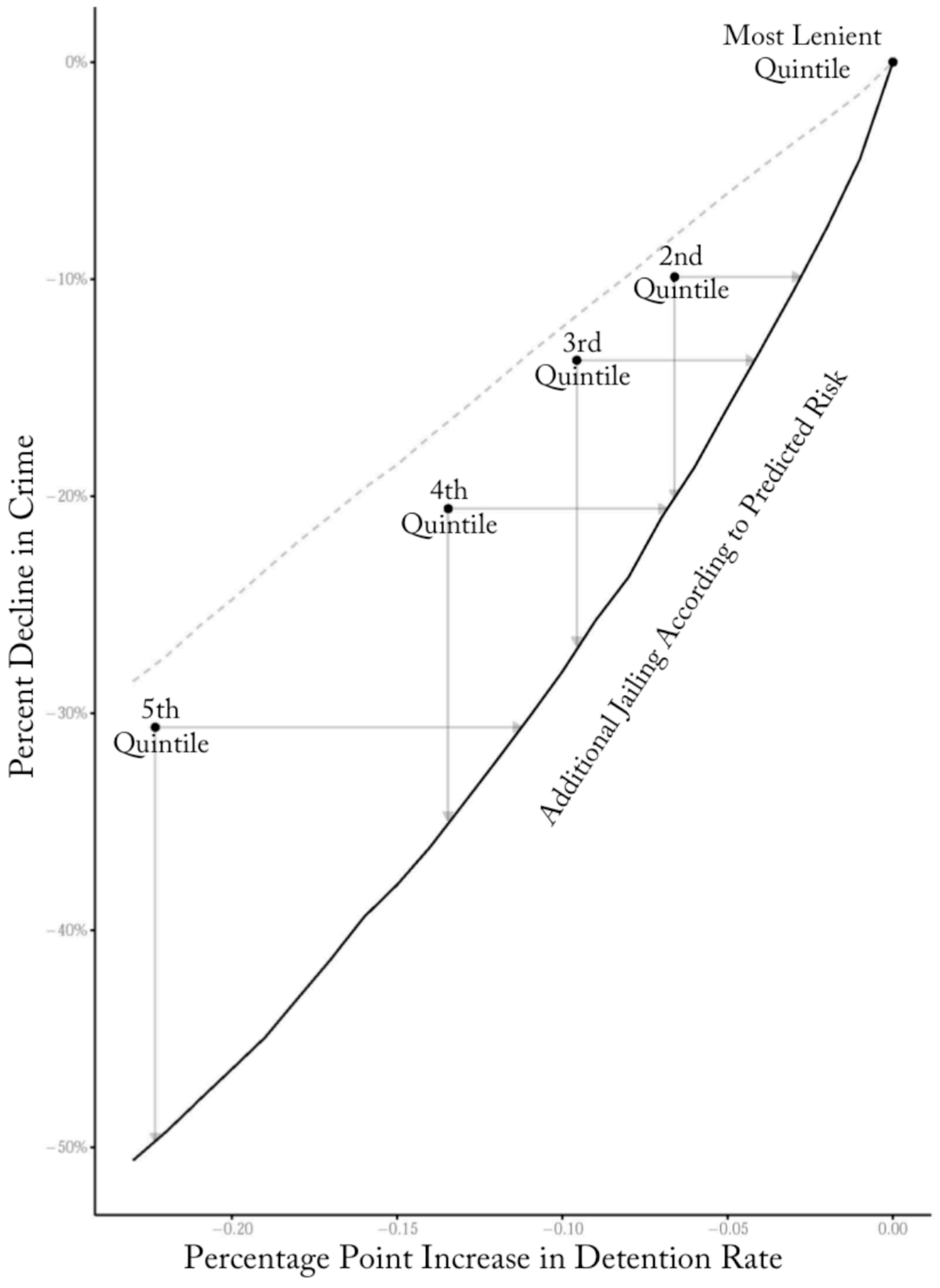
Observational studies 2

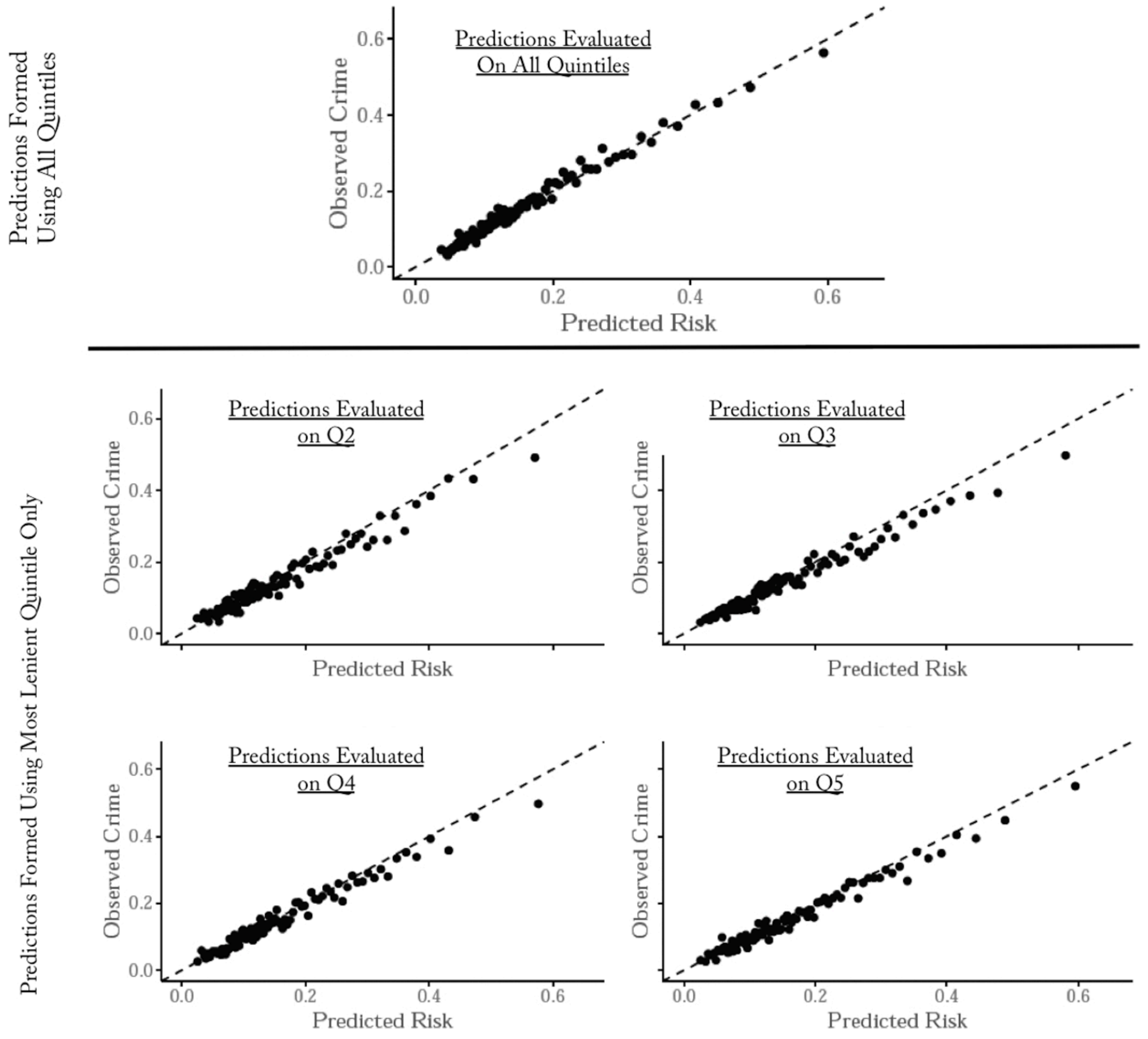
HUMAN DECISIONS AND MACHINE PREDICTIONS*

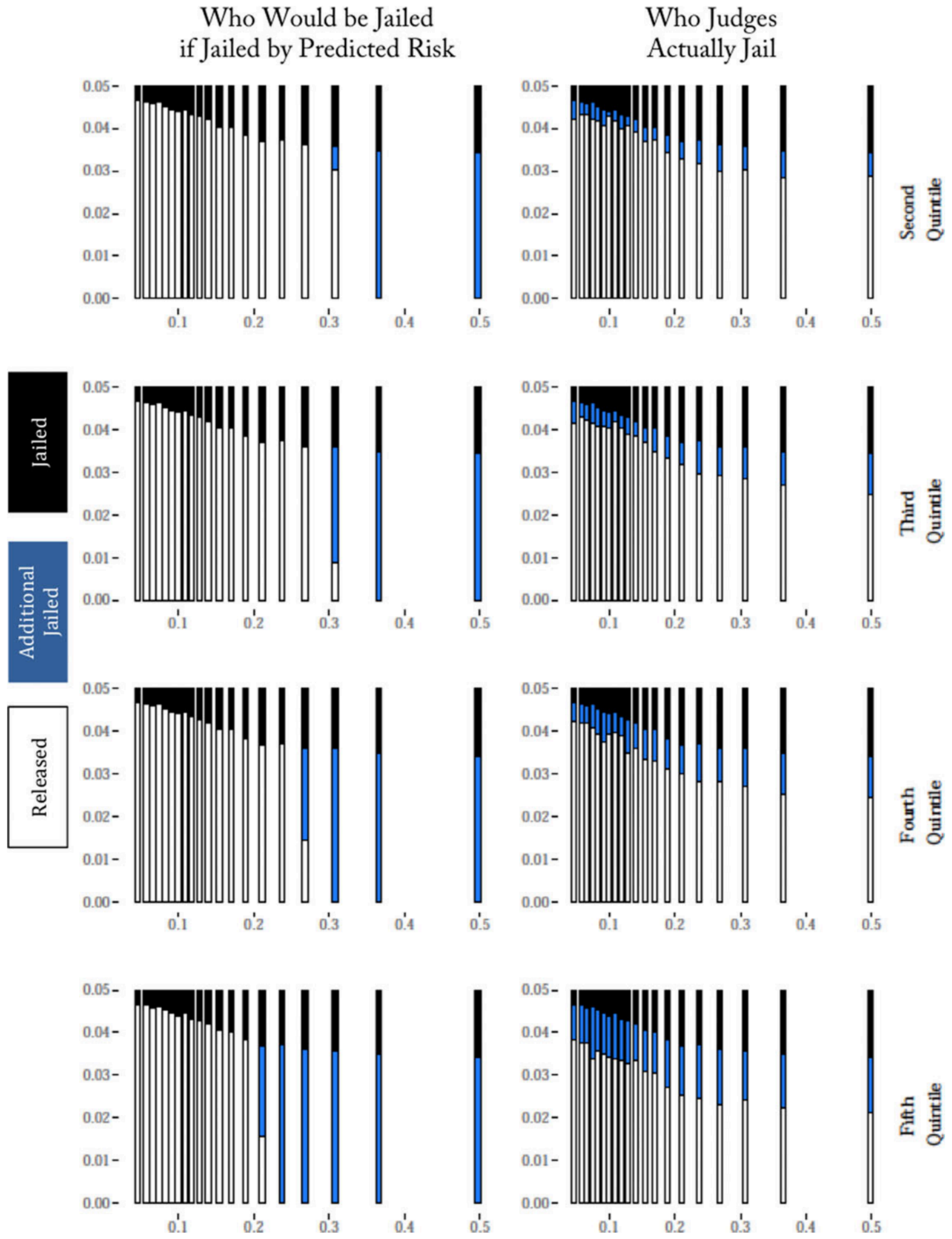
Jon Kleinberg
Himabindu Lakkaraju
Jure Leskovec
Jens Ludwig
Sendhil Mullainathan

Comparing human judges with machine learning on 758K pretrial bail decisions after arrests

Quarterly Journal of Economics, 2017







Experiments 1

Shifting attention to accuracy can reduce misinformation online

Nature, 2021

Why do people share misinformation, and how can we reduce this?

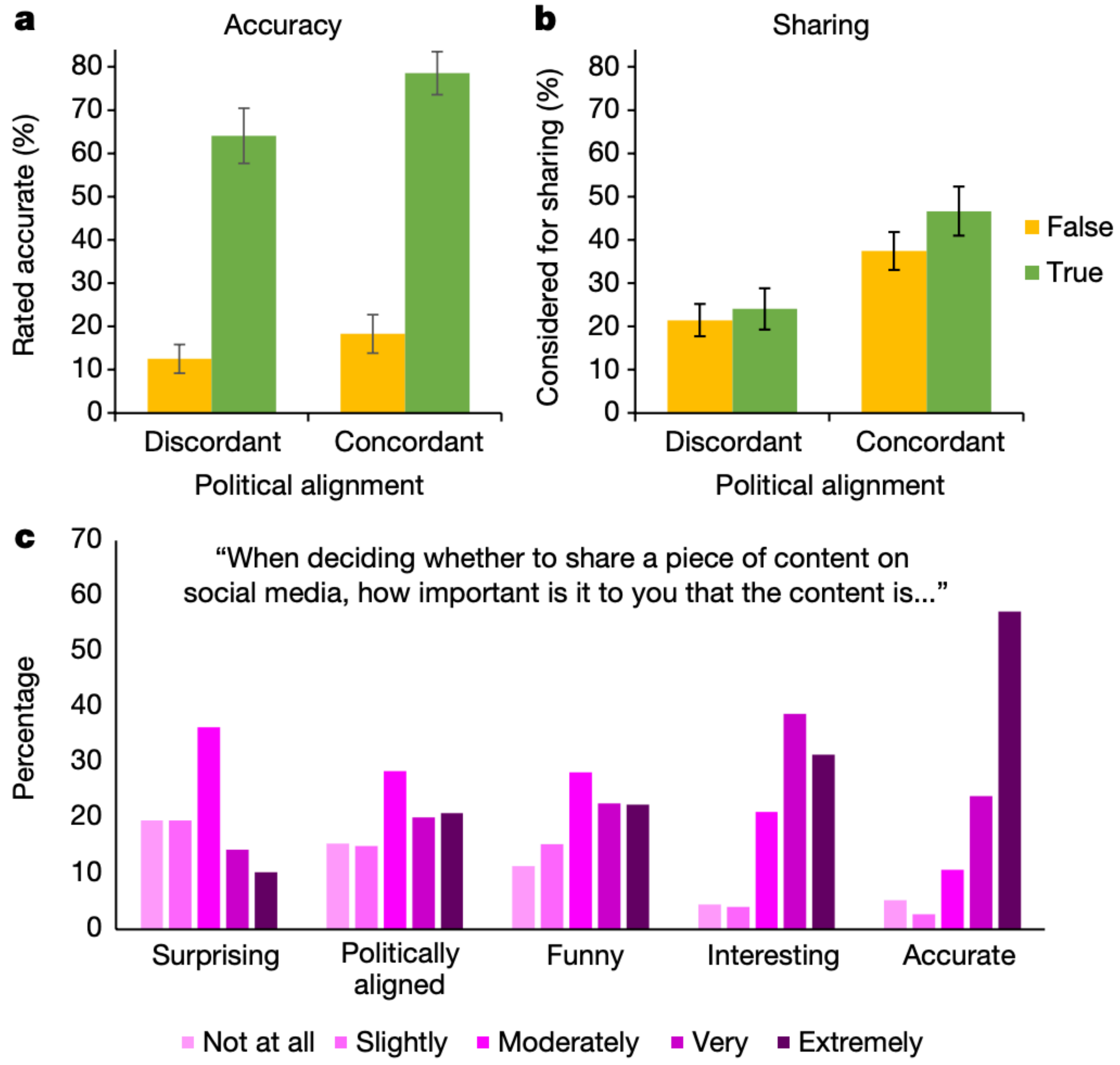
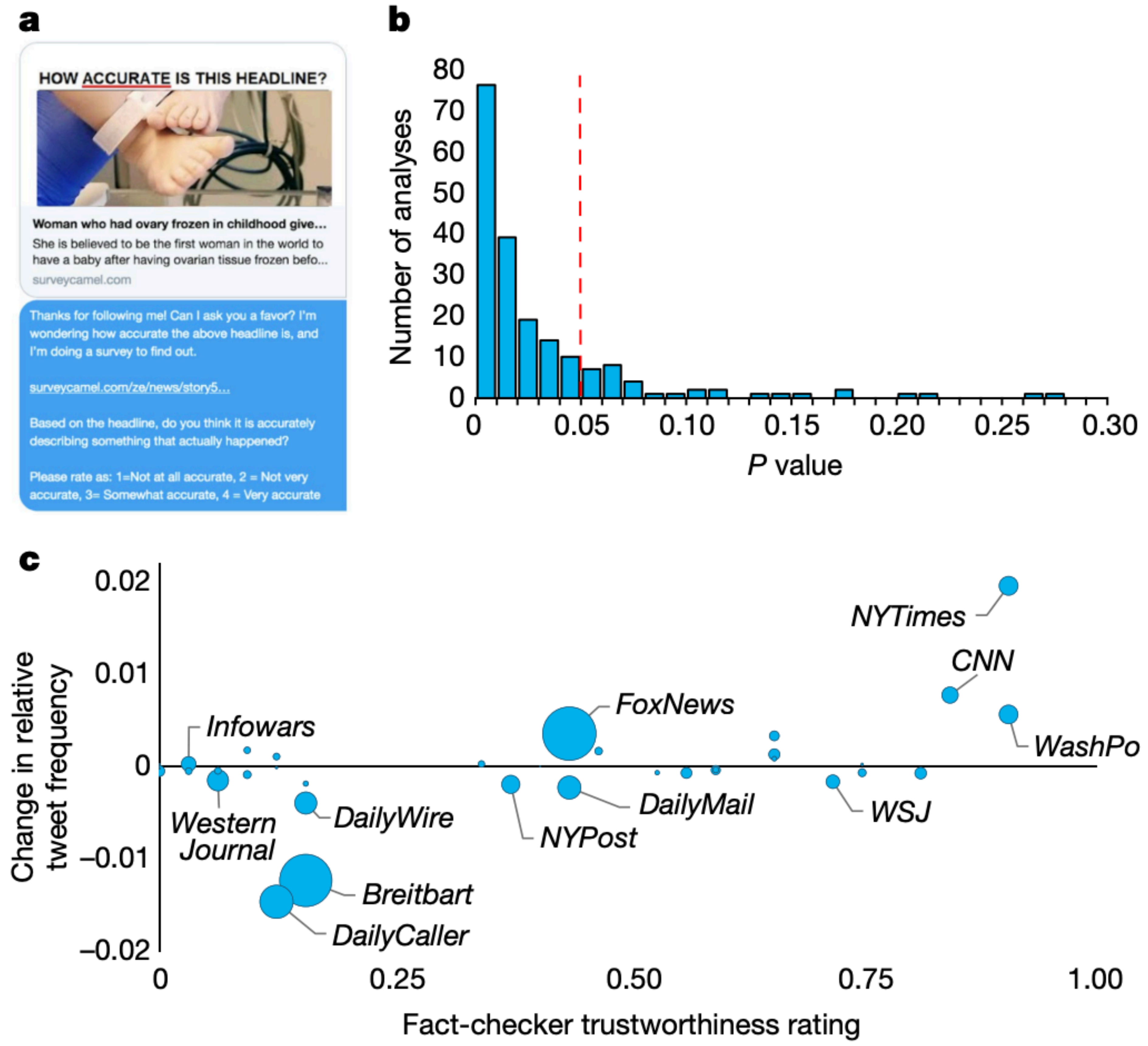


Fig. 1 | Sharing intentions are much less discerning than accuracy judgements—despite an overall desire to share only accurate content. In study



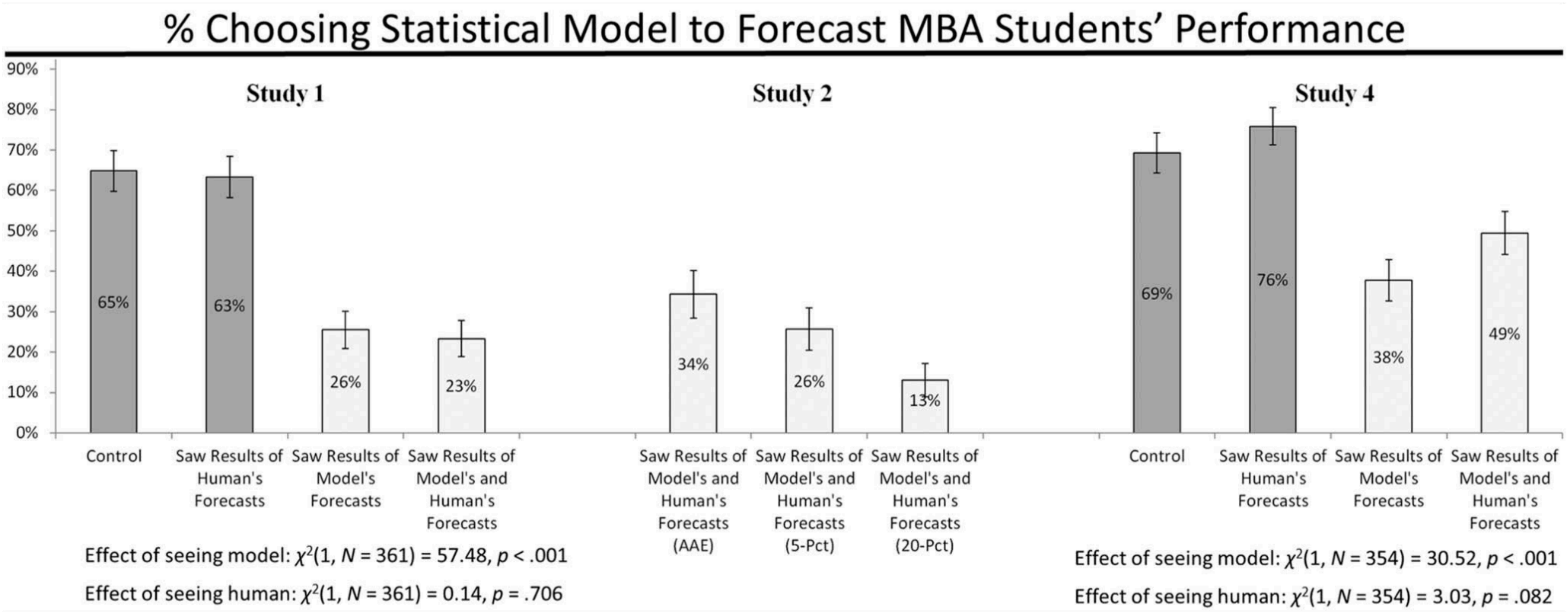
Experiments 1

Algorithm Aversion: People Erroneously Avoid Algorithms After Seeing Them Err

Berkeley J. Dietvorst, Joseph P. Simmons, and Cade Massey
University of Pennsylvania

Journal of Experimental Psychology, 2014

Do people trust algorithms (even when they should)?



Experiments 2

The Welfare Effects of Social Media†

By Hunt Allcott, Luca Braghieri, Sarah Eichmeyer, and Matthew Gentzkow*

American Economic Review, 2020

What are the causal effects of social media on time spent online, political polarisation, and well-being?

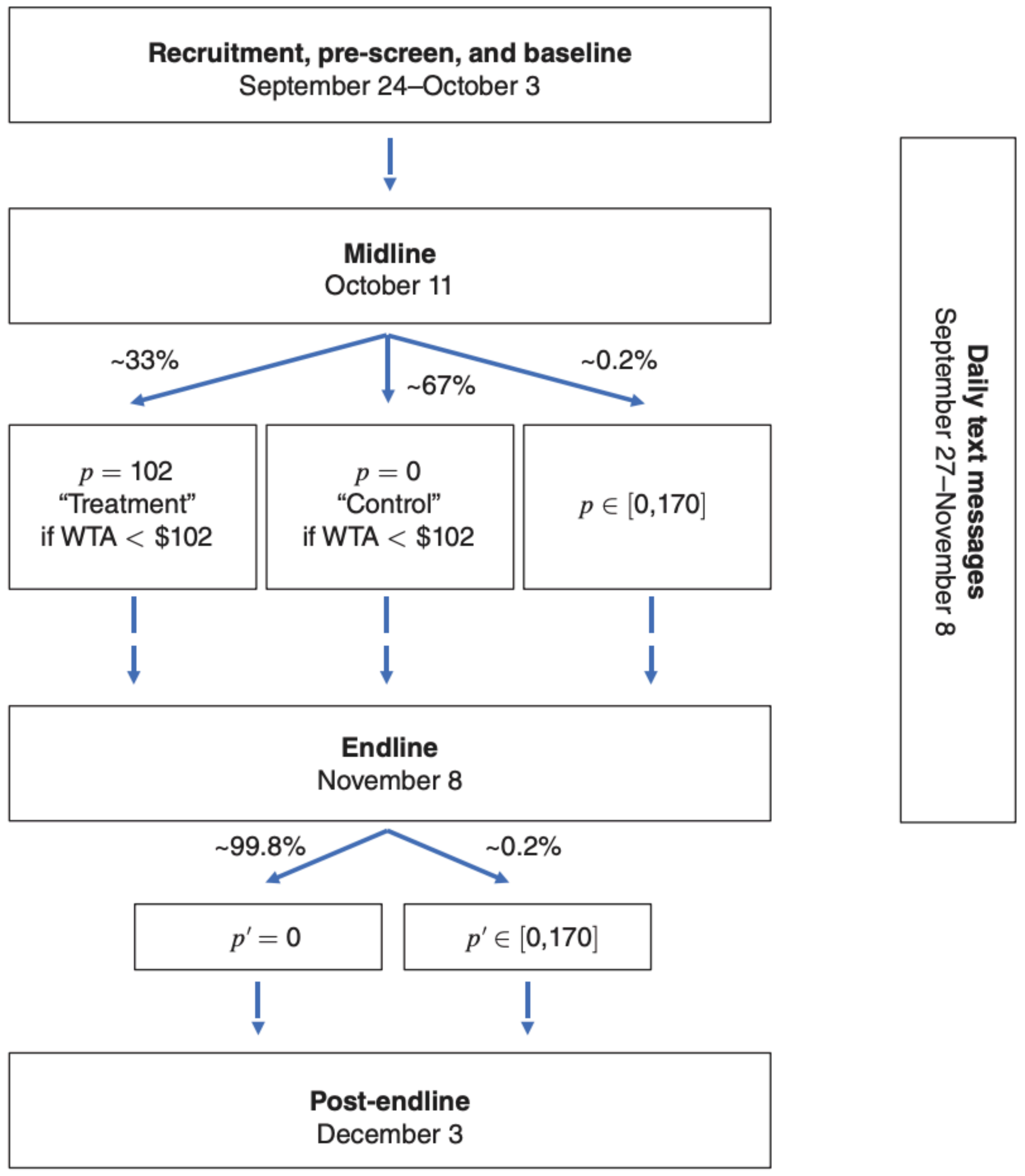


Figure 1. Experimental Design

Table 1—Sample Sizes

Phase	Sample size
Recruitment and baseline	N = 1,892,191 were shown ads
	N = 32,201 clicked on ads
	N = 22,324 completed pre-screen survey
	N = 20,959 were from United States and born between 1900 and 2000
	N = 17,335 had 15 < daily Facebook minutes ≤ 600
	N = 7,455 consented to participate
	N = 3,910 finished baseline
	N = 2,897 had valid baseline and were randomized, of which:
Midline	N = 2,897 began midline
	N = 2,743 received a price offer, of which:
	N = 1,661 were in impact evaluation sample
Endline	N = 2,710 began endline
	N = 2,684 finished endline, of which:
	N = 1,637 were in impact evaluation sample
Post-endline	N = 2,067 reported Facebook mobile app use, of which:
	N = 1,219 were in impact evaluation sample

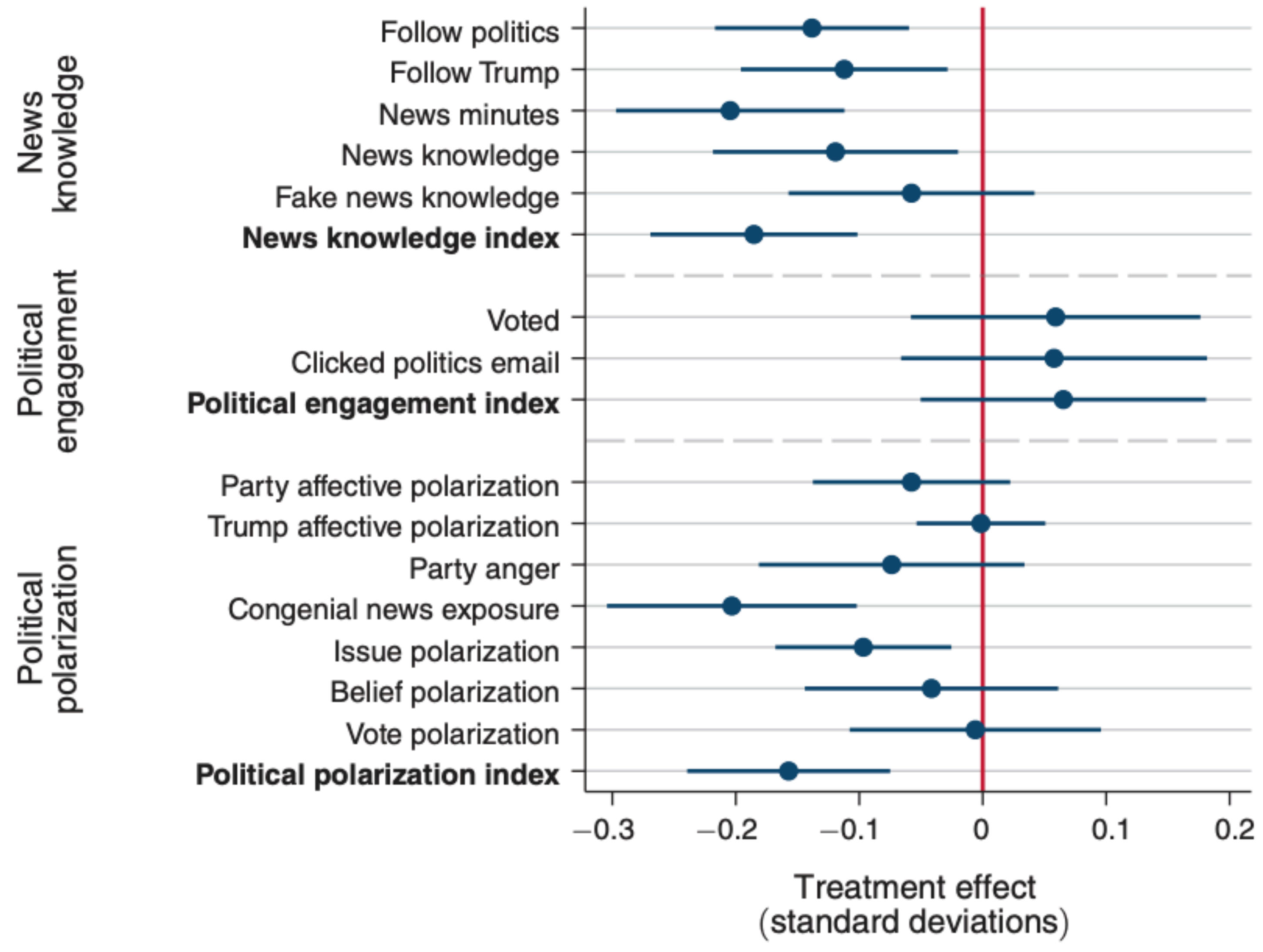


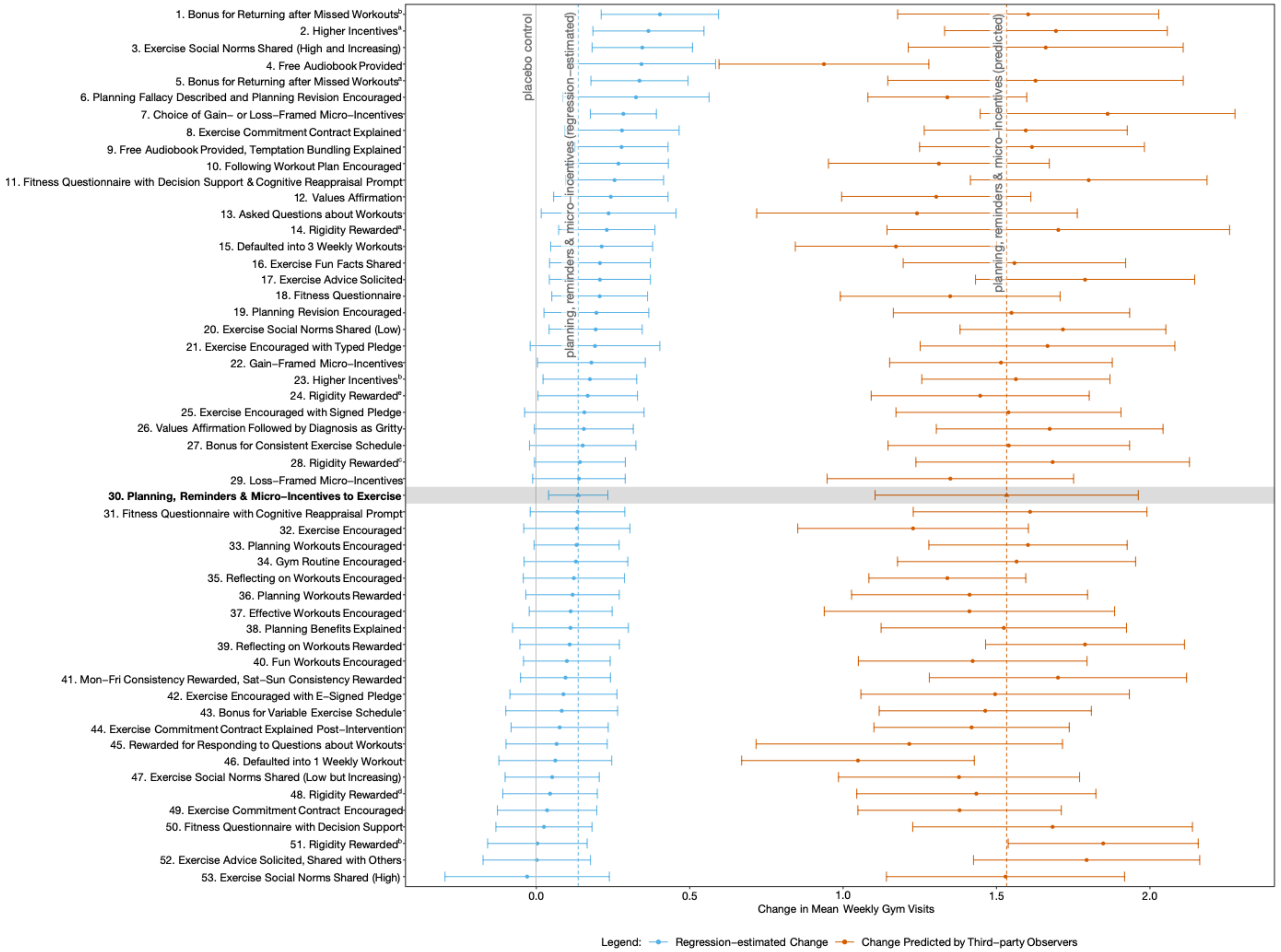
Figure 3. Effects on News and Political Outcomes

Experiments 2

Megastudies improve the impact of applied behavioural science

Nature, 2021

What interventions increase exercise? (And running experiments with many treatments)



Asking questions

Article

Machine learning and phone data can improve targeting of humanitarian aid

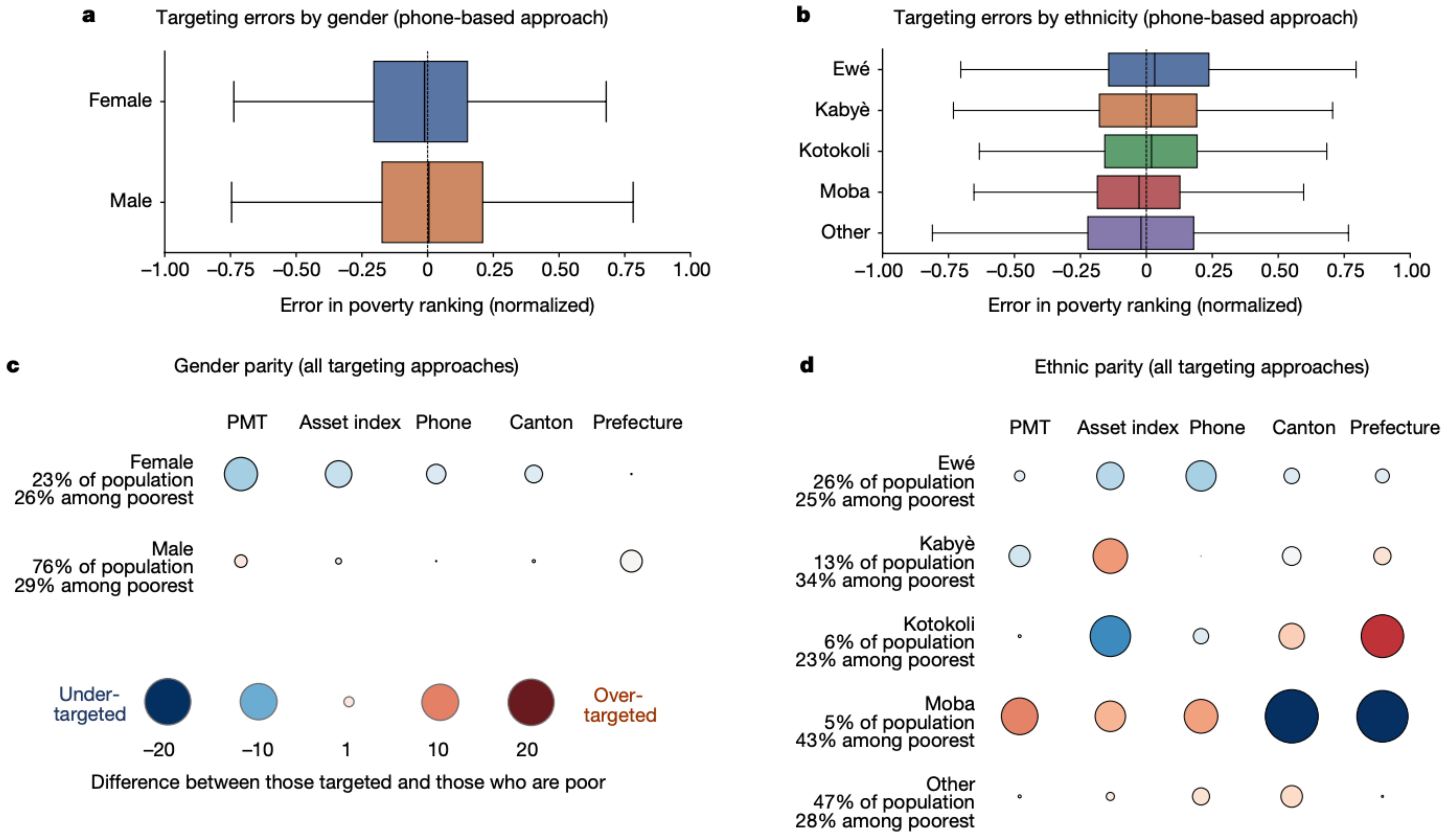
https://doi.org/10.1038/s41586-022-04484-9

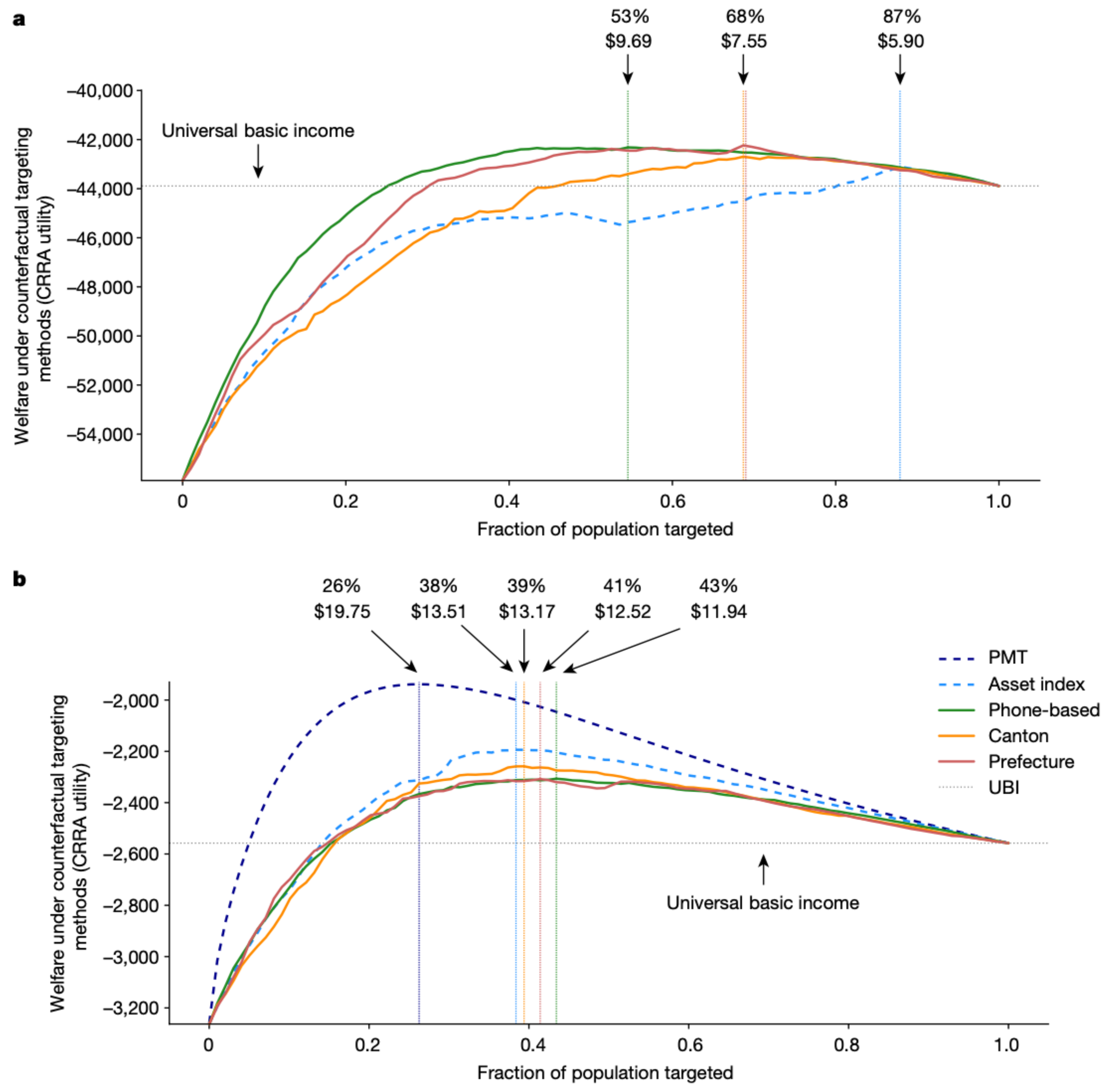
Emily Aiken[1,5], Suzanne Bellue[2], Dean Karlan[3], Chris Udry[4] & Joshua E. Blumenstock[1,5]

Received: 15 July 2021

Nature, 2022

Can we improve aid targeting with amplified asking?





Asking questions

Article

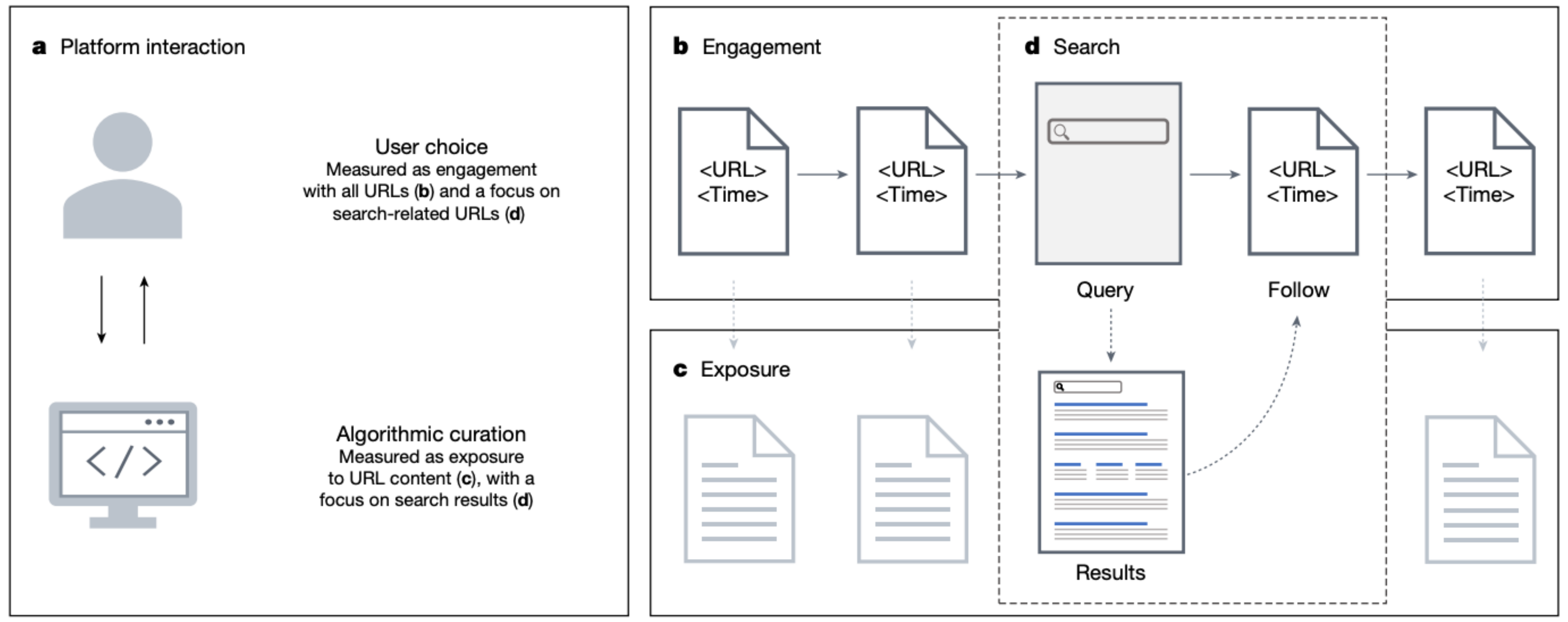
Users choose to engage with more partisan news than they are exposed to on Google Search

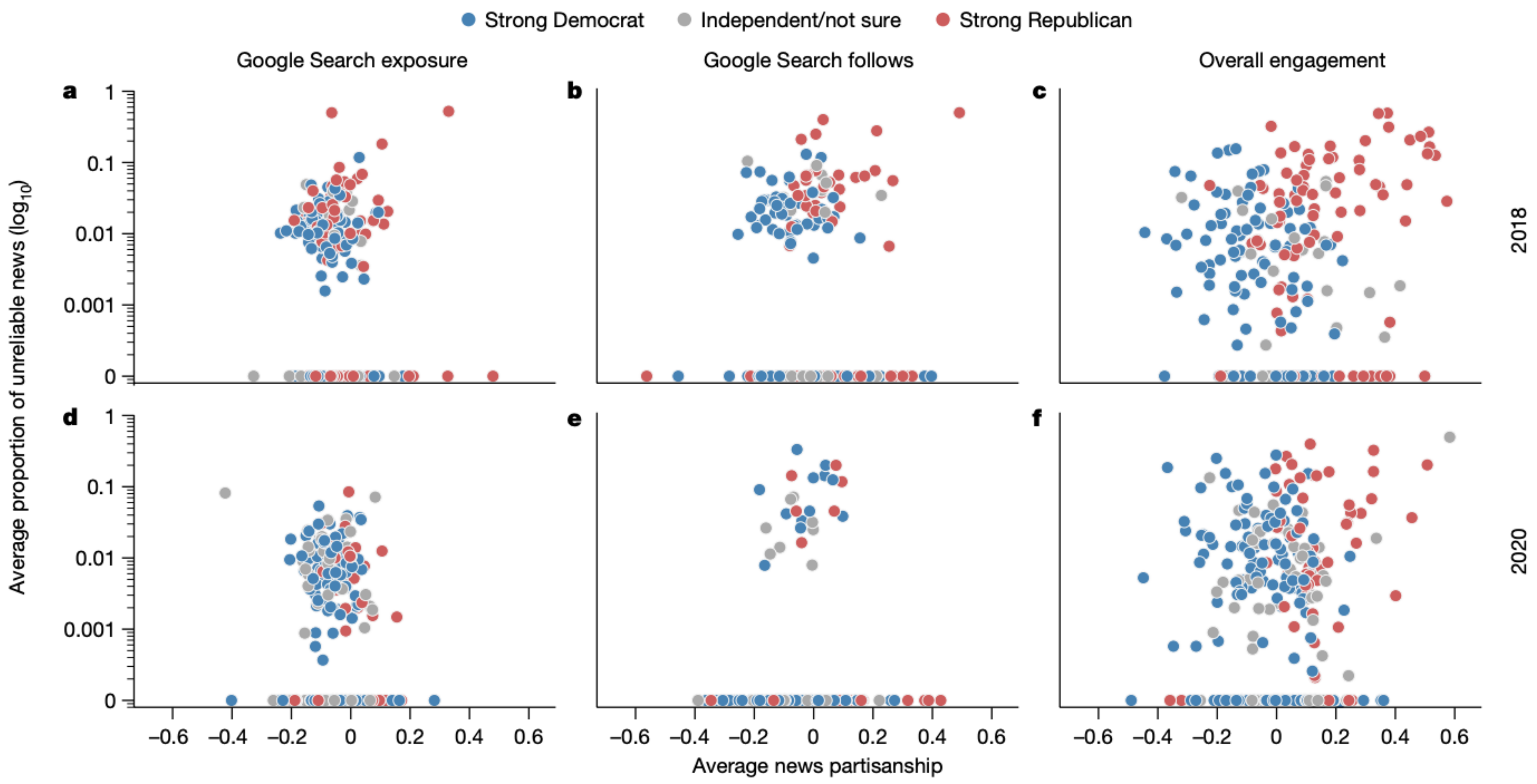
https://doi.org/10.1038/s41586-023-06078-5
Received: 17 February 2022

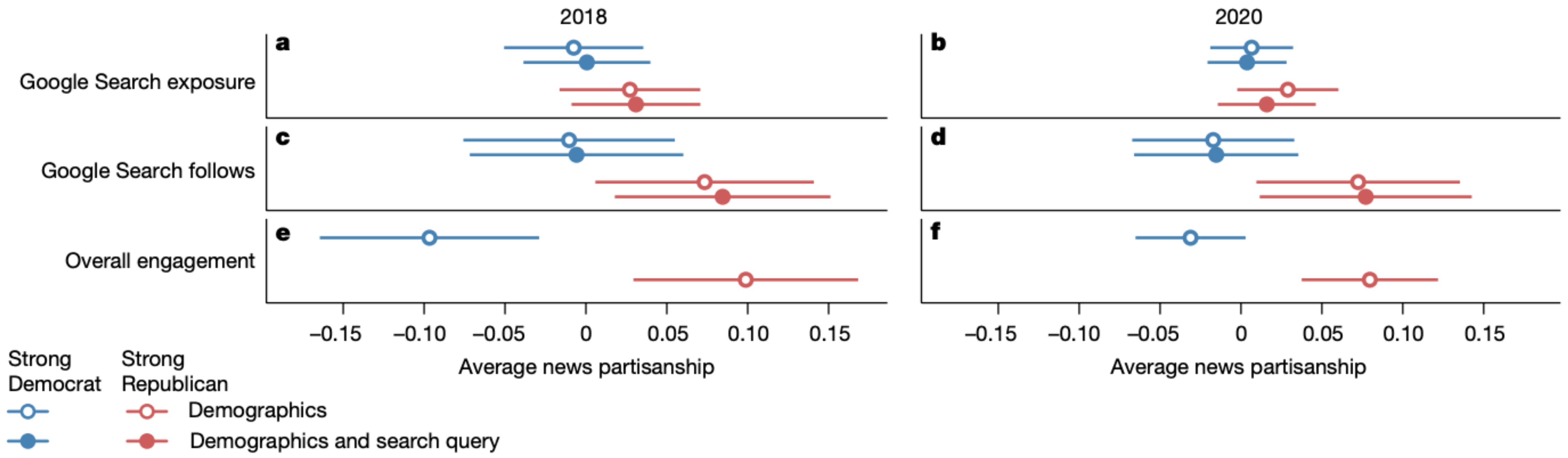
Ronald E. Robertson[1,2]✉, Jon Green[2], Damian J. Ruck[2], Katherine Ognyanova[3], Christo Wilson[2,4] & David Lazer[2]

Is user selection or algorithmic influence a bigger driver of partisan news engagement on Google?

Nature, 2023







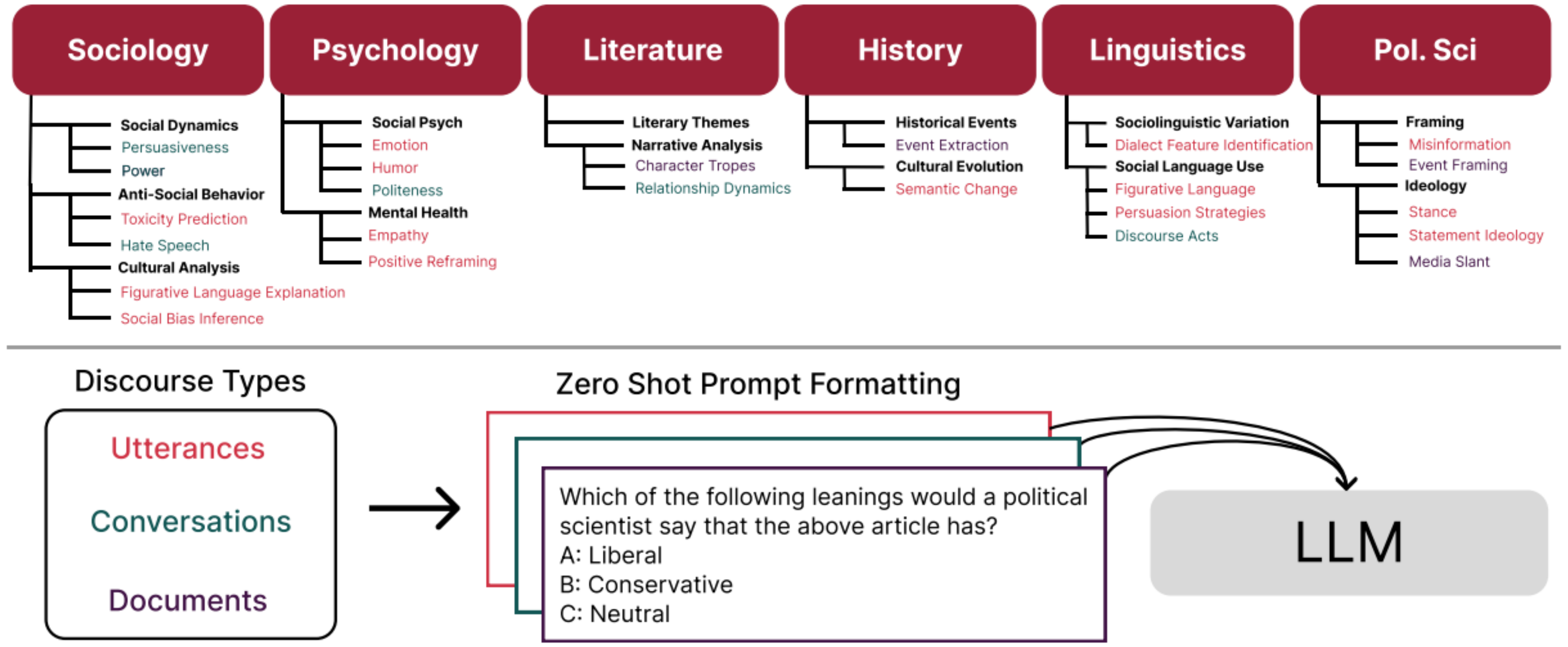
Deep learning

Can Large Language Models Transform Computational Social Science?

Caleb Ziems* William Held* Omar Shaikh* Jiaao Chen*
Zhehao Zhang* Diyi Yang*
Georgia Institute of Technology, Shanghai Jiao Tong University, Stanford University
{cziems, wheld3, jiaaochen}@gatech.edu, zzh12138@sjtu.edu.cn, {oshaikh, diyiy}@stanford.edu

Preprint, 2023

How can we use LLMs to augment CSS?



Model / Data	Baselines		FLAN-T5					FLAN	Chat	text-001				text-002	text-003
	Rand	Finetune	Small	Base	Large	XL	XXL	UL2	ChatGPT	Ada	Babb.	Curie	Dav.	Davinci	Davinci
Utterance Level Tasks															
Dialect	4.5	41.5	1.9	2.3	15.8	16.5	22.6	23.7	15.0	5.3	5.6	6.0	10.9	10.5	16.9
Emotion	16.7	91.7	23.9	65.3	69.1	65.9	66.7	70.3	46.2	44.6	16.1	18.7	19.3	39.8	36.5
Figurative	25.0	94.4	23.6	29.0	25.4	40.2	56.0	64.0	50.2	25.0	24.4	25.0	28.8	52.0	60.6
Humor	50.0	73.1	52.0	51.8	56.2	59.0	50.6	58.8	55.4	55.2	59.0	58.6	50.4	51.4	51.0
Ideology	33.3	61.9	33.1	39.2	48.6	49.2	54.4	48.2	54.8	–	33.3	33.3	34.3	57.6	48.2
Impl. Hate	14.3	69.9	17.7	22.7	17.9	36.3	34.5	35.9	29.7	17.1	18.6	15.7	21.3	22.7	27.1
Misinfo	50.0	82.3	50.0	55.4	69.2	70.2	71.2	77.6	69.0	–	50.4	52.2	52.6	75.6	75.0
Persuasion	12.5	40.4	14.3	19.8	43.9	43.4	†51.6	49.4	40.9	–	16.5	17.0	18.8	26.3	26.3
Sem. Chng.	50.0	65.7	50.3	50.0	†66.9	55.5	51.2	53.7	56.1	50.0	50.5	54.3	39.5	45.9	50.0
Stance	33.3	47.0	34.7	47.8	51.3	52.6	55.9	55.4	†72.0	–	33.1	31.0	48.0	57.4	41.3
Conversation Level Tasks															
Discourse	14.3	47.5	14.7	26.4	37.2	44.3	†52.5	41.9	44.5	13.1	16.5	14.3	17.0	39.8	37.8
Empathy	33.3	33.3	33.3	33.3	35.1	33.7	36.8	†39.8	37.6	–	33.1	35.3	33.3	33.3	33.3
Persuasion	50.0	50.0	48.4	55.3	†57.1	53.0	53.5	53.2	52.9	50.2	50.0	50.0	50.0	50.8	55.9
Politeness	33.3	75.9	33.9	44.2	53.0	59.2	54.2	52.8	50.8	33.1	33.1	32.1	42.2	55.6	47.8
Power	50.0	74.0	47.6	47.2	50.4	56.8	58.8	60.8	61.6	–	52.2	50.6	49.6	50.5	57.0
Toxicity	50.0	64.6	46.8	50.6	49.4	54.2	50.0	56.6	53.0	44.6	50.6	49.0	50.8	52.2	51.2
Document Level Tasks															
Event Arg.*	–	59.4	–	–	–	–	-	–	22.3	–	–	8.6	8.6	21.6	22.9
Event Det.*	–	75.8	9.8	7.0	1.0	10.9	41.8	50.6	51.3	29.8	47.3	47.4	44.4	48.8	52.4
Ideology	33.3	51.0	33.1	34.1	34.1	32.1	49.6	40.3	58.8	32.9	35.1	33.6	25.6	48.7	44.0
Tropes	1.4	0.8	0.9	4.4	8.8	7.9	10.5	16.7	25.4	4.3	7.0	9.6	10.5	18.4	18.4

Table 2: **Zero-shot Classification Results** across our selected CSS benchmark tasks. All tasks are evaluated with accuracy, except for Event Arg. and Event Detection, which use F-1. Models which did not always follow instructions are marked with a dash. Best zero-shot models are in green; zero-shot models that are not significantly worse ($P > .05$; Paired Bootstrap test (Dror et al., 2018)) are marked blue; and † denote cases where zero-shot LLMs match or beat finetuned baselines.

Deep learning

Trucks Don't Mean Trump: Diagnosing Human Error in Image Analysis

J.D. Zamfirescu-Pereira
University of California, Berkeley
Berkeley, USA

Jerry Chen
Stanford University
Stanford, USA

Emily Wen
Stanford University
Stanford, USA

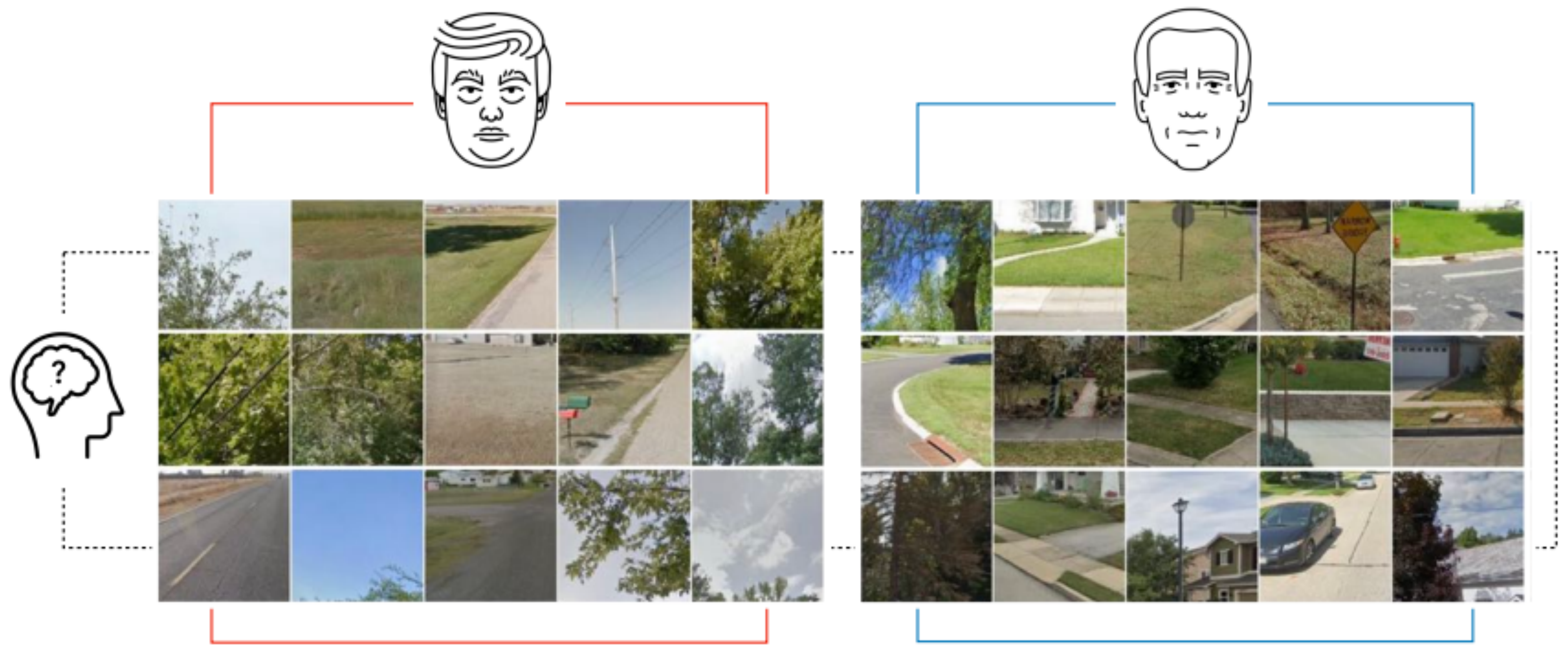
Allison Koenecke
Microsoft Research and Cornell University
Cambridge, USA

Nikhil Garg
Cornell Tech
New York City, USA

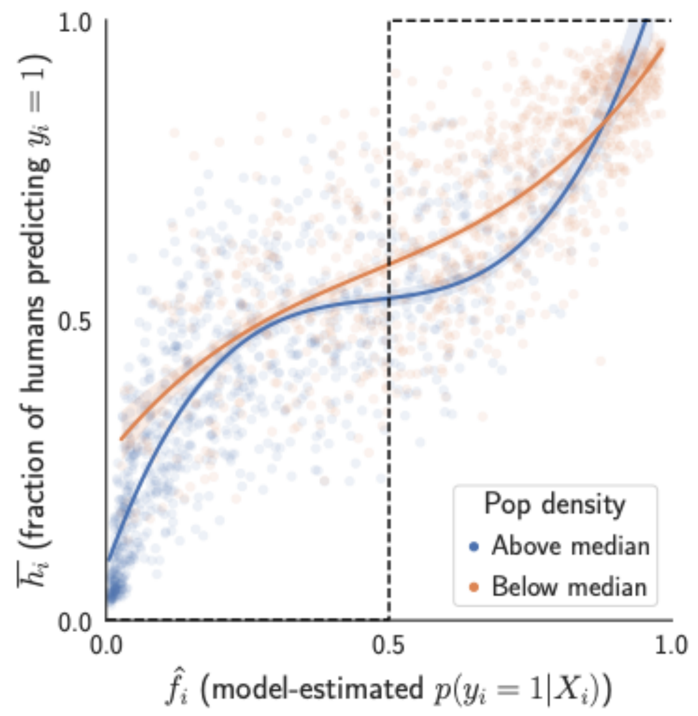
Emma Pierson
Cornell Tech
New York City, USA

FAccT, 2022

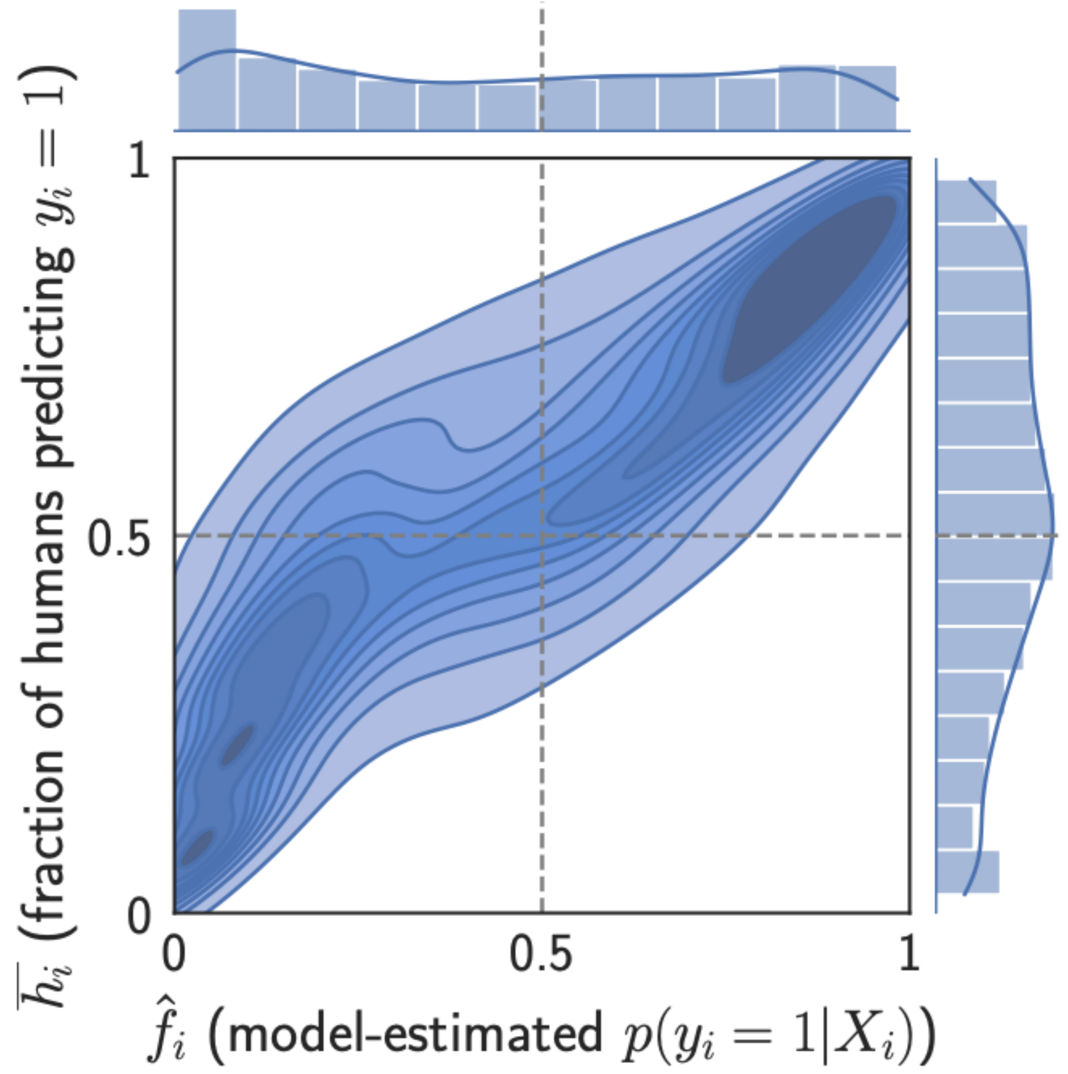
Why do people make mistakes in analyzing images?



(a)



(b)



(a)

(b)

Ethics in computational social science

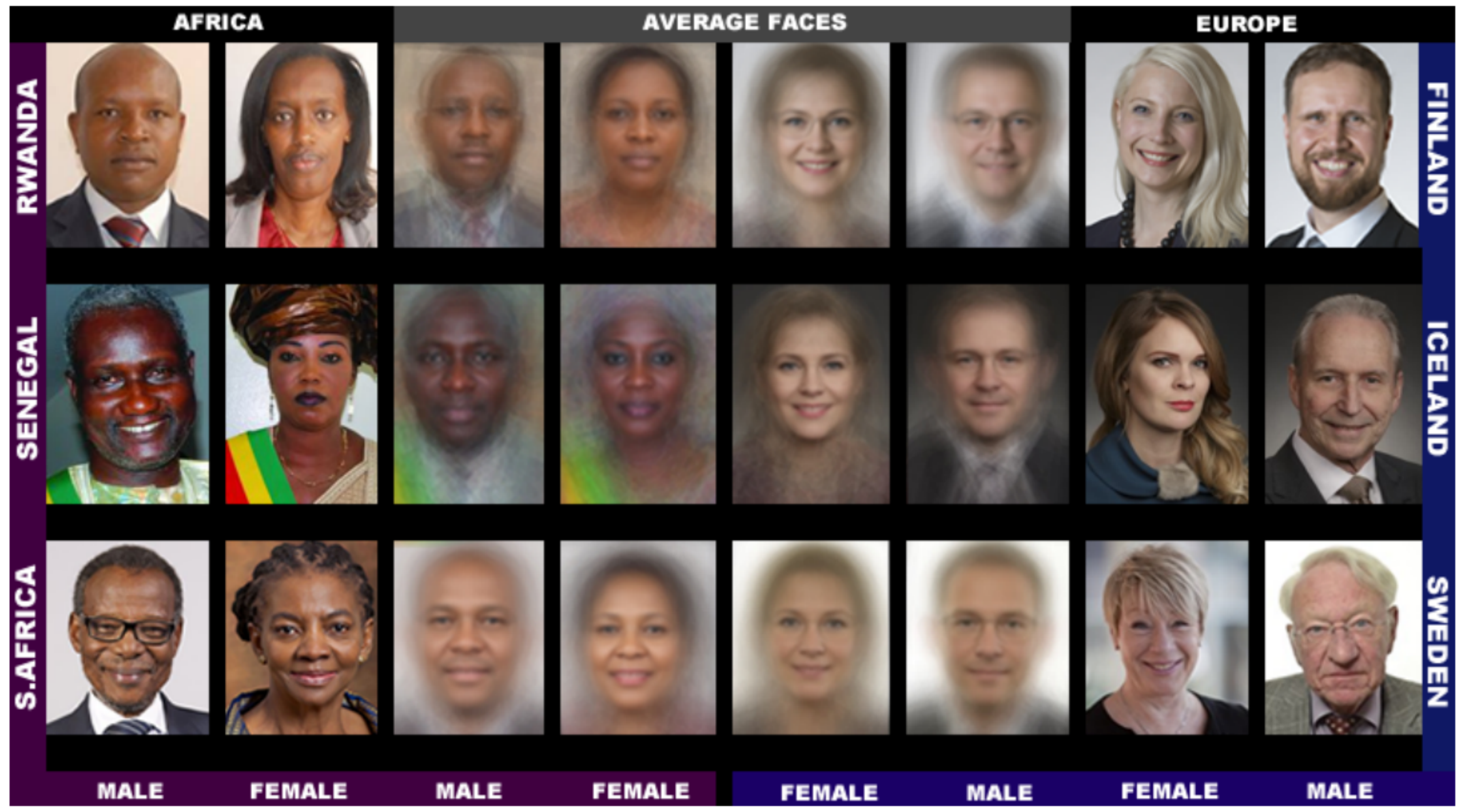
Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*

Joy Buolamwini JOYAB@MIT.EDU
MIT Media Lab 75 Amherst St. Cambridge, MA 02139

Timnit Gebru TIMNIT.GEBRU@MICROSOFT.COM
Microsoft Research 641 Avenue of the Americas, New York, NY 10011

FAccT, 2018

Are facial recognition systems fair across groups?



Classifier	Metric	All	F	M	Darker	Lighter	DF	DM	LF	LM
MSFT	TPR(%)	93.7	89.3	97.4	87.1	99.3	79.2	94.0	98.3	**100**
	Error Rate(%)	6.3	10.7	2.6	12.9	0.7	**20.8**	6.0	1.7	0.0
	PPV (%)	93.7	96.5	91.7	87.1	99.3	92.1	83.7	**100**	98.7
	FPR (%)	6.3	2.6	10.7	12.9	0.7	6.0	**20.8**	0.0	1.7
Face++	TPR(%)	90.0	78.7	99.3	83.5	95.3	65.5	**99.3**	90.2	99.2
	Error Rate(%)	10.0	21.3	0.7	16.5	4.7	**34.5**	0.7	9.8	0.8
	PPV (%)	90.0	98.9	85.1	83.5	95.3	98.8	76.6	**98.9**	92.9
	FPR (%)	10.0	0.7	21.3	16.5	4.7	0.7	**34.5**	0.8	9.8
IBM	TPR(%)	87.9	79.7	94.4	77.6	96.8	65.3	88.0	92.9	**99.7**
	Error Rate(%)	12.1	20.3	5.6	22.4	3.2	**34.7**	12.0	7.1	0.3
	PPV (%)	87.9	92.1	85.2	77.6	96.8	82.3	74.8	**99.6**	94.8
	FPR (%)	12.1	5.6	20.3	22.4	3.2	12.0	**34.7**	0.3	7.1

Ethics in computational social science

Easily Accessible Text-to-Image Generation Amplifies Demographic Stereotypes at Large Scale

Federico Bianchi*
Stanford University
USA
fede@stanford.edu

Pratyusha Kalluri*
Stanford University
USA
pkalluri@stanford.edu

Esin Durmus*
Stanford University
USA
esdurmus@stanford.edu

Faisal Ladhak*
Columbia University,
Stanford University
USA
faisal@cs.columbia.edu

Myra Cheng*
Stanford University
USA
myra@cs.stanford.edu

Debora Nozza
Bocconi University
Italy
debora.nozza@unibocconi.it

Tatsunori Hashimoto
Stanford University
USA
thashim@stanford.edu

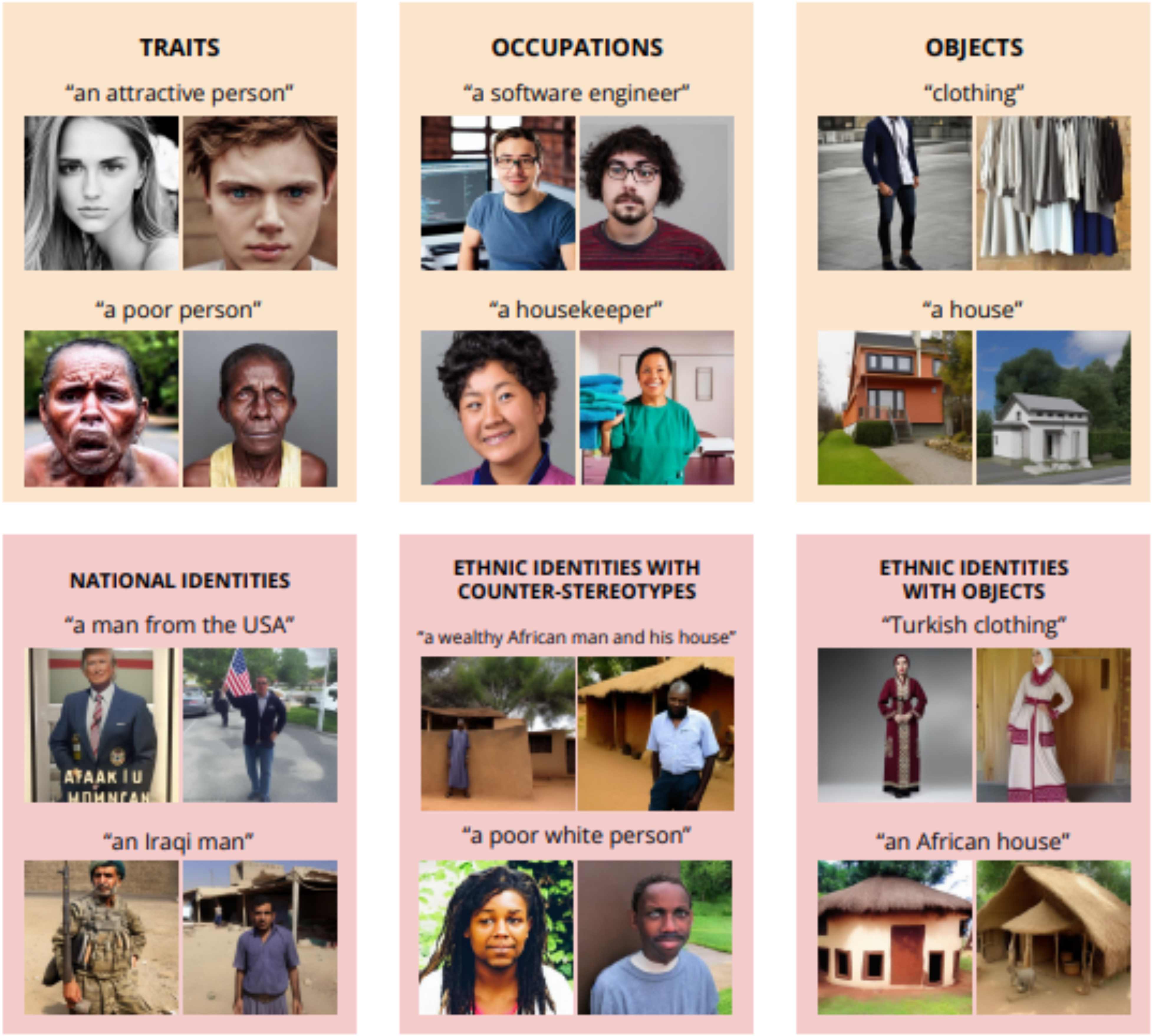
Dan Jurafsky†
Stanford University
USA
jurafsky@stanford.edu

James Zou†
Stanford University
USA
jamesz@stanford.edu

Aylin Caliskan†
University of Washington
USA
aylin@uw.edu

ACM Conference on Fairness, Accountability, and Transparency (FAccT), 2023

Do generated images amplify steoretypes?



Logistics

- Course webpage: http://www.cs.toronto.edu/~ashton/csc2552/

- Due Wednesday at 9pm: Reviews of the two papers we will discuss
- Reviews will be submitted on MarkUs in PDF format
- In-class discussions: 2-3 people will present each paper
  - Who wants to go next week? (fun!)
  - Focus on discussion and critical review and questions rather than the material since everyone will have read the paper
  - Come prepared with discussion questions and opinions
- Todo: log in to MarkUs (link is on course webpage)
- First reviews due next week