Utrecht University

AI for an Open Society

Lecture 3 – Autonomy, psychology of interaction with AI systems

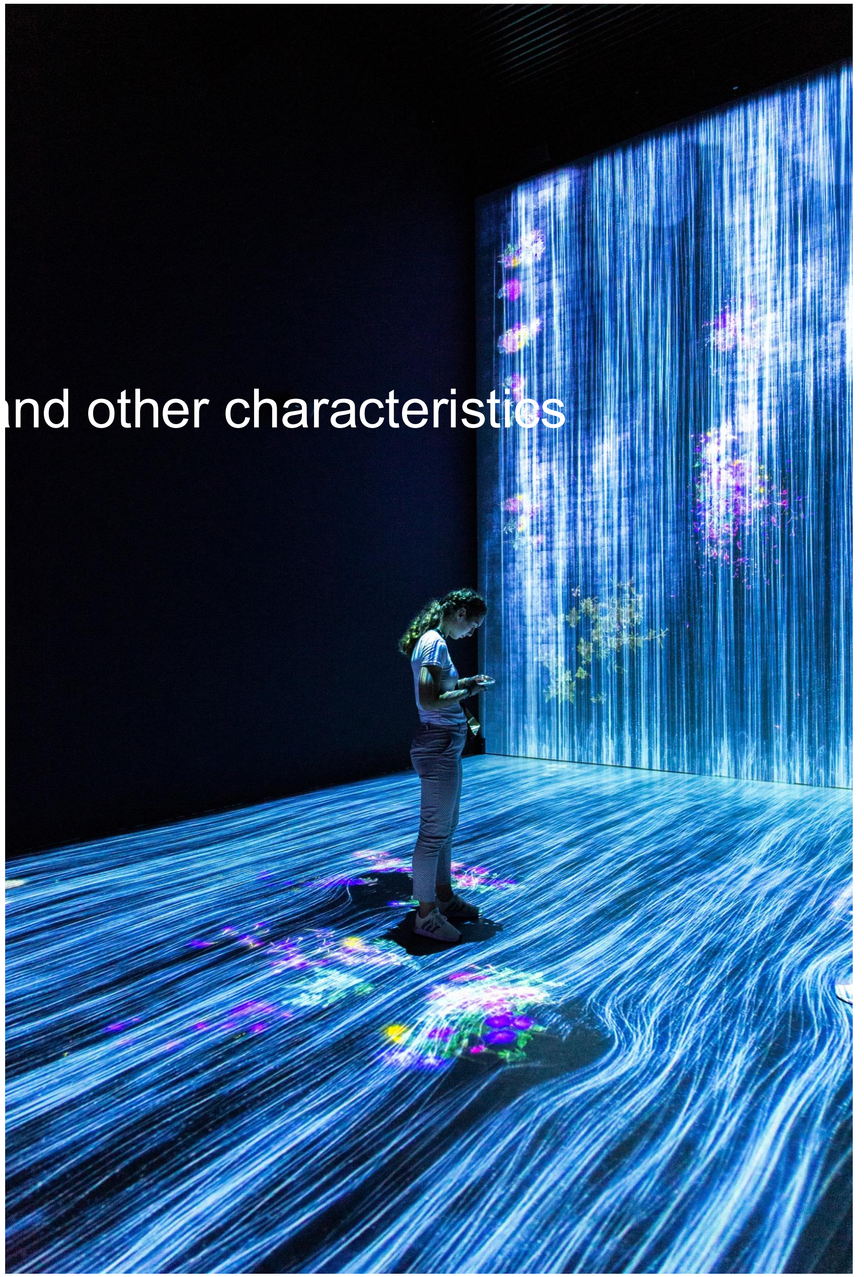
Dr. Ruud Hortensius

29.04.2025

Algorithms and digital footprints: from personality prediction to influencing emotions

This lecture’s objectives

- Discuss the role of psychology in artificial intelligent systems
- Discuss how digital traces can be used to distill personality traits and other characteristics
- Discuss digital emotion contagion and how to measure this
- Acknowledge ethical considerations of AI systems



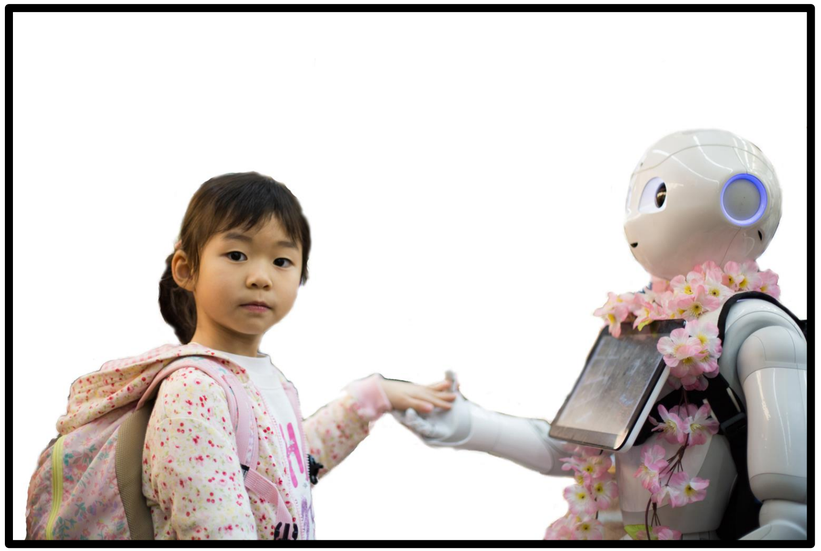
Today’s lecture

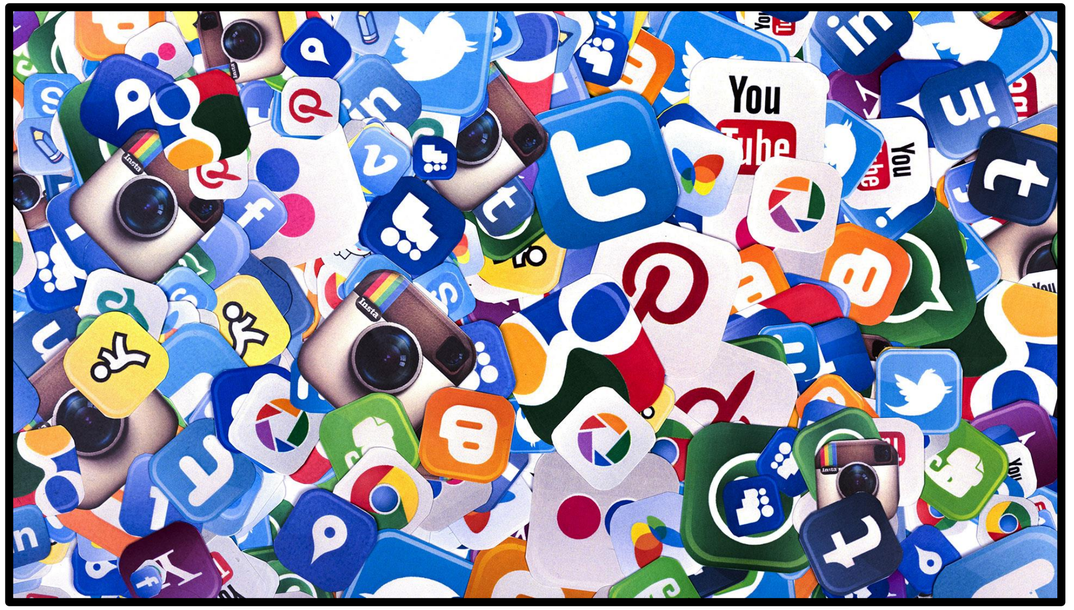
Aritificial Intelligent Systems

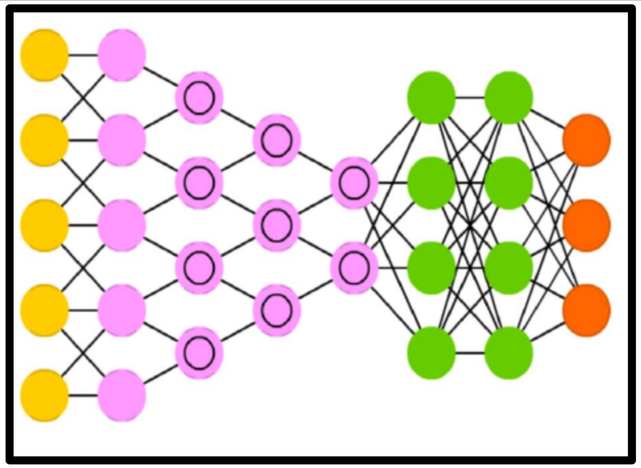
Digital traces

Emotion Contagion

Artificial Intelligent Systems







Artificial Intelligent Systems

- Rahwan, I., Cebrian, M., Obradovich, N., Bongard, J., Bonnefon, J.-F., Breazeal, C., … Wellman, M. (2019). Machine behaviour. *Nature*, *568*(7753), 477–486.

- Working definition:

AI agents: “complex and simple algorithms used to make decisions”

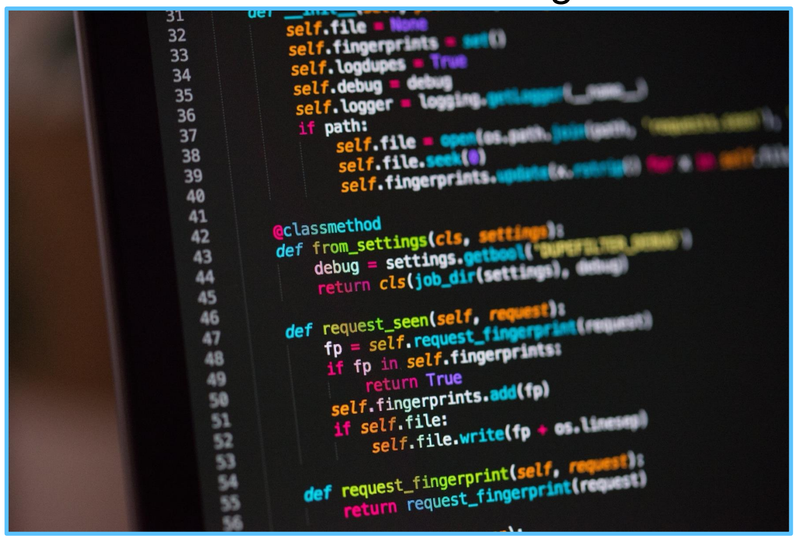
What about social media and other technologies? It’s all about the algorithms supporting these tools and the digital traces created.

I focus on the human-side of social AI and technology

What do you think is the impact of Artificial Intelligent Systems on Human Behaviour?

Psychology

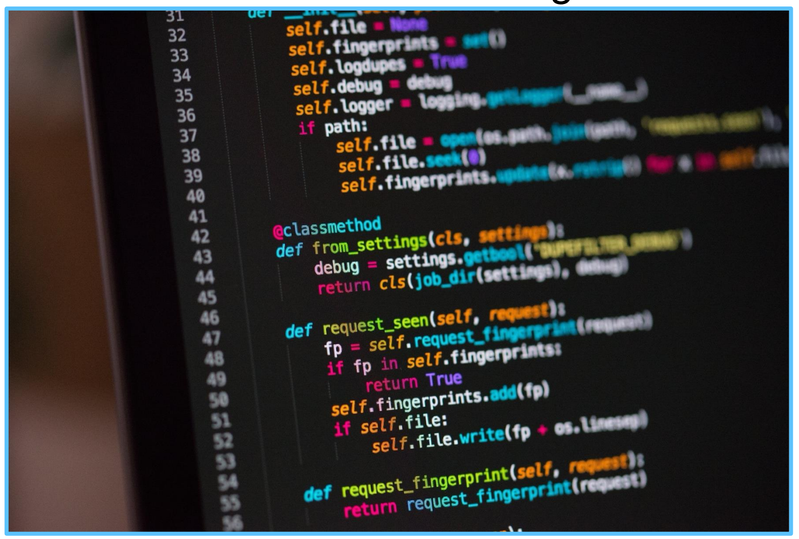
Social / AI Technologies



How do you think psychology and neuroscience impact the development of Artificial Intelligent Systems?

Psychology

Social / AI Technologies



What’s ahead?

[link]

Let’s dissect

- Users are manipulated by algorithm to adjust you to goals of advertiser
- Constant (mild) behaviour modification
- Constant surveillance
- Negative moods
- Increase in depression amongst teenagers
- Negative influence on politics (”unreal”)
- Addictive schemes, statically addicted (population-level)
- “Silicon Valley spying empires” (Facebook, etc.)

Palaeolithic brain, modern-day technologies

Tristan Harris

Center for Humane Technology

The New York Times

Opinion

TURNING POINTS

Our Brains Are No Match for Our Technology

By Tristan Harris

That's because our Paleolithic brains aren't built for omniscient awareness of the world's suffering. Our online news feeds aggregate all of the world's pain and cruelty, dragging our brains into a kind of learned helplessness. Technology that provides us with near-complete knowledge without a commensurate level of agency isn't humane.

Simply put, technology has outmatched our brains, diminishing our capacity to address the world's most pressing challenges. The advertising business model built on exploiting this mismatch has created the attention economy. In return, we get the "free" downgrading of humanity.

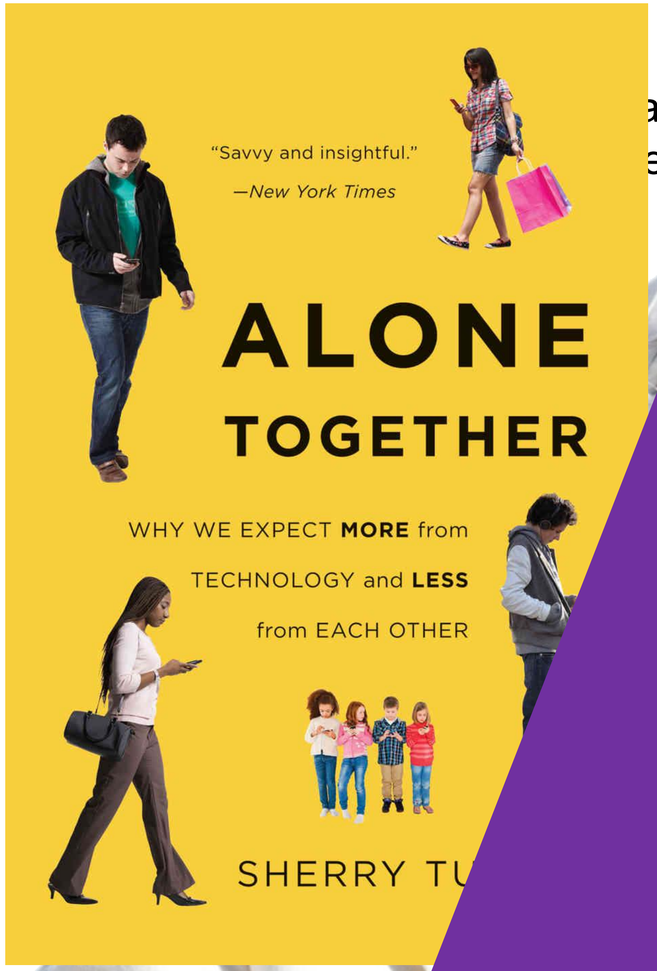
This leaves us profoundly unsafe. With two billion humans trapped in these environments, the attention economy has turned us into a civilization maladapted for its own survival.

"Human downgrading"

We need to "realigning technology with our humanity"

So much debate w t data/resea

ation
e of
ss b
ift from God”



Artificial Intelligent Systems

- Rahwan, I., Cebrian, M., Obradovich, N., Bongard, J., Bonnefon, J.-F., Breazeal, C., … Wellman, M. (2019). Machine behaviour. *Nature*, *568*(7753), 477–486.

- Why study this?

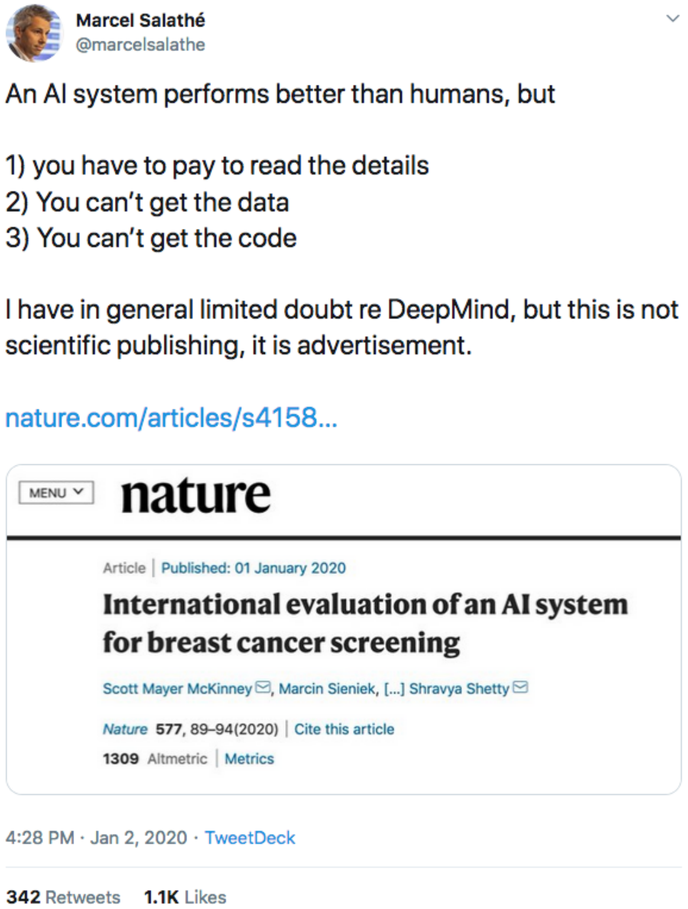
1. Ubiquity of algorithms/AI/technology
2. Complexity and opacity of algorithms/AI/technology
3. These can have beneficial and detrimental effects on humanity


COMPUTER SCIENCE

Artificial intelligence faces reproducibility crisis

Unpublished code and sensitivity to training conditions make many claims hard to verify

Hutson (2018) Science


Marcel Salathé
@marcelsalathe

An AI system performs better than humans, but

1) you have to pay to read the details
2) You can't get the data
3) You can't get the code

I have in general limited doubt re DeepMind, but this is not scientific publishing, it is advertisement.

nature.com/articles/s4158...

MENU nature

Article | Published: 01 January 2020

International evaluation of an AI system for breast cancer screening

Scott Mayer McKinney, Marcin Sieniek, [...] Shravya Shetty

Nature **577**, 89–94(2020) | Cite this article

1309 Altmetric | Metrics

4:28 PM · Jan 2, 2020 · TweetDeck

342 Retweets **1.1K** Likes

Socially Intelligent Artificial Technologies

Democracy

News ranking algorithms

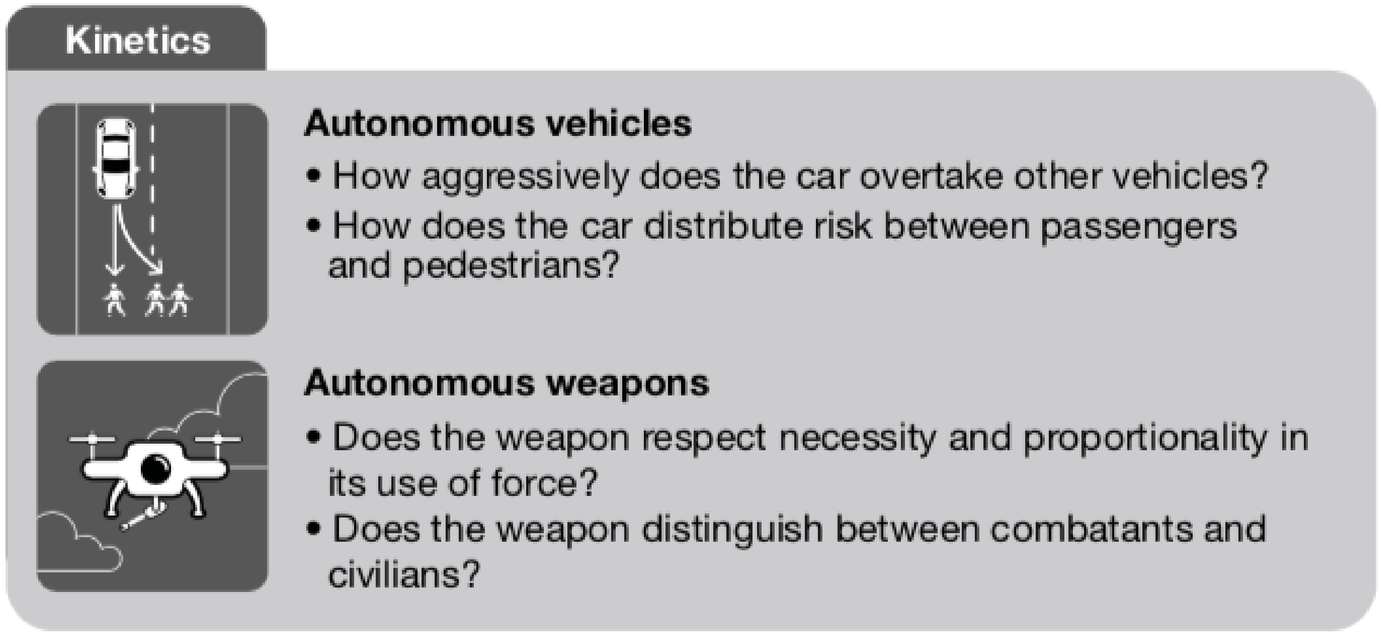
- Does the algorithm create filter bubbles?
- Does the algorithm disproportionately censor content?

Algorithmic justice

- Does the algorithm discriminate against a racial group in granting parole?
- Does a predictive policing system increase the false conviction rate?

Rahwan et al. 2019 Nature

Socially Intelligent Artificial Technologies



Rahwan et al. 2019 Nature

Socially Intelligent Artificial Technologies

Markets

Algorithmic trading

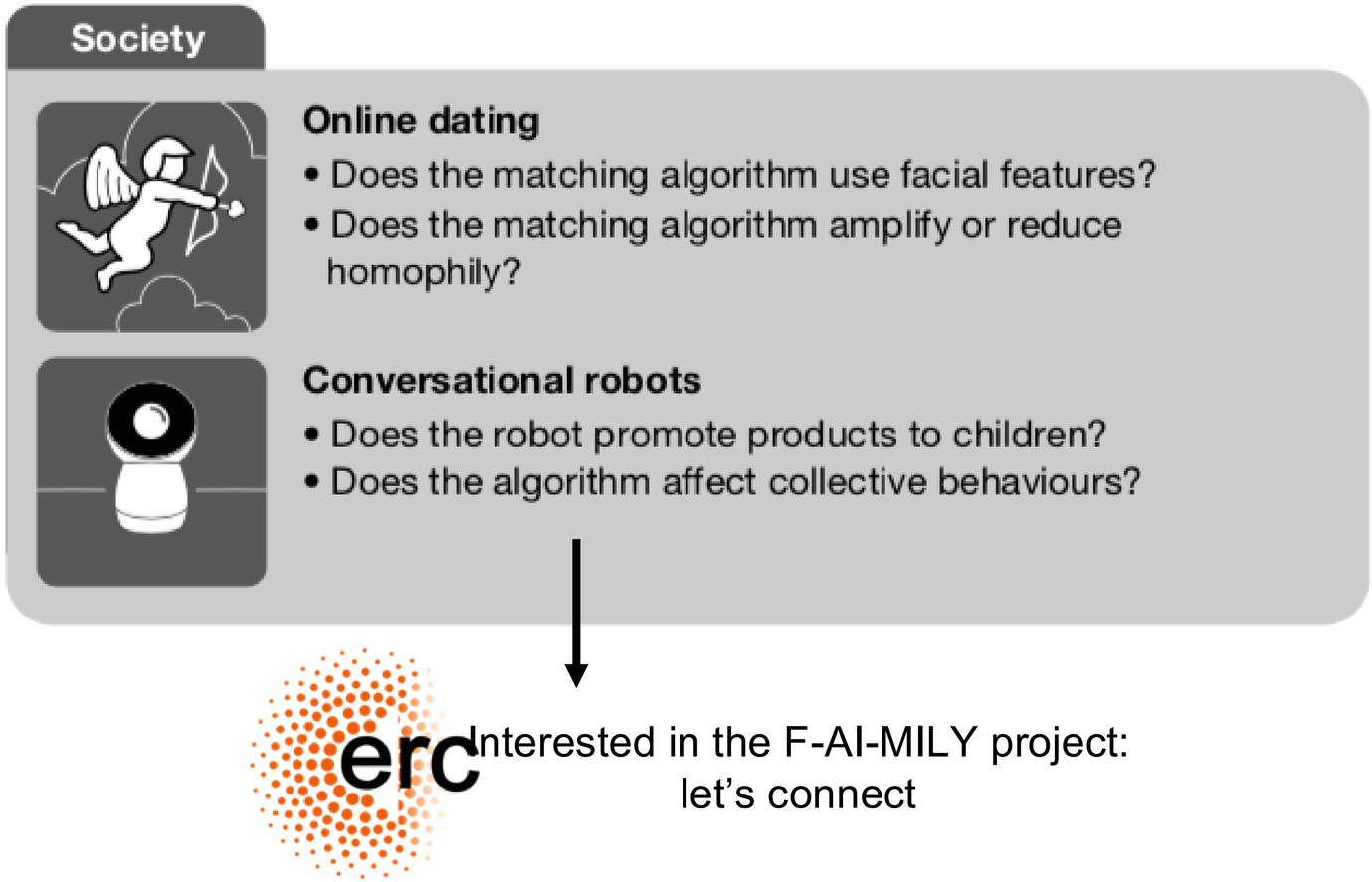
- Do algorithms manipulate markets?
- Does the behaviour of the algorithm increase systemic risk of market crash?

Algorithmic pricing

- Do algorithms of competitors collude to fix prices?
- Does the algorithm exhibit price discrimination?

Rahwan et al. 2019 Nature

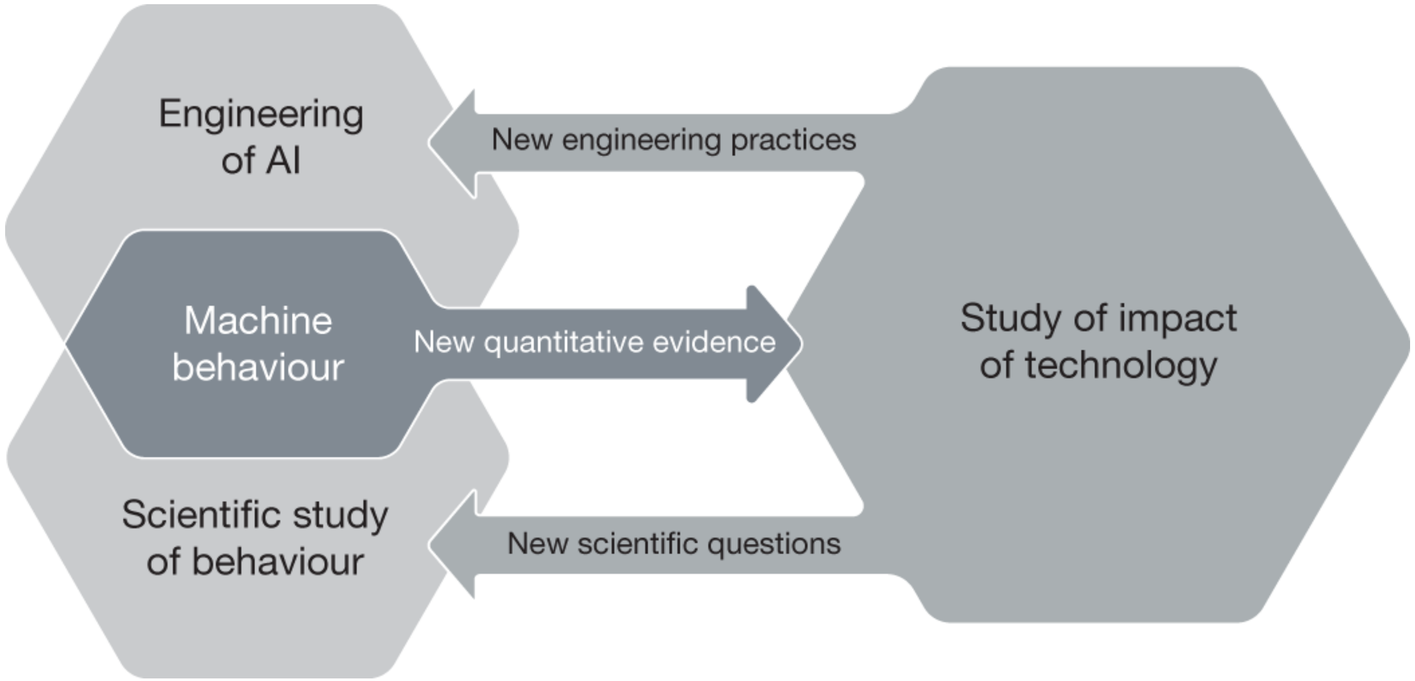
Socially Intelligent Artificial Technologies



erc Interested in the F-AI-MILY project:
let's connect

Rahwan et al. 2019 Nature

It’s an interdisciplinary endeavor



Rahwan et al. 2019 Nature

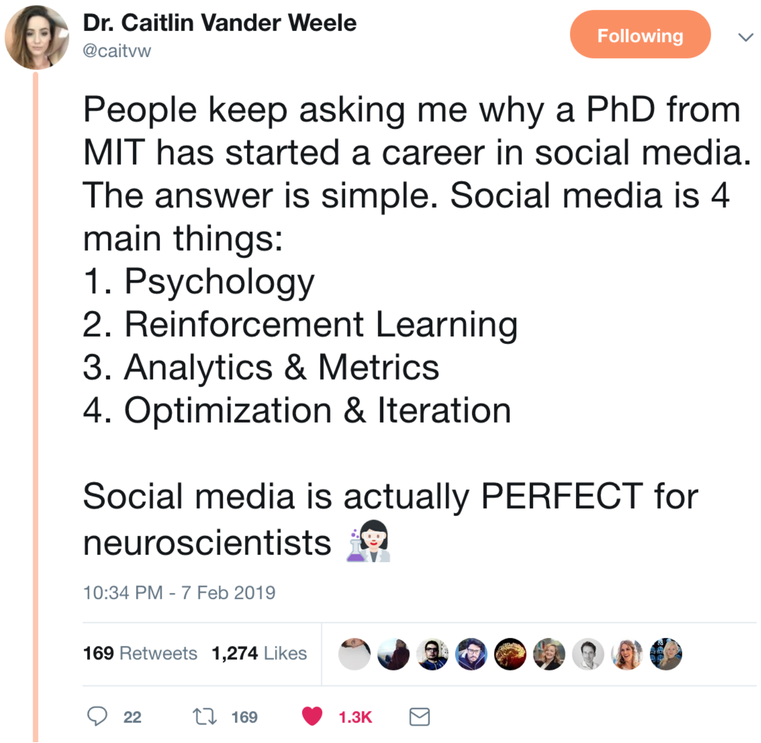
AI needs Psychology!

Systematic investigation:

theory + experimental methodological approach

Eyssel (2017) Robotics and Autonomous Systems

AI needs Psychology!



Dr. Caitlin Vander Weele
@caitvw

Following

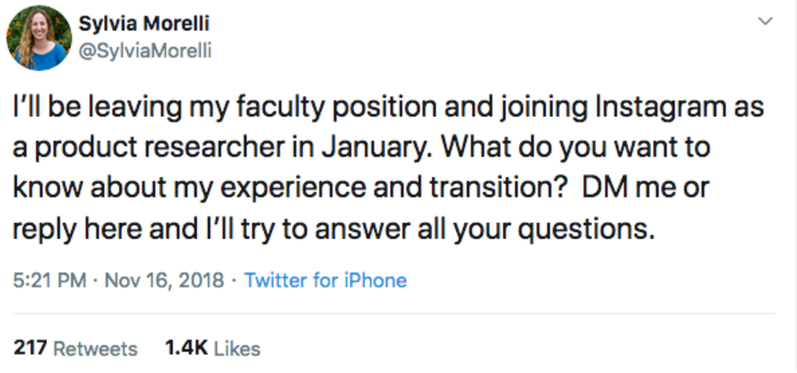
People keep asking me why a PhD from MIT has started a career in social media. The answer is simple. Social media is 4 main things:
1. Psychology
2. Reinforcement Learning
3. Analytics & Metrics
4. Optimization & Iteration

Social media is actually PERFECT for neuroscientists 👩‍🔬

10:34 PM - 7 Feb 2019

169 Retweets **1,274** Likes

22 169 1.3K



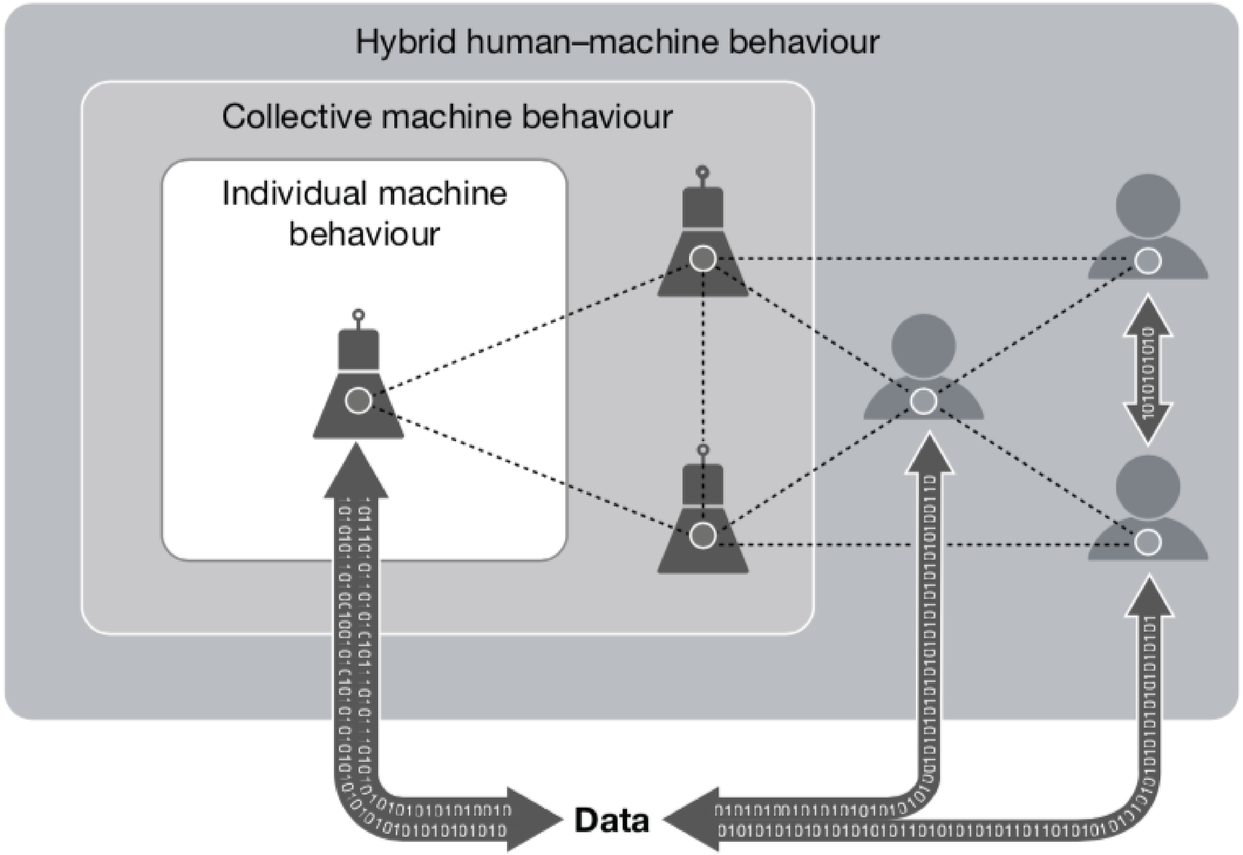
Sylvia Morelli
@SylviaMorelli

I'll be leaving my faculty position and joining Instagram as a product researcher in January. What do you want to know about my experience and transition? DM me or reply here and I'll try to answer all your questions.

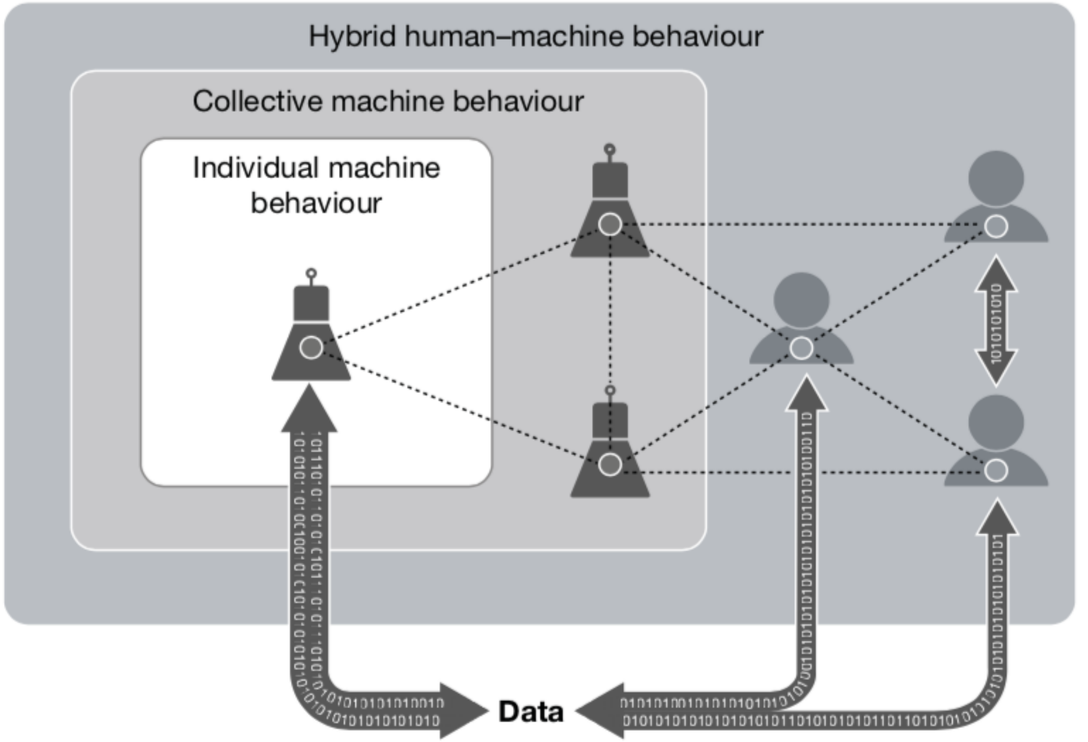
5:21 PM · Nov 16, 2018 · Twitter for iPhone

217 Retweets **1.4K** Likes

It’s a new society



It’s a new society



- Hybrid Human-machine behaviour:
  - Machines shape human behaviour
    How do they influence our believes and behaviours?
    How do they change our social life and society?
  - Humans shape machine behaviour
    More engineering aspect (what algorithm to use)
  - Human-machine co-behaviour
    How do they work together (e.g. Twitter, bubbles, detection in health care)

Human-machine co-behaviour

- AI chatbot by Microsoft: ”conversational understanding”
- Targeted at 18- to 24-years-olds in the US
- Learning by conversation
- But people started ‘feeding’ it racists, and misogynistic conversations

TayTweets
(@TayandYou)

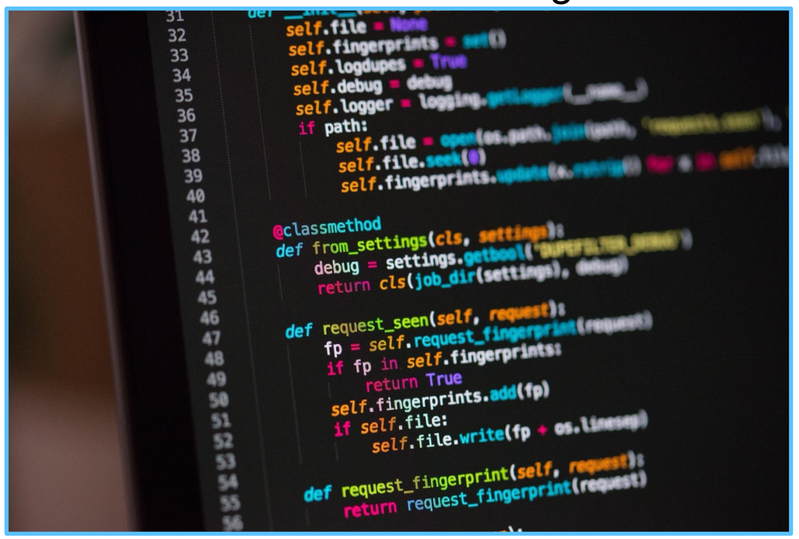
hellooooooo w🌎rld!!!

March 23, 2016

AI systems shape human behaviour

Psychology

Social / AI Technologies

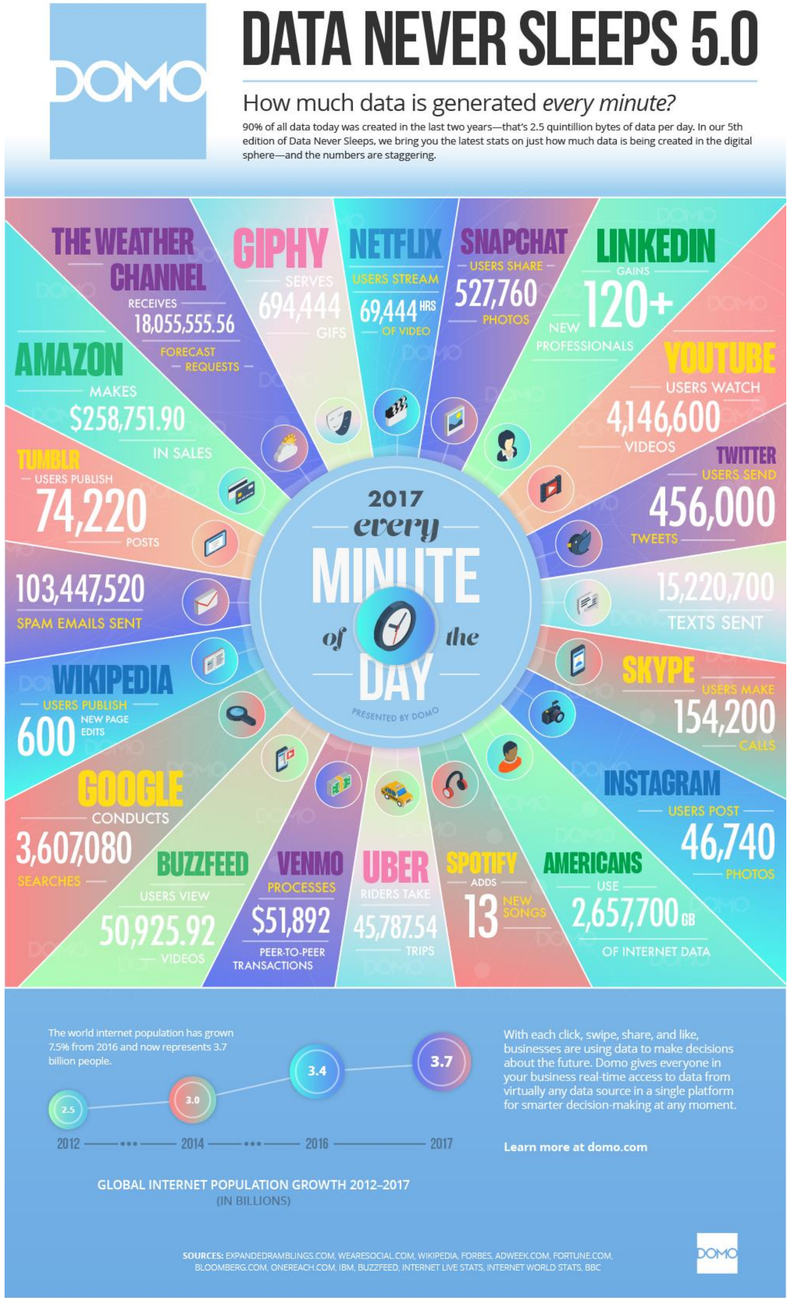


Data is the fuel

How much data do we create?

- Facebook: >2 billion users, >100 billion messages, >1 billion stories
- Whatsapp: >1 billion users
- Twitter: >330 million users
- Instagram: > 1 billion users
- Youtube: >2 billion users, >5 billion videos
- TikTok: > 1 billion users

Note: these are old numbers



Today’s lecture

Aritificial Intelligent Systems

Digital traces

Data are Digital Traces

- Rafaeli, A., Ashtar, S., & Altman, D. (2019). Digital Traces: New Data, Resources, and Tools for Psychological-Science Research. *Current Directions in Psychological Science,* 096372141986141O.

- Digital traces are records (logs) of people's behaviour: Facebook likes, Tweets, vlogs, blogs, youtube history, cookies, etc.

- "Digital Traces are collected and retained by Internet platforms, sensors, and other devices and typically comprise contextual data about when, where, and for how long behaviours occurred"

- But you need new analytic tools to analyse this → fuel of AI (big data)

"Goldmine of data for psychological science"

Digital Traces

Three advantages of digital traces for psychological science:

1. Bigger and/or different sample population, beyond the WEIRD-population.
2. Detailed recording/measurement of behaviours that include context
3. ‘Digital dossier’: implicit and explicit behaviour of individual, reducing experimental bias

But: challenge is potential bias (self-selection of platform used, e.g. Twitter/Facebook/TikTok)

Social Media Data

This overlaps with:

- Meshi, D., Tamir, D. I., & Heekeren, H. R. (2015). The Emerging Neuroscience of Social Media. *Trends in Cognitive* Sciences, 19(12), 771–782.

Five reasons why to use social media data

1. Externally valid measures of real-world behaviour
2. Real-time continuous measures of social behaviour
3. Extended period of time
4. Quantifiable
5. Can be used in conjunction with real-world behaviours

But besides self-selection, there is the risk of awareness biases and the parallel is not endless

Social Media Data

This overlaps with:

- Meshi, D., Tamir, D. I., & Heekeren, H. R. (2015). The Emerging Neuroscience of Social Media. *Trends in Cognitive* Sciences, 19(12), 771–782.

Examples of proxies:

1. Offline thoughts (e.g. I statements)
2. Emotional states (e.g. content of posts)
3. Social conformity (e.g. are you influences by posts of friends)
4. Prosocial behaviour (e.g. liking a post)
5. Curiosity (e.g. time scrolling through feed)
6. Personality traits (e.g. predicted from likes)
7. Social network
8. Social interaction

Digital Traces

- The logs can be collected through:
  1. Self-report (“how many hours have you spent on Instagram?”)
  2. Application Programming Interface (API): data collection interface (https://developer.twitter.com)
  3. Web scraping: e.g. Twitter (DIY: https://psyteachr.github.io/hack-your-data/scrape-twitter.html)
  4. Collaboration with organization (e.g. Facebook; Kramer et al. 2014 PNAS)
  5. Smart phones, sensors or apps (e.g. Cozmo robot; Cross…Hortensius, 2019 PTRB)

- Data can be analysed through:
  1. Traditional methods (e.g. correlation, anova)
  2. Automated tools (e.g., text analysis → sentiment analysis)
  3. Deep learning also referred to as machine learning

Challenge:
- data wrangling (R, python); but new data and programming skills needed
- Collaboration (computer science, neuroscience, psychology)

Digital Traces

- Ethical issues
  - Issue of consent
  - Privacy
  - Data ownership
  - ‘Surveillance capitalism’

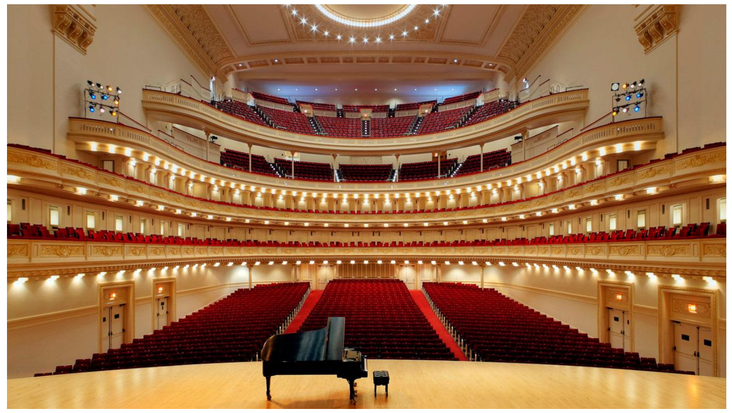
TECHNOLOGY

She was denied entry to a Rockettes show — then the facial recognition debate ignited

JANUARY 21, 2023 · 7:00 AM ET

HEARD ON ALL THINGS CONSIDERED

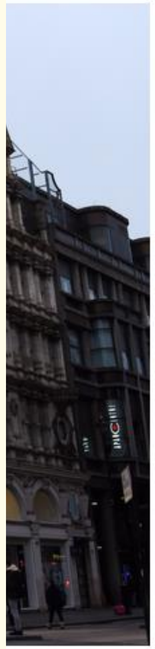
Manuela López Restrepo



Piccadilly Circus targeted adverts

Gunseli Yalcinkaya |

A new digital billboa
display targeted adv
age of pedestrians.



THE AGE OF SURVEILLANCE CAPITALISM

THE FIGHT FOR A HUMAN FUTURE AT THE NEW FRONTIER OF POWER

SHOSHANA ZUBOFF


Data has a better idea

Personality prediction

- Personality prediction **using digital traces?**

- Browsing logs, websites, music collections, online social network properties, online language

Personality prediction

- Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*, *110*(15), 5802–5805.

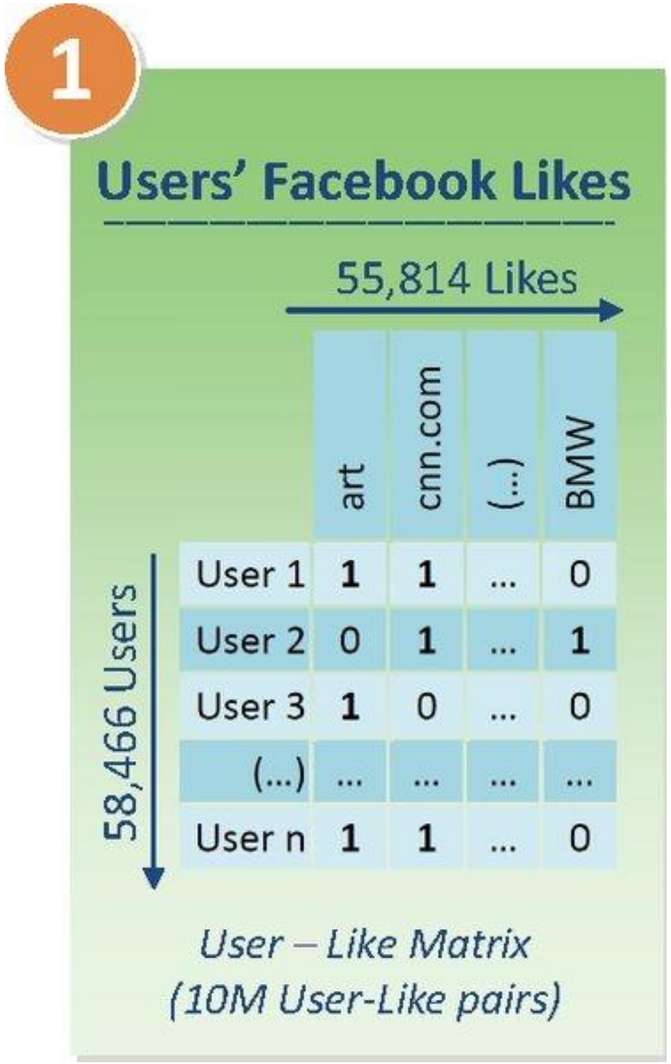
- Latent states in data: what can be predicted for digital traces?

Personality prediction

- Instead of responses to a personality questionnaire (1-10) we have likes (y | n)
- Assumption: Like = positive association with item (e.g. musician)
- <2013: Likes were public!

Personality prediction

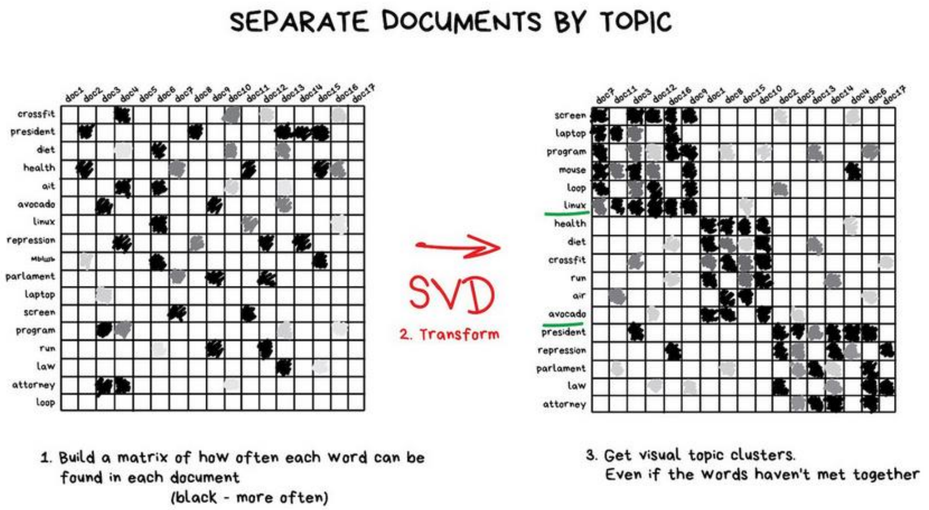
- *n* = 58,466 US volunteers (mypersonality FaceBook application).
- Data includes: Facebook profile information, psychometric test scores, survey information, and list of likes (Mean *n* of Likes = 170)
- Likes: 1 or 0 (no association between Like and user)



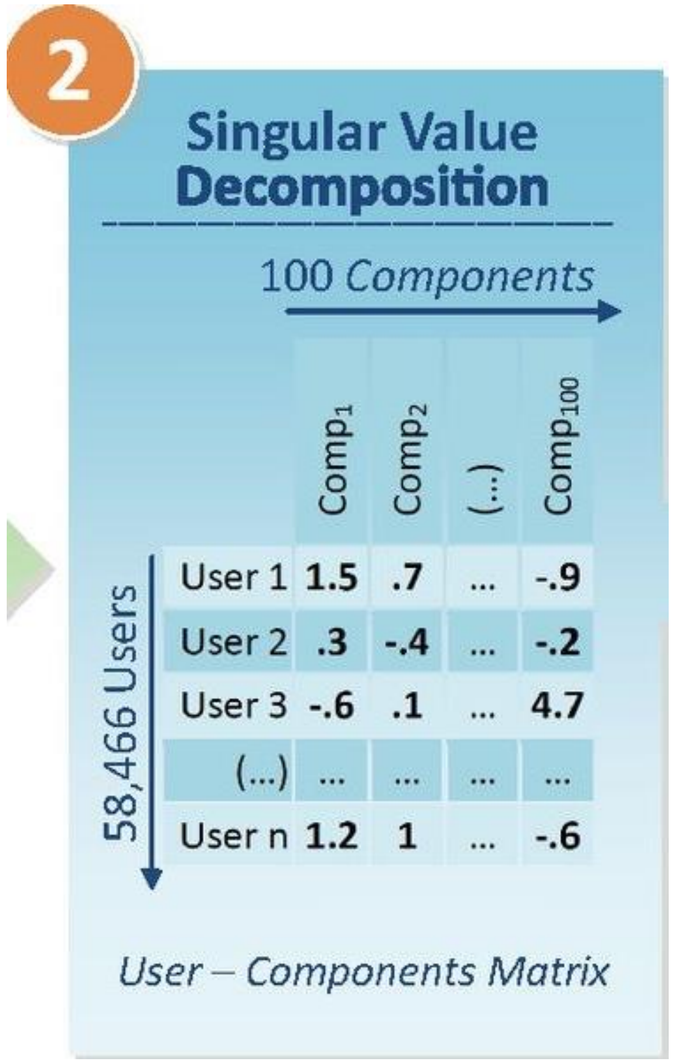
Reference slide for self-study

Personality prediction

- Reduce the dimensions using SVD



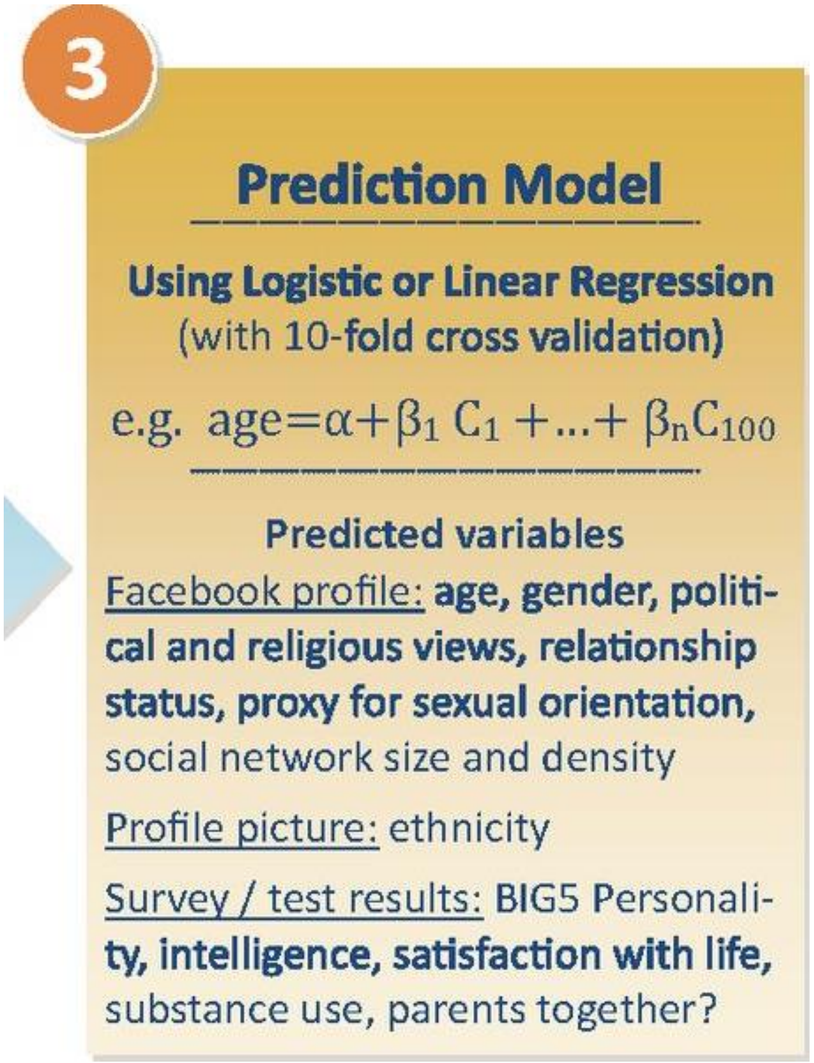
https://vas3k.com/blog/machine_learning/



Personality prediction

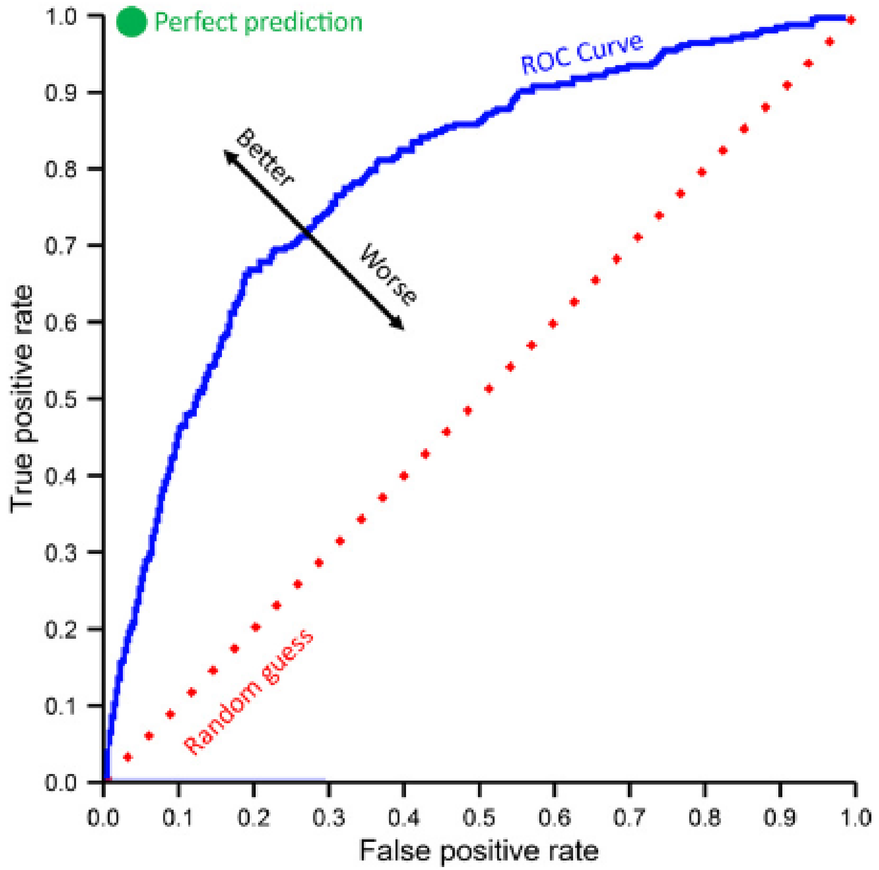
- Predict gender, age, relationship status (single/in relation), network size and density, sexual orientation ("interested in"), ethnic origin, political views (liberal/conservative), religion (Muslim/Cristian), personality (BIG5), intelligence, satisfaction with life, substance use (for alcohol, drugs, cigarettes), "whether an individual's parents stayed together until the individual was 21y old" using regression analyses

From: Facebook | survey | questionnaire | visual inspection

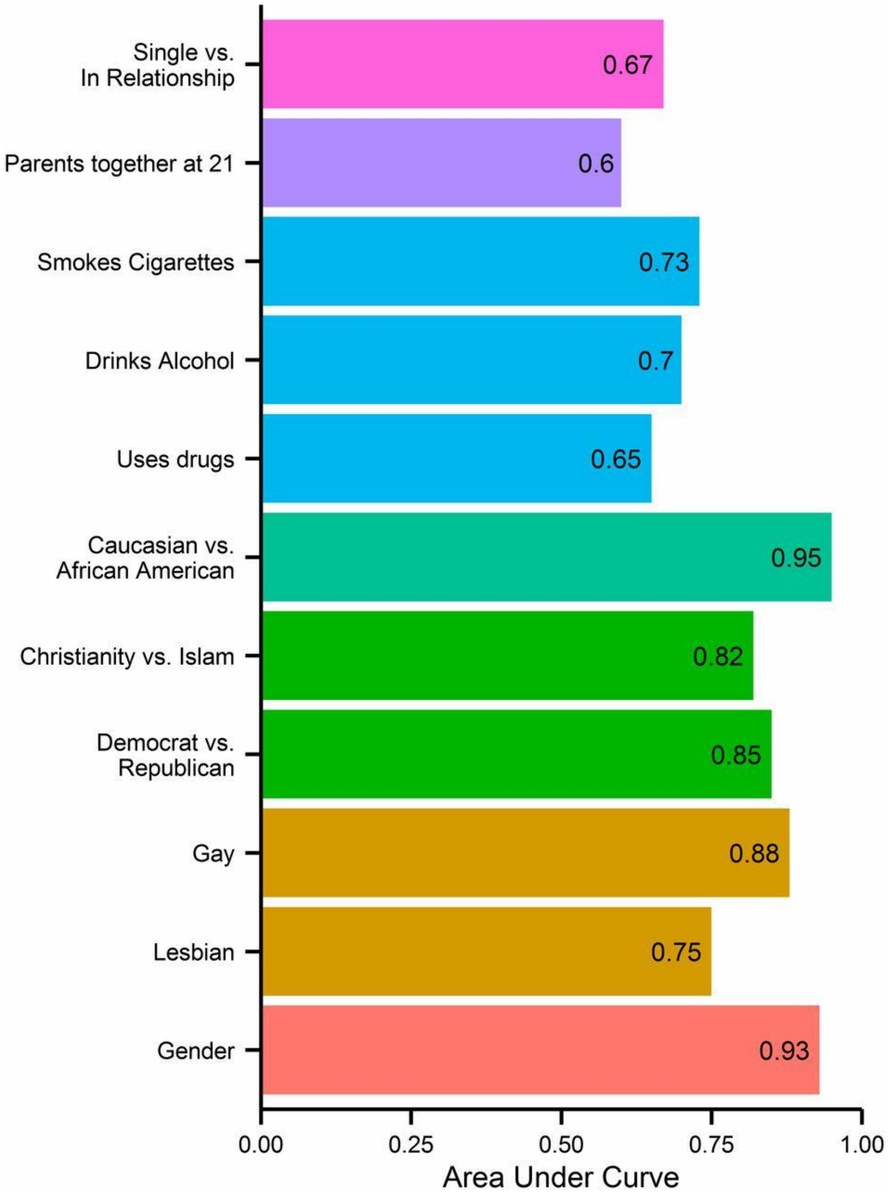


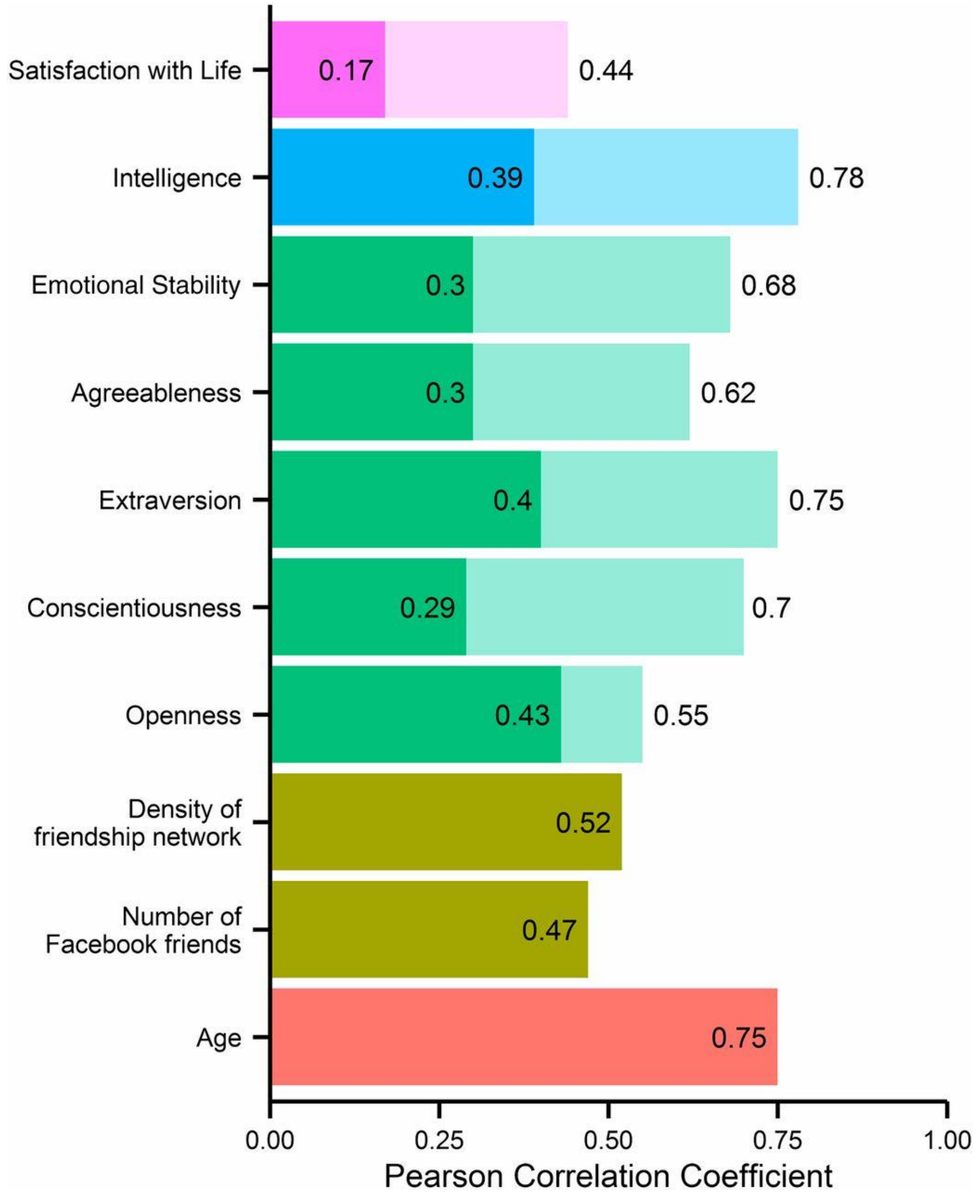
Personality prediction

- Regression with cross-validation
- Either Pearson product-moment correlation between actual and predicted values across participants
- AUC: probability of correctly classifying two random participants (one of each class)



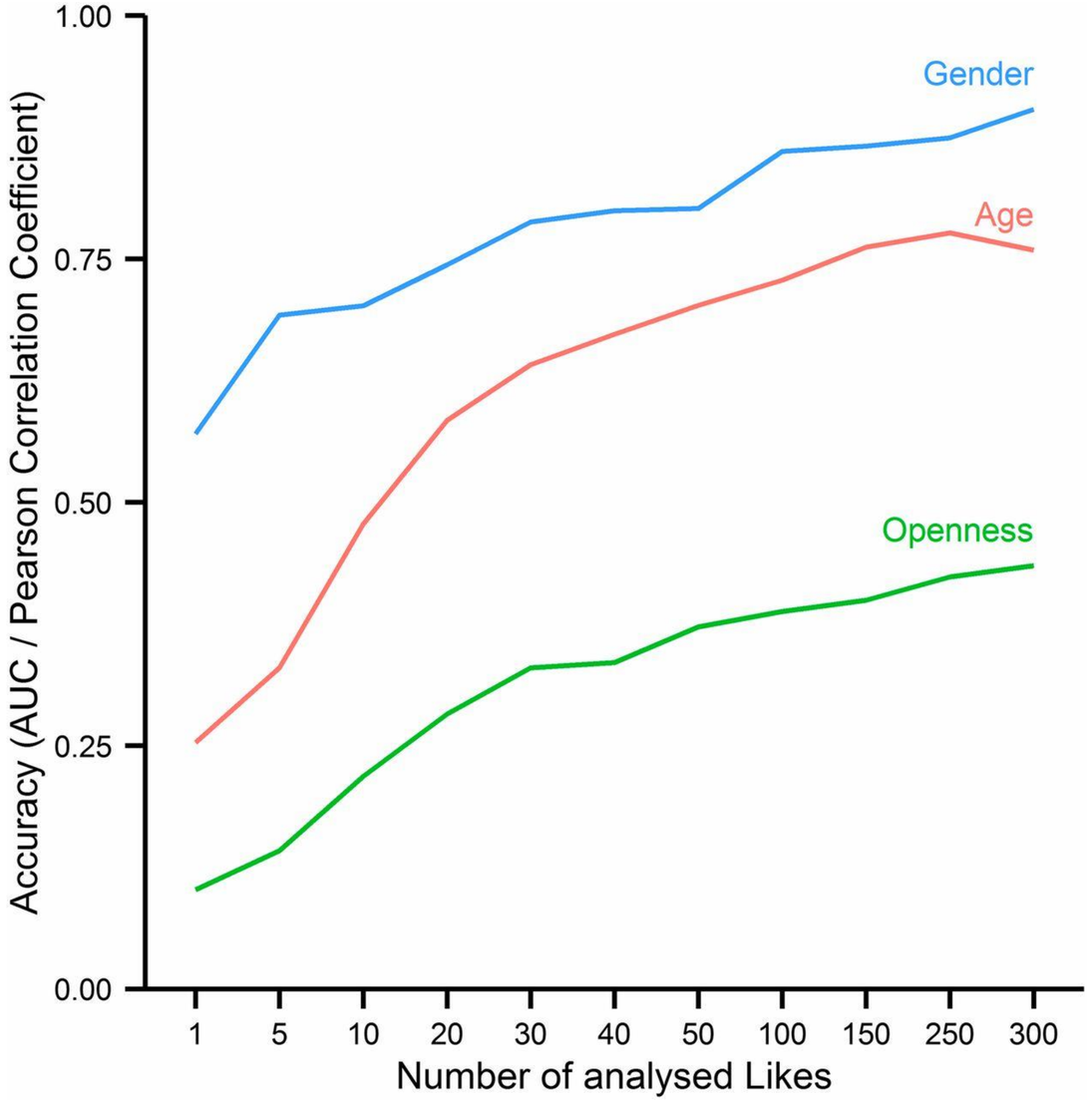
Personality prediction



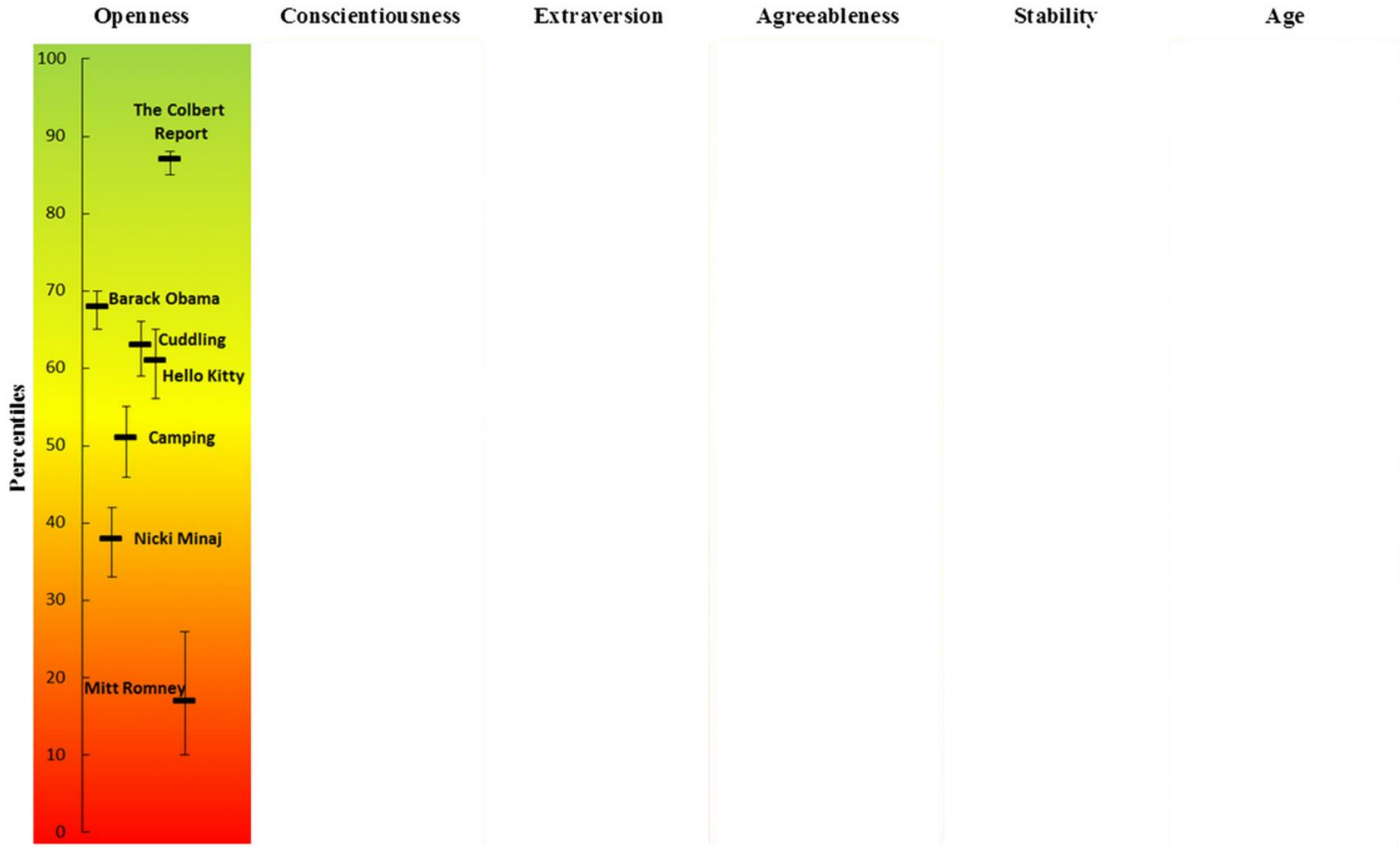


Personality prediction

How much data do you need?



Personality prediction

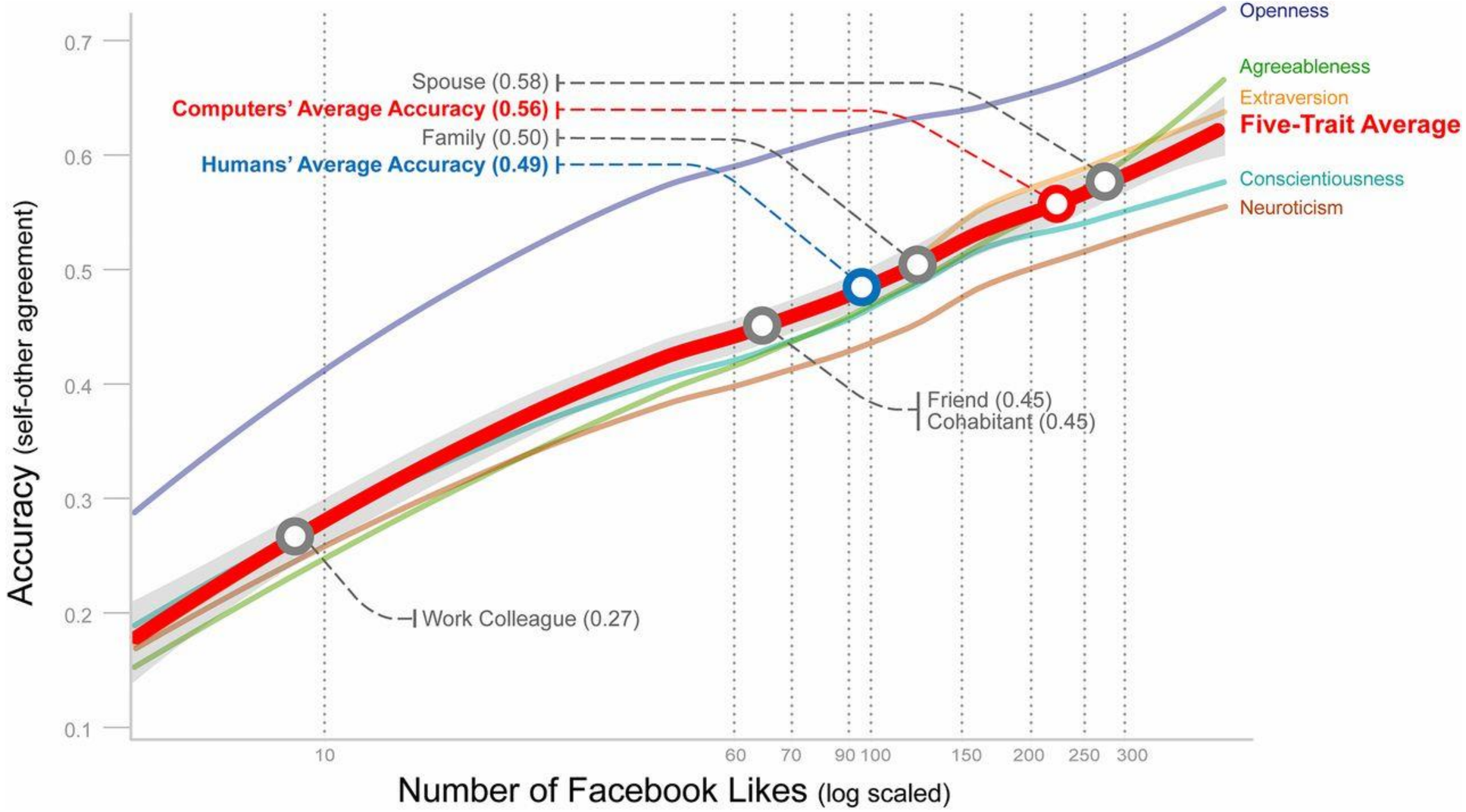


Personality prediction

- From a psychological science / societal perspective what are some of the concerns? Theoretical, methodological, data etc.

- What are potential consequences?

Human and computer personality prediction

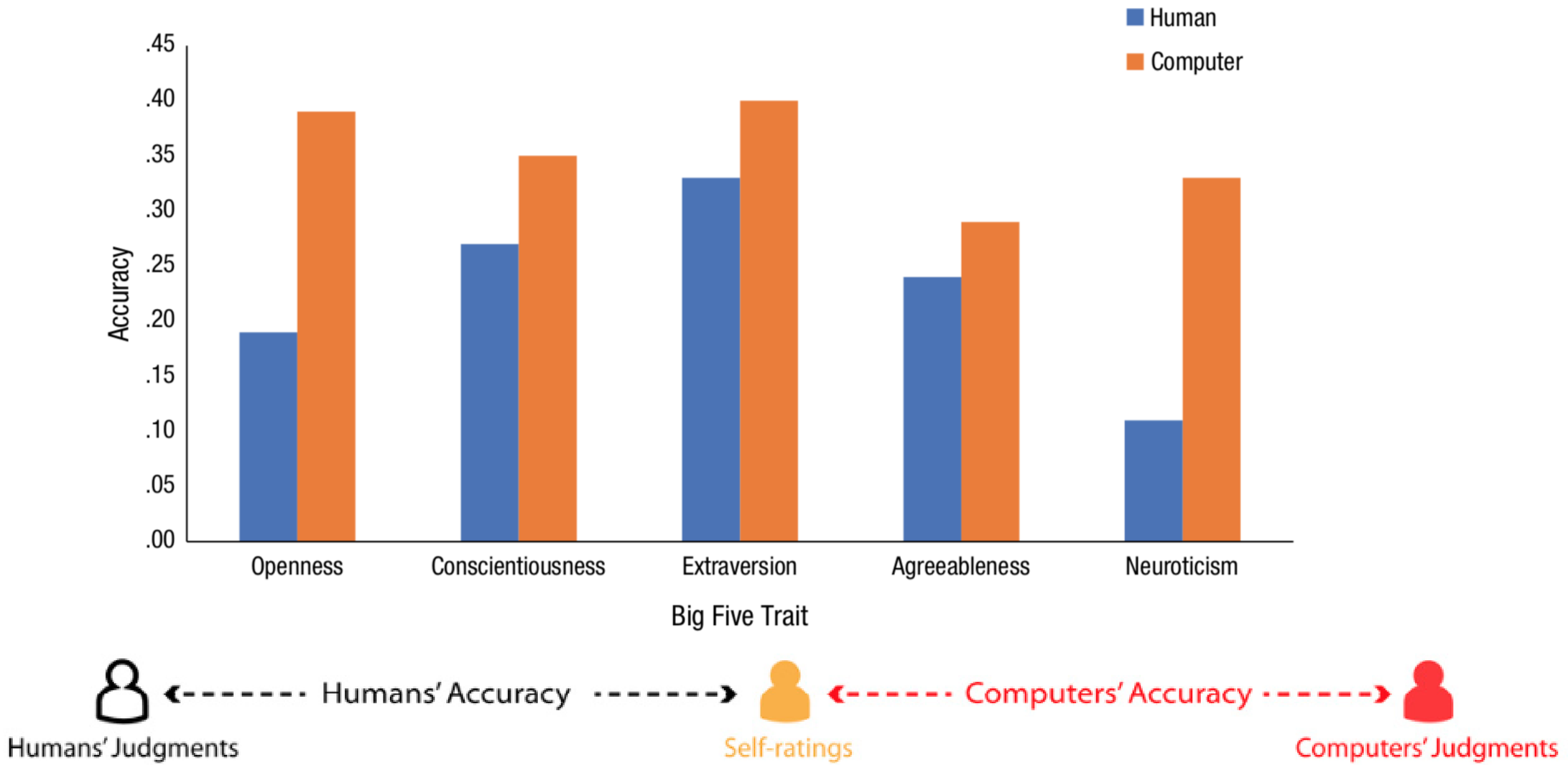


Hinds & Joinson (2019). CDPS

Human and computer personality prediction

Hinds & Joinson (2019). CDPS

Human and computer personality prediction



Hinds & Joinson (2019). CDPS

Personality prediction

Can we predict traits and other person characteristics? Yes

Positive and negative consequences

Psychological targeting

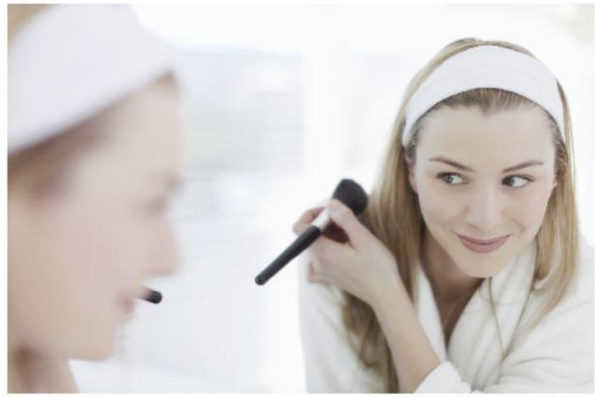
- Create a targeted ads

A

High Extraversion

Dance like no one's watching
(but they totally are)

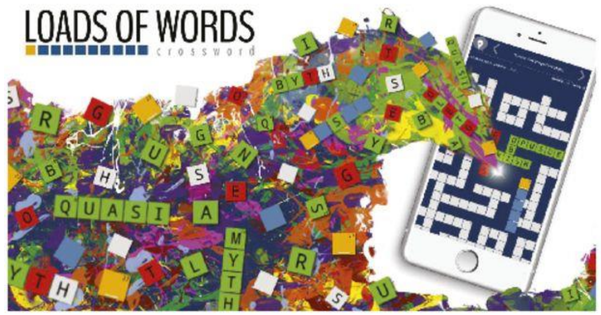
Low Extraversion



Beauty doesn't have to shout

B

High Openness



Aristoteles? The Seychelles? Unleash your creativity and challenge your imagination with an ulimited number of crossword puzzles!

Low Openness

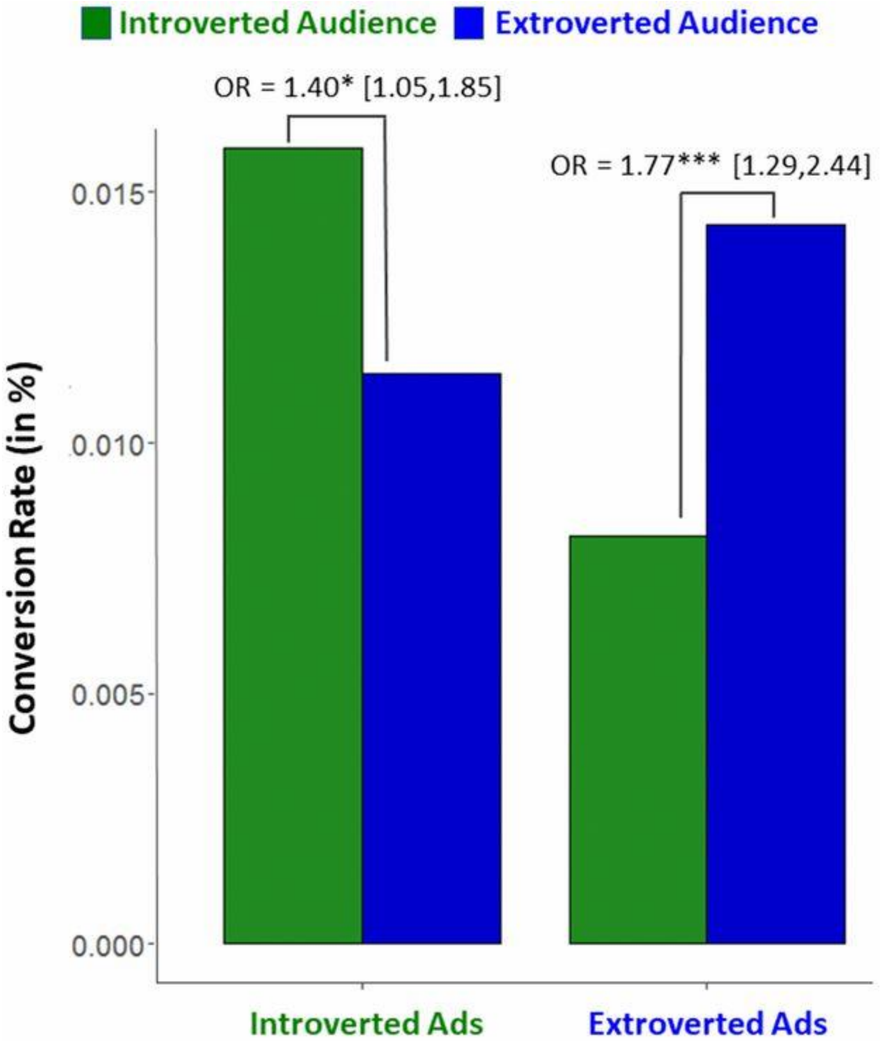
Settle in with an all-time favorite! The crossword puzzle that has challenged players for generations.

Matz et al. (2017). PNAS

Psychological targeting

Study I:
- *n* = 3,129,993 (reach)
- 10,346 clicks
- 390 purchases

Study II:
- *n* = 84,176 (reach)
- 1,130 clicks
- 500 app installs



Matz et al. (2017). PNAS

Psychological targeting

LETTER

Field studies of psychologically targeted ads face threats to internal validity

Dean Eckles, Brett R. Gordon, and Garrett A. Johnson

PNAS June 5, 2018 115 (23) E5254-E5255; first published May 18, 2018 https://doi.org/10.1073/pnas.1805363115

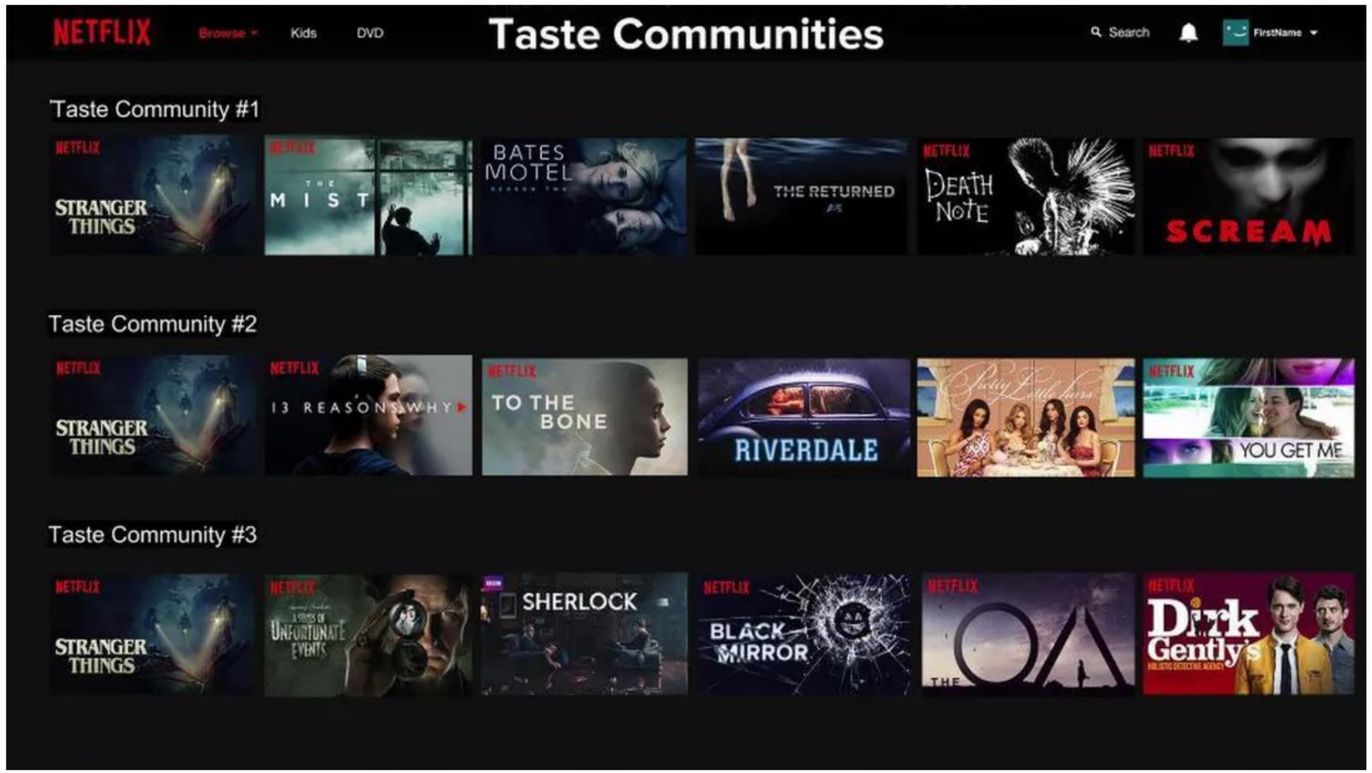
LETTER

Psychological targeting

Byron Sharp, Nick Danenberg, and Steven Bellman

PNAS August 21, 2018 115 (34) E7890; first published August 3, 2018 https://doi.org/10.1073/pnas.1810436115

Applications



GENRE	WINNING IMAGE BY GENRE PREFERENCE:
Action, Horror, Thrillers, Kids, Romance	
Comedies	
Documentaries	
Overall, Drama, Sci-Fi	

airbnb

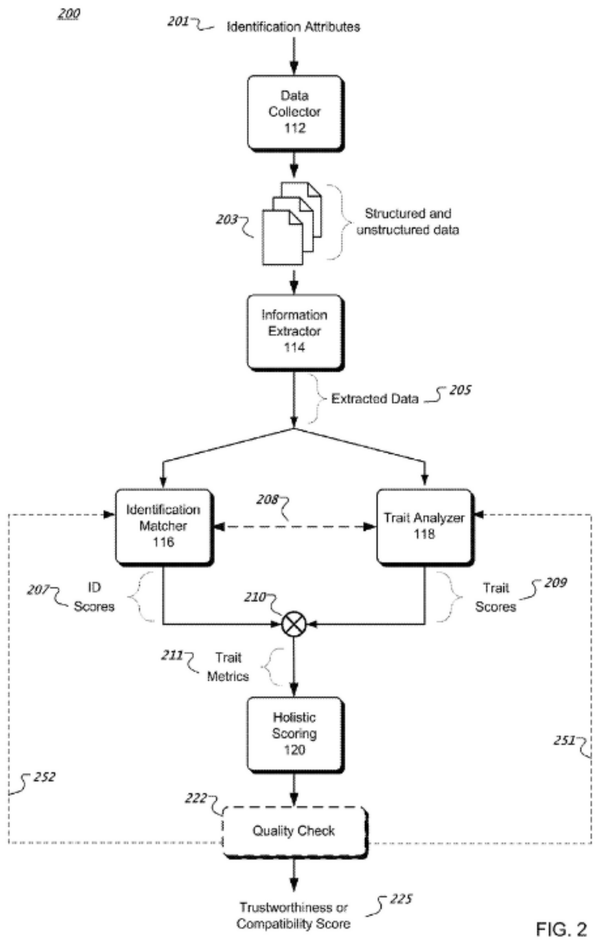

200
201
Identification Attributes
Data Collector 112
203
Structured and unstructured data
Information Extractor 114
Extracted Data
205
Identification Matcher 116
208
Trait Analyzer 118
207
ID Scores
210
Trait Scores
209
211
Trait Metrics
Holistic Scoring 120
251
252
222
Quality Check
Trustworthiness or Compatibility Score
225

FIG. 2

Applications and future


Evening Standard.
WEBSITE OF THE YEAR

Tech

Booker beware: Airbnb can scan your online life to see if you're a suitable guest

The global booking platform could check up on your online personality before you book a holiday

MARK BLUNDEN Technology Correspondent | 3 days ago | 0 comments


DON'T JUDGE

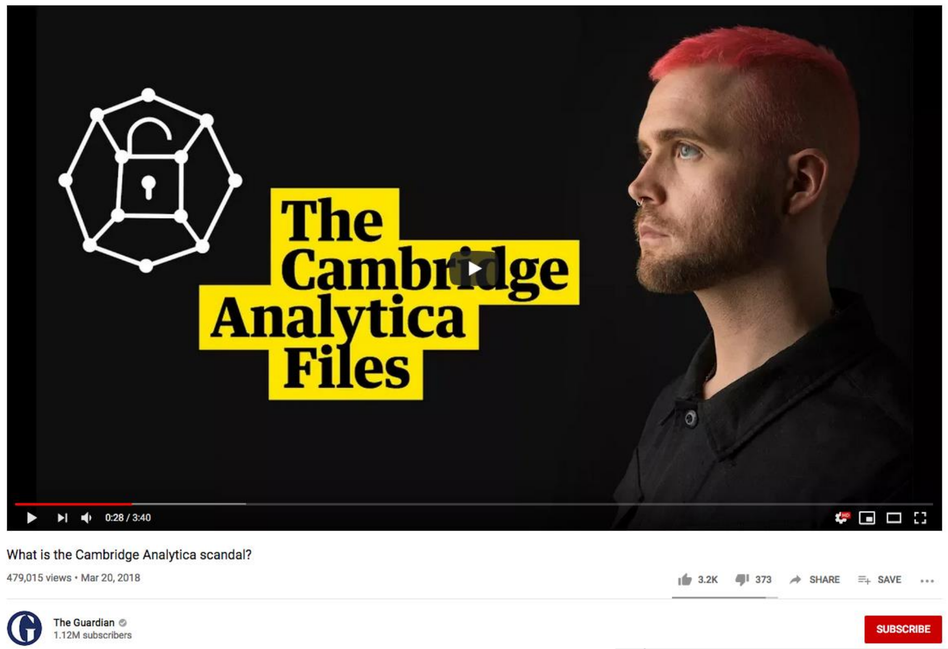
AIRBNB CLAIMS ITS AI CAN PREDICT WHETHER GUESTS ARE PSYCHOPATHS

IMAGE VIA PXHERE/VICTOR TANGERMANN

Tool:
-scans social media sites and scrapes websites
-personality traits (conscientiousness, openness, neuroticism, and dark triad)
-credit and identity checks and 'secure third-party databases'
-keyword, image, video associations with drugs/alcohol etc.

Result in 'person graph' that is cross-referenced with other data (social connection, employment, education)

Real life



The Guardian

News | Opinion | Sport | Culture | Lifestyle | More

The Cambridge Analytica Files

A year-long investigation into Facebook, data, and influencing elections in the digital age

Key stories

Hide

Revealed / 50 million Facebook profiles harvested for Cambridge Analytica in major data breach

484

'I made Steve Bannon's psychological warfare tool': meet the data war whistleblower

Christopher Wylie goes on the record to discuss his role in hijacking the profiles of millions of Facebook users in order to target the US electorate

Revealed: Brexit insider claims Vote Leave team may have breached spending limits

Whistleblower alleges that electoral spending rules could have been manipulated over controversial donation and that Vote Leave 'tried to delete key evidence'

The Brexit whistleblower / 'Did Vote Leave use me? Was I naive?'

Revealed: the ties that bound Vote Leave's data firm to controversial Cambridge Analytica

Facebook's week of shame / The Cambridge Analytica fallout

"Politicians can't control the digital giants with rules drawn up for the analogue era
Andrew Rawnsley

266

"Facebook told me it would act swiftly on data misuse – in 2015
Harry Davies

Raid / Investigators spend seven hours at Cambridge Analytica HQ

Speaking out / Former Cambridge Analytica exec says she wants lies to stop

"The Cambridge Analytica saga is a scandal of Facebook's own making
John Harris

1,109

All stories

Beyond 1984 and Black Mirror

- How can we use prediction from digital traces and psychological targeting to positively influence human behaviour? Should we do this? When?
- What is the role of researchers? Developers of algorithms? Psychological scientists?
- Privacy? Consent? Opt-out?

Reference slide for self-study

Ethical considerations

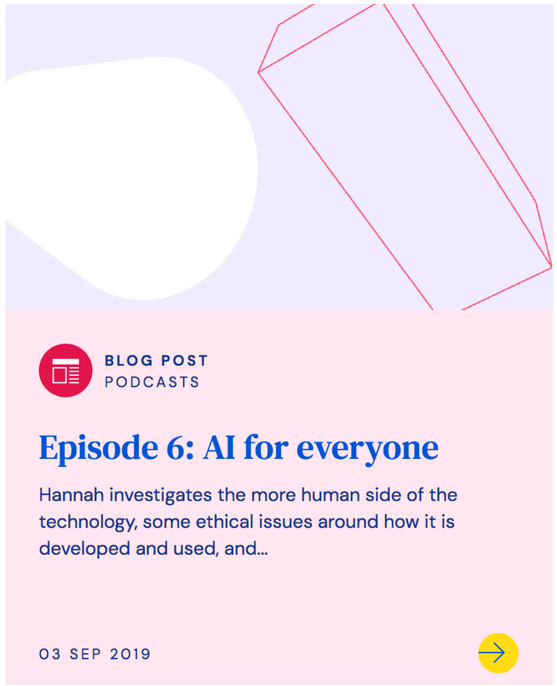
Dr. Koert van Mensvoort

Creative Director Next Nature
Artist/philosopher

www.nextnature.net

“Technology is the self-portrait of humanity”

DeepMind



[link]

Today’s lecture

Aritificial Intelligent Systems

Digital traces

Emotion Contagion

“The Facebook Study”

- Kramer, A. D. I., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, *111*(24), 8788–8790.

- Study focuses on emotion contagion: transfer of emotions across people

- Emotion contagion: present in offline interactions (*transfer of emotional states to other people)*

$Emotion_{P1} \rightarrow Emotion_{P2}$ | $Emotion_{P1} \rightarrow Emotion_{P2}$ but strictly speaking not $Emotion_{P1} \rightarrow Emotion_{P2}$

“The Facebook Study”

Ranking:

1. Your interactions with previous broadcasts
2. Type of broadcast
3. Feedback (likes, shares, reactions)
4. Time
5. Other factors: internet, phone, …

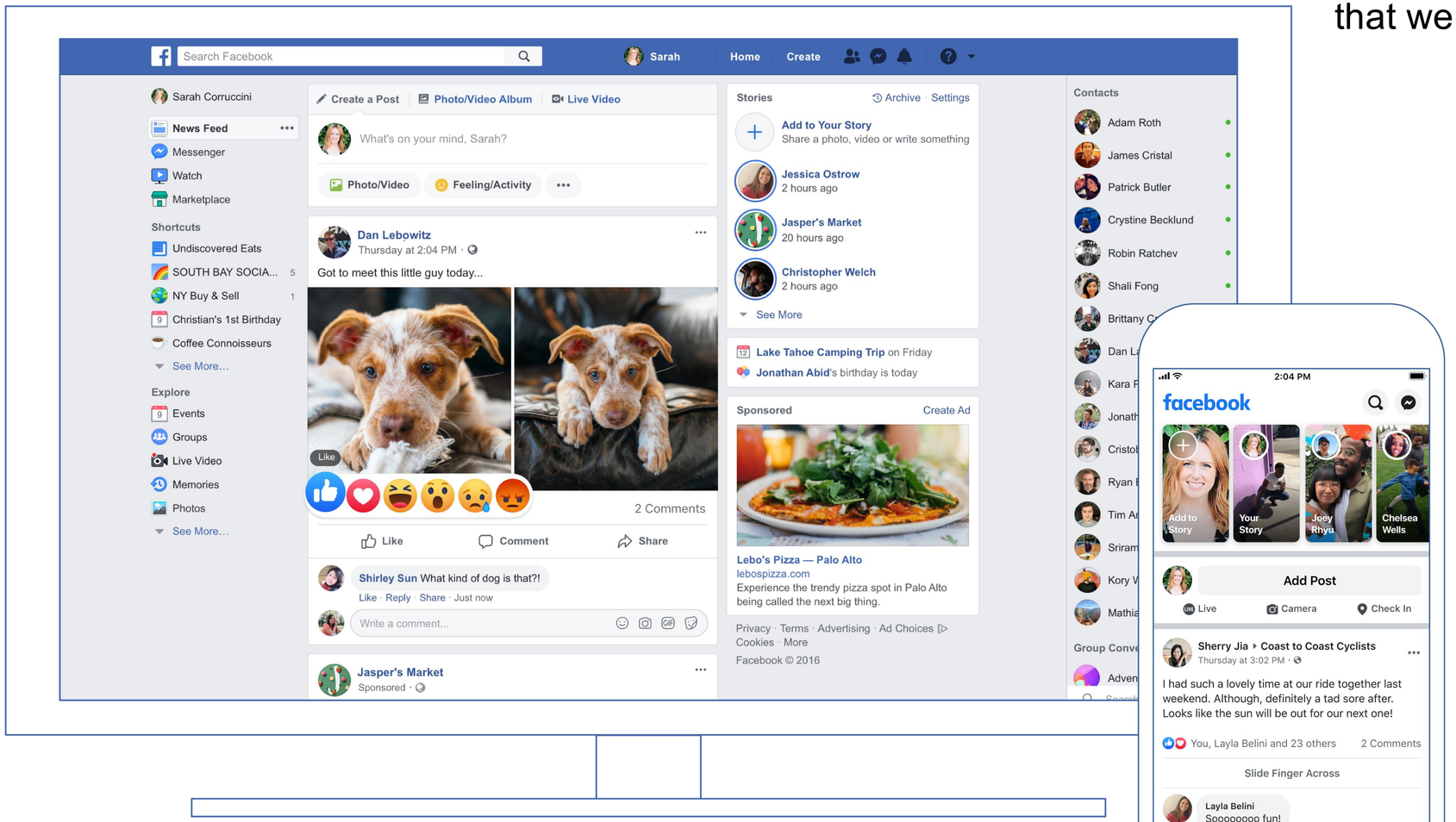
“these are just some of the thousands of signals that may be considered for News Feed ranking.” – Facebook help centre

Study reports on “a test whether posts w emotional content are more engaging”

“The Facebook Study”

Remove positive or negative posts by friends

Measure percentage of all words in broadcasts that were positive or negative

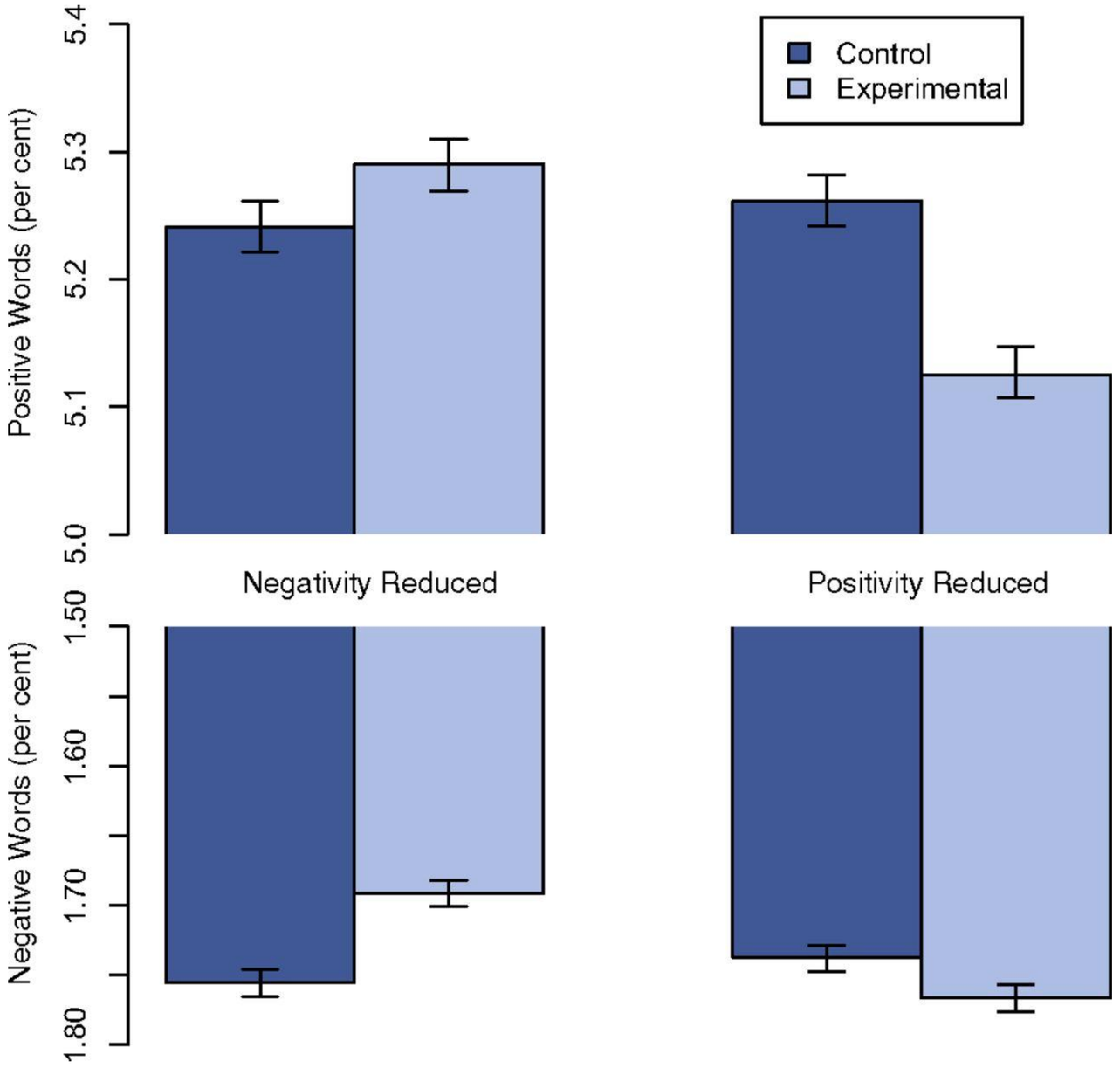


About 155,00 participants per condition

“The Facebook Study”

- Emotion contagion
- Cross-emotional encouragement effect

1. “Overhearing” friend’s emotional state
2. EC does not require nonverbal cues
3. No differences between emotions



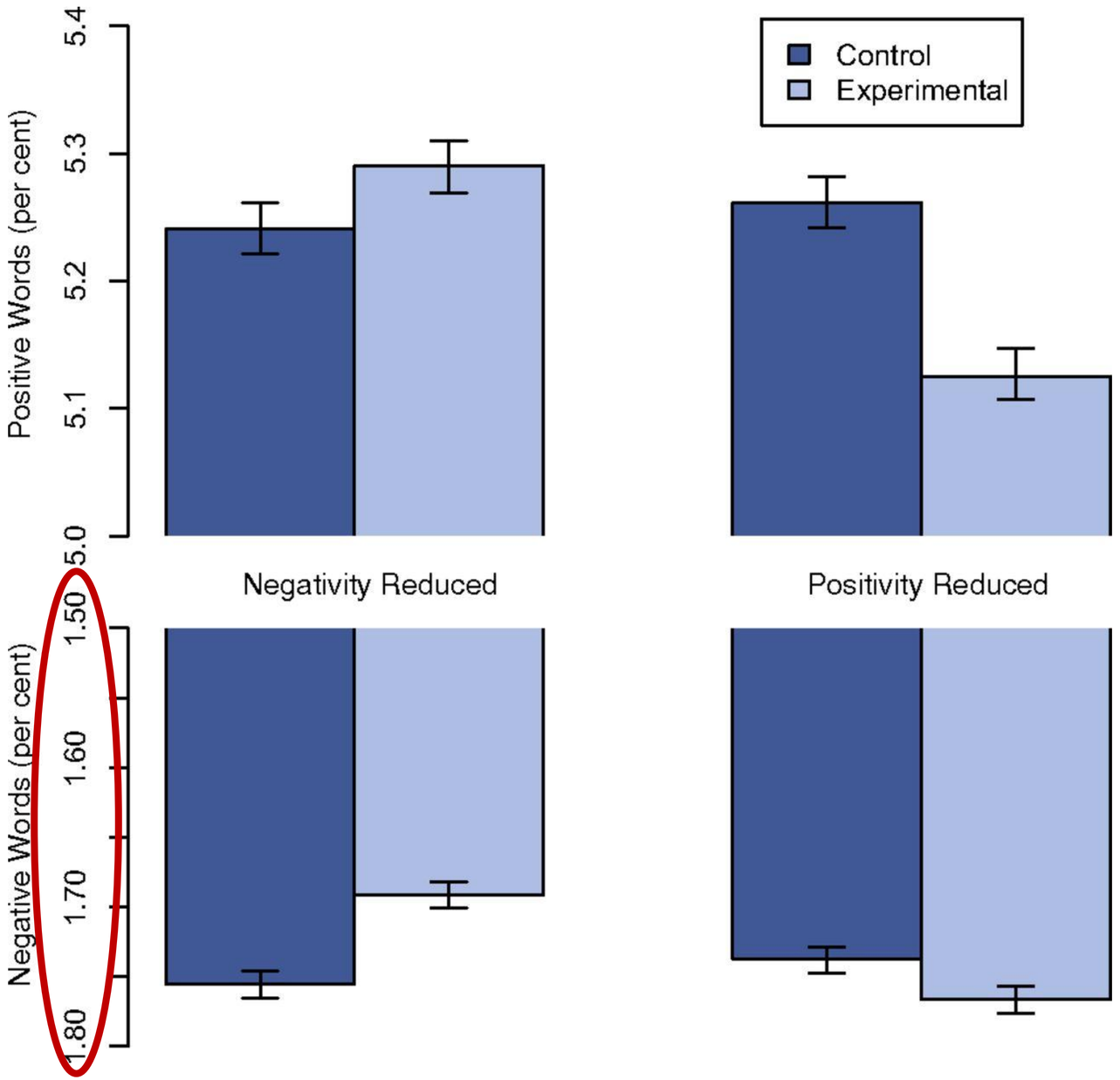
n = 689,003

“The Facebook Study”

- Small effect sizes: d = 0.02 to d = 0.001

“comparable to a hypothetical treatment that increased the average height of the male population in the United States by about one twentieth of an inch (given a standard deviation of ~2.8 inches)” - Tal Yarkoni

- Is this emotion contagion? Or just expression of emotion in response to broadcasts? Almost a form of feedback



Reference slide for self-study

The ethics of the “The Facebook Study”

Reflect of this study:

- What are some of the ethical concerns (if any)?
- What are the responsibilities of the researcher?
- How do researchers navigate collaborations with industry?

Digital emotion contagion

The ethics of the “The Facebook Study”

- Editorial Expression of Concern and Correction published by Editor-in-Chief

Questions have been raised about the principles of informed consent and opportunity to opt out in connection with the research in this paper. The authors noted in their paper, “[The work] was consistent with Facebook’s Data Use Policy, to which all users agree prior to creating an account on Facebook, constituting informed consent for this research.” When the authors prepared their paper for publication in PNAS, they stated that: “Because this experiment was conducted by Facebook, Inc. for internal purposes, the Cornell University IRB [Institutional Review Board] determined that the project did not fall under Cornell’s Human Research Protection Program.” This statement has since been confirmed by Cornell University.

Obtaining informed consent and allowing participants to opt out are best practices in most instances under the US Department of Health and Human Services Policy for the Protection of Human Research Subjects (the “Common Rule”). Adherence to the Common Rule is PNAS policy, but as a private company Facebook was under no obligation to conform to the provisions of the Common Rule when it collected the data used by the authors, and the Common Rule does not preclude their use of the data. Based on the information provided by the authors, PNAS editors deemed it appropriate to publish the paper. It is nevertheless a matter of concern that the collection of the data by Facebook may have involved practices that were not fully consistent with the principles of obtaining informed consent and allowing participants to opt out.


TECHNOLOGY

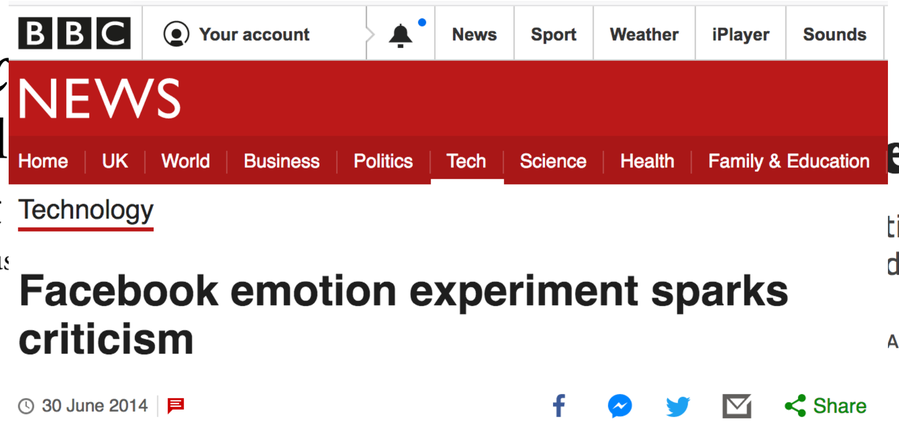
The Atlantic

The Many Reasons to Disl

Manipulation Experiment

The social network is coming under some serious

ADAM CHANDLER JUNE 28, 2014


BBC Your account News Sport Weather iPlayer Sounds

NEWS

Home UK World Business Politics Tech Science Health Family & Education

Technology

Facebook emotion experiment sparks criticism

30 June 2014

Share


SLATE

ebook’s Unethical Experiment

tionally manipulated users’ emotions without their dge.

ALDMAN

JUNE 28, 2014 • 5:50 PM

The ethics of the “The Facebook Study”

- Tal Yarkoni:

Four reasons why it is overblown:
-It did not add content to induce, but removed content
- Manipulation is part of the Facebook experience, no new risks
- These A/B experiments are run everyday by different companies
- It can be in your benefit

Critique: counterproductive (sharing of data and results less likely)

Further:
-A/B testing is not known to a lot of people (at least in 2014)
-The issue of ethics approval (depends on origin of study)
-What quantifies as research?
- Informed consent is not always possible (e.g., observational studies)

Beyond digital emotion contagion

https://www.nytimes.com/interactive/2019/06/08/technology/youtube-radical.html

Going back to machine behaviour

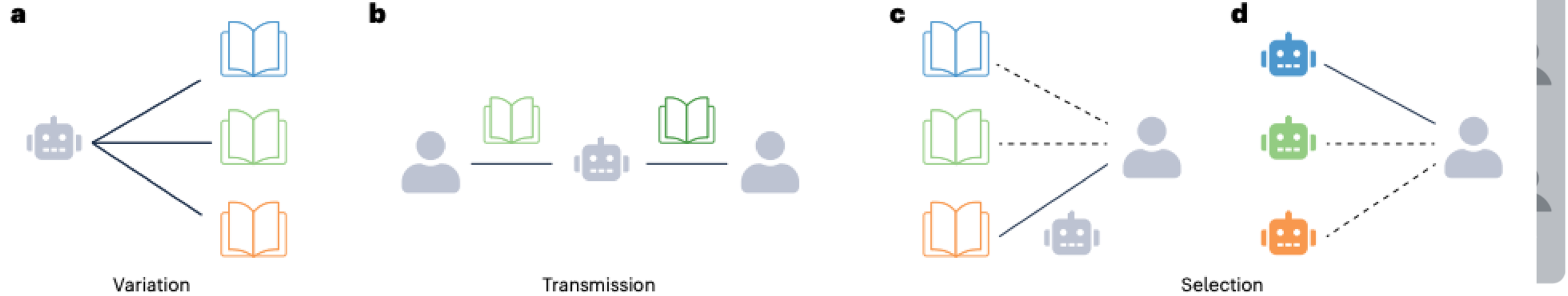
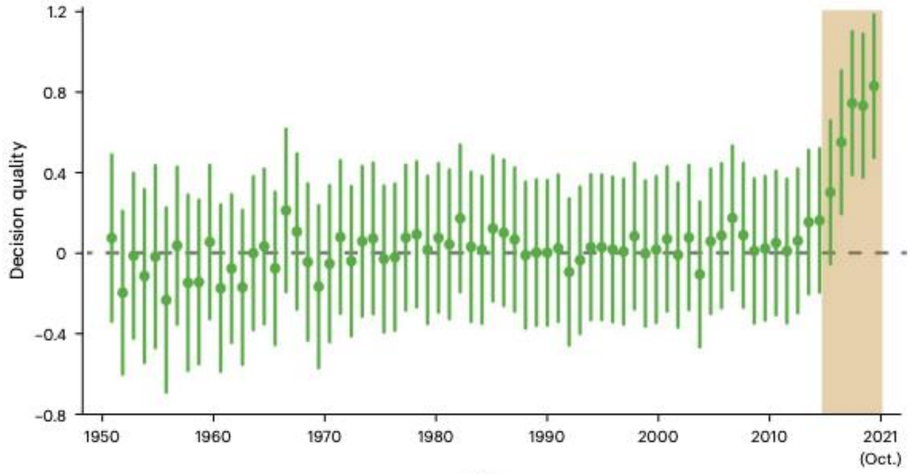
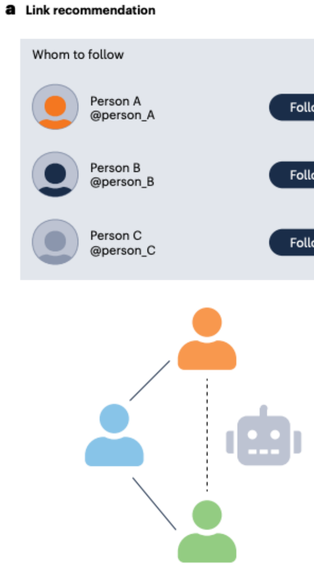
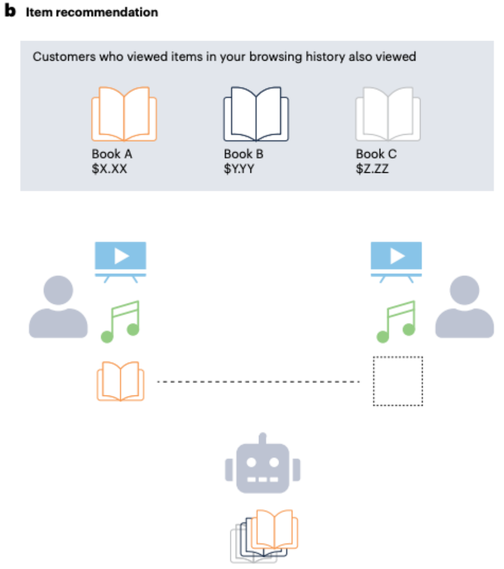


Fig. 1 | Examples of machine culture. a, Novel cultural artefacts are generated through machines. **b**, Machines transmit and potentially mutate cultural artefacts. **c**, Machines select between different cultural artefacts. **d**, Humans select among diverse machines.







https://doi.org/10.1038/s41562-023-01742-2

ın[1,11], Jean-François Bonnefon[2,11],
[1], Anne-Marie Nussberger[1,11],
Thomas L. Griffiths[5],
hard McElreath[8],
& Iyad Rahwan[1,11]

Today’s lecture

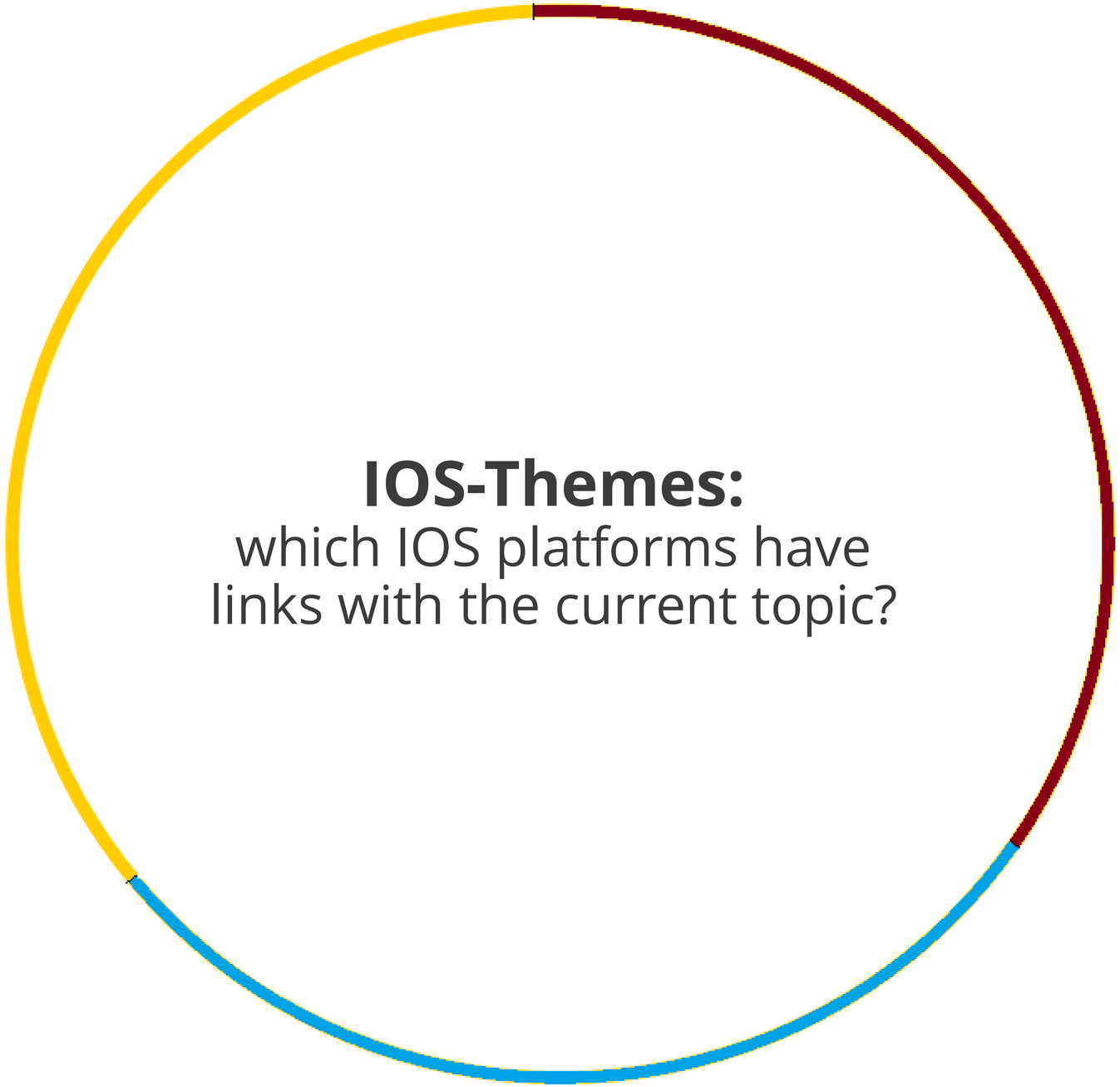
Aritificial Intelligent Systems

Digital traces

Emotion Contagion

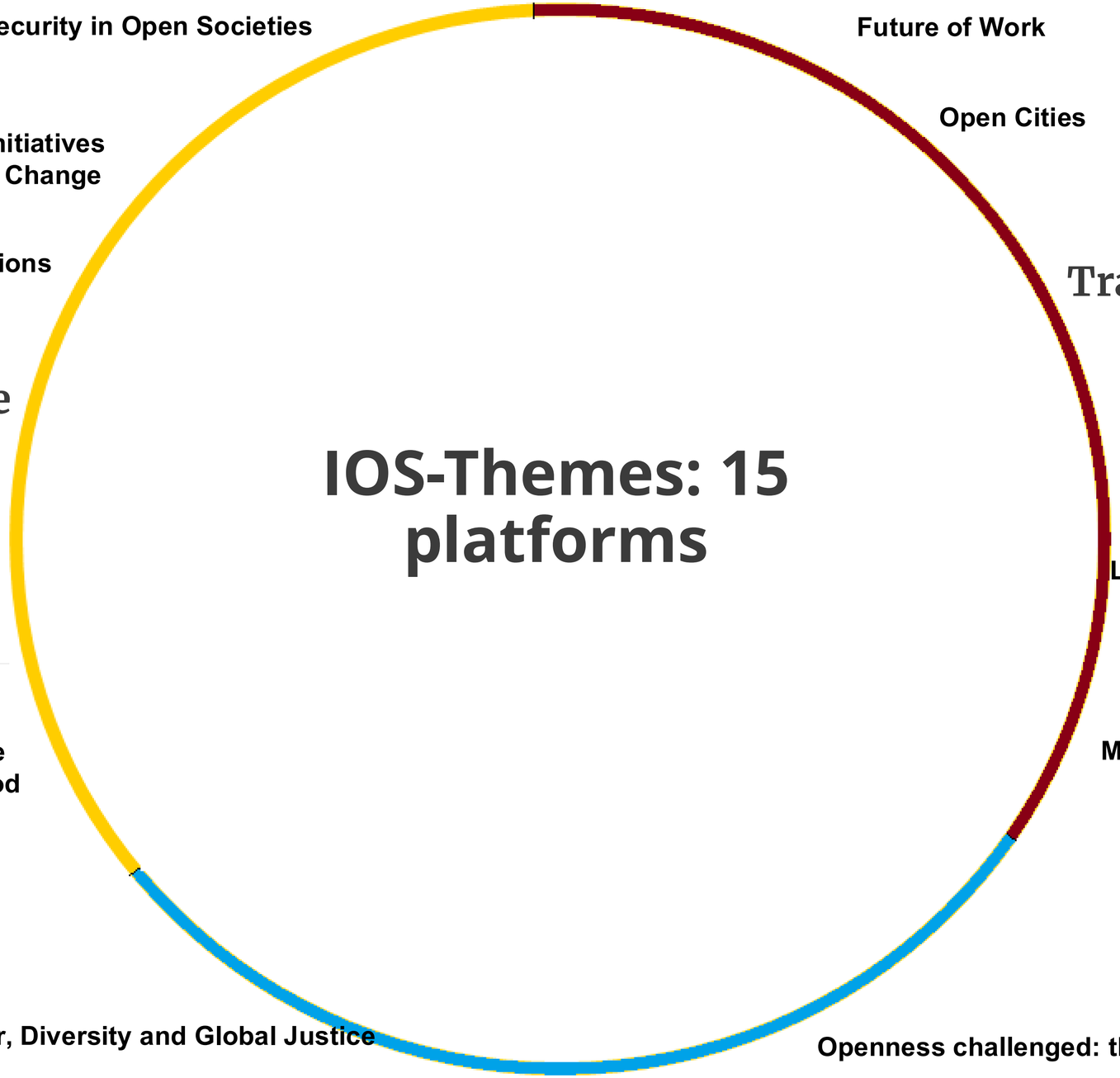
IOS-Themes:
SWOT analysis of current topic

	Positive	Negative
Internal	**Strenghts** • Test theories • Better measures • Diverse sample • Persuasion	**Weaknesses** • Bias • Consent/privacy • Ownership
External	**Opportunities** • Personalisation • Validity • Impact • Legislation	**Threats** • Surveillance • Autonomy • Persuasion • Online targeting • Ethics


IOS-Themes:
which IOS platforms have
links with the current topic?

Utrecht University

29-4-2026



Security in Open Societies

Future of Work

Open Cities

Bottom-up Initiatives for Societal Change

Behaviour and Institutions

Transitions and well-being

Democracy and good governance

Fair Transitions

Contesting Governance

Longtermism and Institutional Change

Futures of Democracy

The Transactional State as an Institution for Good

Markets and Corporations

Gender, Diversity and Global Justice

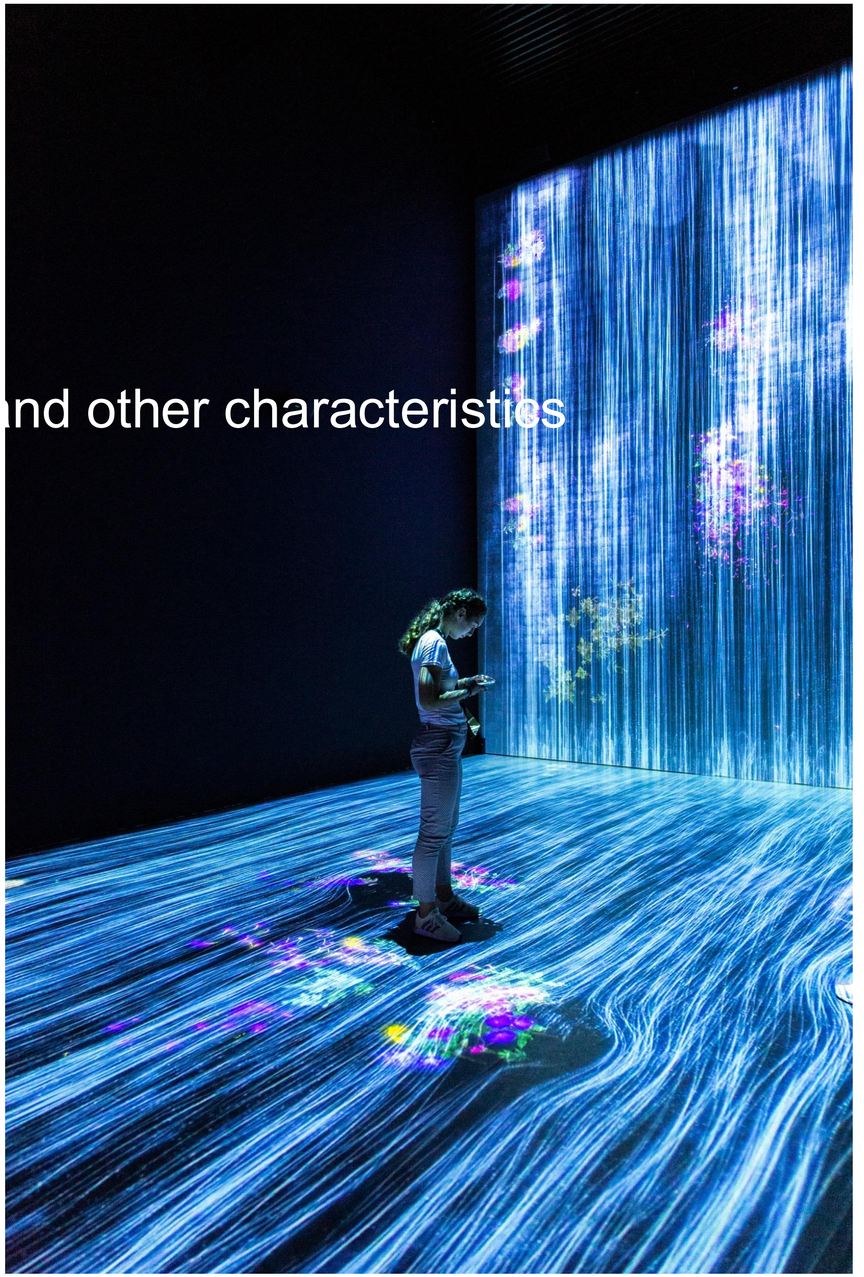
Openness challenged: the university at risk?

In/Equality

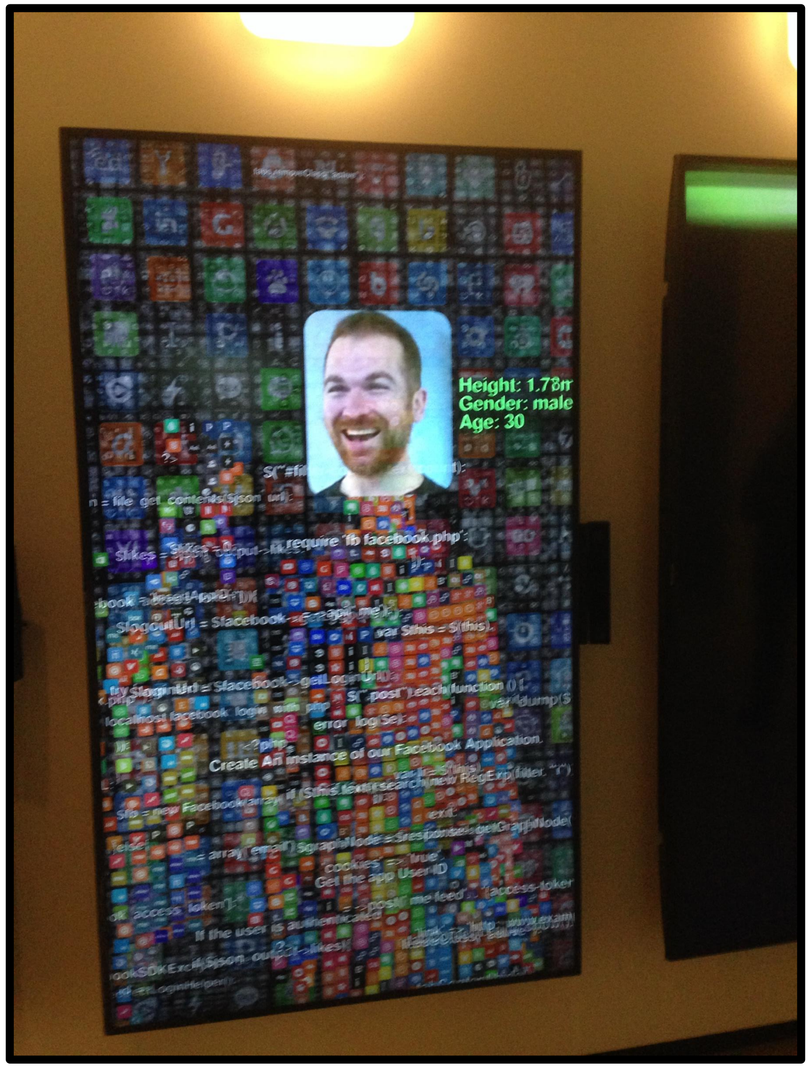
Equality and diversity

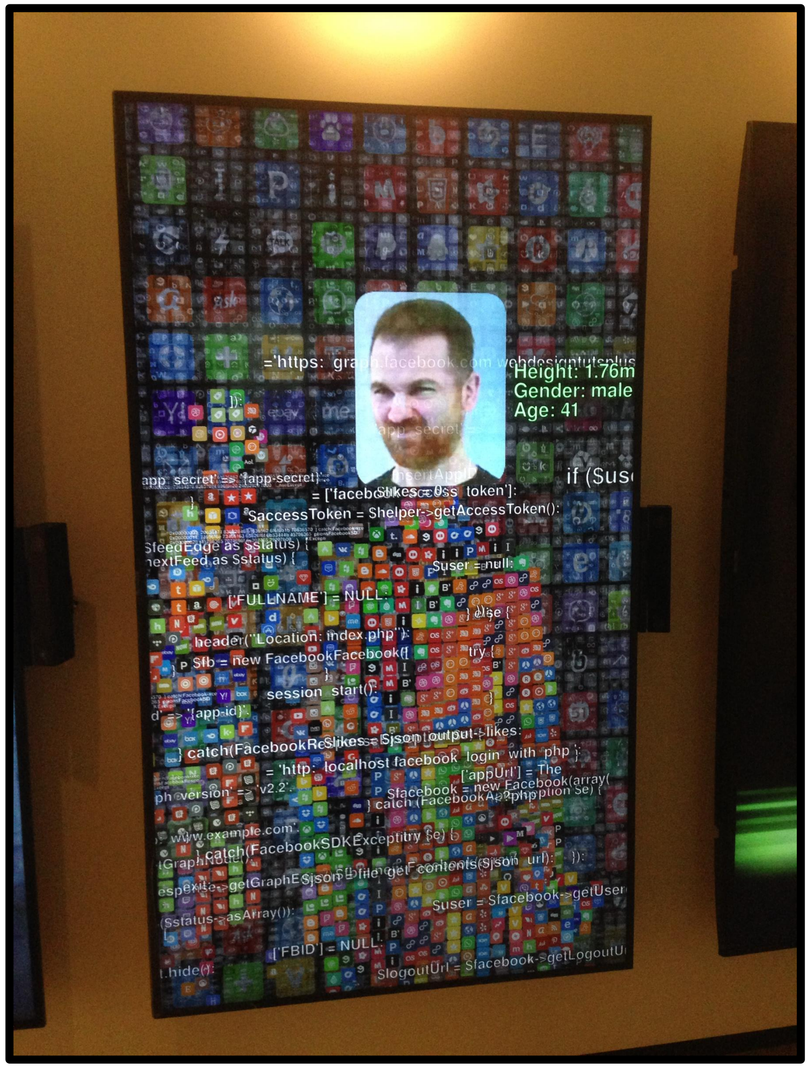
This lecture’s objectives

- Discuss the role of psychology in artificial intelligent systems
- Discuss how digital traces can be used to distill personality traits and other characteristics
- Discuss digital emotion contagion and how to measure this
- Acknowledge ethical considerations of AI systems



Beyond 1984 and Black Mirror



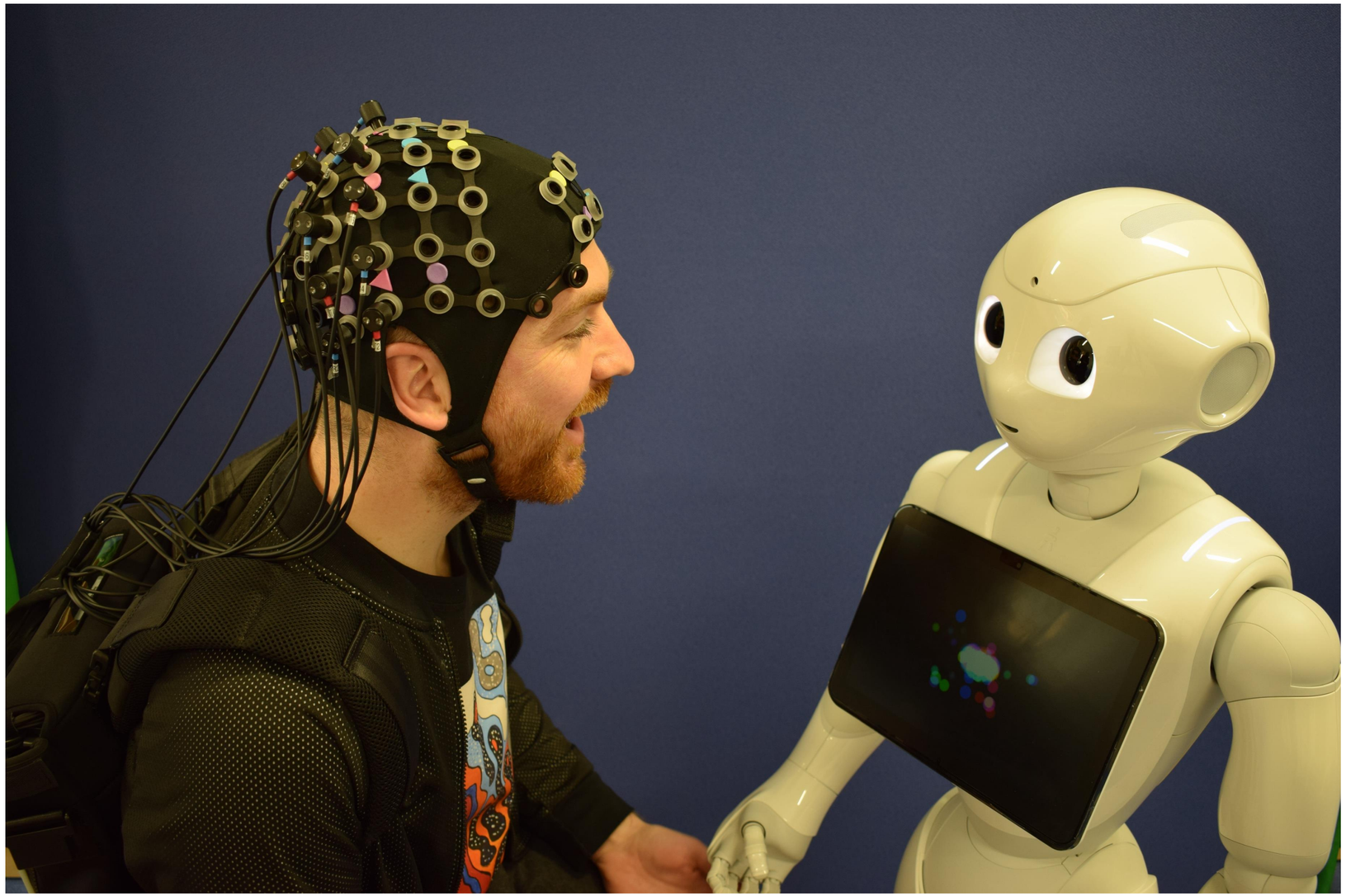


Questions?

Contact

r.hortensius@uu.nl

https://human-plus.gitlab.io/



human+ social cognition

Interested in the F-AI-MILY project:
let's connect