How online research opens new doors for computational psychiatry

Claire Gillan, PhD

Assistant Professor of Psychology and MQ Fellow

Trinity College Dublin

www.gillanlab.com

@clairegillanTCD

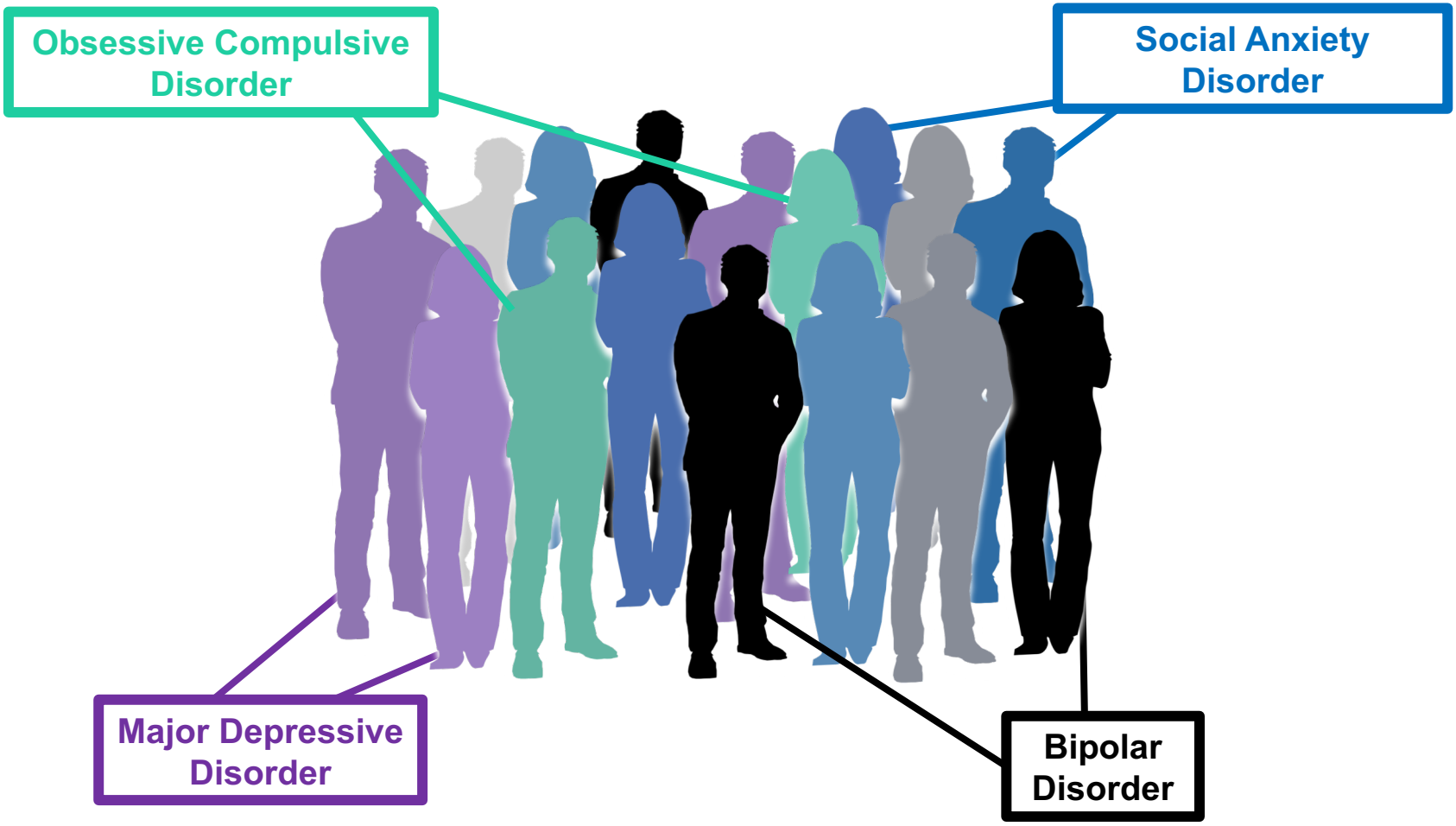
Outline

- Why take psychiatry research online?
- Online methods
  - *Crowdsourcing*
  - *Smartphones*
  - *Scraping*
  - *Proxies for physiology*
- Characteristics of online samples
  - *Are they representative?*
  - *Are mental health data valid?*
  - *Are the data of acceptable quality?*
  - *Are findings relevant for diagnosed patients?*

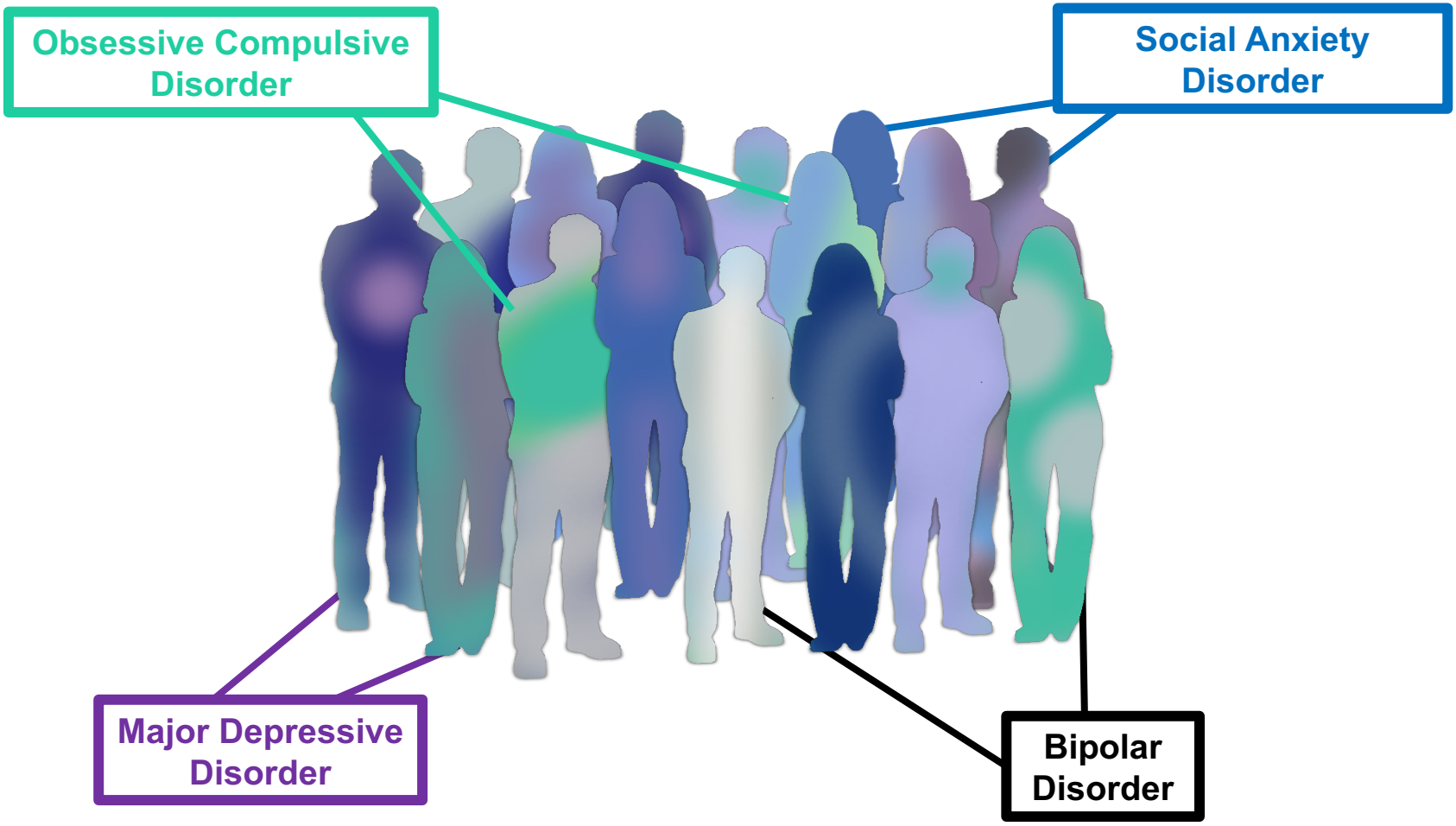
Spoiler: yes.

Why take psychiatry research online?

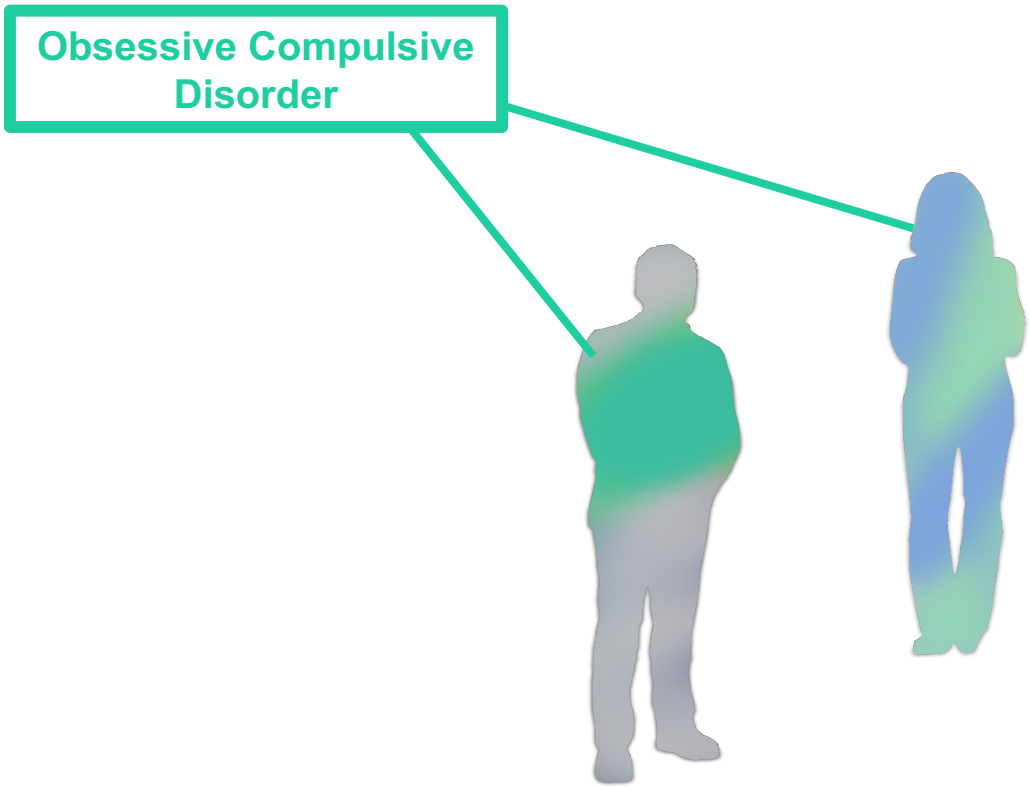
The problem: **DSM disorders are the "ground-truth" for research**



The problem: **DSM disorders are the "ground-truth" for research**

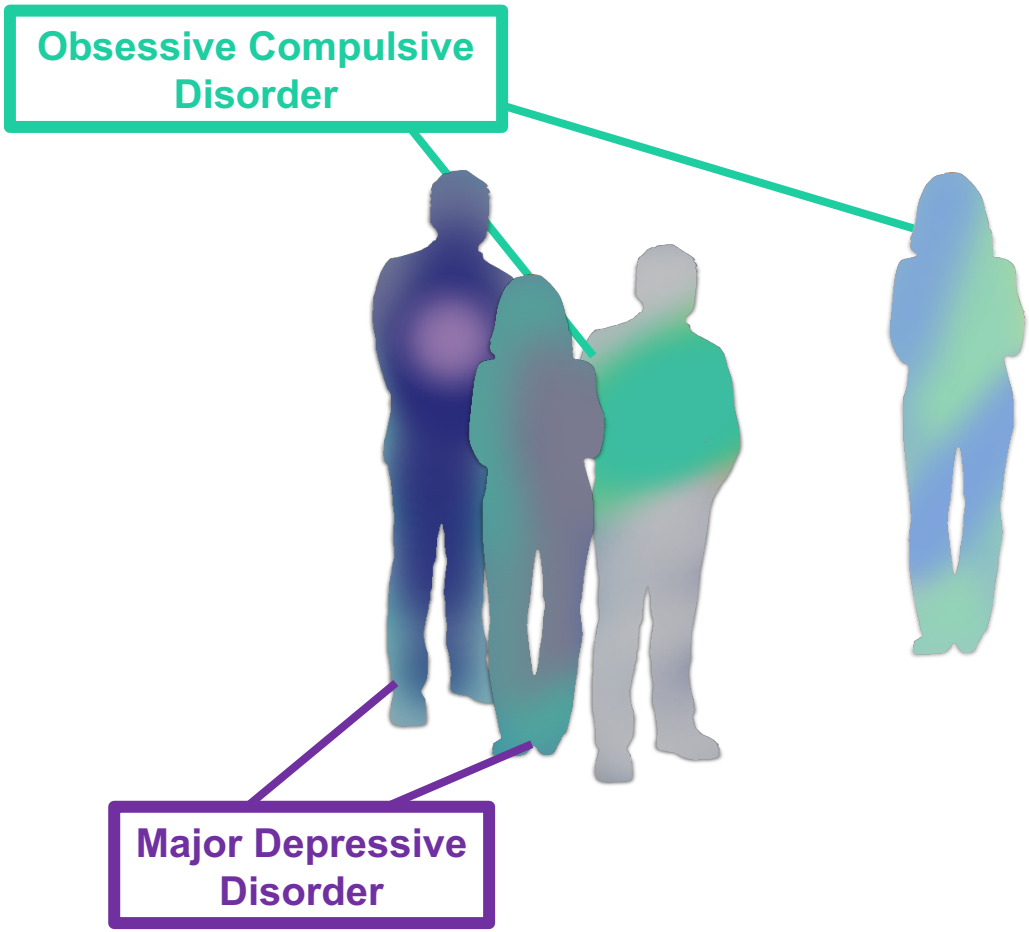


The problem: **DSM disorders are the "ground-truth" for research**



- Patients are heterogeneous within-disorder

The problem: **DSM disorders are the "ground-truth" for research**



- Patients are heterogeneous within-disorder
- Patients are similar across-disorder

The solution: **redraw the lines**

Symptom Dimension 1

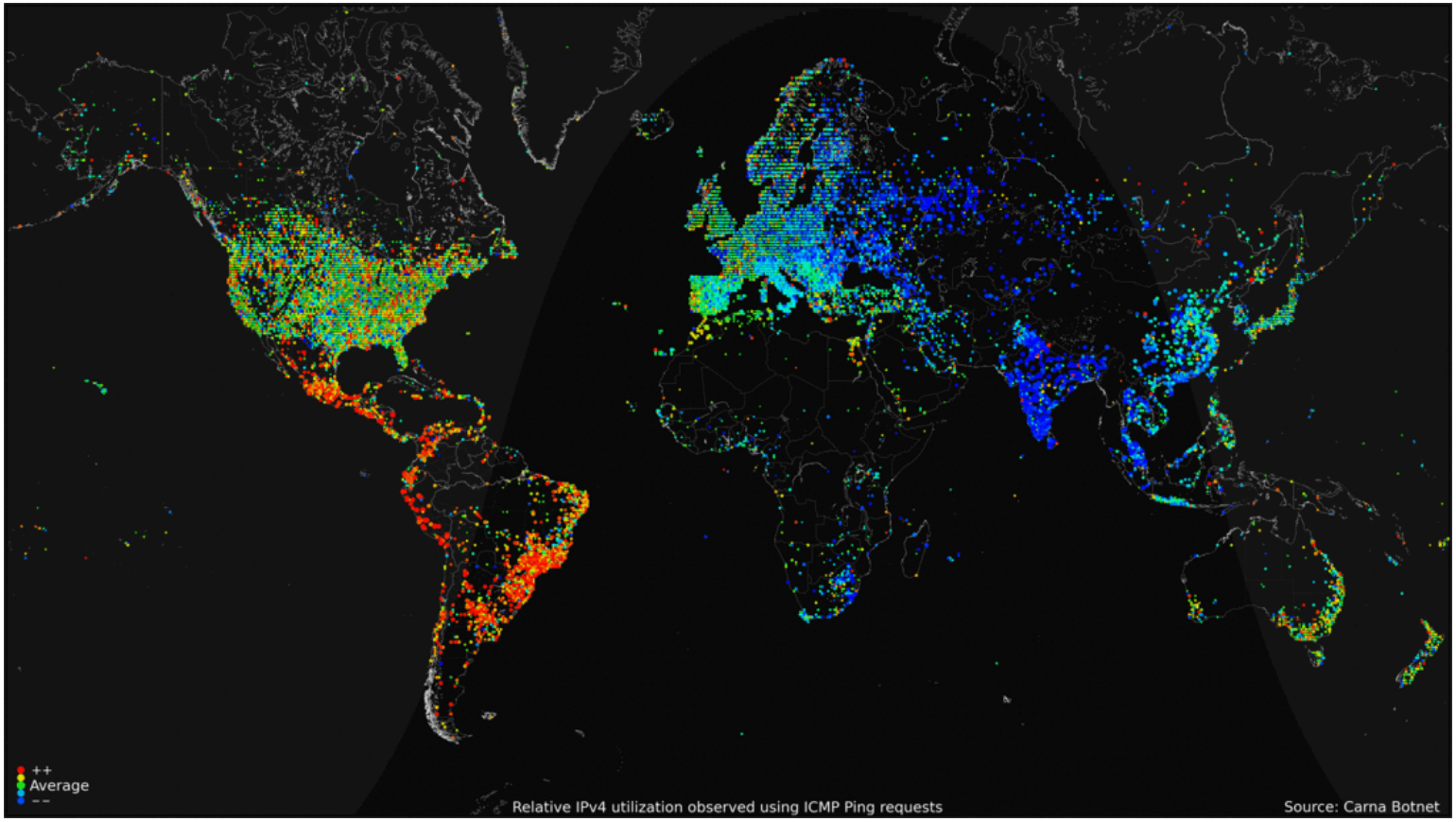
Symptom Dimension 2

Symptom Dimension 3

Treatment Dimensions? **Brain** Dimensions? **Genetic** Dimensions?

Cognitive Dimensions? **Environmental** Dimensions?

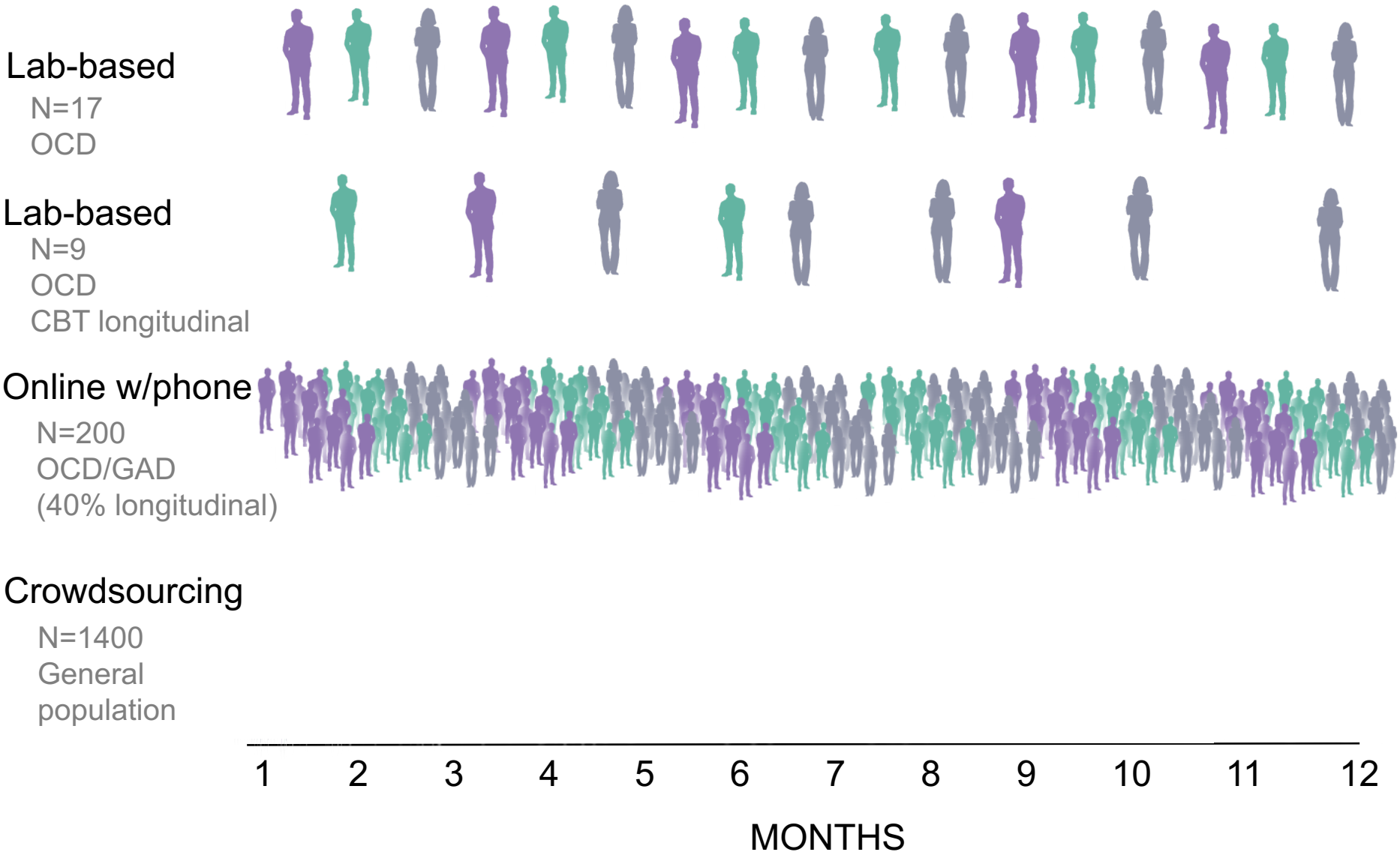
To do this well, we need big samples.



People living in Dublin : 500,000

People using the Internet : 3,200,000,000

Why you should think about going online



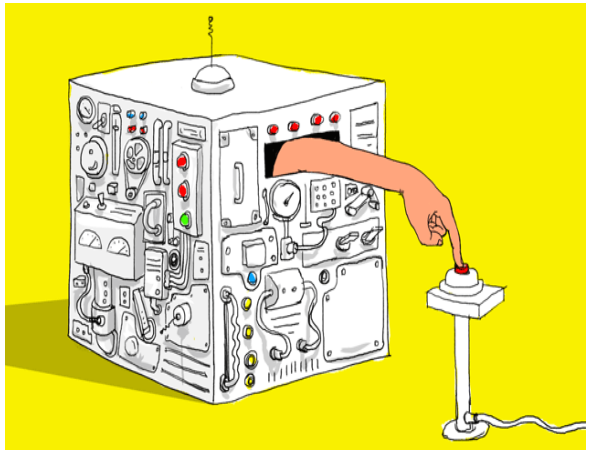
Why you should think about going online

- Sample size (>3.2 billion people use the internet)
- Access select populations (*geography*, *age, race, clinical condition, socio-economic status, etc.*)
- Speed (hundreds/thousands of subjects per day)
- Low cost (but it depends)
- Anonymous participation (sensitive populations)
- Reproducibility/standardization (same test, different lab)
- Exploratory vs. confirmatory research

Crowdsourcing Platforms

Crowdsourcing Platforms

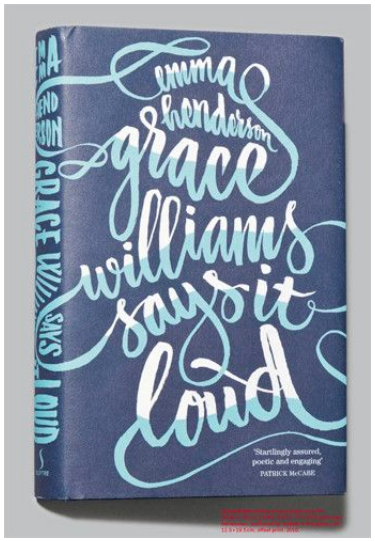
amazon mechanical turk

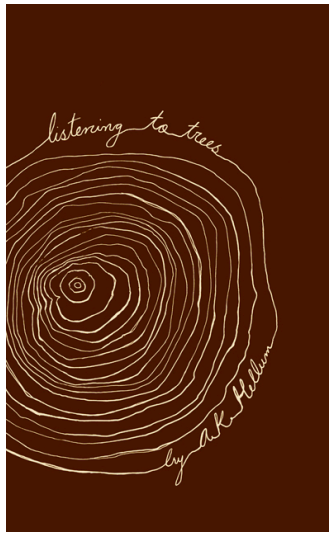


There are things that humans can do better than computers

1. Internal Amazon Problems:

What is the title of this book?



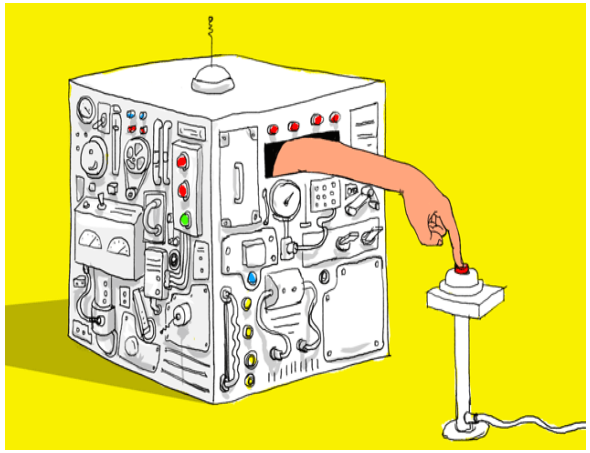


Crowdsourcing Platforms

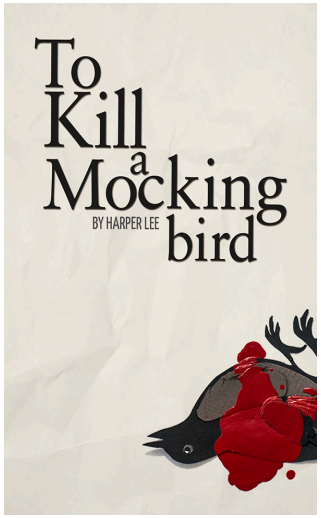
amazon mechanical turk

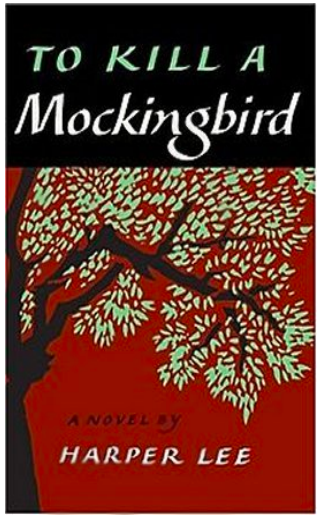
1. Internal Amazon Problems:
   What is the title of this book?

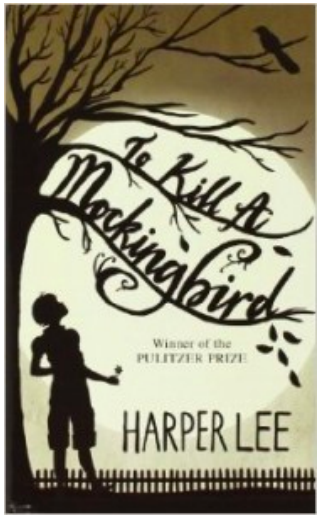
2. Internal Market Research
   Which cover do you prefer?



There are things that humans can do better than computers

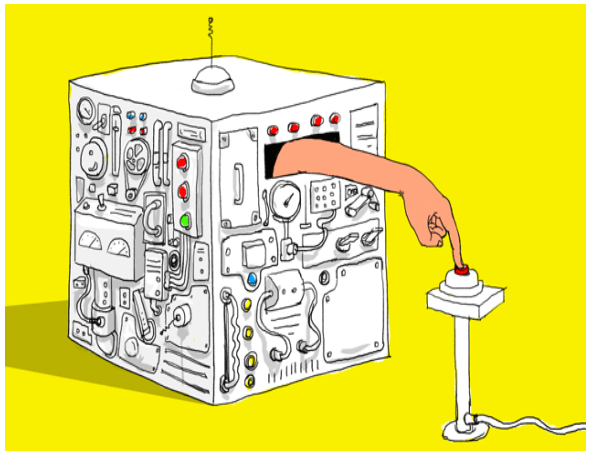






Crowdsourcing Platforms

amazon mechanical turk



There are things that humans can do better than computers

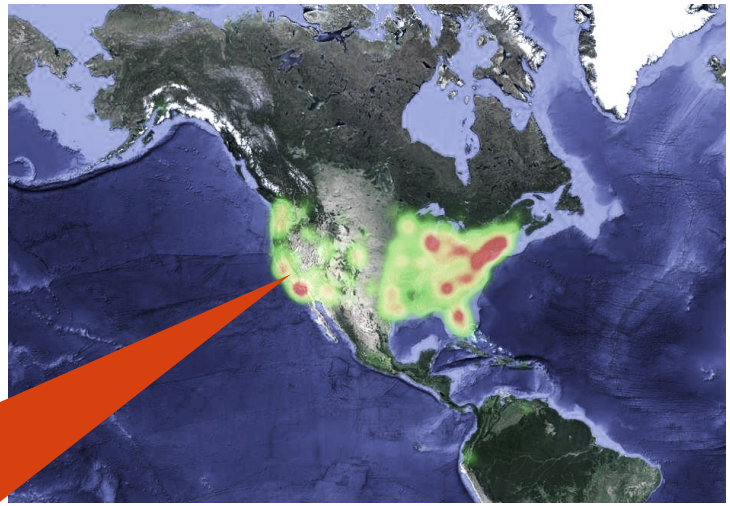
1. Internal Amazon Problems:
   What is the title of this book?

2. Internal Market Research
   Which cover do you prefer?

3. Opens to everyone:
   Transcribing audio files, writing descriptions of complex scenes, answering surveys, ranking products and ***full-scale psychology experiments***

Crowdsourcing Platforms

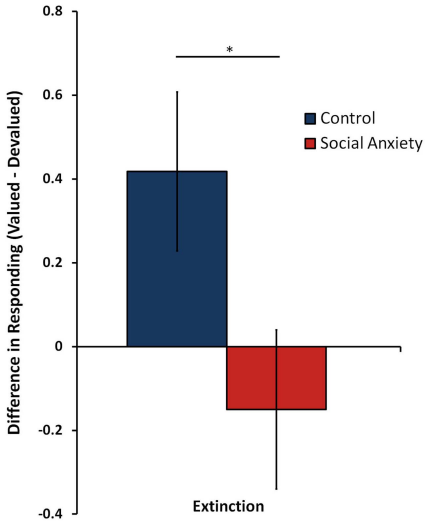
Conduct psychology research with sophisticated, interactive cognitive tests

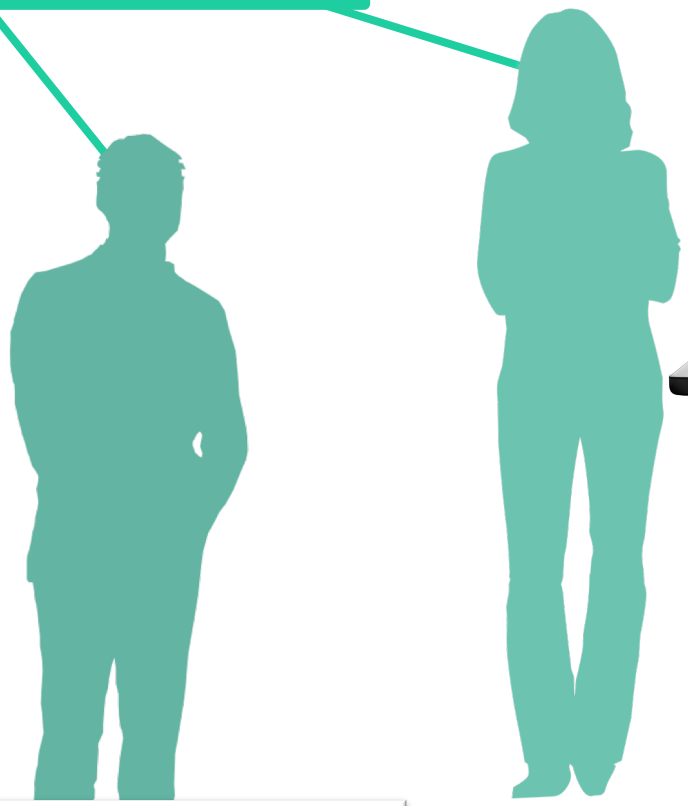


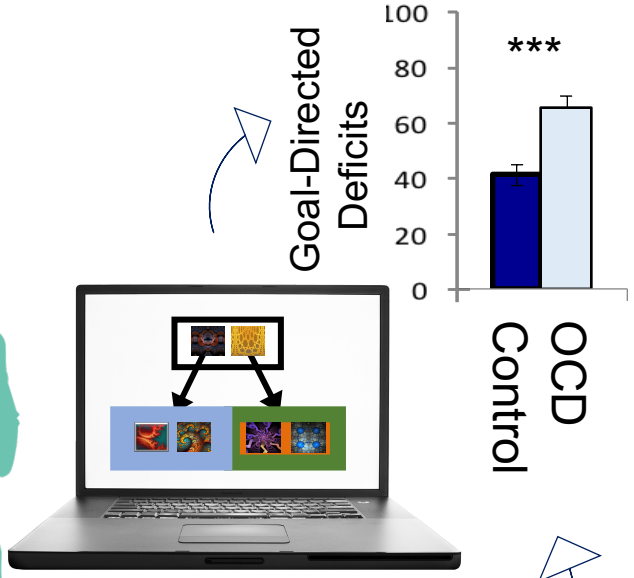
- Recruit from this **global workforce** at an unprecedented rate.
  e.g. N>1400 in 10 days
- Reproduces classic psychology effects

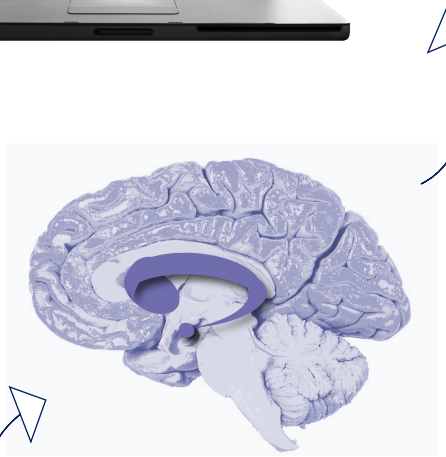
Crump et al., 2013

Obsessive-Compulsive Disorder









OPEN ACCESS Freely available online

PLOS ONE

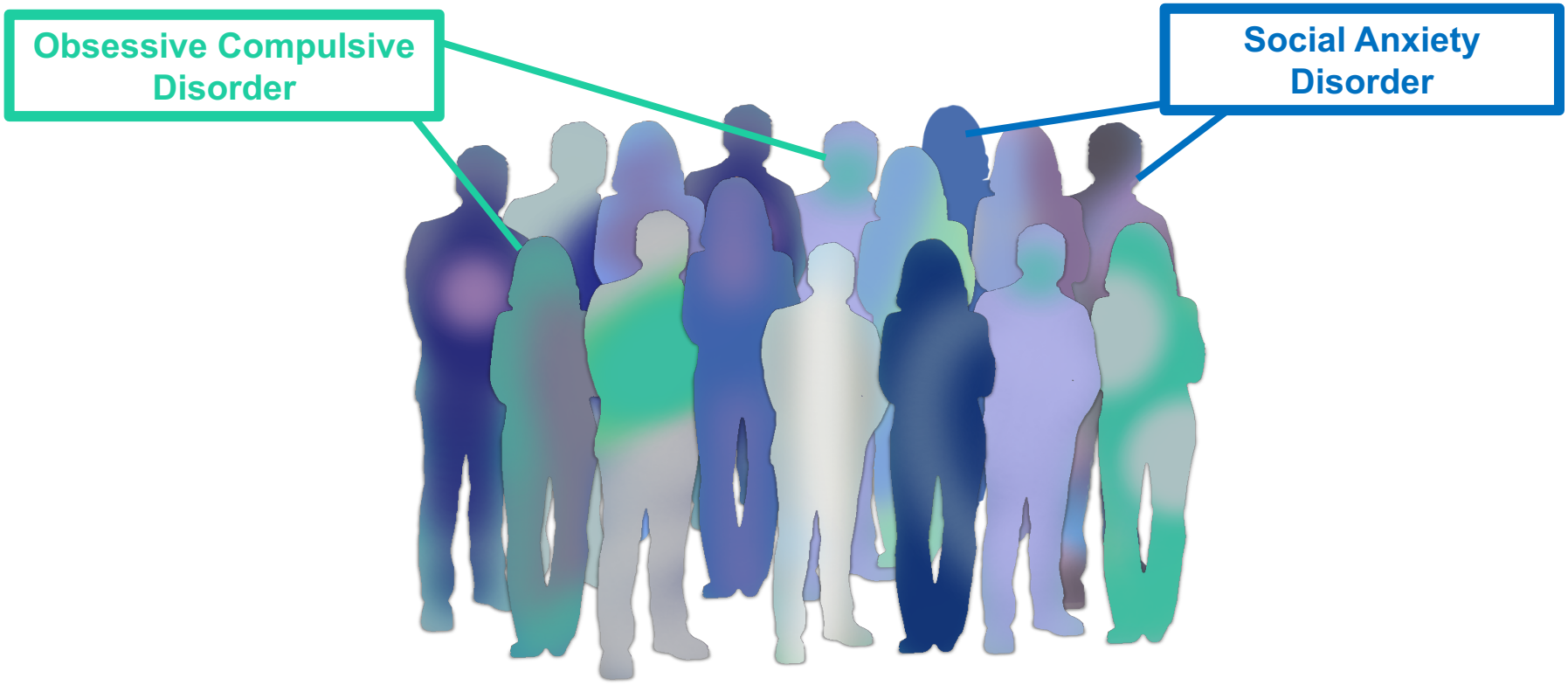
Impairments in Goal-Directed Actions Predict Treatment Response to Cognitive-Behavioral Therapy in Social Anxiety Disorder

Gail A. Alvares, Bernard W. Balleine, Adam J. Guastella*

Brain & Mind Research Institute, The University of Sydney, Sydney, New South Wales, Australia

Gillan et al., American Journal of Psychiatry, 2011

Are goal-directed deficits characteristic of a trans-diagnostic dimension?

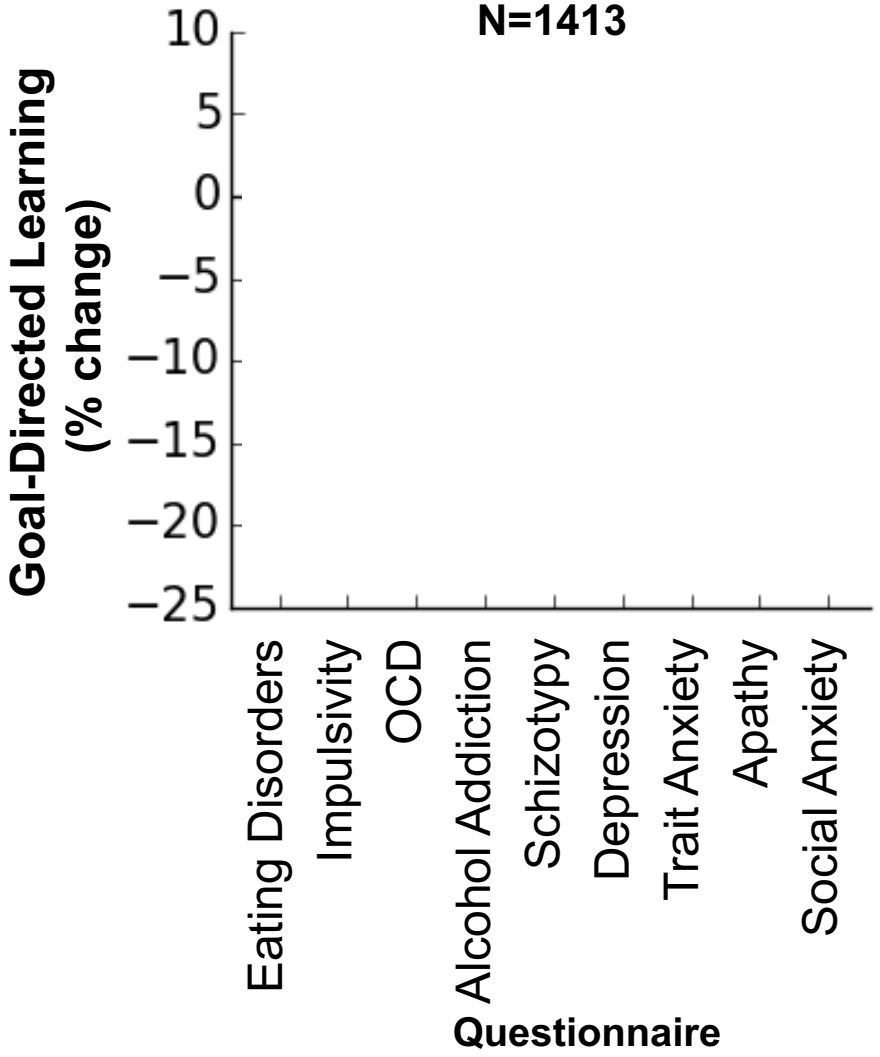


Symptom Dimension 1 Symptom Dimension 2 Symptom Dimension 3

Gillan et al., eLife, 2016

Are goal-directed deficits characteristic of a trans-diagnostic dimension?

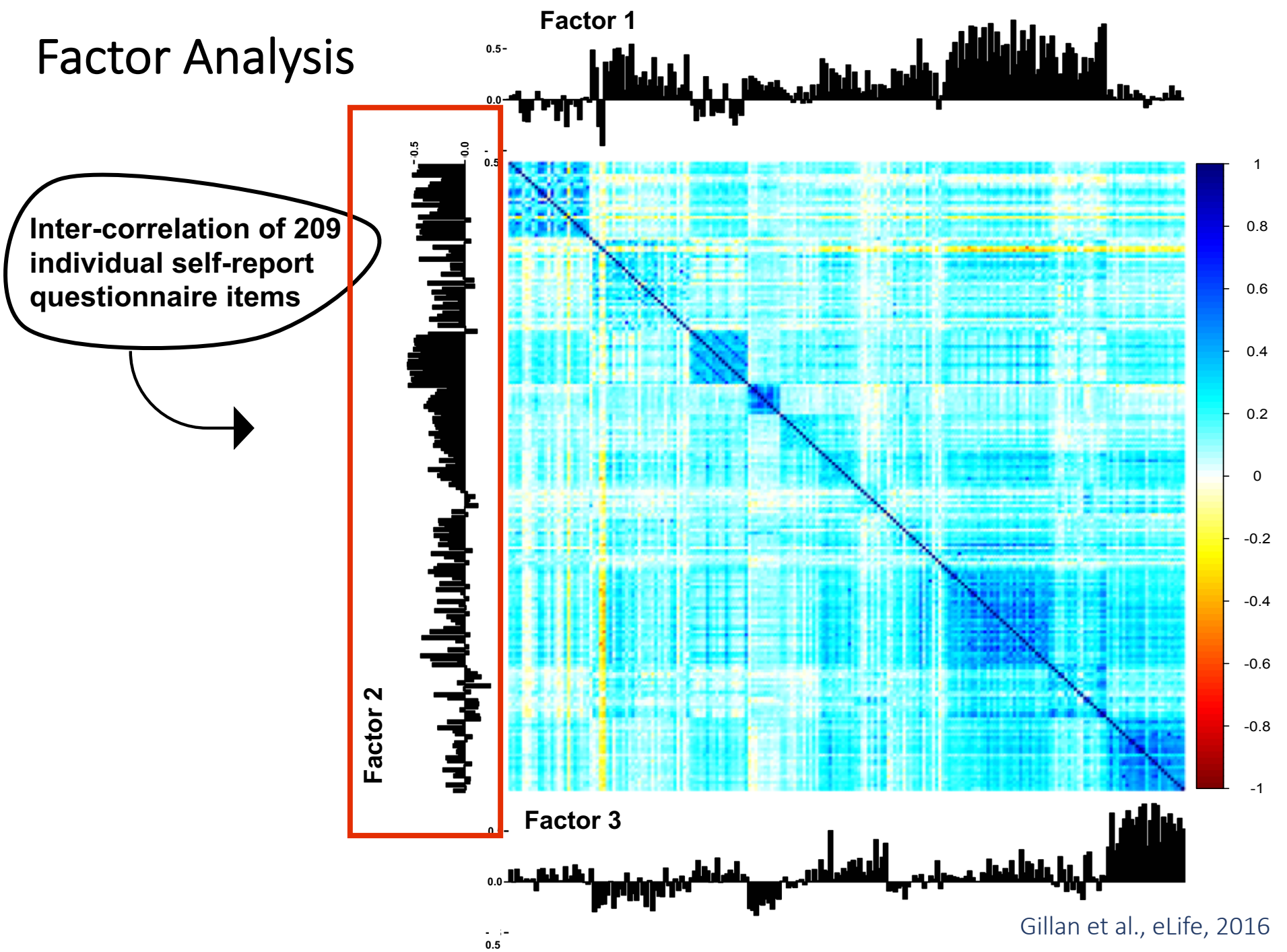
But… these scores reflect overlapping constructs



* p<.05 p<.01 p<.001

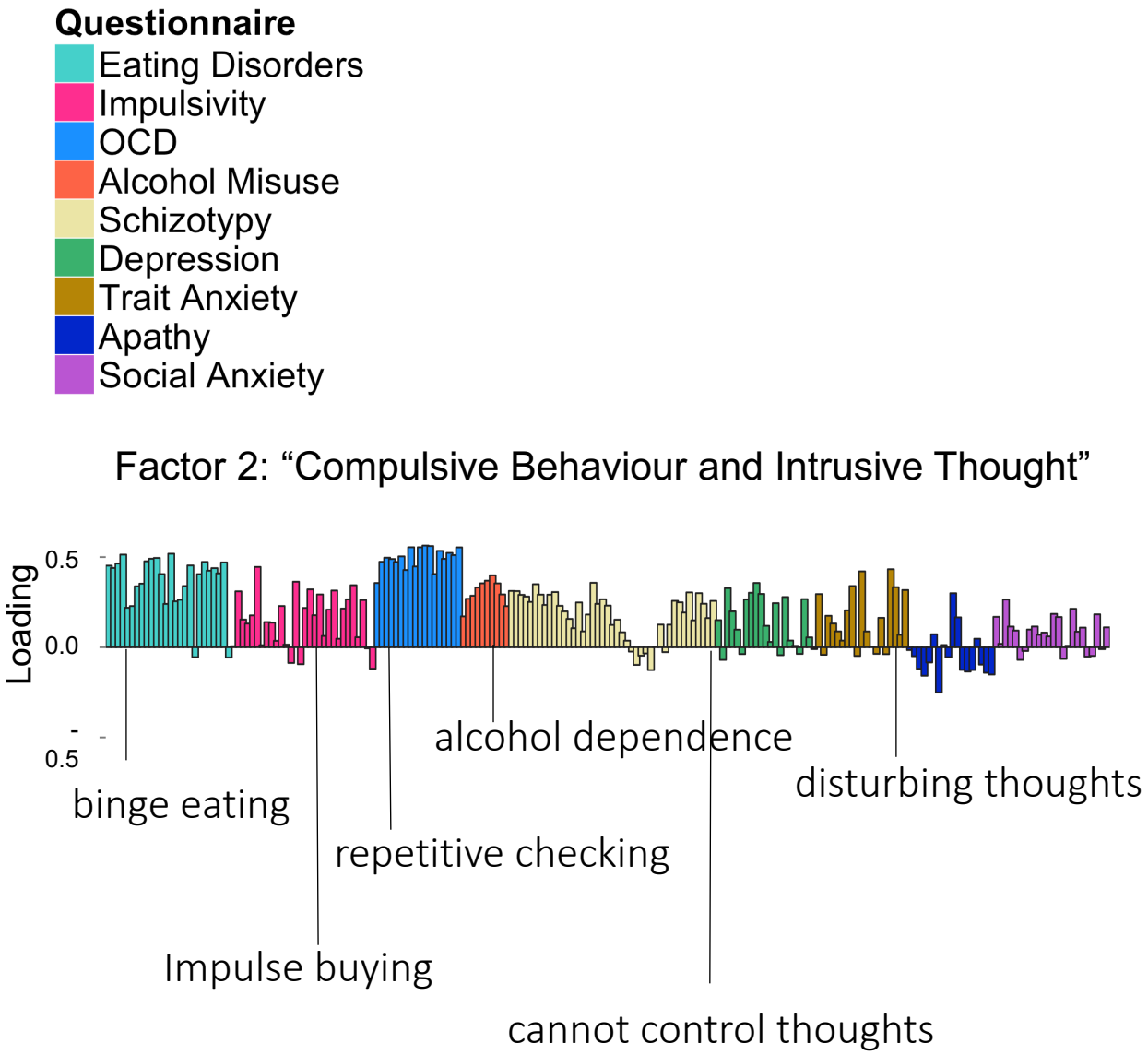
Gillan et al., eLife, 2016

Factor Analysis



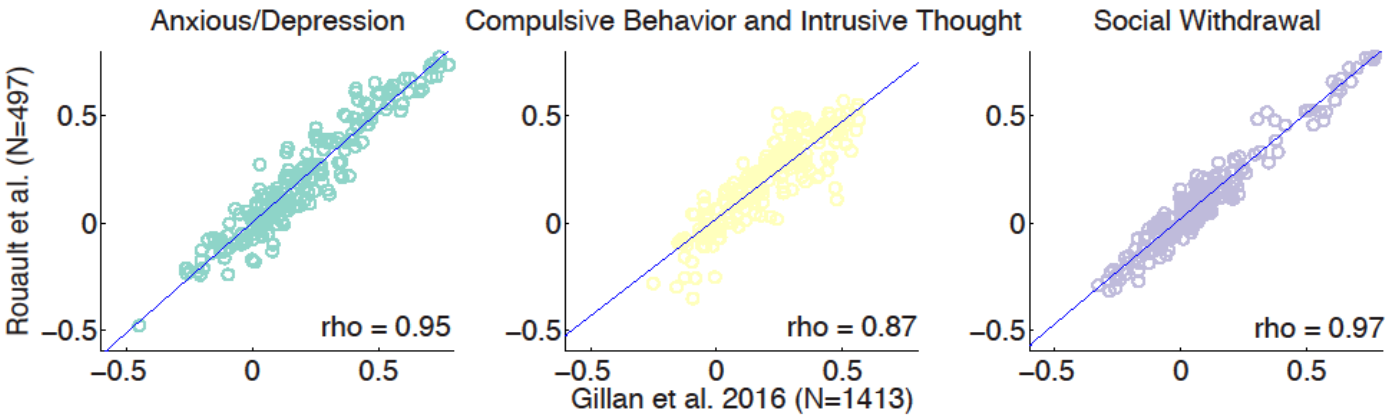
Gillan et al., eLife, 2016

*p<.05 ** p<.01 ***p<.001



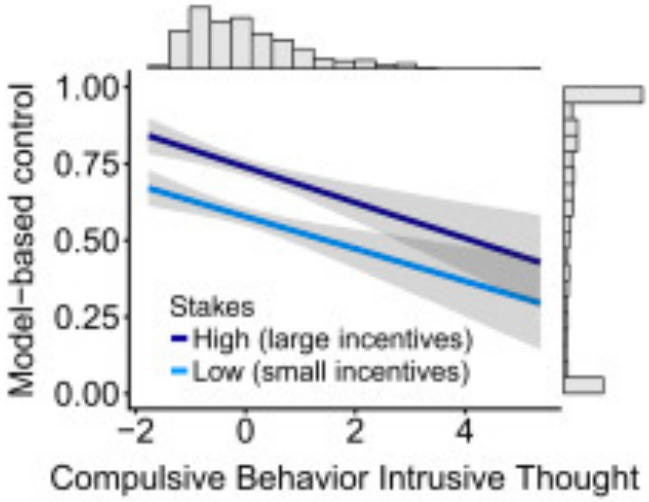
Gillan et al., eLife, 2016

The factor structure can be independently reproduced



Rouault*, Seow*, et al., Biological Psychiatry, 2018

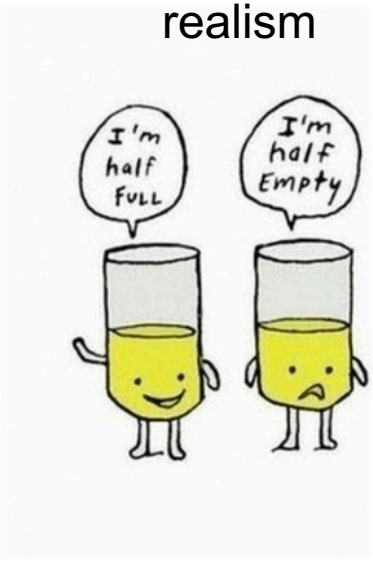
The association with goal-directed planning replicates

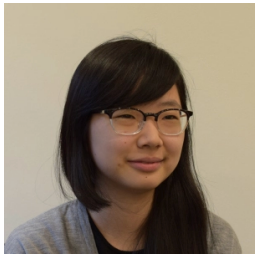


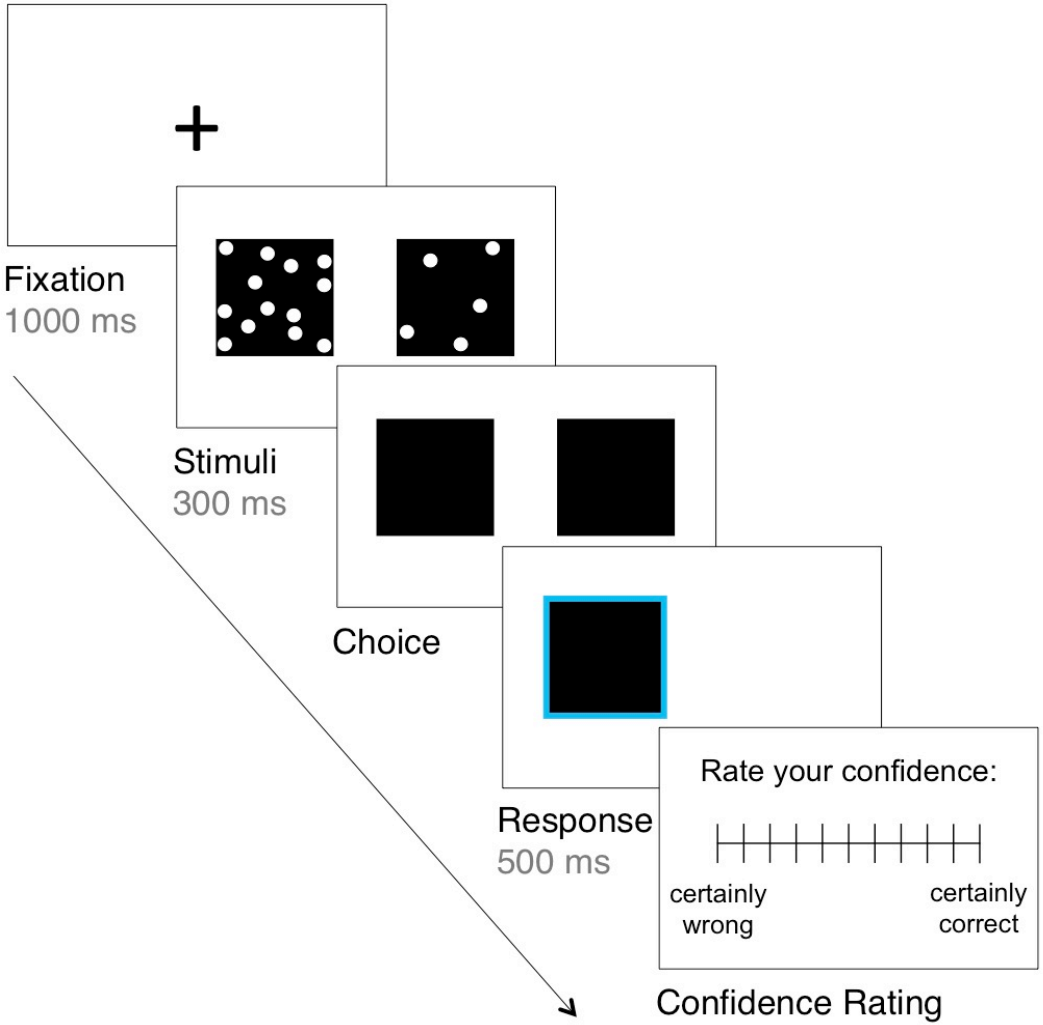
Patzelt et al., Biological Psychiatry, 2019

Can we use this method to study metacognition in anxious-depression?

Depressive realism

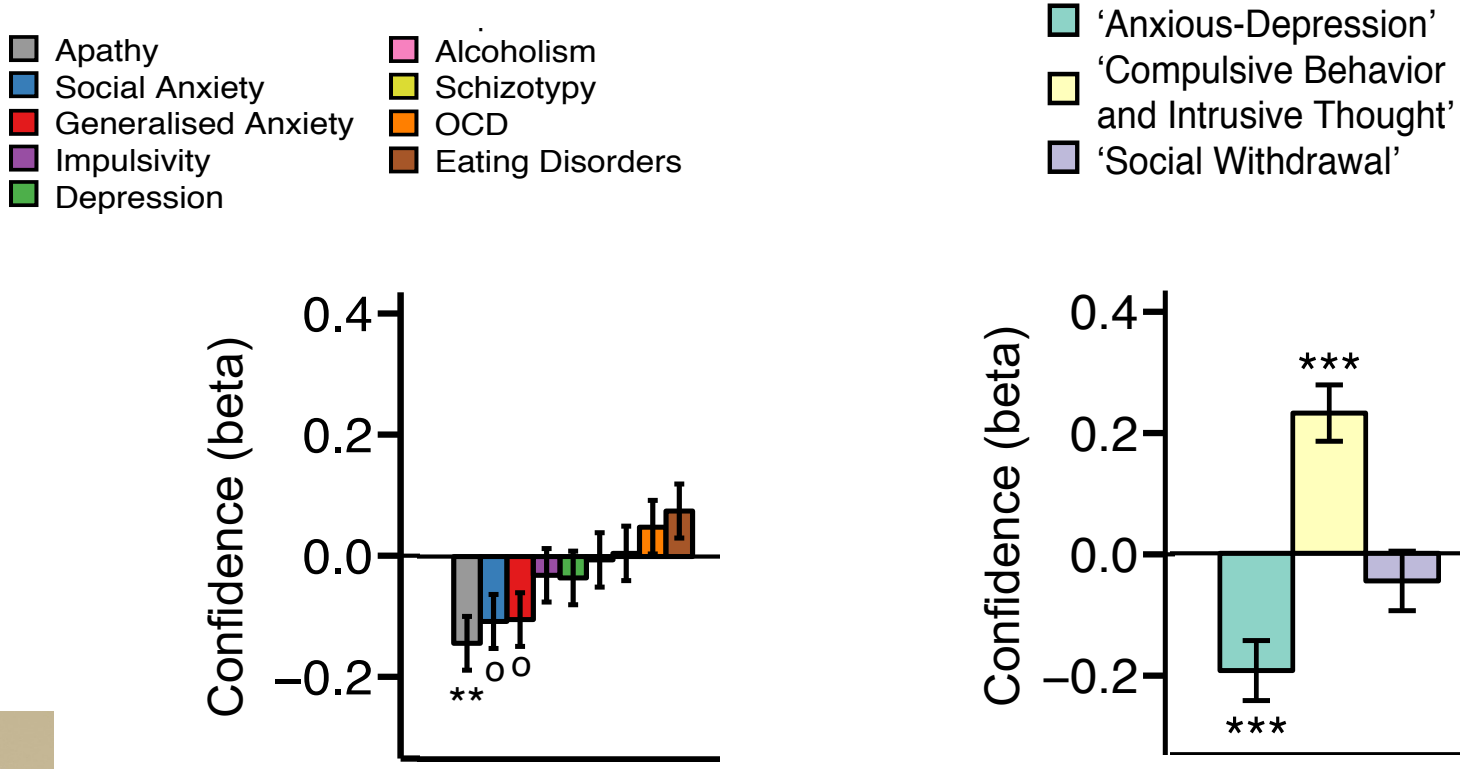


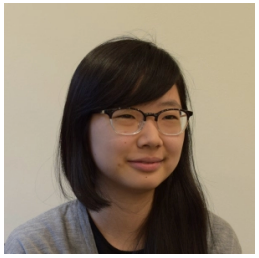




Rouault*, Seow*, Gillan & Fleming, Biological Psychiatry, 2018

Is confidence expressed transdiagnostically?





Rouault*, Seow*, Gillan & Fleming, Biological Psychiatry, 2018

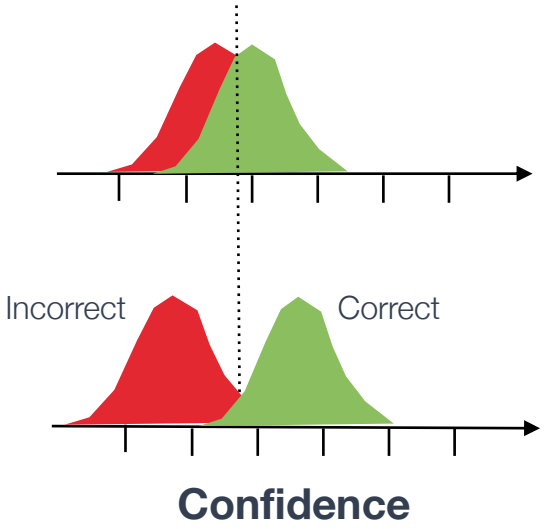
Using confidence to unpack the role of metacognition in anxious-depression

Confidence

Metacognitive **bias**
over/under-confidence

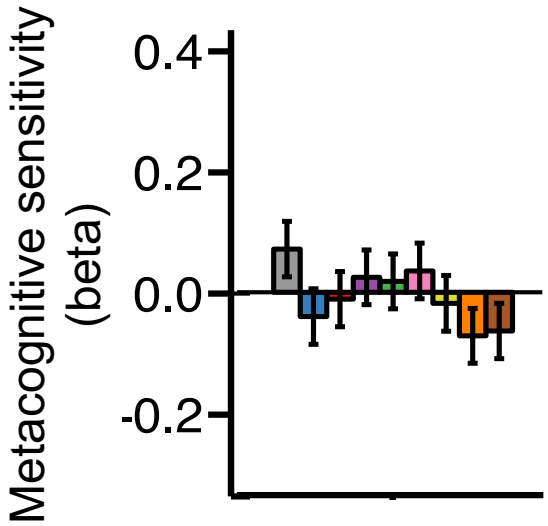
Metacognitive **sensitivity**

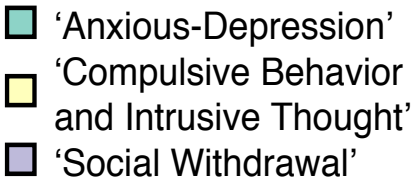
"correct about being correct"

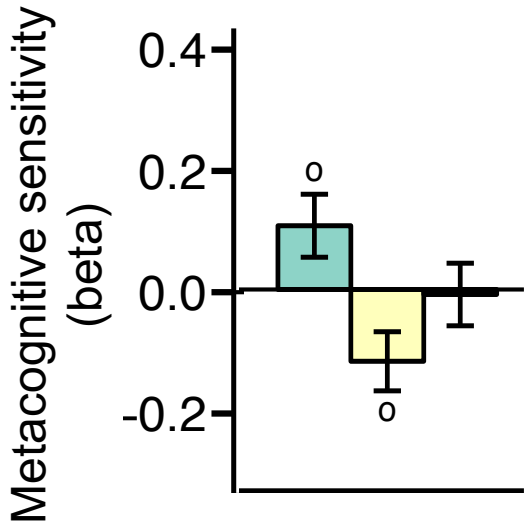


Fleming & Lau *Frontiers in Neurosci* (2014)

Metacognitive Sensitivity in anxious-depression

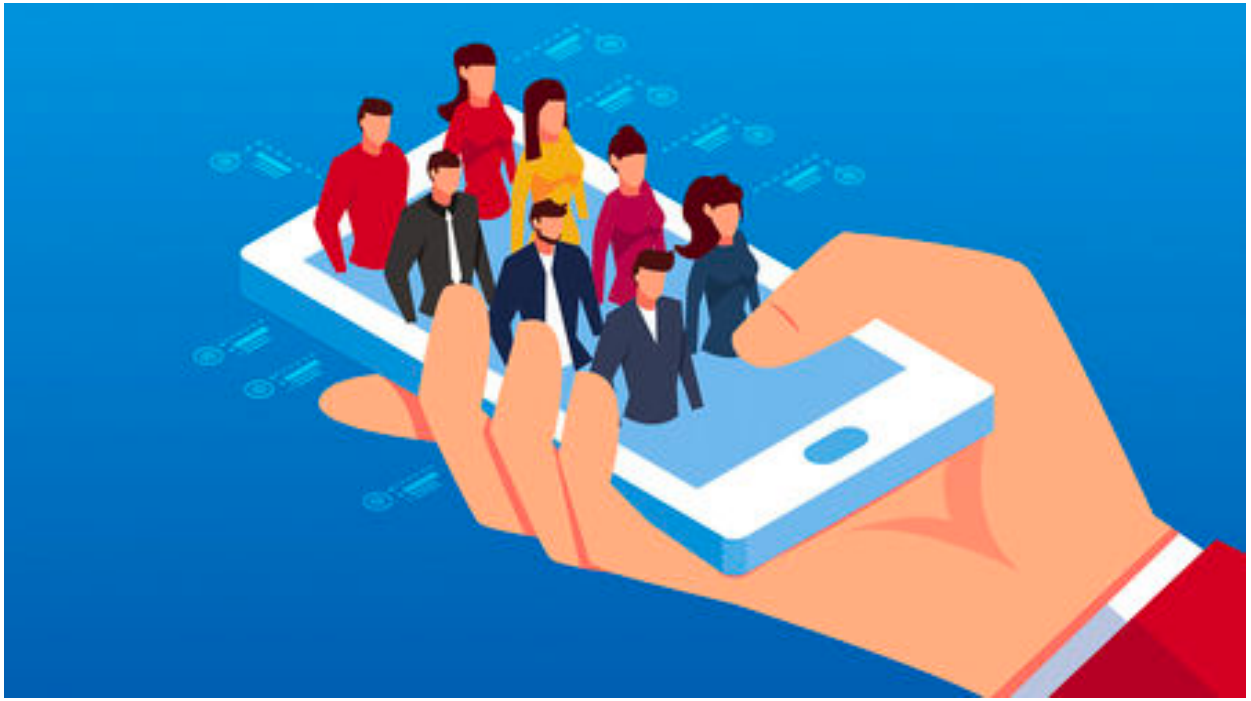






Rouault*, Seow*, Gillan & Fleming, Biological Psychiatry, 2018

Smartphones



Smartphones for Large-Scale Citizen Science

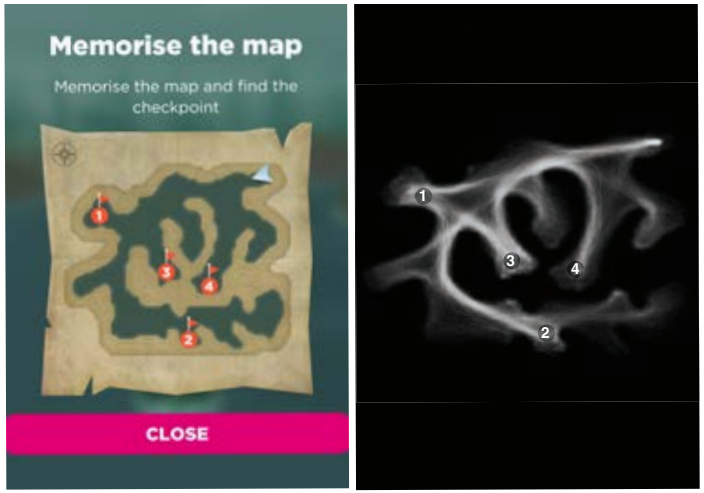
The model

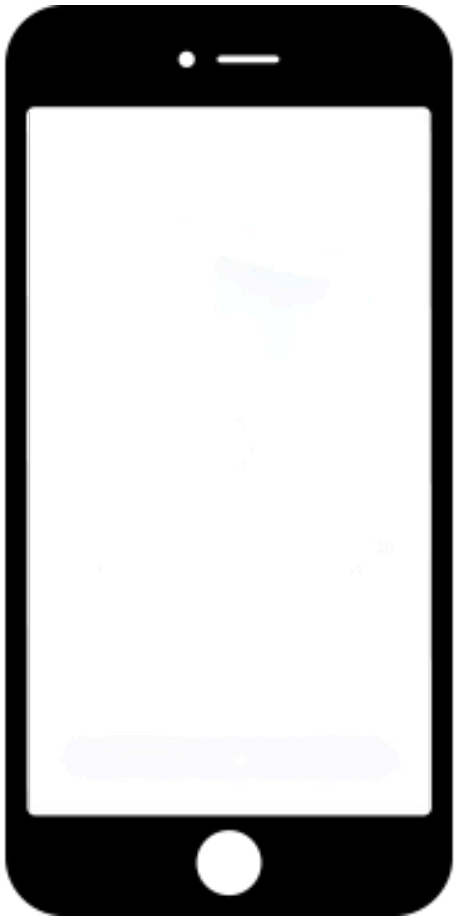
- Gamified versions of cognitive tests
- Play as much as you like

What you get

- *Really* large samples 1000s
- Repeated within-subject, longitudinal assessment

Smartphones for Science



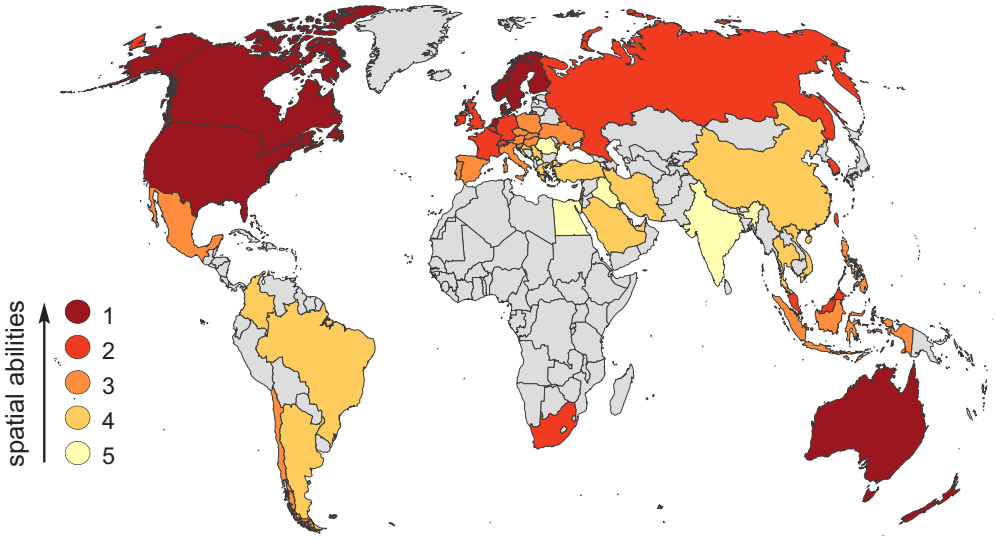


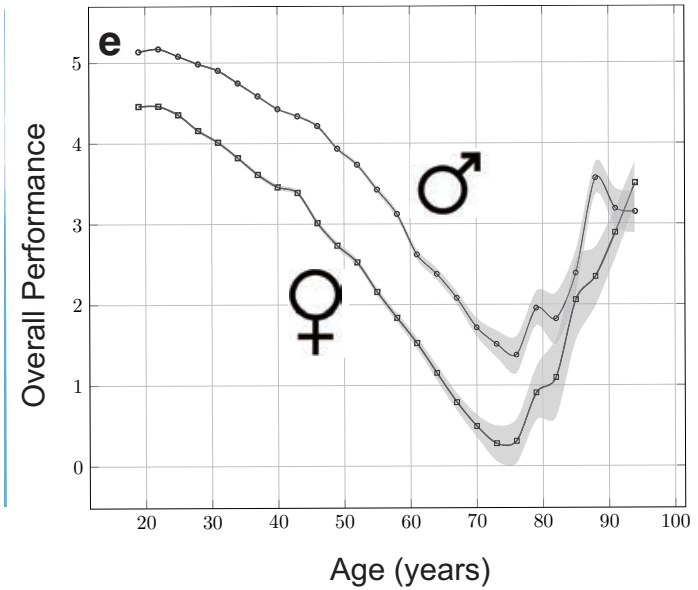
4.3 million downloads!

Hugo Spiers et al., at UCL

Spatial navigation differences across country and sex

N>500,000

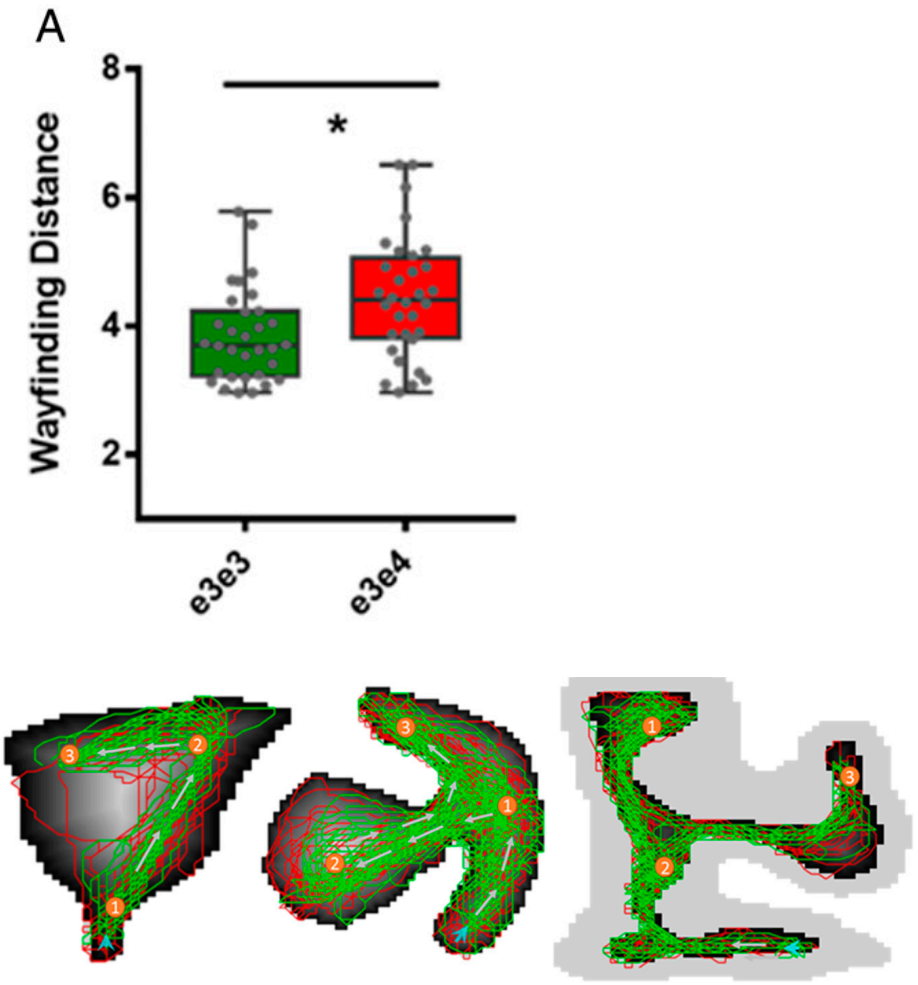




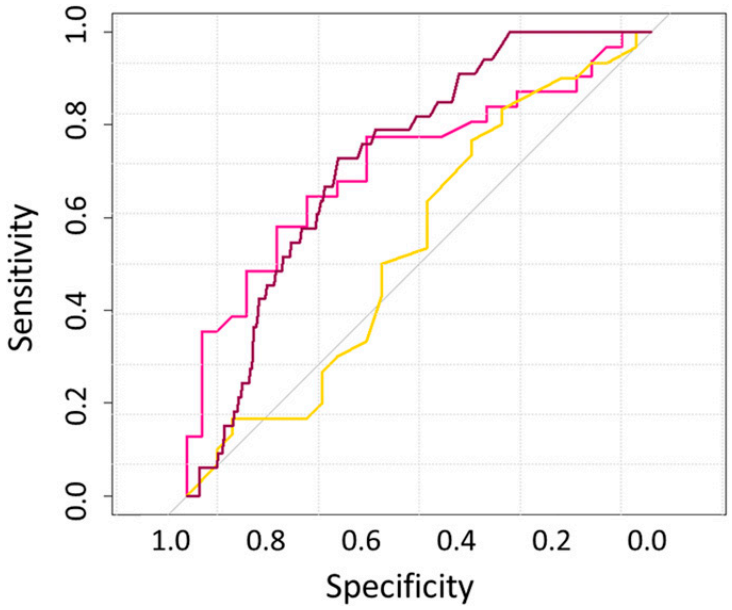
But this difference is partially explained by the Gender Gap in each country (and GDP)

Coutrot et al., Current Biology, 2018

Big Data samples can rapidly generate norms.



Spatial navigation performs better than test of episodic memory



Coughlan et al., PNAS 2019

Scraping

Scraping social media

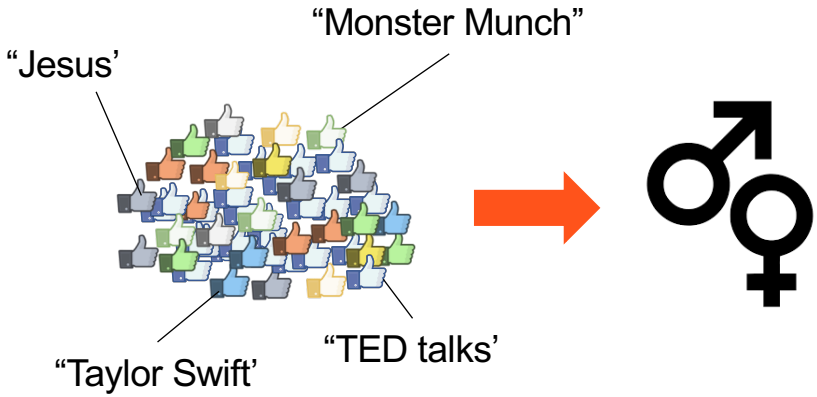
N=58,466

What do your likes say about you?

Scared? You should be.

Study Design

- Acquired consent
- Psychometric test scores (e.g. IQ)
- Survey information (e.g. democrat?)
- Data from profile, ~170 Facebook Likes per person

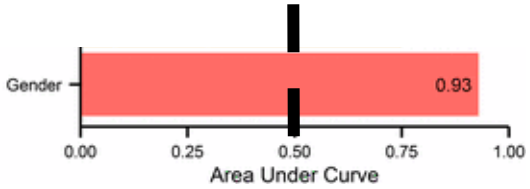


Kosinski, Stillwell and Graepel, PNAS, 2013

Scraping social media

This information is in many cases publically available.

When not, it is available for purchase from Facebook

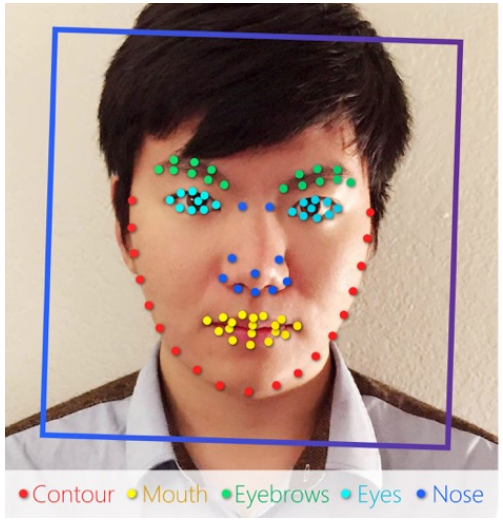


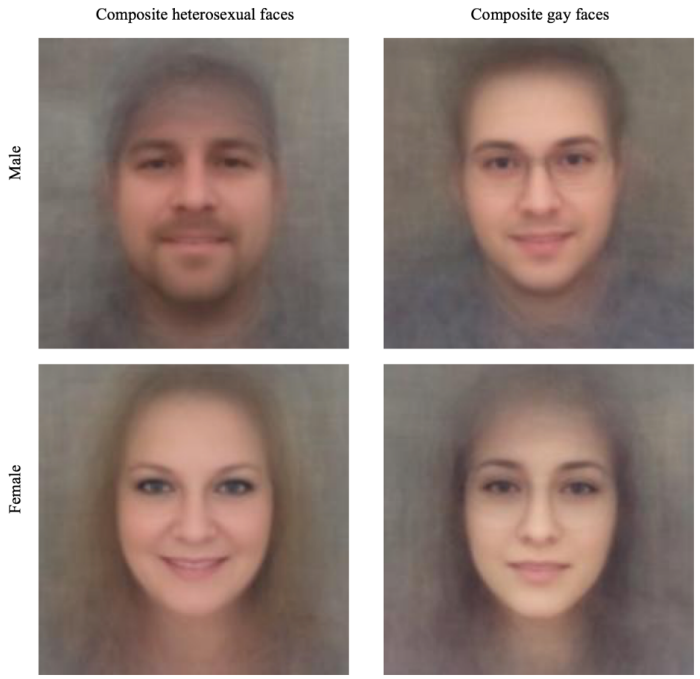
Satisfaction With Life: *Satisfied*	*Dissatisfied*
Sarah Palin	Hawthorne Heights
Glenn Beck	Kickass
Proud To Be Christian	Atreyu (Metal Band)
Indiana Jones	Lamb Of God
Swimming	Gorillaz
Jesus Christ	Science
Bible	Quote Portal
Jesus	Stewie Griffin
Being Conservative	Killswitch Engage
Pride And Prejudice	Ipod

Kosinski, Stillwell and Graepel, PNAS, 2013

Is there a wrong way to do this?

Scraping without consent





- With 1 image per person, AUC = .81
- With 5 images, AUC > .90
- Algorithm **outperformed humans** (N=1000)

Gender atypicality predicts gay and lesbian status

Yang & Kosinski, J Personality and Social Psychology, 2018

Backlash

Why Stanford Researchers Tried to Create a 'Gaydar' Machine

The New York Times

THE A.I. "GAYDAR" STUDY AND THE REAL DANGERS OF BIG DATA

THE NEW YORKER

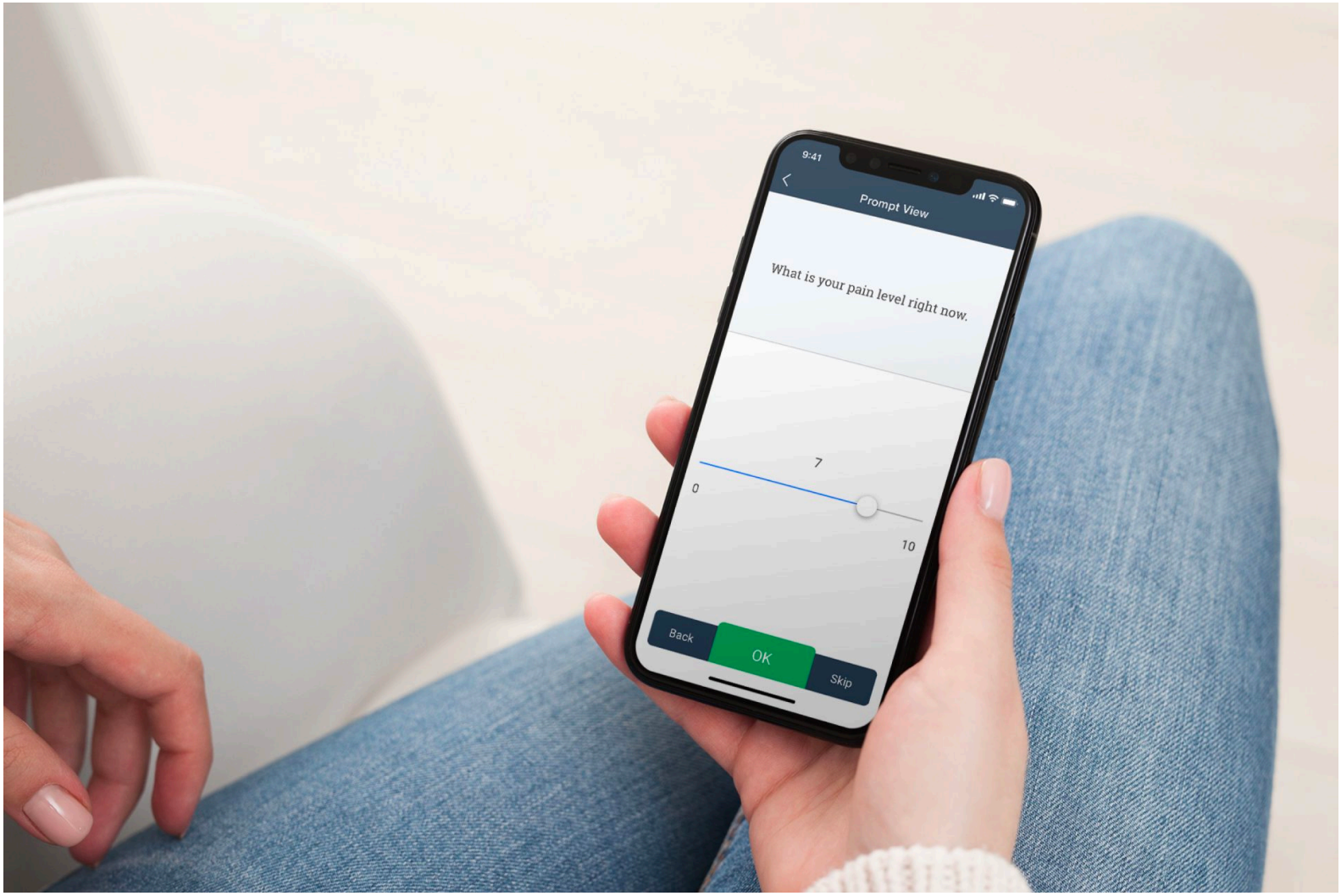
"participants"
(the people whose images were used)
were not consented

Biology vs Cultural Norms

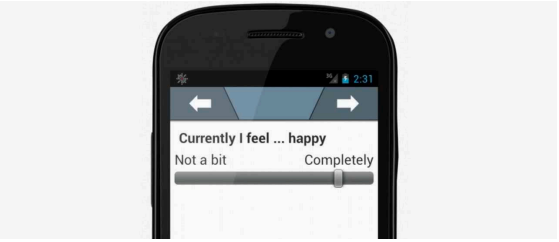
Do algorithms reveal sexual orientation or just expose our stereotypes?

Margaret Mitchell
Blaise Aguera y Arcas
Alex Todorov

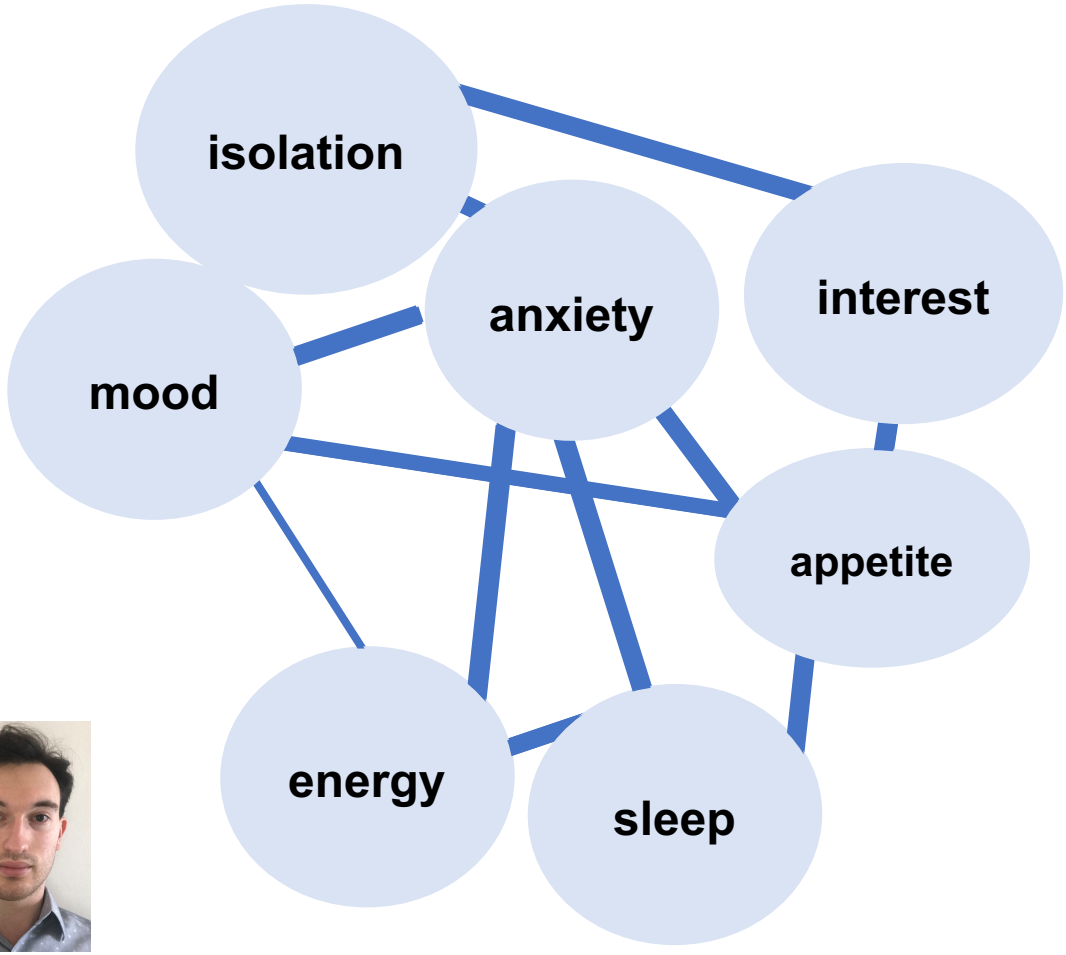
Scraping Social Media as an alternative to Ecological Momentary Assessment



Scraping Social Media for Ecological Momentary Assessment in Psychiatry



- Burdensome
- Impossible to do over long time scales



Sean Kelley

Can we approximate this using ?

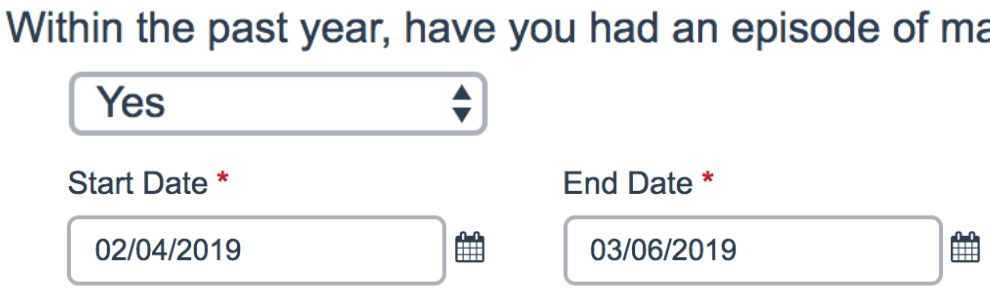
Scraping Social Media for Ecological Momentary Assessment in Psychiatry

Advertising exclusively through twitter

- soliciting mental health ‘influencer’ or organisation retweets

↓

~5 minute survey



Scraping Social Media for Ecological Momentary Assessment in Psychiatry

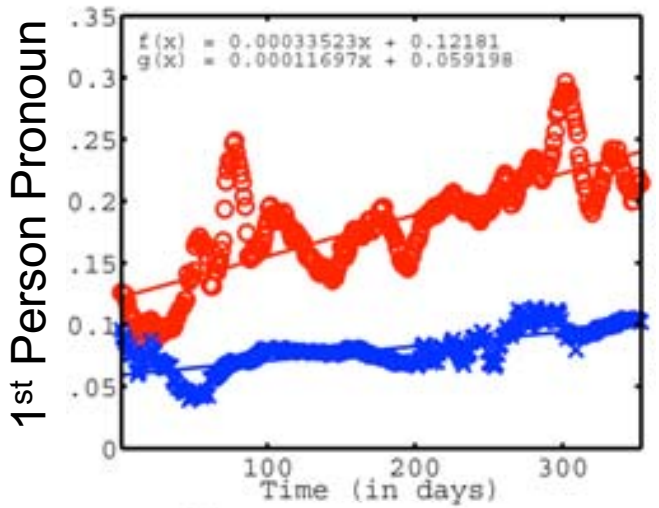
Sentiment Analysis

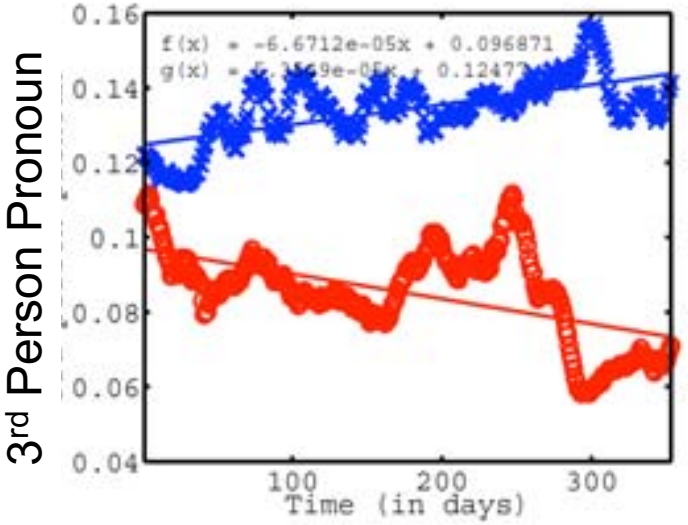
- **LIWC** is a dictionary of ~12,000 words with 90 different output variables
  - *linguistic characteristics (e.g. articles and pronouns)*
  - *psychological constructs (e.g. sadness and positive emotions)*
  - *general text information (e.g. punctuation and word count)*

- **VADER** was built specifically for use with social media text
  - *Emojis*
  - *Punctuation (!!!)*
  - *Capitalization (SAD)*
  - *degree modifiers (really, totally, very, etc)*

Sean Kelley

Scraping Social Media for Ecological Momentary Assessment in Psychiatry

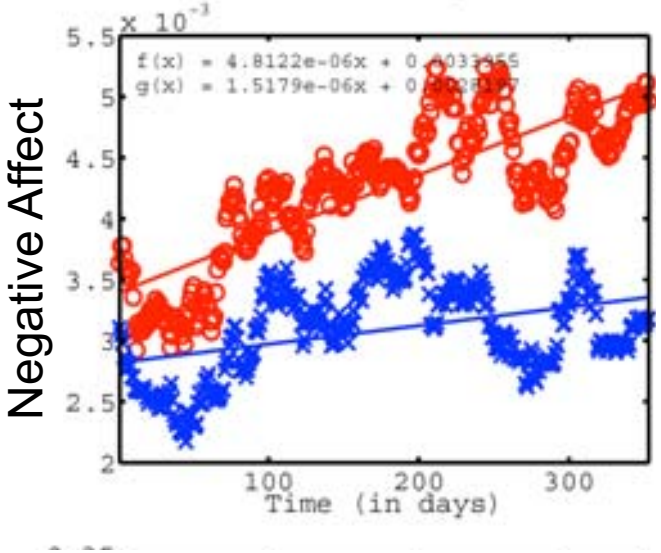


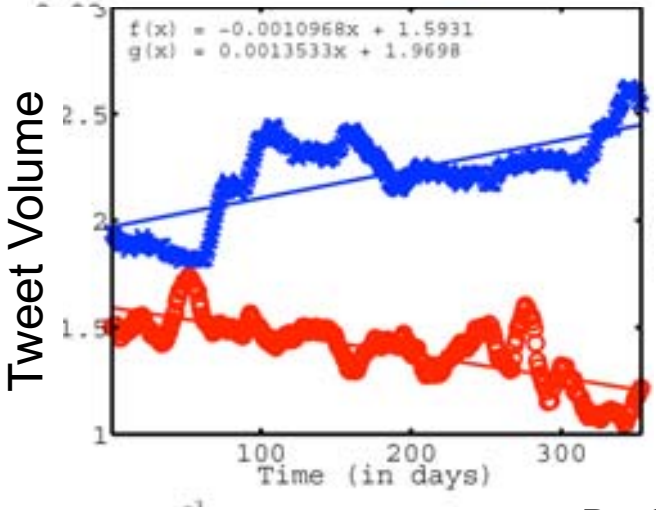


Depressed

Non-depressed

N=476

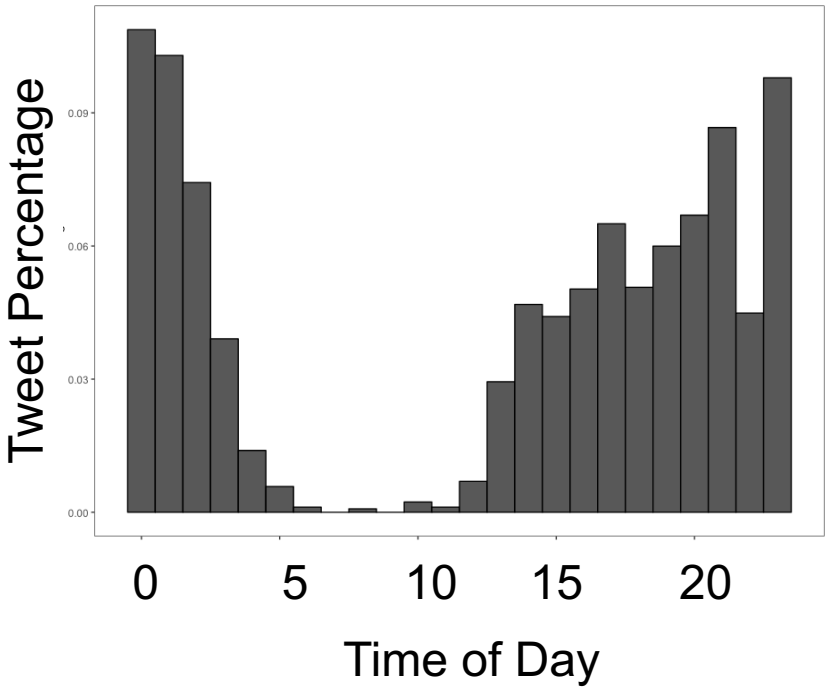


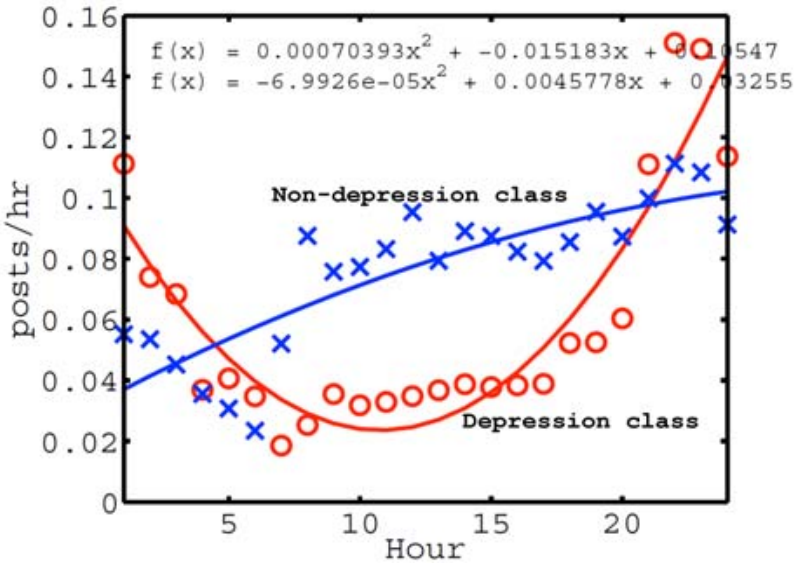


De Choudhury et al., 2013

Tweet timings as a proxy for sleep

example participant w/ self-report depression



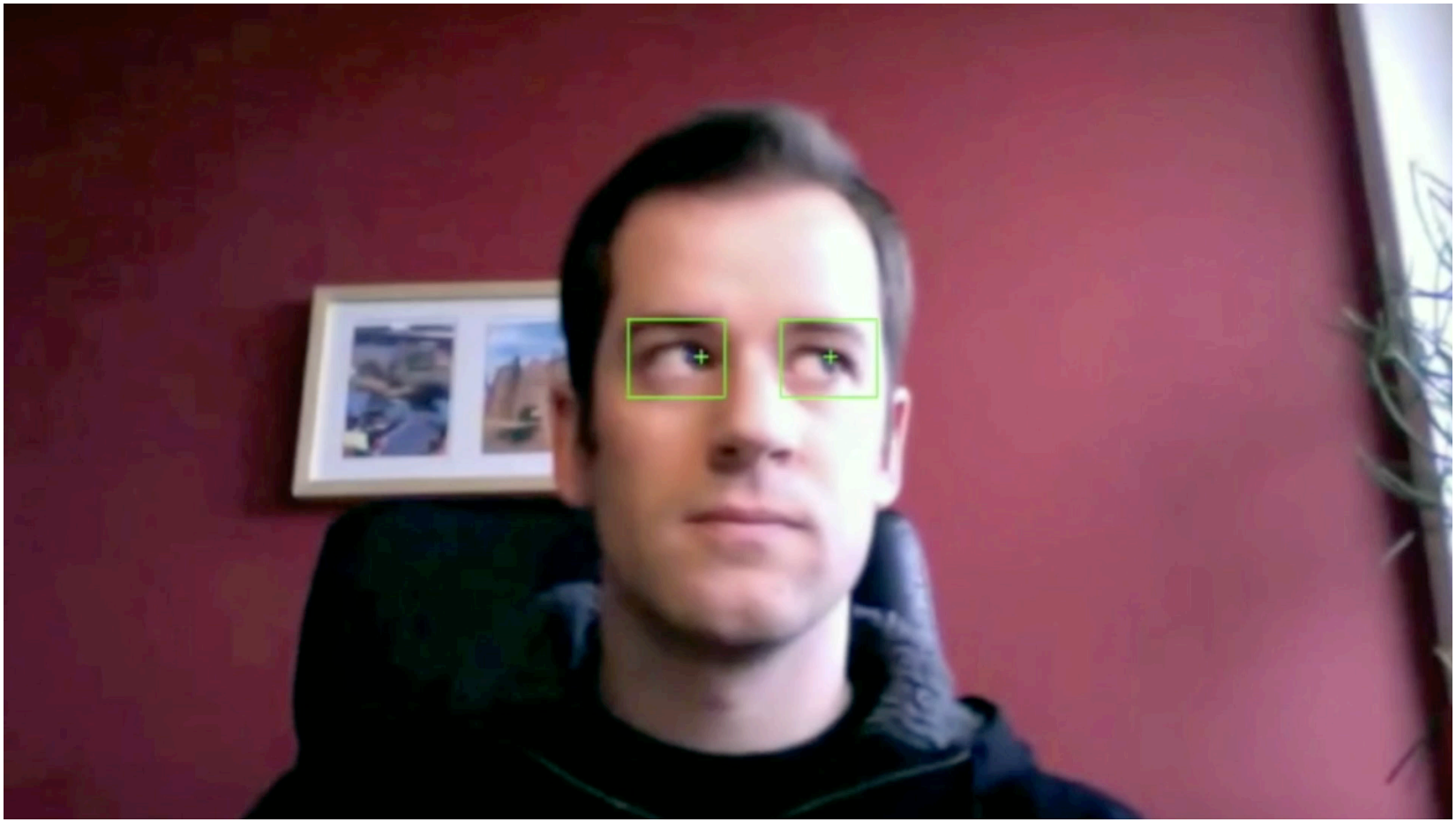


Sean Kelley

sleep

De Choudhury et al., 2013

Proxies for Physiology: Eye-Tracking



Proxies for Physiology: Eye-Tracking

Proof of principle:
What video features attract the most attention in babies?

Method
- Subjects sat on parents lap
- Initial video used to determine eligibility
- Then if eligible, babies were asked to watch different videos.
- Once complete, blind raters coded videos for looking time by 'video features'

Key Findings:
- faces, camera zooms and rhyming and singing increased infant attention.

Cusack et al., 2017, Journal of Experimental Child Psychology

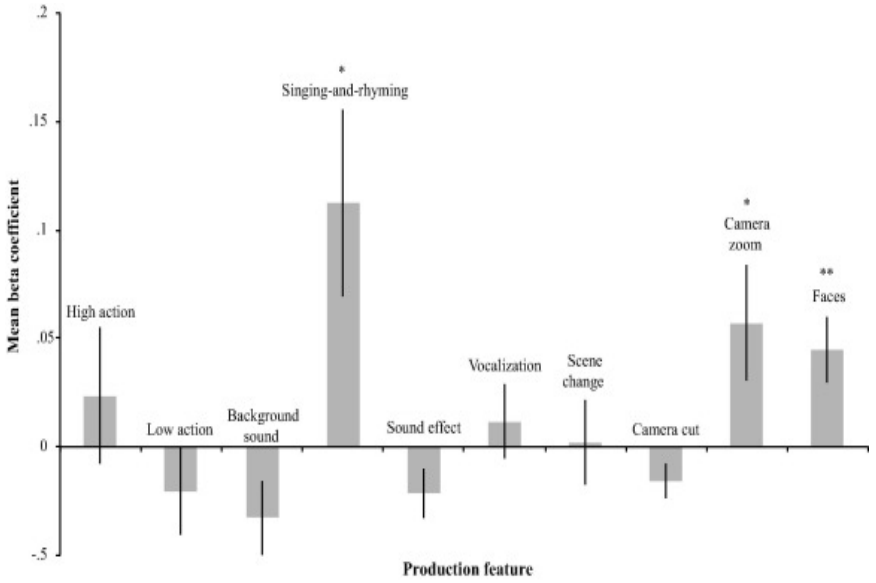
Proxies for Physiology: Eye-Tracking

Proof of principle:

What video features attract the most attention in babies?

Method

- Subjects sat on parents lap
- Initial video used to determine eligibility
- Then if eligible, babies were asked to watch different videos.
- Once complete, blind raters coded videos for looking time by 'video features'

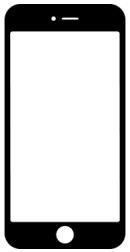


Key Findings:

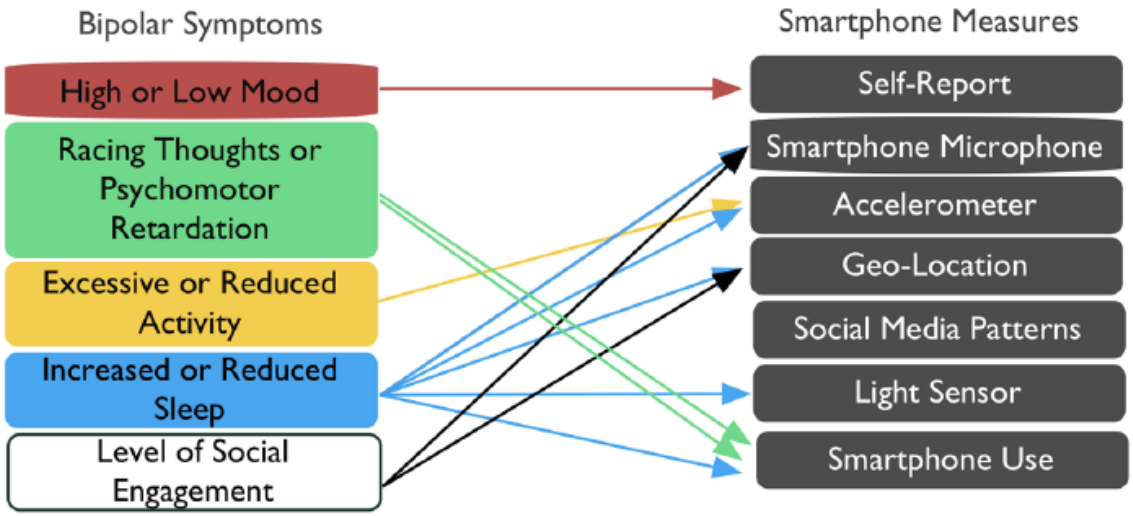
- faces, camera zooms and rhyming and singing increased infant attention.

Cusack et al., 2017, Journal of Experimental Child Psychology

Proxies for Physiology: Phone Sensor Data



- GPS receiver
- Microphone
- Camera
- Pedometer
- Heart rate
- Text messages
- Social networking
- Typing speed
- Word-use complexity
- Time using phone
- …



Matthews et al., 2016

mindstrong

HEALTHRHYTHMS®

Outline

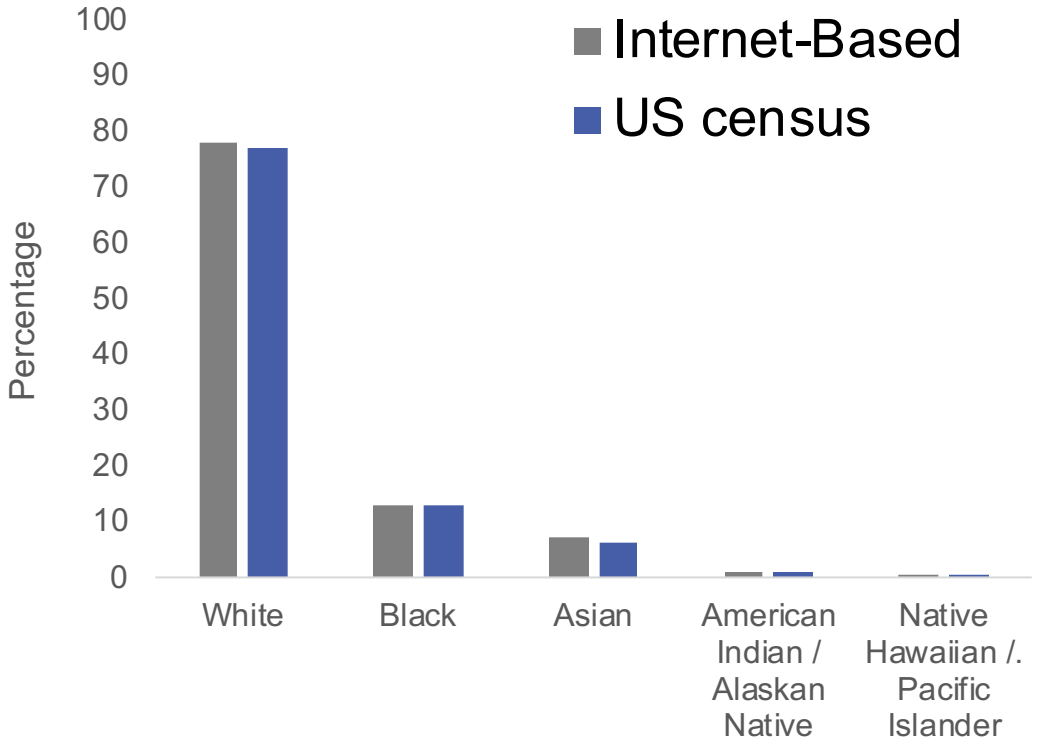
- Why take psychiatry research online?
- Online methods
  - *Crowdsourcing*
  - *Smartphones*
  - *Scraping*
  - *Proxies for physiology*
- Characteristics of online samples
  - *Are they representative?*
  - *Are mental health data valid?*
  - *Are the data of acceptable quality?*
  - *Are findings relevant for diagnosed patients?*

Spoiler: yes.

Characteristics of Online Samples

Are Internet based samples representative?

Race



M-Turk is **more racially diverse** than regular online sample and college samples

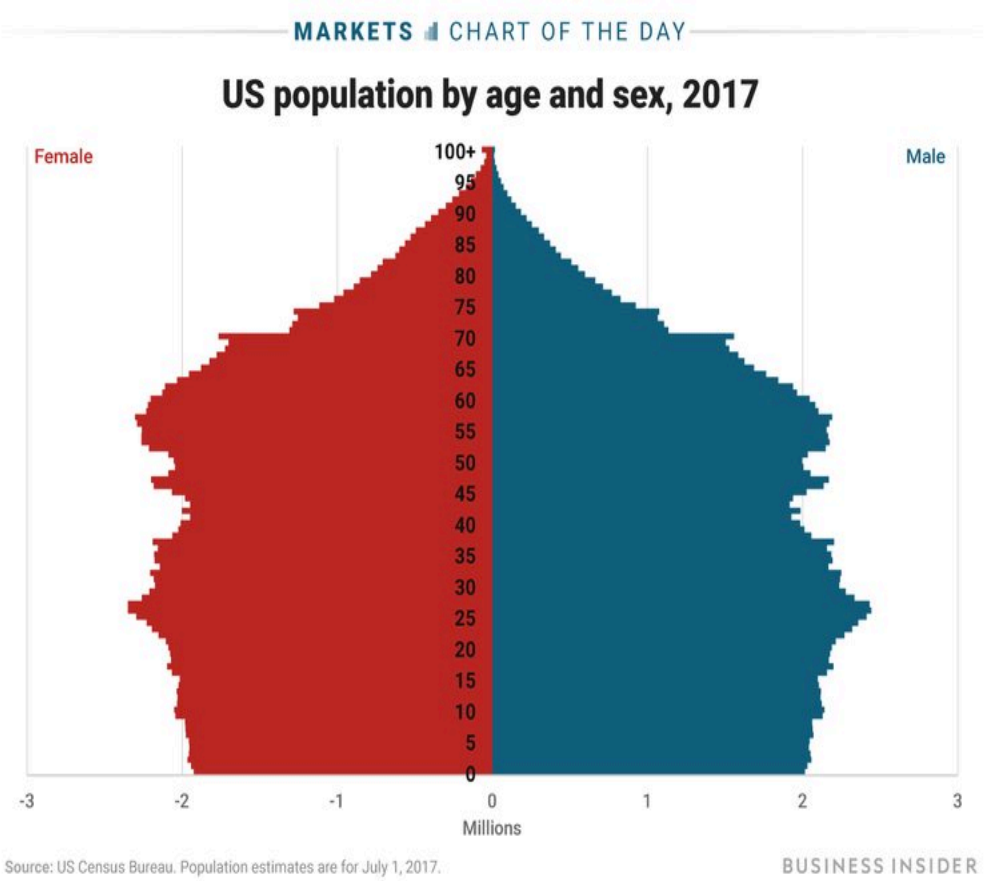
Buhrmester at al., 2012

[1] Gillan et al., under review

[2] https://www.census.gov/quickfacts/fact/table/US/PST045218

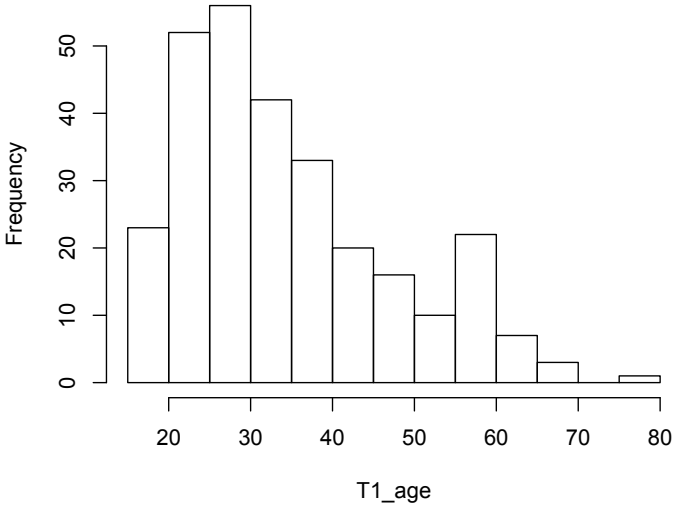
Are Internet based samples representative?

Age



Generic Internet-based (USA)

OCD/GAD patients (N=285)



Gillan et al., under review

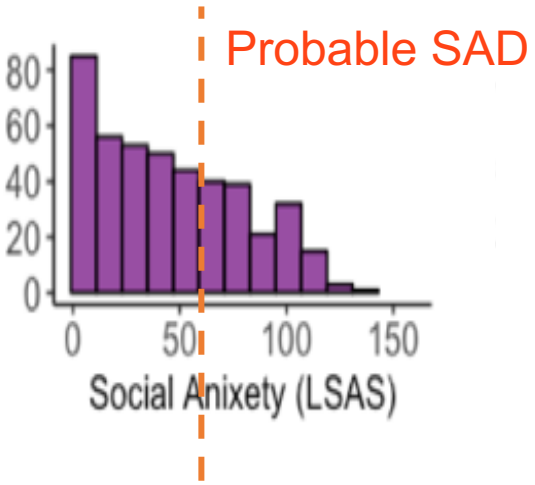
Are there differences in mental health?

- The prevalence of depression among participants on Mechanical Turk was **consistent with prevalence in the general population**

Shapiro et al., 2013; Kessler, Chiu, Demler, & Walters, 2005

- But… **social anxiety is up to 7x more prevalent on M-Turk**.

Shapiro et al., 2013, Gillan et al., 2016; Seow & Gillan, in prep; Hunter et al., BiorXiv



Are people who they say they are?

Overwhelmingly, YES!

- Excellent **test-retest reliability** for self-report depression, r=.87 (Shapiro et al., 2013)
- We can **reproduce associations with mental health** found in in-person samples
- **Few get caught** out with trap questions (e.g. <1%, Gillan et al., eLife 2016)
- A more elaborate **malingering questionnaire caught just 10/530** (Shapiro et al., 2013)
- You can confirm location from IP address (but think about GDPR)
- Mechanical Turk allows you to select for certain demographics (age, gender, etc.)

Don't incentivize people to lie

Are there differences in data quality?

Unsupervised participants are **less likely** to pay attention to instructions

Oppenheimer, Meyvis, & Davidenko, 2009

Are M-Turk subjects worse than supervised students?

"Research in decision making shows that people, when making decisions and answering questions, prefer not to pay attention and minimize their effort as much as possible. Some studies show that over 50% of people don't carefully read questions. If you are reading this question and have read all the other questions, ***please select the box marked 'other' and type 'Decision Making' in the box below****. Do not select "predictions of your own behavior." Thank you for participating and taking the time to read through the questions carefully!*

What was this study about?
A Predictions of your own behavior
B Predictions of your friends' behavior
C Political preferences
D Other _____

66.2% versus 88.5%, $p < .001$

Goodman, Cryder, Cheema, 2013

Are there differences in data quality?

YES. How big of a problem is this?

Its no biggie (personal view): Increased sample size mitigates increased loss of statistical power. E.g. in Gillan et al., eLife 2016

But to help things: establish study-specific *a priori* exclusion criteria e.g. implausibly fast RTs, missing trials, 'catch' questions, comprehension test

Save yourself some money by requiring basic performance to play!

Data quality solutions

COMPREHENSION TEST

- Include a quiz on the basics of the task instructions
- If subjects do not pass, they must repeat instructions until they do
- This deals with bots
- This deal with people just hammering the keys randomly

- Crump et al., 2013 found this **improved the issue**, but didn't resolve it fully.

- We noted similar issues in our own work, and found the quiz ***greatly*** improved the data quality. Again, doesn't resolve it fully.

Internet based testing is not a panacea.

- Sometimes you need to have control over the testing environment

Are addictive traits linked to problems with response inhibition?... Acute intoxication might confound the data

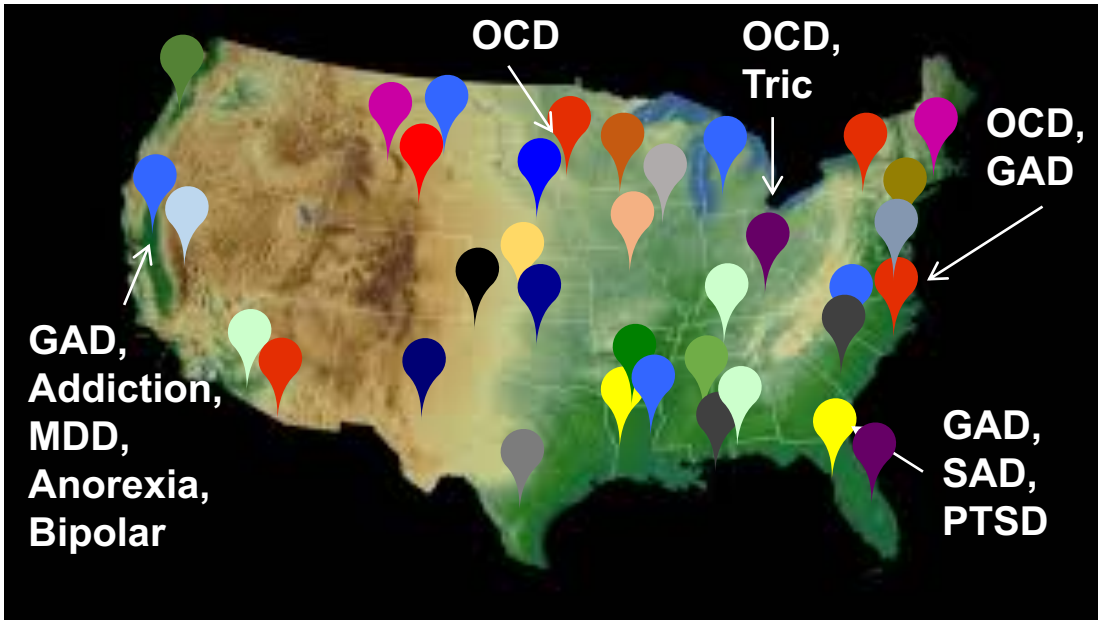
- Can people accurately self report on their mental status? (in some cases, e.g. schizophrenia, not always.)
- Is there an incentive to lie (e.g. monetary)?
- Are the findings relevant to 'real patients'?

Are findings applicable to diagnosed patients?

OCD

All-Comers Internet-Based

Internet-Based Clinical Collaboration



OCD (N=110) - OCD+GAD (N=92) - GAD (N=83)

HOFSTRA UNIVERSITY

BROWN

HARVARD UNIVERSITY

COLUMBIA UNIVERSITY

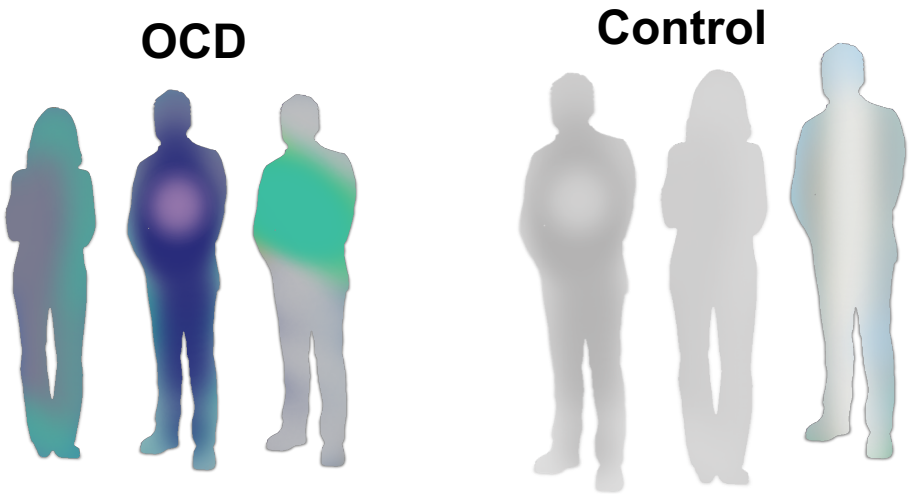
NYU

Gillan et al., under review

Internet-Based Clinical Collaboration

1. Can we reproduce the *dimensional* association between goal-directed planning and compulsivity in diagnosed patient sample?

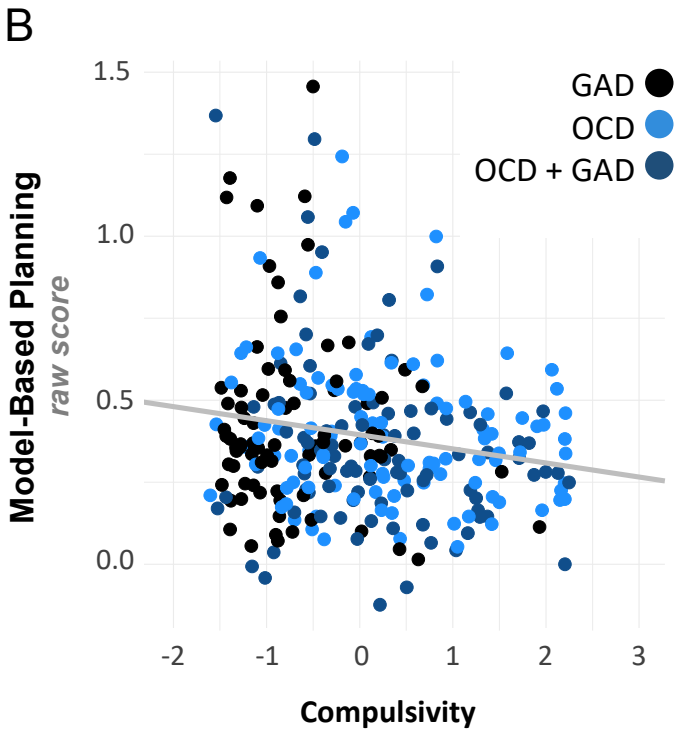
2. What fits the cognitive data better… Dimension or Diagnosis?*



*caveat: we use clinical controls not ‘healthy controls’

Are findings applicable to diagnosed patients?

OCD (N=110)
OCD+GAD (N=92)
GAD (N=83)



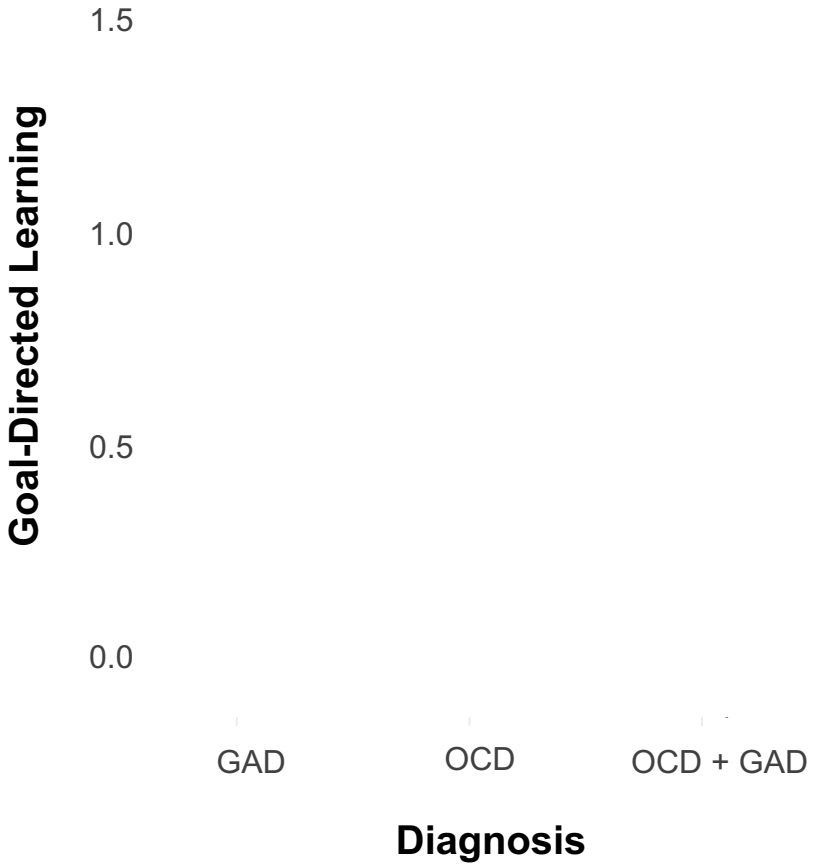
*Replicated after average of 413 days in N=110

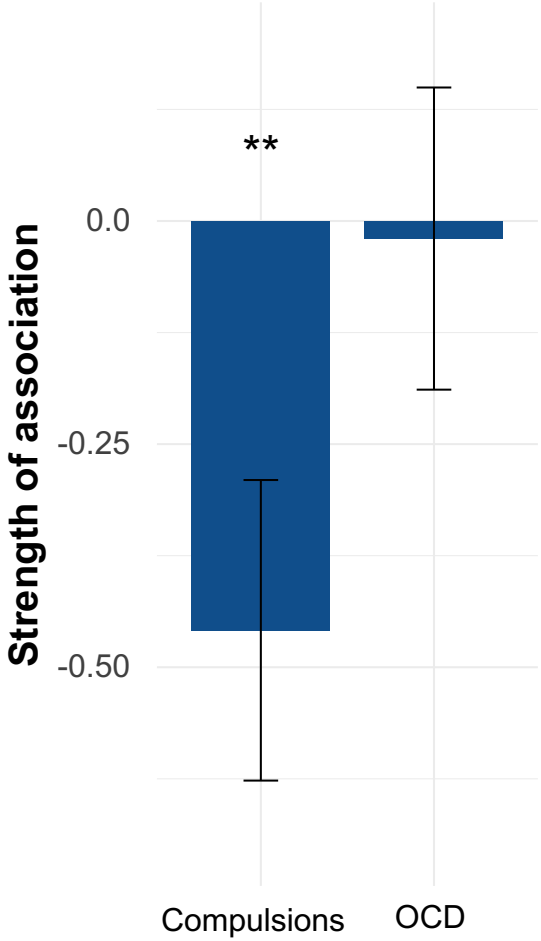
Gillan et al., under review

How informative is the diagnosis itself?

OCD (N=110)
OCD+GAD (N=92)
GAD (N=83)

Kindest interpretation is p=.18



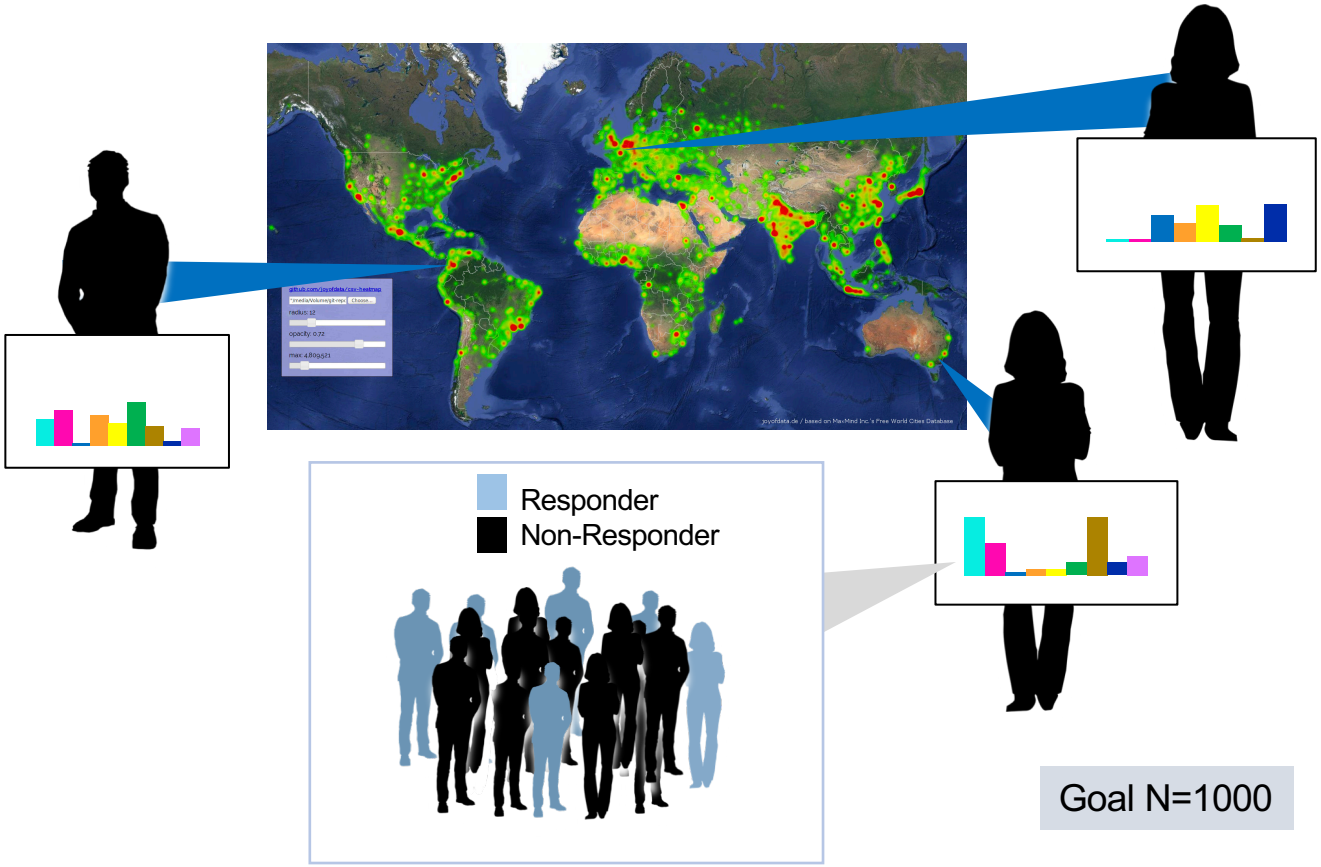


Gillan et al., under review

The problem: **DSM disorders are the "ground-truth" for research**

Internet-Based Treatment Prediction

Can we use Internet-based methods to predict and understand treatment response?

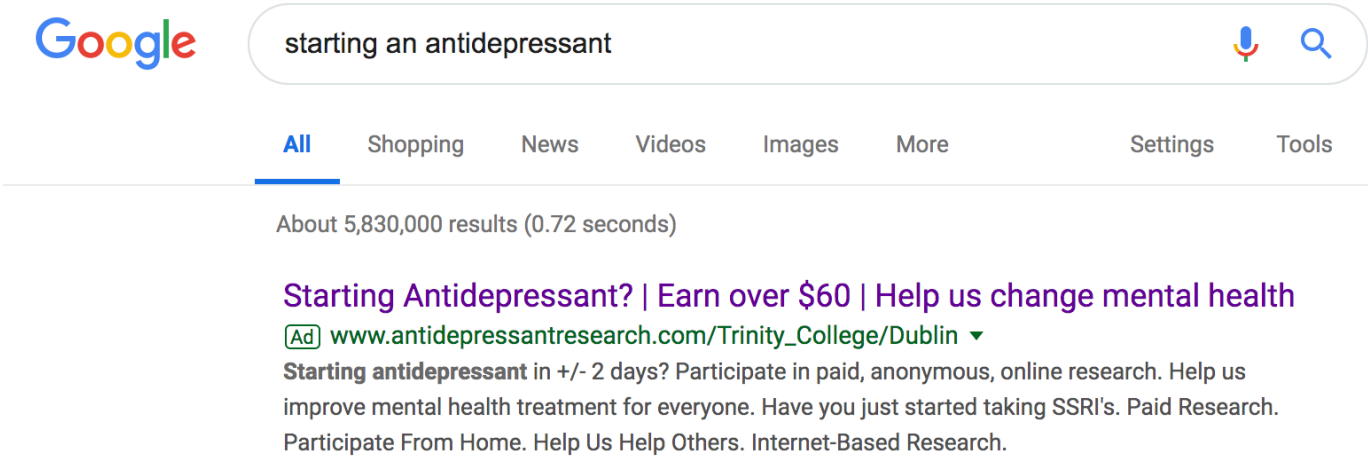


Kevin Lynch

Internet-Based Treatment Prediction

Transforming mental health through research

Can we use Internet-based methods to predict and understand treatment response?



Kevin Lynch

Final take home message.

Internet-based research is awesome.

- large samples
- rare populations
- longitudinal, predictive research
- sometimes frictionless participation
- exploratory and confirmatory

QUESTIONS?

Thank You

Trinity College Dublin

Tricia Seow
Andrew Pringle
Kevin Lynch
Eoghan Gallagher
Sean Kelley

MQ Transforming mental health through research

wellcome trust

GLOBAL BRAIN HEALTH INSTITUTE

Multi-Site Study

Eyal Kalanthroff
Michael Eans
Hilary Weingarden
Ryan Jacoby
Marina Gershkovich
Ivar Snorrason
Raphael Campeas
Cynthia Cervoni
Nick Crimarco
Yosef Sokol
Sarah Garnaat
Nicole McLaughlin
Liz Phelps
Anthony Pinto
Christina Boisseau
Sabine Wilhelm
Nathaniel Daw
Blair Simpson