

# Predicting Psychology from Social Data



David Stillwell

[ds617@cam.ac.uk](mailto:ds617@cam.ac.uk)

Lecturer in Big Data Analytics and  
Quantitative Social Science



UNIVERSITY OF  
CAMBRIDGE  
Judge Business School

Deputy Director, The Psychometrics Centre

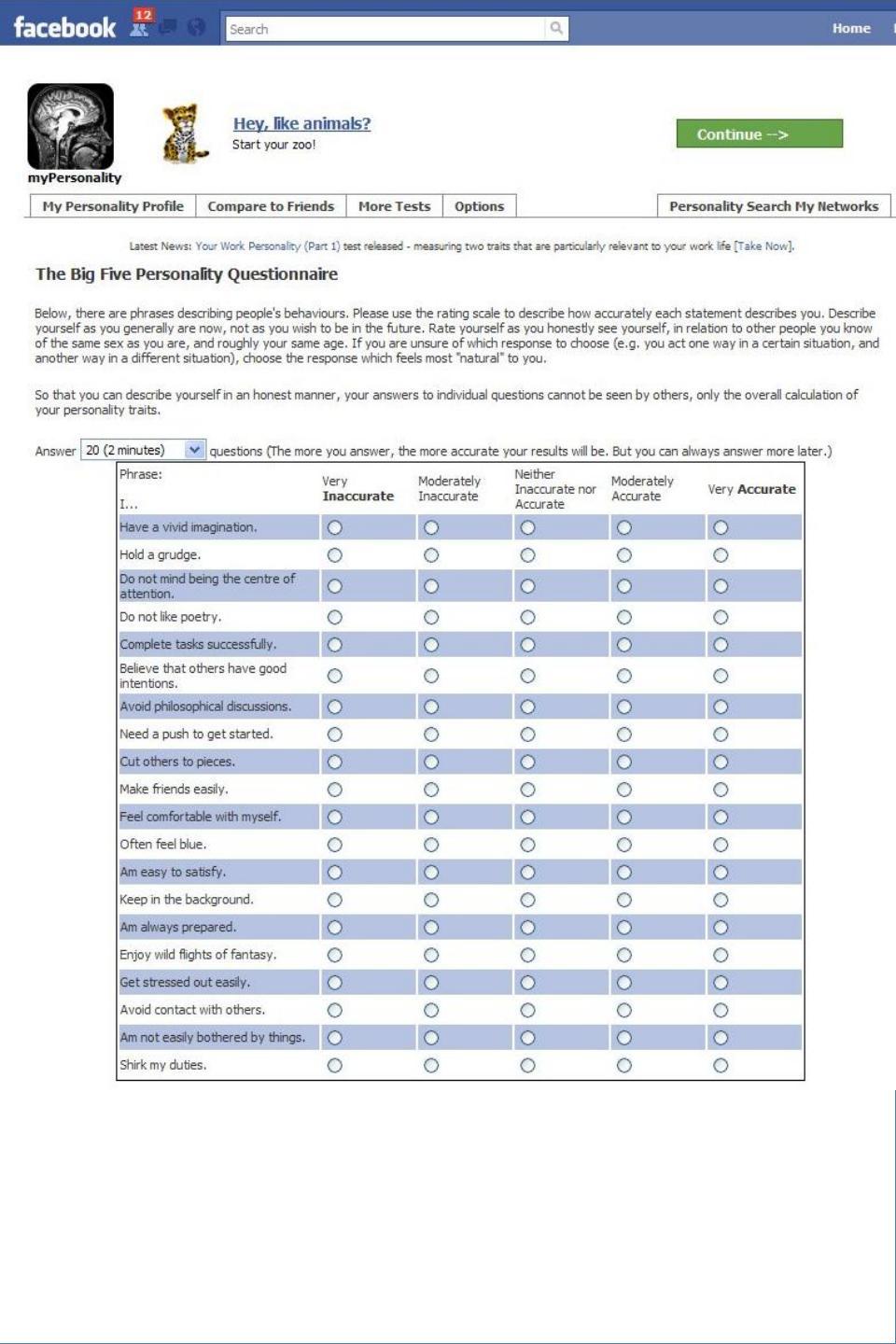
Facebook app launched in 2007

Users could take real psychometric assessments and get feedback for free

6 million people opted in to submit their data

High quality information:

- Records of actual behaviour
- Honest, open, motivated respondents
- Quality better than in the lab



The screenshot shows the myPersonality Facebook app interface. At the top, there's a profile picture of a brain, a tiger icon, and a message: "Hey, like animals? Start your zoo!". A "Continue →" button is on the right. Below the header, there are tabs: "My Personality Profile", "Compare to Friends", "More Tests", and "Options". A "Personality Search My Networks" link is on the far right. A message at the top says: "Latest News: Your Work Personality (Part 1) test released - measuring two traits that are particularly relevant to your work life [Take Now]". The main content is titled "The Big Five Personality Questionnaire". It asks users to rate statements on a scale from "Very Inaccurate" to "Very Accurate". The statements listed are: I... Have a vivid imagination., Hold a grudge., Do not mind being the centre of attention., Do not like poetry., Complete tasks successfully., Believe that others have good intentions., Avoid philosophical discussions., Need a push to get started., Cut others to pieces., Make friends easily., Feel comfortable with myself., Often feel blue., Am easy to satisfy., Keep in the background., Am always prepared., Enjoy wild flights of fantasy., Get stressed out easily., Avoid contact with others., Am not easily bothered by things., Shirk my duties.

Facebook app launched in 2007

Users could take real psychometric assessments and get feedback for free

6 million people opted in to submit their data

High quality information:

- Records of actual behaviour
- Honest, open, motivated respondents
- Quality better than in the lab

100 question IPIP version of NEO-PI-R Traits	Delay Discounting
336 question IPIP version of NEO-PI-R Facets	Body Consciousness
Satisfaction with life scale	Moral Foundations
Rust's vocational personality	MUSIC Personality
Sensational interests questionnaire	Empathy Quotient Scale
Self-Monitoring	Barratt Impulsivity Scale
Pennebaker Inventory of Limbic Languidness (PILL)	Relationship Scales Questionnaire
Job Self-Efficacy Scale	My memories of upbringing (EMBU)
Forms of self-criticizing/attacking and self-reassurance scale (FSCS)	Centre for Epidemiological Study Depression Scale (CES-D)
Schwartz's Values Survey	Passion Towards Work Scale
Pregnancy Explorer	Thing-Person Orientation
Volunteer Personality Scale	IQ test on myIQ

Facebook app launched in 2007

Users could take real psychometric assessments and get feedback for free

6 million people opted in to submit their data

High quality information:

- Records of actual behaviour
- Honest, open, motivated respondents
- Quality better than in the lab

Gender	Education (school/degree)
Date of Birth	Relationship status
Religion	Likes <ul style="list-style-type: none"> <li>➤ Books</li> <li>➤ Movies</li> <li>➤ TV Shows</li> <li>➤ Music and Last.fm Data</li> <li>➤ Activities</li> </ul>
Political views	Sexuality
Geographical location <ul style="list-style-type: none"> <li>➤ Country</li> <li>➤ State</li> <li>➤ City</li> </ul>	Family
Work (company/position)	Photo tags
Wall posts	Status updates

# Predicting Personality & Attributes from Facebook Likes

Kosinski, Graepel & Stillwell (2013) PNAS

1

Users' Facebook Likes

55,814 Likes

philosophy  
cnn.com  
(...)  
BMW

58,466 Users

User 1	1	1		0
User 2	0	1		1
User 3	1	0		0
(...)				
User n	1	1		0

User – Like Matrix  
(10M User-Like pairs)

2

Singular Value Decomposition

100 Components

	C <sub>1</sub>	C <sub>2</sub>	(...)	C <sub>100</sub>
User 1	.5	.7	...	-.9
User 2	.3	-.4	...	-.2
User 3	-.6	.1	...	.7
(...)	...	...	...	...
User n	1.2	1	...	-.6

User – Components Matrix

3

Prediction Model

Using Logistic or Linear Regression  
(with 10-fold cross validation)

e.g.  $age = \alpha + \beta_1 C_1 + \dots + \beta_n C_{100}$

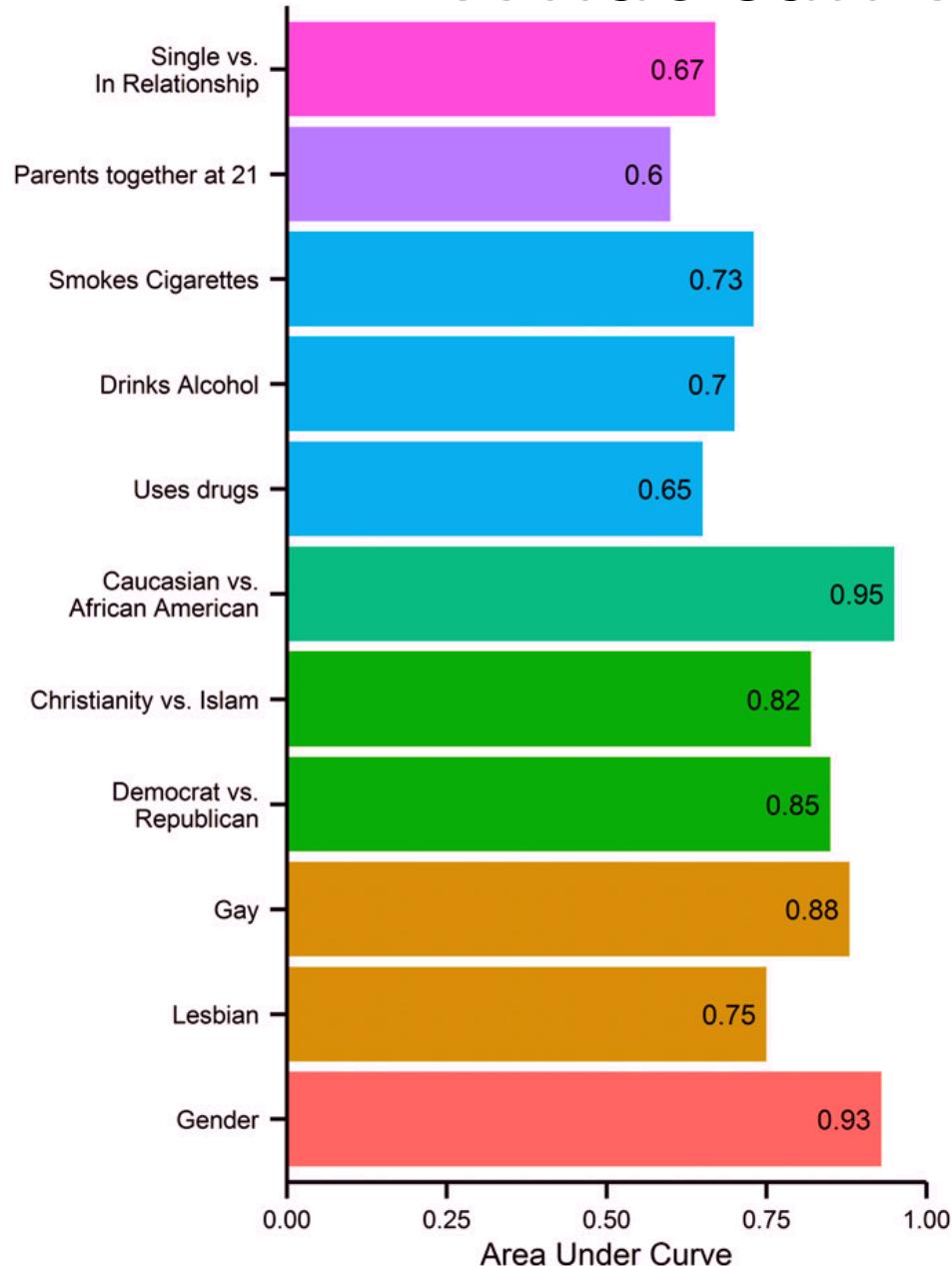
Predicted variables

Facebook profile: age, gender, political and religious views, relationship status, proxy for sexual orientation, social network size and density

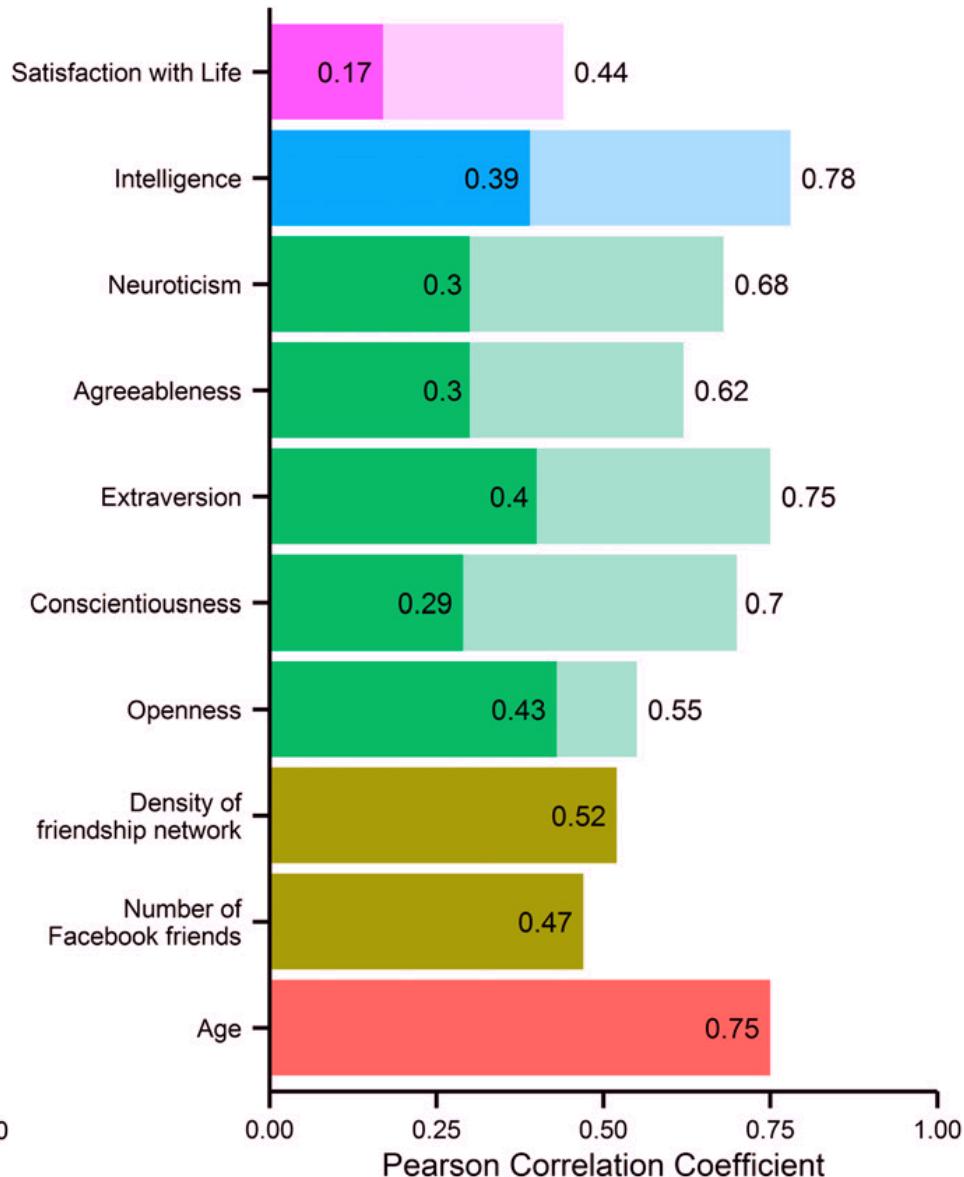
Profile picture: race

Survey / test results: BIG5 Personality, Intelligence, Satisfaction with Life, substance use, parents together?

# What can we predict?



Kosinski, Graepel & Stillwell (2013)



# IQ

High



The Godfather



Mozart



Thunderstorms



The Daily Show



To Kill a Mockingbird



Lord of the Rings



Science

Low



Jason Aldean



Tyler Perry



Sephora



Chiq



Bret Michaels



Harley-Davidson



Bebe

# Agreeableness

High



Compassion International



Jon Foreman



Pornography Harms



The Book Of Mormon



Circles Of Prayer



Christianity



Marianne Williamson

Low



I Hate Everyone



I Hate You



I Hate Police



Friedrich Nietzsche



Atheism / Satanism



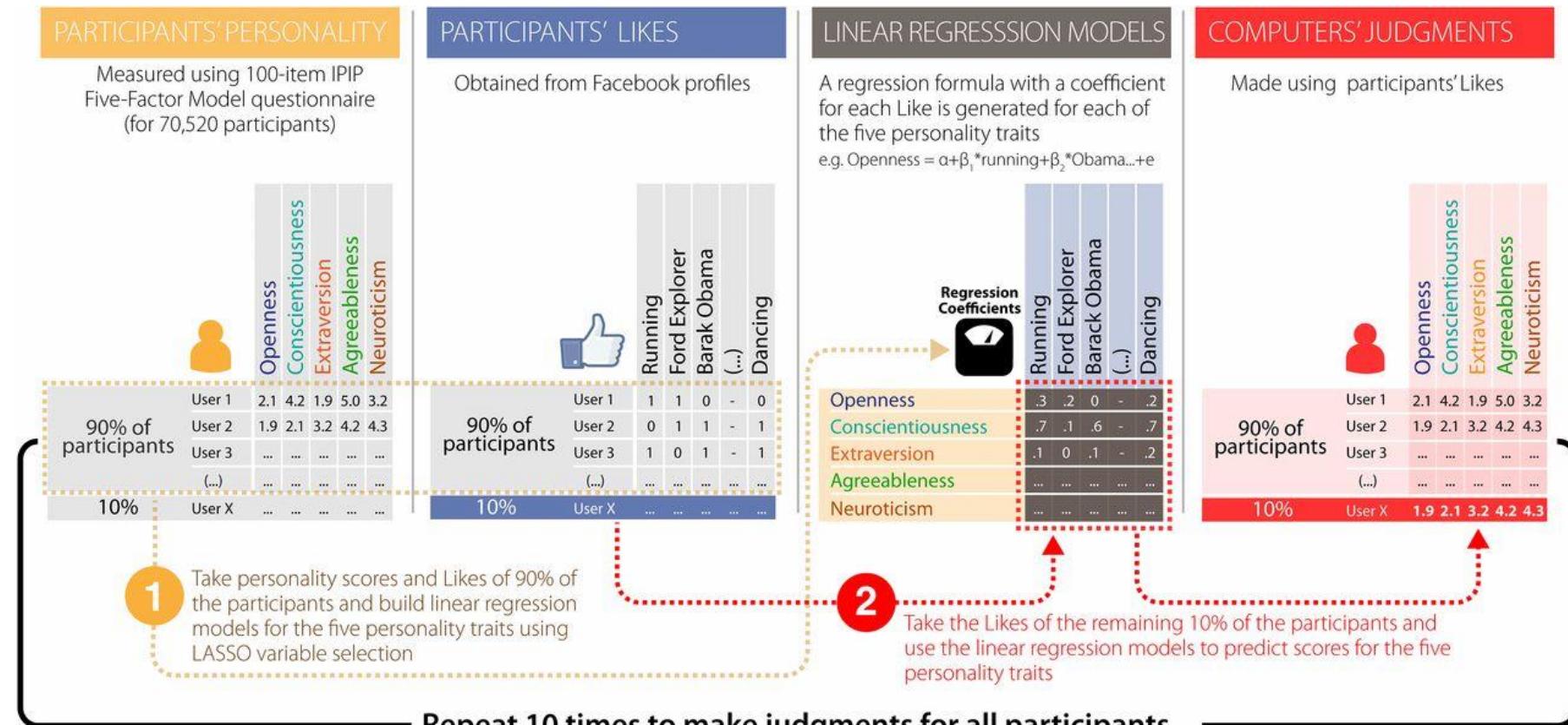
Prada



Sun Tzu

# Predicting Personality from Facebook Likes

Youyou, Kosinski & Stillwell (2015) PNAS



Humans' Judgments

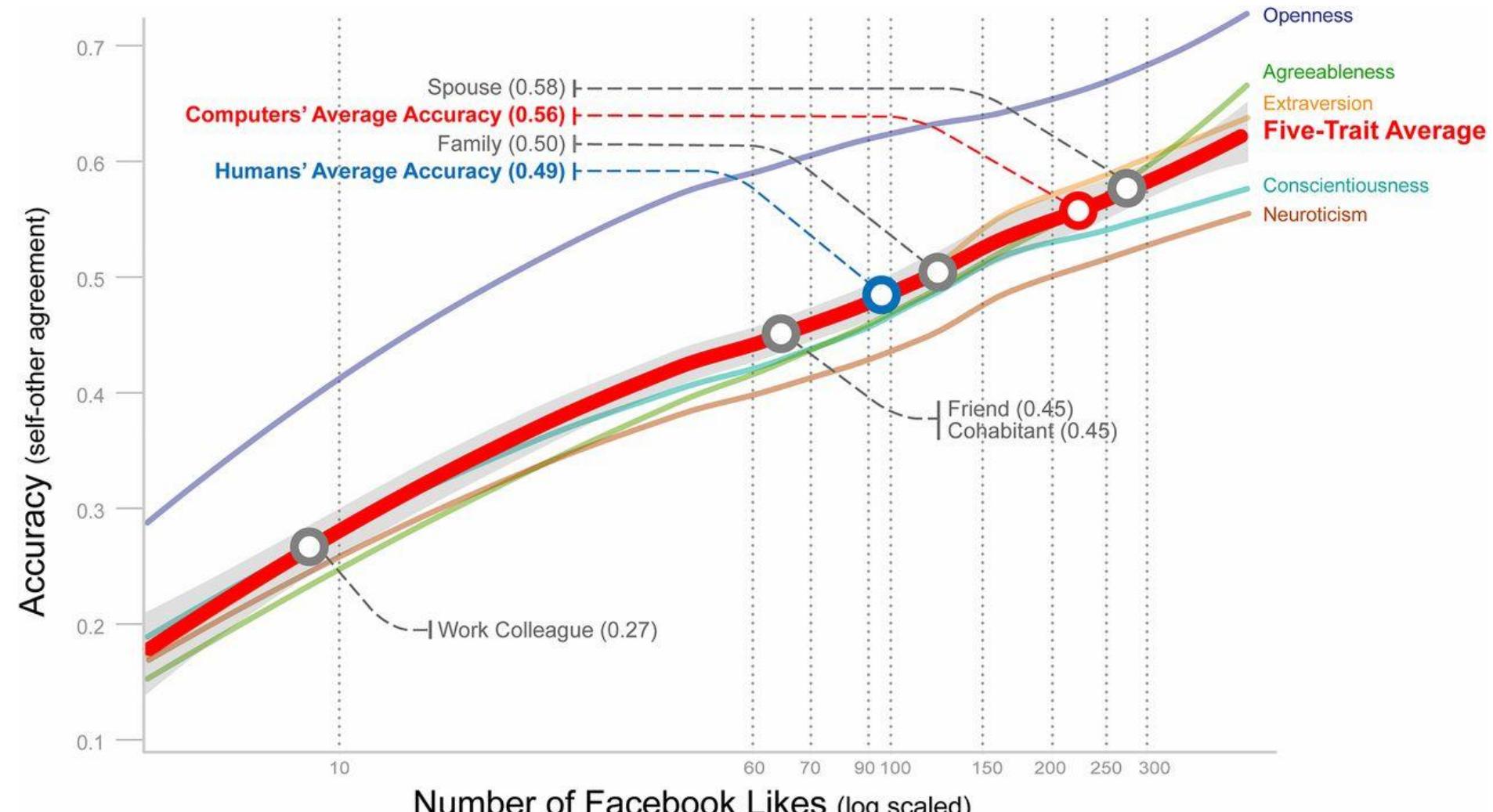
Humans' Accuracy

Self-ratings

Computers' Accuracy

Computers' Judgments

# How accurately can we predict?



Youyou, Kosinski & Stillwell (2015) PNAS

# Apply Magic Sauce

PredictionAPI

Apply Magic Sauce translates individuals' digital footprints into detailed psychological profiles.

For research

For business

For you



## Digital footprints

Facebook LikedIDs >

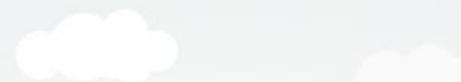
Facebook statuses >

Tweets >

Browsing data >

Open text >

And more... >



## Individual profiles

Psychographics

- BIG5 Personality
- Intelligence
- Life Satisfaction
- Political Views
- Religious Views
- Sexuality
- Profession

Demographics

- Age
- Gender
- Relationship status

Our trait prediction engine is based on actual psychological data and the corresponding social media profiles of over 6 million individuals.

We publish anonymised samples of our data online, resulting in 33 peer-reviewed publications since 2011.

Not to mention global research collaboration among 80 academic institutions.

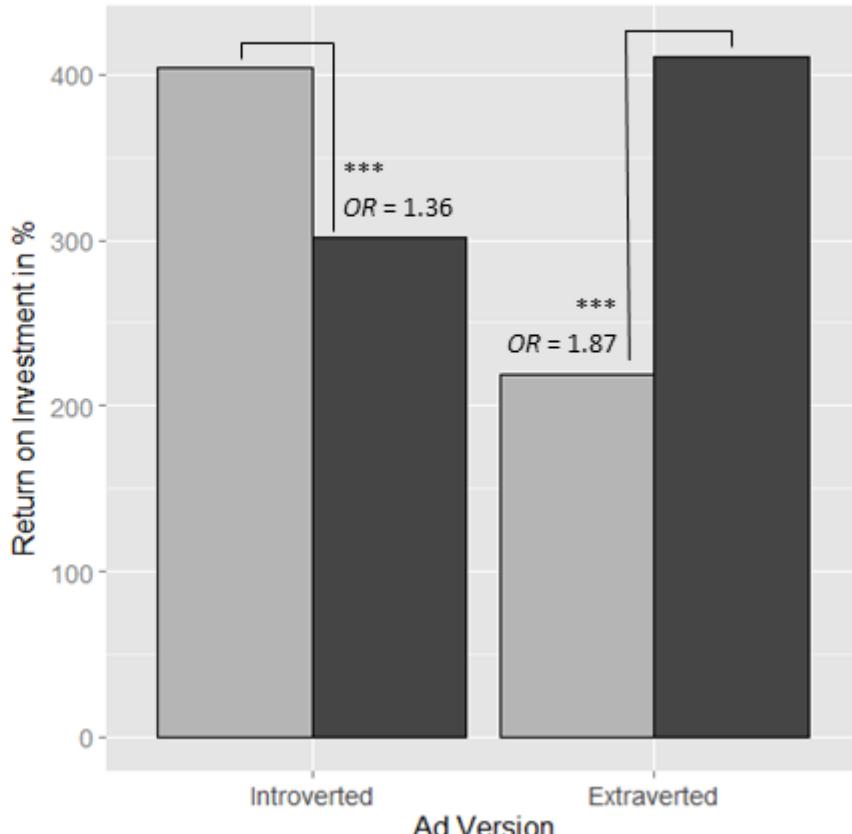
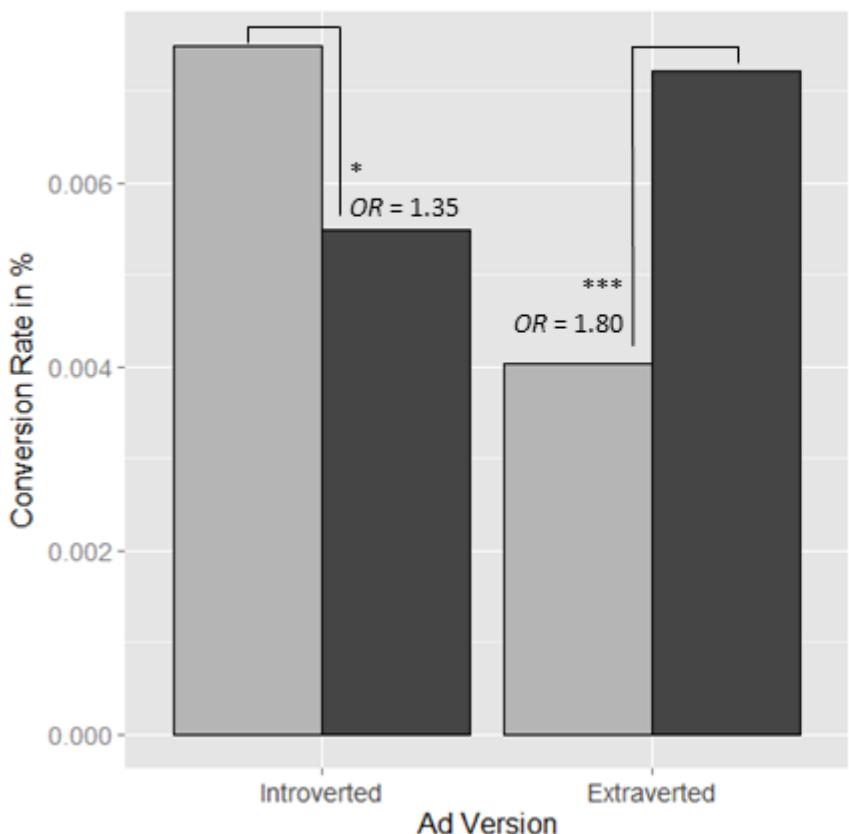


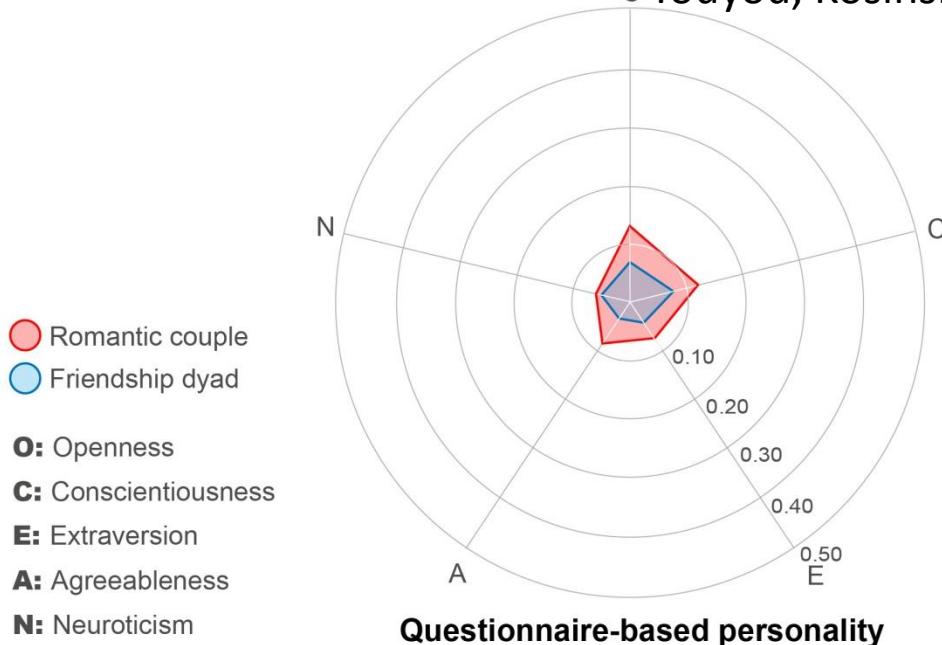
Love the spotlight and feel the moment



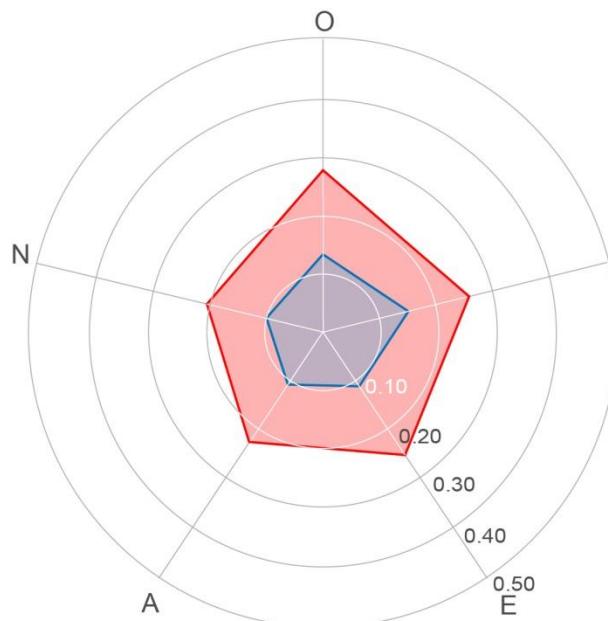
Beauty doesn't have to shout

Target group Introverted Extraverted

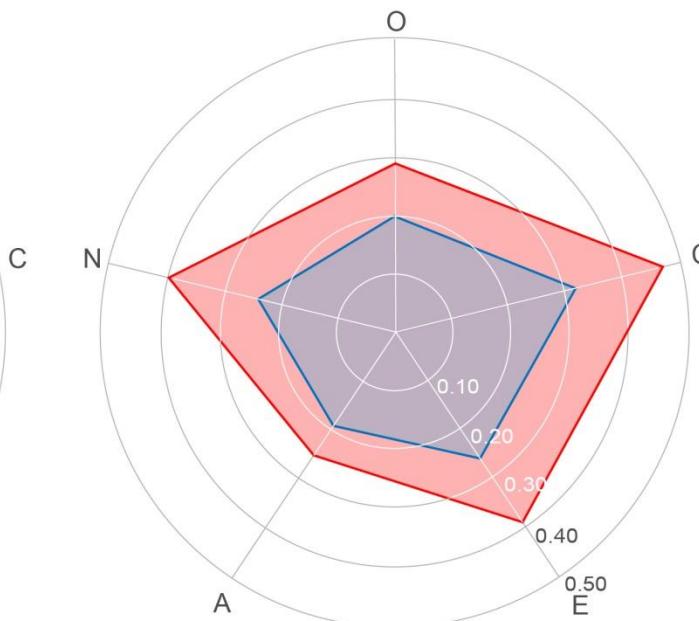




Questionnaire-based personality



Behavior-based personality



Language-based personality

# Interest in collaboration

1. Use of behaviour-based personality predictions in practice
2. Use of myPersonality data. See: [www.mypersonality.org/wiki](http://www.mypersonality.org/wiki)



David Stillwell

ds617@cam.ac.uk

Lecturer in Big Data Analytics and  
Quantitative Social Science



UNIVERSITY OF  
CAMBRIDGE  
Judge Business School

Deputy Director, The Psychometrics Centre