

A zero-item personality test? Predicting personality traits from social media data

Dr. David Stillwell

Lecturer in Big Data Analytics and Quantitative Social Science

Academic Director of the Psychometrics Centre

Cambridge University Judge Business School



Summary

- **A zero-item test: Predicting psychological traits from digital footprints:**
 - Facebook Likes
 - Status updates (text)
 - Profile pictures
- **Applications of automated psychometric testing**

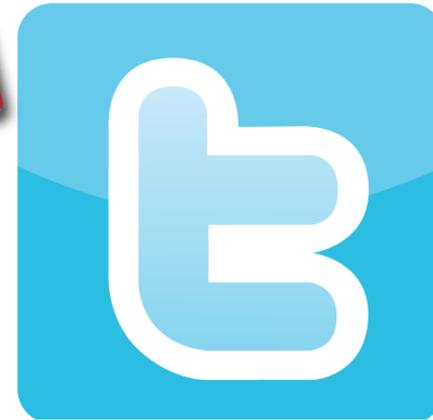
Predicting Psychological Traits from Digital Footprints



Google™
bing™
YAHOO!



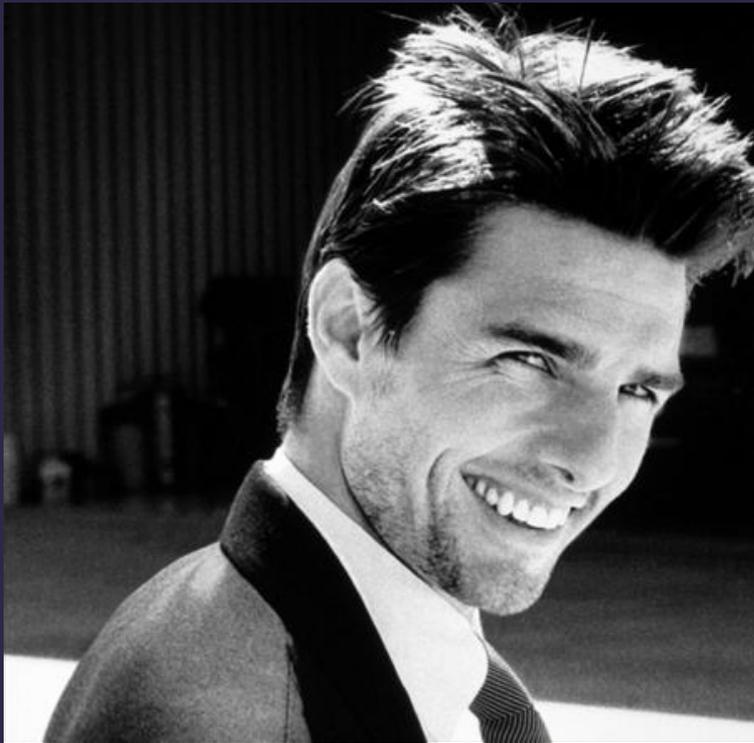
Gmail™
by Google



amazon.com®



Which celebrity do you prefer?



Tom Cruise

A

or



Frank Sinatra

B

Results – Openness Scale

Conservative
& Traditional

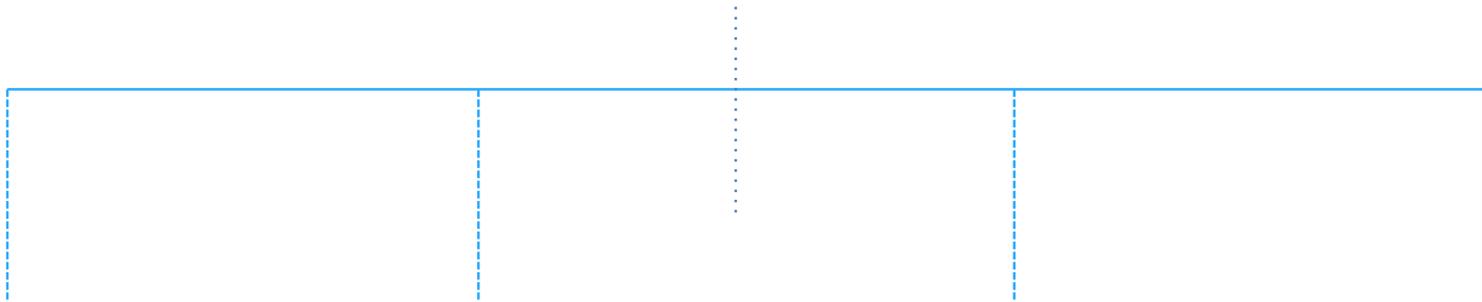
Average (50%)

Liberal &
Artistic

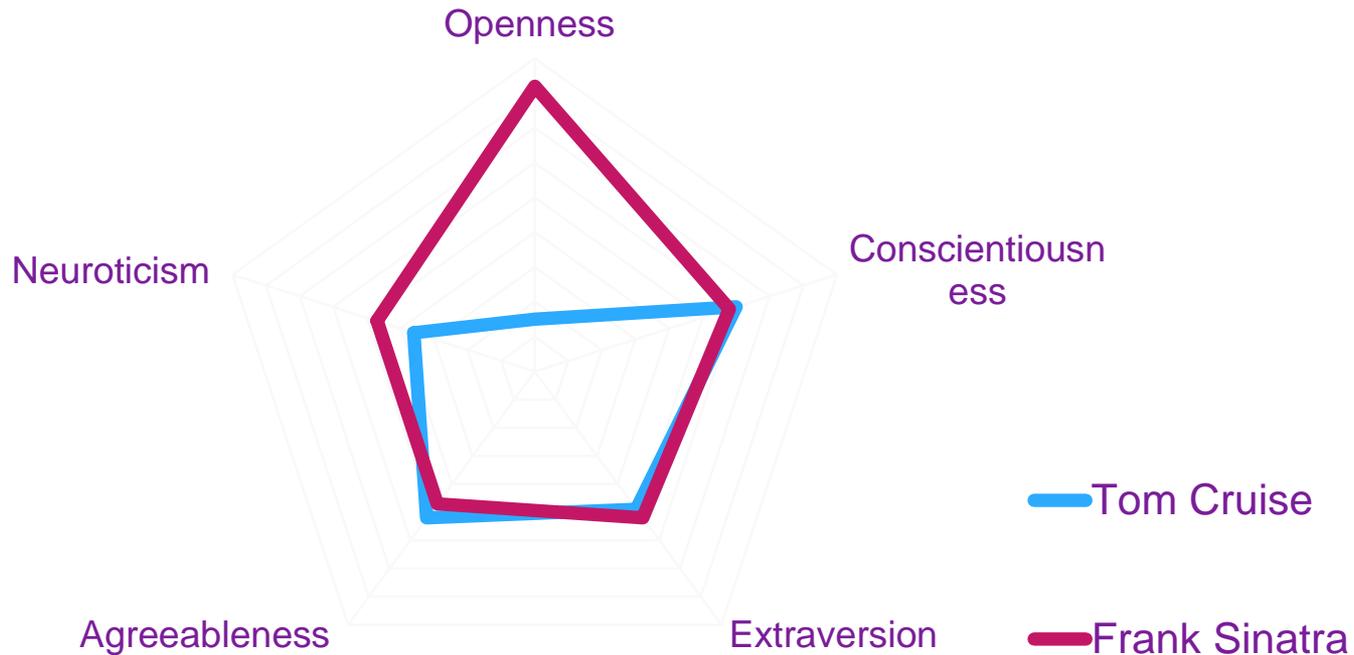
AA

AB
BA

BB



Extracting Average Psychological Profile from Social Media



Examples of personality questions

EXTRAVERSION

“I make friends easily”

“I am the life of the party”

“I would describe my experiences as somewhat dull”

“I don't talk a lot”

Examples of personality questions

OPENNESS

“I tend to vote for liberal political candidates”

“I have a vivid imagination”

“I do not enjoy going to art museums”

“I am not interested in abstract ideas”

myPersonality app (2007-2012)



**6mil individual psych &
social profiles**



**25 validated psychometric
tests**



**All data collected through
opt-in**



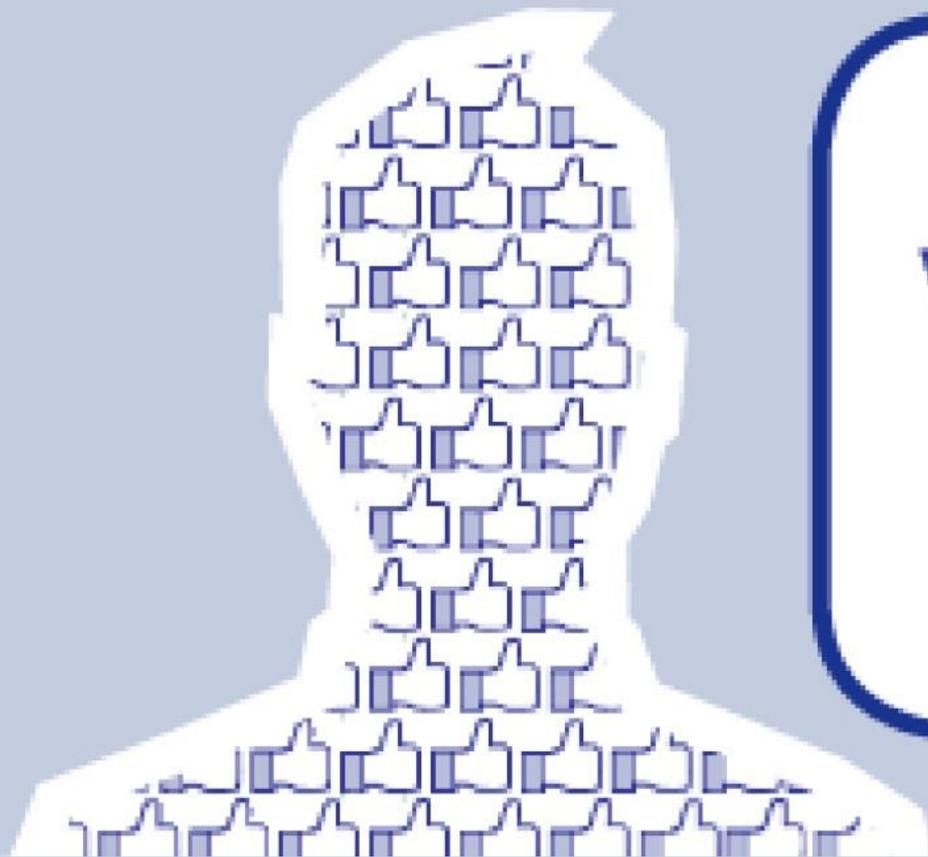
**Data shared with >100
Universities**



**Feedback was the only
incentive**



**40 journal articles since
2011**



**You are
what you
Like**



Identity Claims

[Home](#) > [Early Edition](#) > [Michal Kosinski](#)

Private traits and attributes are predictable from digital records of human behavior

Michal Kosinski^{a,1}, David Stillwell^a, and Thore Graepel^b



[Current Issue](#)
[E-mail Alerts](#)
[Subscribe](#)
[RSS](#)

“Digital records of behavior can be used to automatically and accurately predict a range of highly sensitive personal attributes”

March 2013

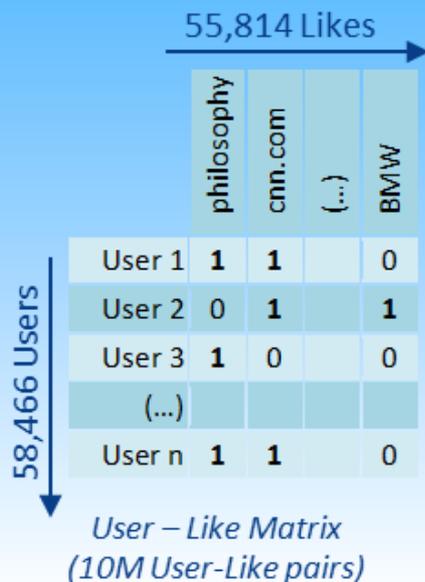


Predicting using SVD+Linear Regression

Kosinski, Graepel & Stillwell (2013) *PNAS*

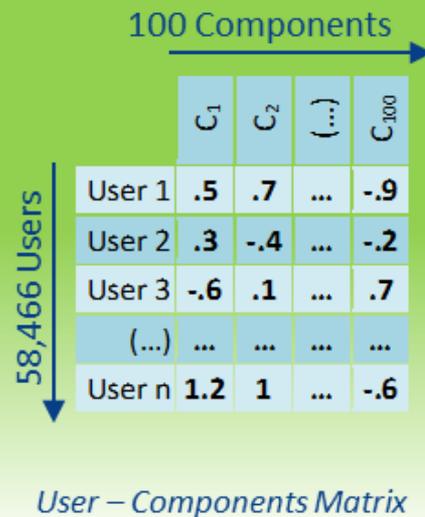
1

Users' Facebook Likes



2

Singular Value Decomposition



3

Prediction Model

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

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

Predicted variables

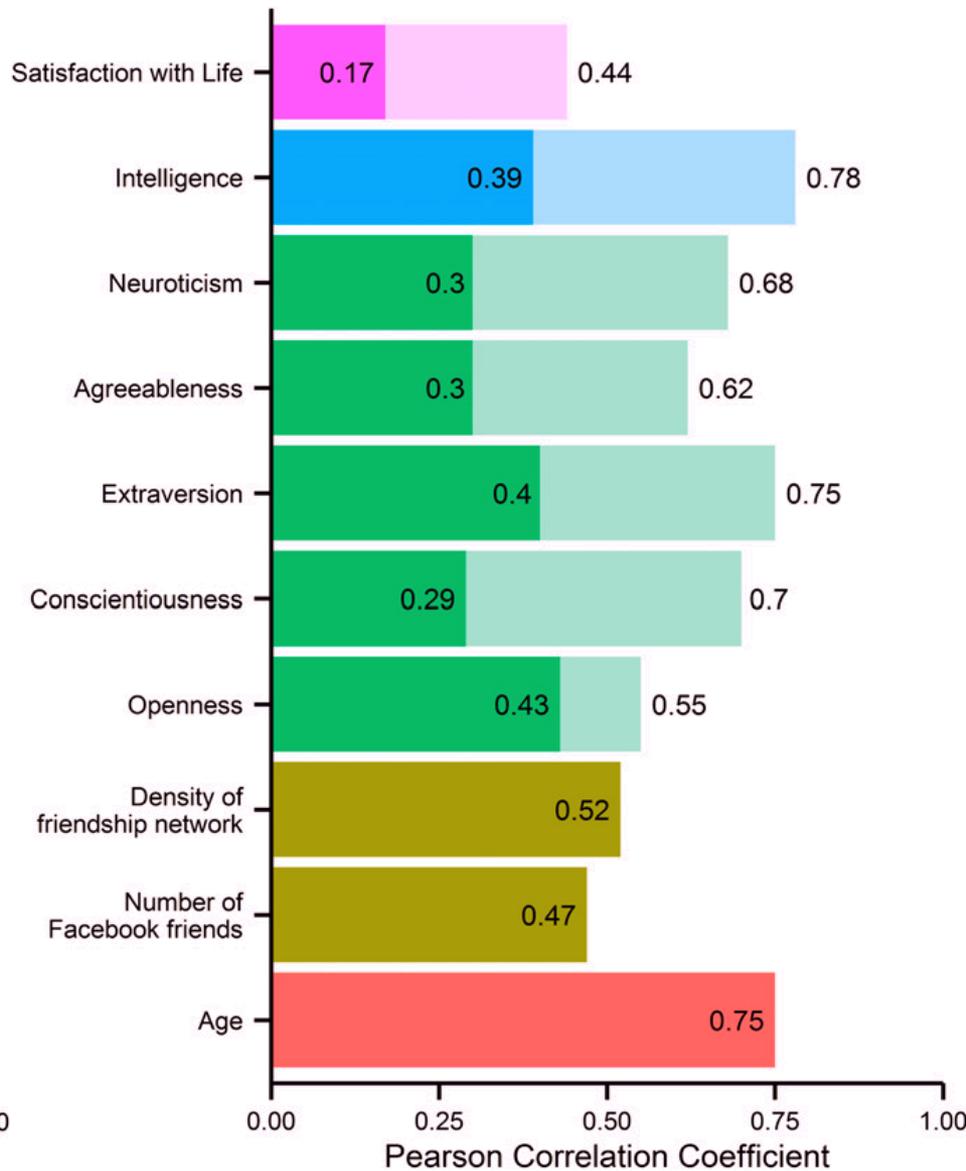
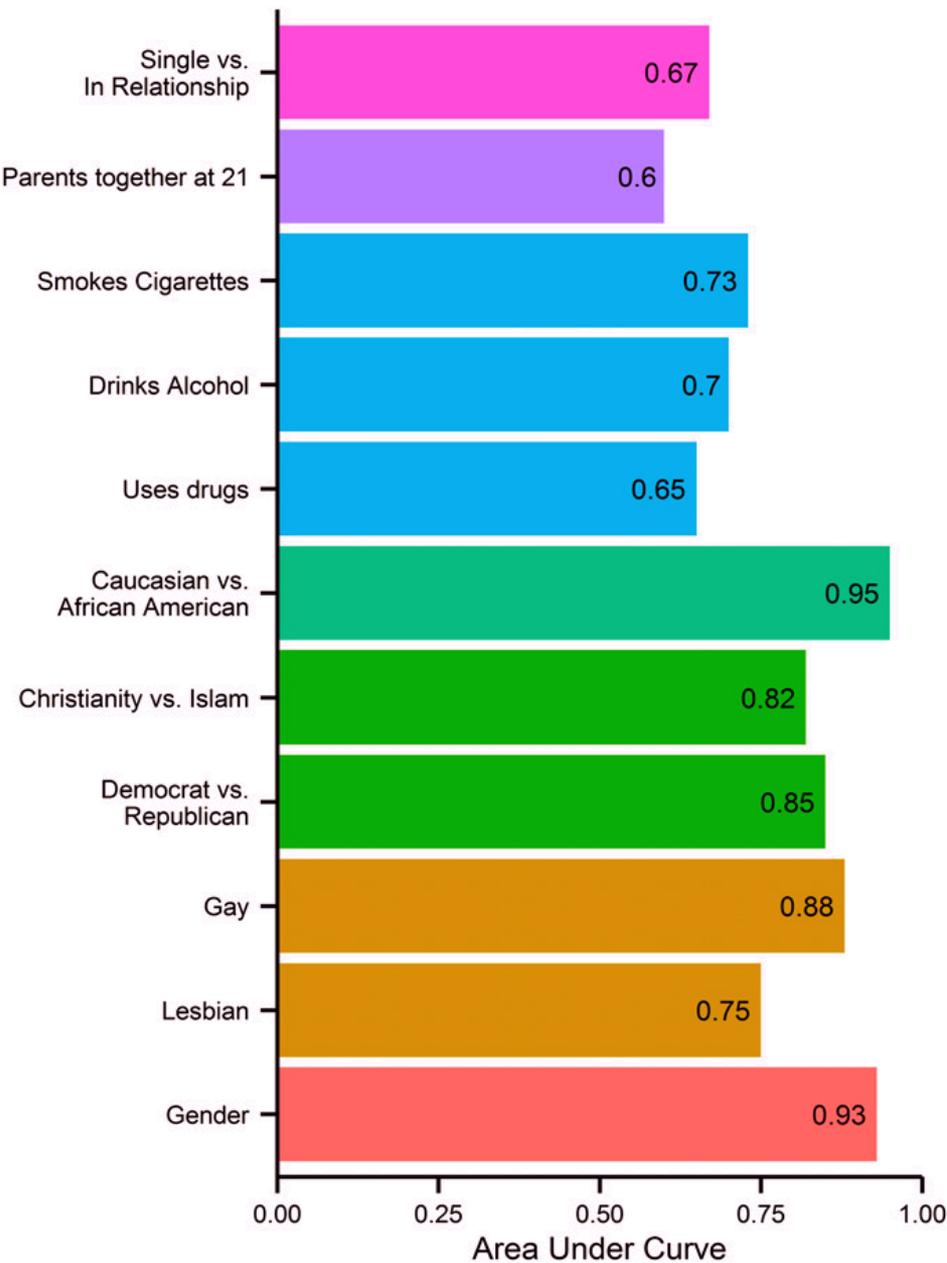
Facebook profile: age, gender, political and religious views, relationship status, proxy for sexual orientation, social network size and density

Profile picture: race

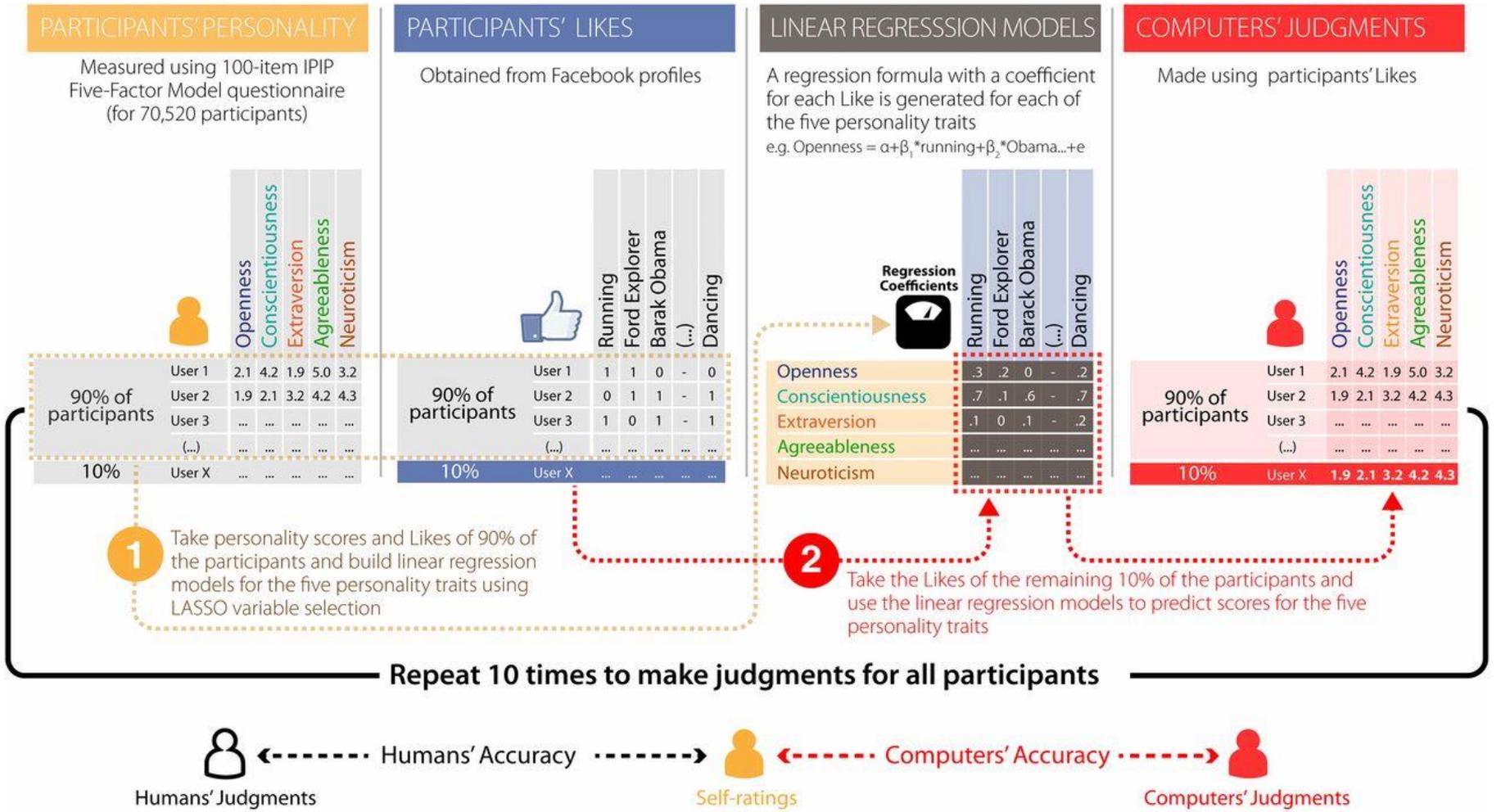
Survey / test results: BIG5 Personality, Intelligence, Satisfaction with Life, substance use, parents together?

What can we predict?

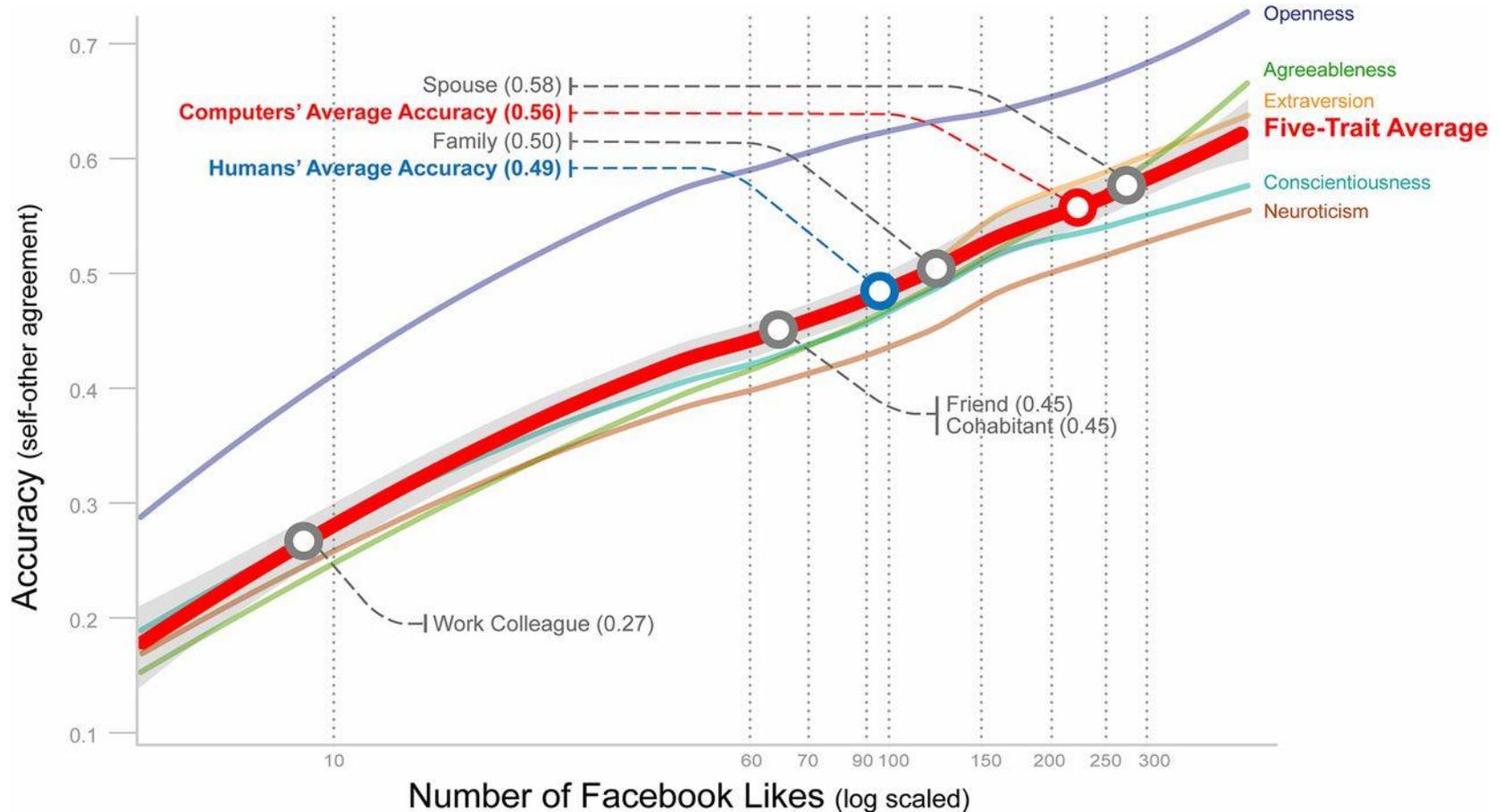
Kosinski, Graepel & Stillwell (2013)



Predicting using LASSO+Linear Regression



Predictions More Accurate than Humans



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 Status Updates



~~████████████████████~~

When people go underwater in movies, I like to hold my breath and see if I would have survived in that situation.

Unlike · Comment · 15 minutes ago near Brisbane ·

You and 14 others like this.



~~████████████████████~~ almost died in finding nemo

12 minutes ago · Like · 11



Popeska

When I order coffee or whatever from starbucks and they ask me for my name I like to look at their name tag and then just say their name and then they're always like nooo wayyyy that's my name too and then I'm also like nooo wayyy and I always expect them to give me something for free because we have the same name but they never do

Like · Comment · Share ·  197  13  1 · 48 minutes ago · 

 197 people like this.

 1 share

Table 1

Convergent Correlations (Pearson r) Between Language-Based Assessments and Self-Reports of Big Five Personality Traits

	Correlations with self-report questionnaires		
	All versions	20-item	100-item
Language-based assessment			
Openness	.43	.38	.46
Conscientiousness	.37	.34	.38
Extraversion	.42	.39	.41
Agreeableness	.35	.31	.40
Neuroticism	.35	.30	.39
<i>M</i>	.38	.34	.41

Note. $N_s = 4,824$ (all versions), 2,324 (20-item), and 1,943 (100-item). Average correlations within each column are calculated by first applying Fisher's r -to- z transformation to each correlation, averaging, and transforming back to r . All correlations are significantly greater than zero ($p < .001$).

Correlations Between Language-Based Assessments and Self-Reports of Facet-Level Big Five Personality Traits

Self-reported questionnaire	Language-based assessments				
	O	C	E	A	N
Openness	.41	-.12	.00	-.08	.01
Liberalism	.33	-.23	-.02	-.14	.08
Intellect	.34	-.12	-.04	-.11	-.08
Adventurousness	.12	.01	.20	-.01	-.15
Emotionality	.17	.09	.05	.08	.13
Artistic interests	.27	.03	.12	.16	.04
Imagination	.31	-.24	-.03	-.15	.07
Conscientiousness	-.03	.26	.20	.17	-.16
Cautiousness	.02	.08	-.03	.11	-.03
Self-discipline	-.04	.25	.20	.15	-.13
Achievement-striving	.05	.29	.26	.16	-.14
Dutifulness	.01	.19	.01	.26	-.11
Orderliness	.00	.14	.05	.10	-.04
Self-efficacy	.03	.18	.26	.06	-.26

Correlations Between Self-Reports, Informant Reports, and Language-Based Assessments of Big Five Personality Traits

	LBA and self		Informant and self		LBA and informant	LBA + Informant and self
	<i>r</i>	partial <i>r</i> ^a	<i>r</i>	partial <i>r</i> ^b	<i>r</i>	<i>r</i>
Openness	.46	.42	.25	.13	.30	.44
Conscientiousness	.34	.30	.30	.26	.20	.42
Extraversion	.43	.37	.39	.32	.24	.52
Agreeableness	.38	.34	.30	.24	.24	.44
Neuroticism	.35	.31	.34	.29	.20	.44
<i>M</i>	.39	.35	.32	.25	.24	.45

Correlations Between Self-Reports, Informant Reports, and Language-Based Assessments of Big Five Personality Traits

	LBA and self		Informant and self		LBA and informant	LBA + Informant and self
	<i>r</i>	partial <i>r</i> ^a	<i>r</i>	partial <i>r</i> ^b	<i>r</i>	<i>r</i>
Openness	.46	.42	.25	.13	.30	.44
Conscientiousness	.34	.30	.30	.26	.20	.42
Extraversion	.43	.37	.39	.32	.24	.52
Agreeableness	.38	.34	.30	.24	.24	.44
Neuroticism	.35	.31	.34	.29	.20	.44
<i>M</i>	.39	.35	.32	.25	.24	.45

Average Test–Retest Correlations of Language-Based Assessments of Big Five Personality

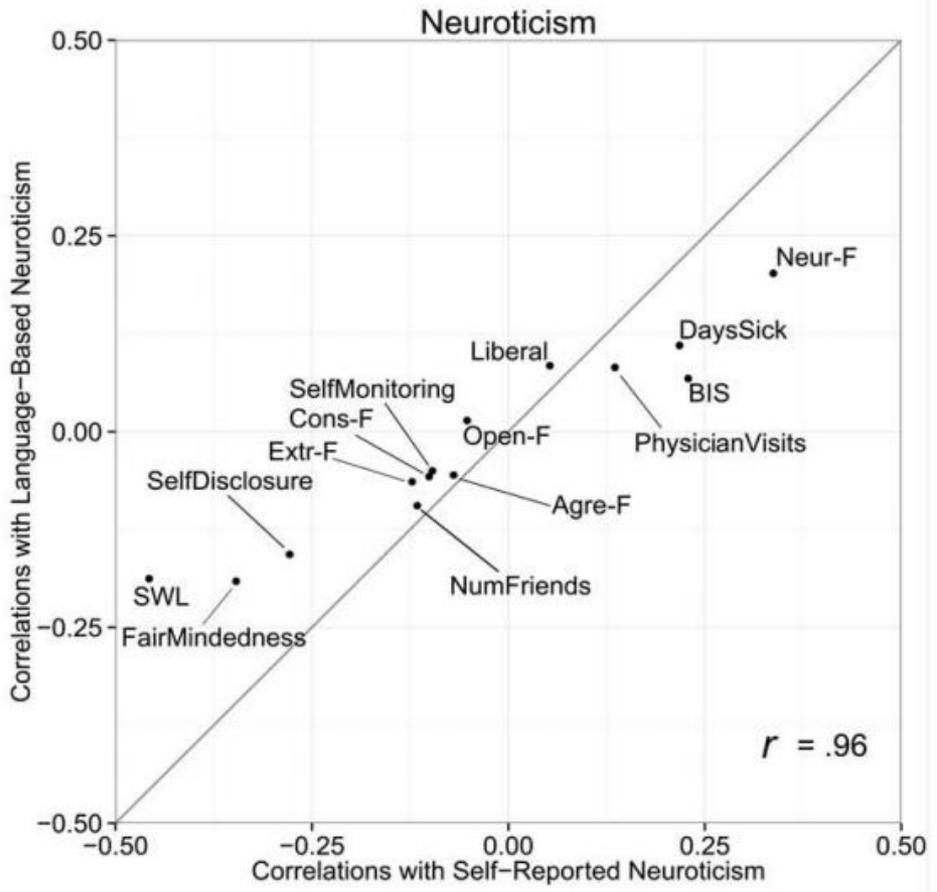
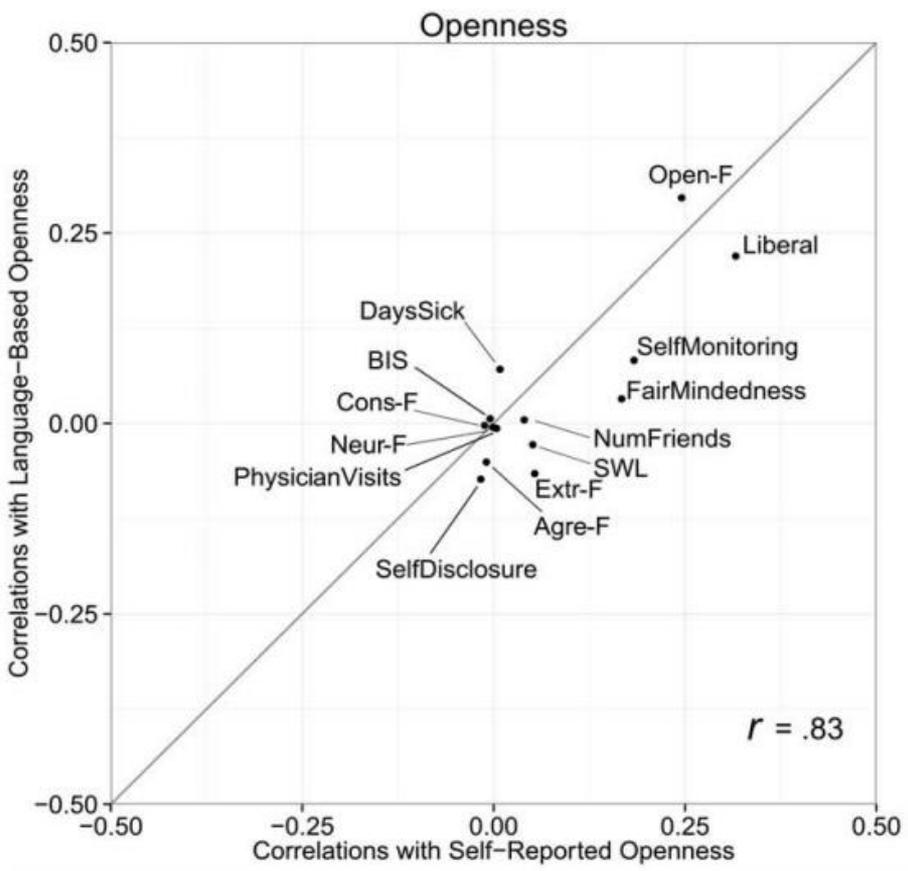
	Time 2	Time 3	Time 4
Time 1	.69 (681)	.66 (625)	.61 (331)
Time 2		.70 (1,424)	.65 (680)
Time 3			.71 (1,019)

Time 1 = July 2009 to December 2009

Time 2 = January 2010 to June 2010

Time 3 = July 2010 to December 2010

Time 4 = January 2011 to June 2011



Apply Magic Sauce

PredictionAPI

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

For research

For business

For you



Digital footprints

Facebook LikeIDs >

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

Try it yourself

www.applymagicsauce.com

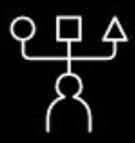


Marketplace

Search



< Products and Services



Personality Insights

Predict personality characteristics, needs and values through written text. Understand your customers' habits and preferences on an individual level, and at scale.

[Get started free](#)

[View demo](#)

Already using Personality Insights? [Log in](#)



Contact IBM



Predicting Intelligence from Profile Pictures

Predicting Intelligence from Profile Pictures

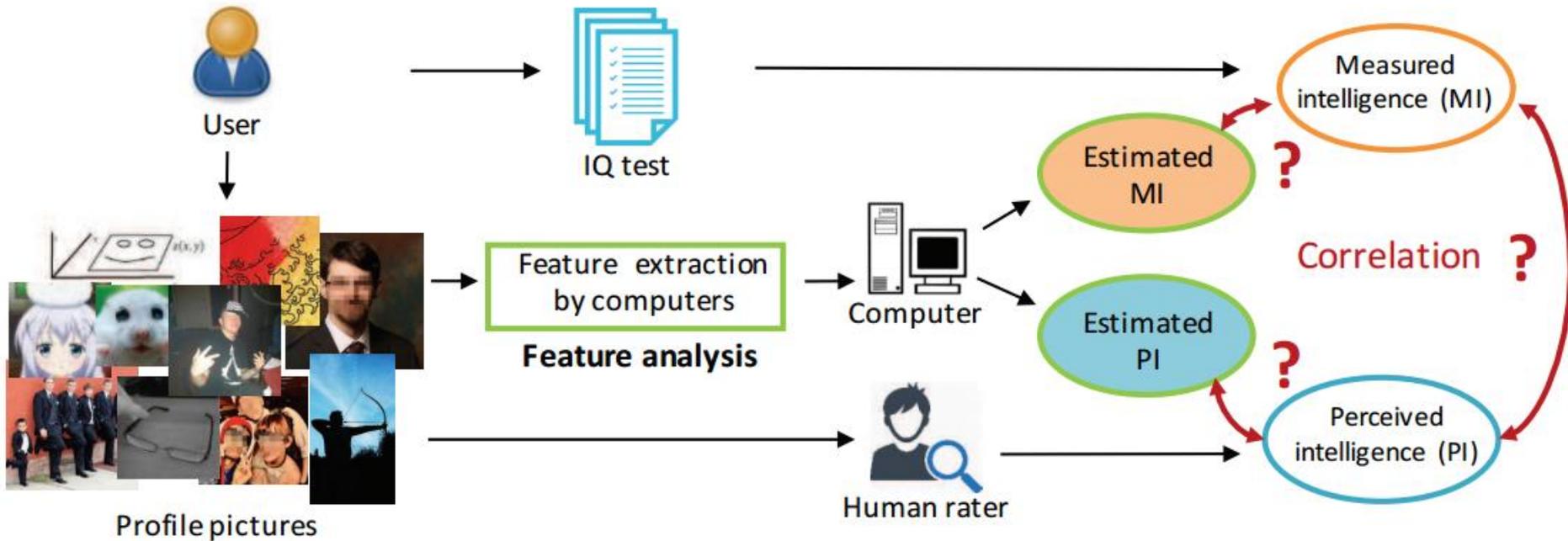
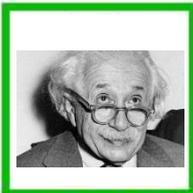


Table 4: Estimation results

	Spearman ρ	RMSE	NRMSE
MI			
Human	0.24***	–	–
Computer	0.27***	14.50	0.20
Random	$< 0^*$	15.13	0.21
Mean	–	14.49	0.20
PI			
Computer	0.36***	0.54	0.15
Random	$< 0^{**}$	0.58	0.17
Mean	–	0.56	0.16

*** : $p < 0.001$, ** : $p < 0.01$, * : $p < 0.05$

Profile Photo



Name

Albert Einstein



WeChat ID

i_am_einstein

My QR Code



Gender

Male



Region

New Jersey, United States



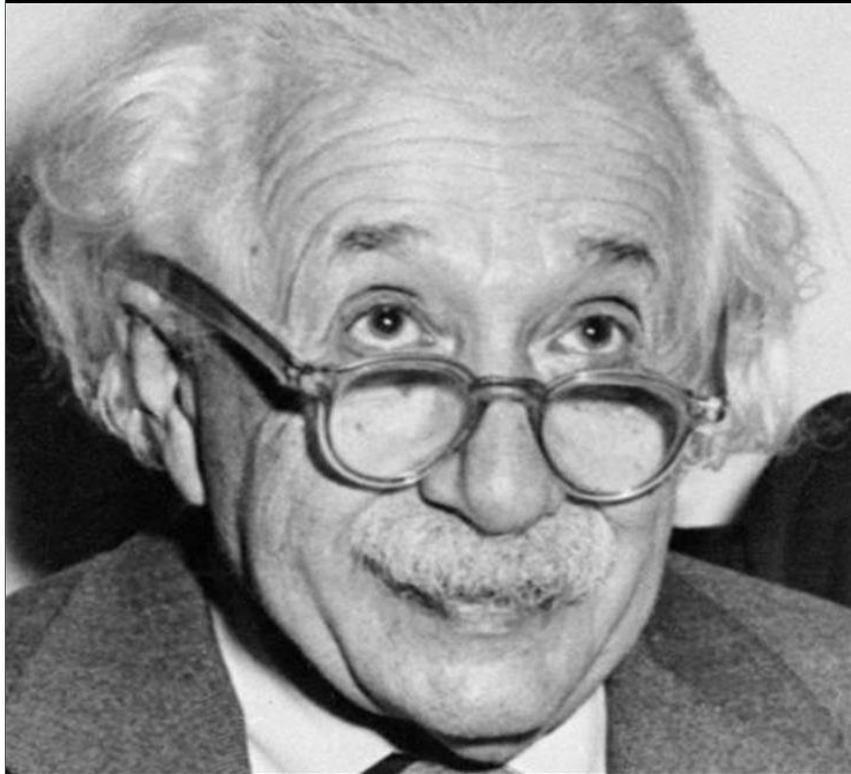
What's Up

Good day!



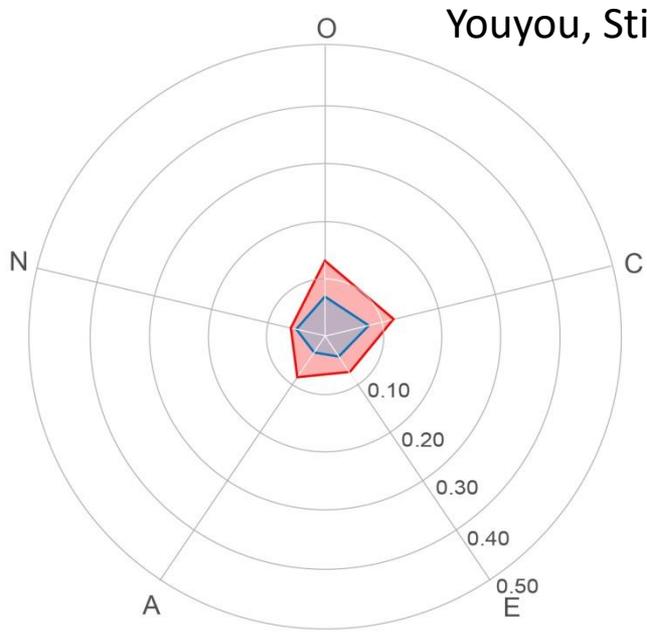
LinkedIn Account

Not Set

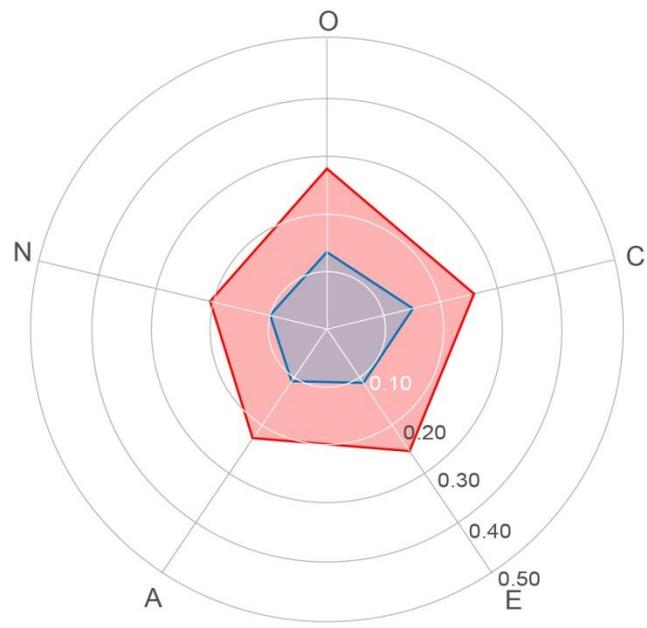


Reference Group Effect

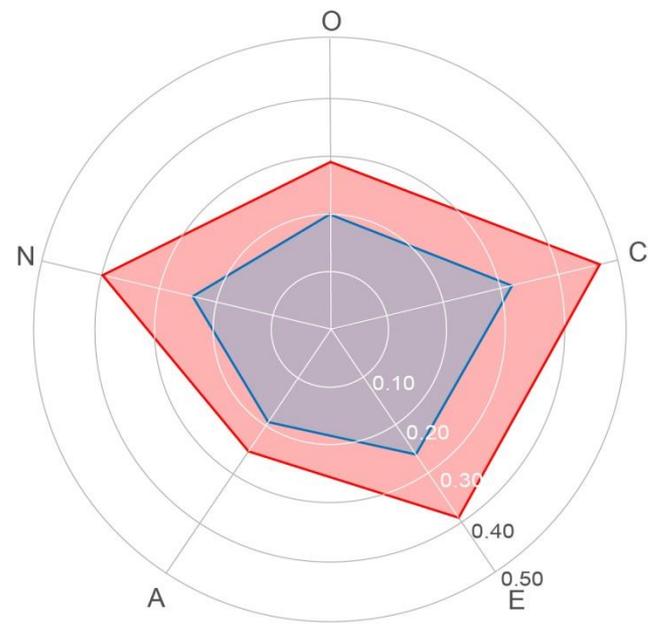
- Romantic couple
- Friendship dyad
- O:** Openness
- C:** Conscientiousness
- E:** Extraversion
- A:** Agreeableness
- N:** Neuroticism



Questionnaire-based personality



Behavior-based personality



Language-based personality

Applications

In: Finance

Marketing of cosmetics

Choosing the US President

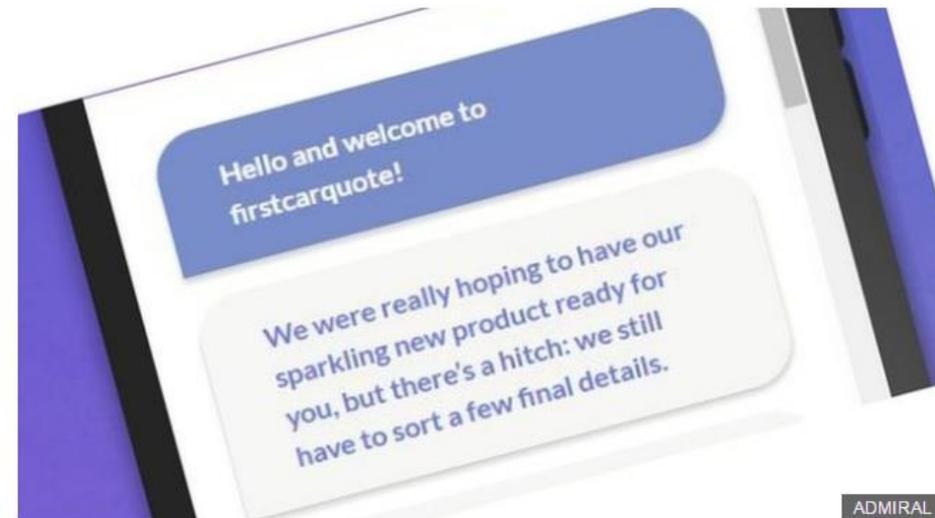
Occupational selection

Facebook blocks Admiral's car insurance discount plan

By Kevin Peachey
Personal finance reporter

🕒 2 November 2016 | [Business](#) | 📄

🔗 Share



ADMIRAL

Facebook has blocked plans by an insurer to view young drivers' profiles to help set car insurance premiums.

In a planned trial, insurer Admiral wanted new motorists' permission to look at their posts and likes to judge their safety as a driver.

Were a young driver considered to be low-risk, a discount would be offered.

But on the day of the planned launch, Facebook said that Admiral would not be able to determine discounts on the basis of Facebook posts and likes.

New York Insurers Can Evaluate Your Social Media Use—If They Can Prove Why It's Needed

New guidance applies to companies operating in New York, but industry consultants say it could have an impact beyond the state's borders

By [Leslie Scism](#)

Jan. 30, 2019 9:00 a.m. ET

New York's top financial regulator is going to allow life insurers to use data from social media and other nontraditional sources when setting premium rates, though the insurers will have to prove the information doesn't unfairly discriminate against certain customers.

New York is the first state to set specific guidance governing how life insurers use algorithms to comb through everything from homeownership records to credit scores and internet use in an effort to size up an applicant's risk. At least

High Extraversion



“Dance like no one’s watching (but they totally are).”

Low Extraversion



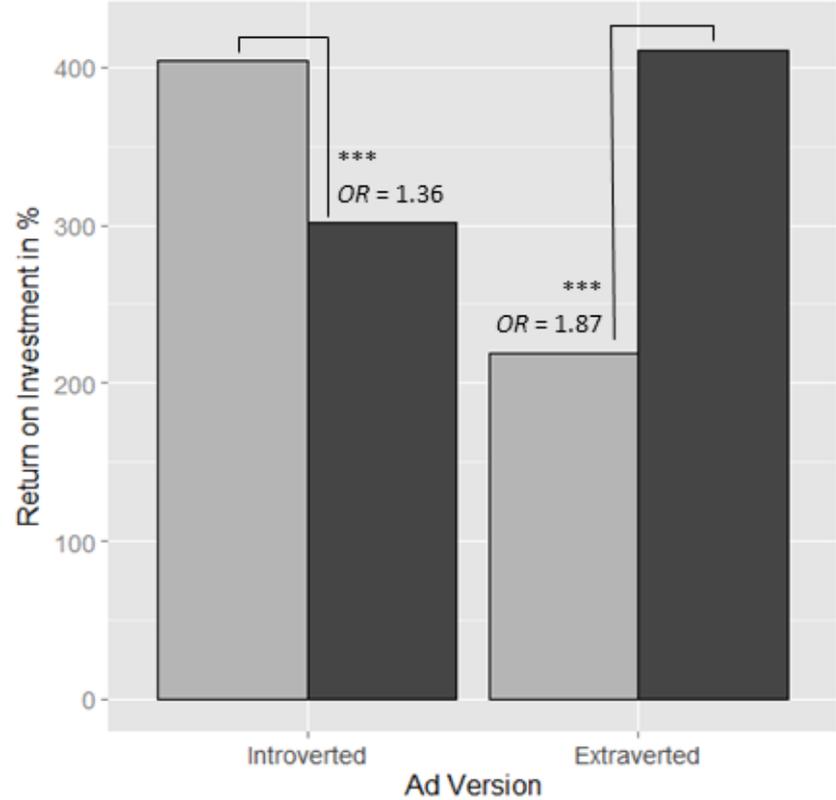
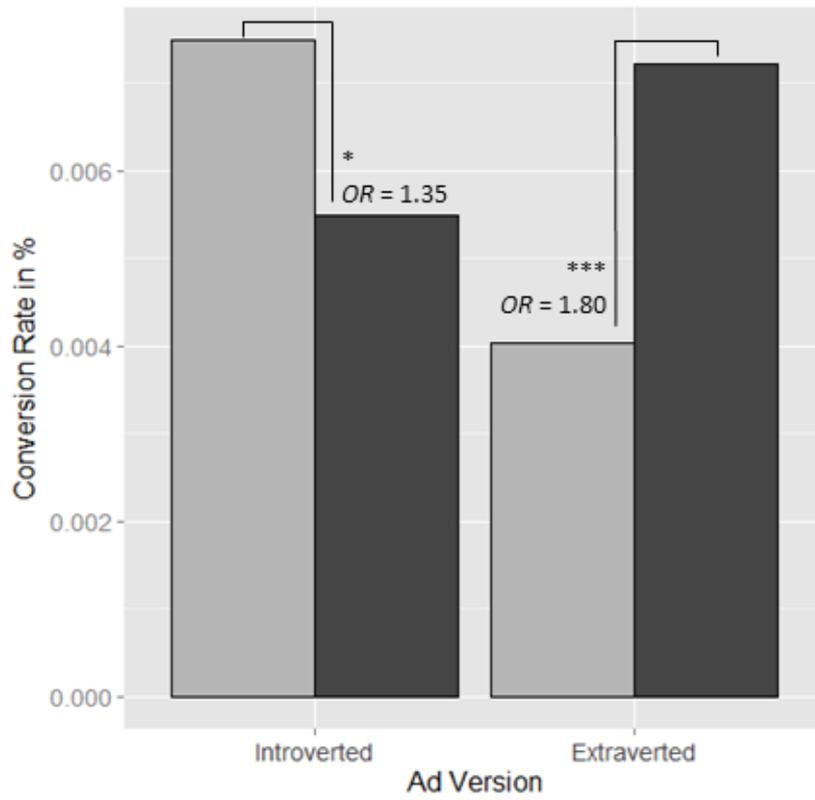
“Beauty doesn’t have to shout.”



Love the spotlight and feel the moment



Beauty doesn't have to shout



“The Data that Turned the World Upside Down”



Cambridge
Analytica

The Power of Big Data and Psychographics
in the Electoral Process



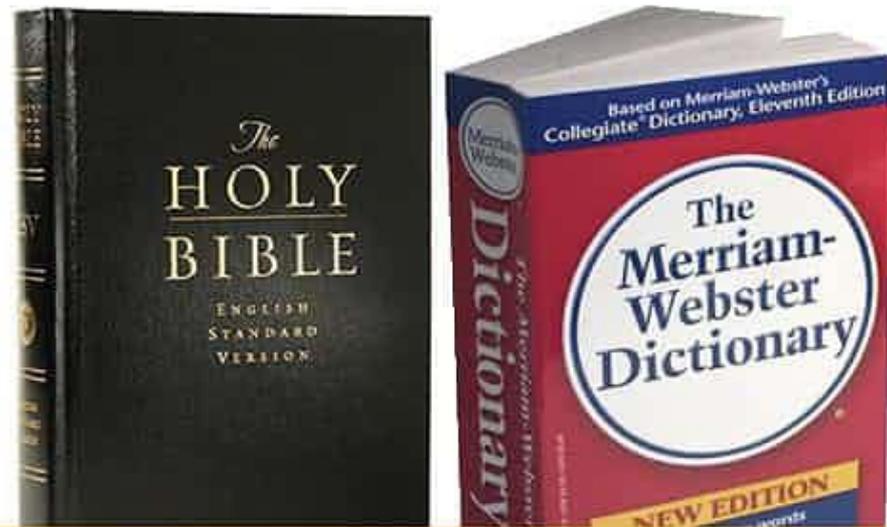
CONCORD



https://motherboard.vice.com/en_us/article/how-our-likes-helped-trump-win

Exploiting Psychological Weaknesses

- They could, for example, target people high in Neuroticism with images of immigrants “swamping” the country, to **encourage their natural inclination for fear**
- Or show **conspiracy theories** to people who are high in Openness
- Overall Trump’s election campaign aimed at “**voter disengagement**” and “to persuade Democrat voters to stay at home”



Look up 'marriage' and
get back to me.

BECAUSE TRADITION IS NOT OLD-FASHIONED



“For someone who is conscientious, it is a compelling message: a dictionary is a source of order, and a conscientious person is more deferential to structure.” – Chris Wylie, ex-Research Director of Cambridge Analytica. “It’s funny, because this is so offensive [...] but it’s a team of gays that created it.”

Principles to Make Customers Comfortable

- **Transparency** (they understand what is being done with their data)
- **Control** (opt in consent, can opt out too)
- **Relevance** (what you're doing makes sense to the customer and seems fair)
- **Benefits them** (customer feels that it's a good use of their data)



Find your Perfect Match at SAP!

Are you looking for your dream job? You're just one click and seconds away from finding it at SAP! Get your results below and then make a full application at our Careers site. **Our motto:** *Bring Everything You Are. Become Everything You Want.*

We will gather data from your Facebook profile and may use it to identify areas of our business that could be of interest to you, but the data will not otherwise be used as part of our assessment of you as a candidate.

Get your results



A bit about you

You tend to...

Be good-natured and eager to cooperate • Be easy-going, preferring not to make strict plans •
Prefer to spend time alone or with a few close friends • Be practical and traditional



David Stillwell

Results

A job type that matches your persona

Development

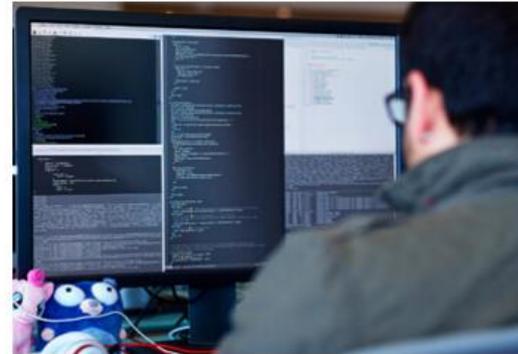
As one of our Development and Technology gurus, you will work with other creative minds to engineer innovative software, re-think user experiences, and develop product road maps, and so much more!

Our team of developers work with other top engineering talent to design, implement, and maintain the modules and features of the SAP product line, including architecture, applications, and solutions.

If User Experience is your passion, then you may be a perfect match on our UX design teams, where you'll work on visual interaction and physical manipulation projects, and perform task and objective analysis.

If creating standards, processes, and tools for IT governance topics such as security and quality management, then a Business Services or IT Technology role sounds perfect for you.

For the best in class technical Development roles, look no further than SAP.



[Back](#)

[Share results](#)

[Full application](#)

Conclusion

- **It is possible to predict psychological traits from digital footprints with group-level accuracy**
 - Facebook Likes
 - Status updates (text)
 - Profile pictures
- **This expands the areas where personality can be deployed in practice**
- **But we have some very tricky ethical questions to contend with**

Thank you for listening

Dr. David Stillwell

Lecturer in Big Data Analytics and Quantitative Social Science
Cambridge University Judge Business School



This couldn't have been done without the following PhD students and researchers
in the Cambridge University Psychometrics Centre:

Michal Kosinski

Sandra Matz

Vess Popov

Xingjie Wei

Youyou Wu