Cambridge Analytica & populist movements
What happened?

Paul-Olivier Dehaye
MyData, September 1st 2017
Dec 2015: Cambridge Analytica works for Ted Cruz

Ted Cruz using firm that harvested data on millions of unwitting Facebook users

Exclusive: Documents reveal donor-funded US startup embedded in Republican's campaign paid UK university academics to collect psychological profiles on potential voters

Ted Cruz campaigns in Sioux City at Briar Cliff University, a private Catholic school. Photograph: Zuma Wire/Rex Shutterstock
# OCEAN model of personality

<table>
<thead>
<tr>
<th>Trait</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Openness</td>
<td>Curious, original, intellectual, creative, and open to new ideas.</td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>Organized, systematic, punctual, achievement oriented, and dependable.</td>
</tr>
<tr>
<td>Extraversion</td>
<td>Outgoing, talkative, sociable, and enjoys being in social situations.</td>
</tr>
<tr>
<td>Agreeableness</td>
<td>Affable, tolerant, sensitive, trusting, kind, and warm.</td>
</tr>
<tr>
<td>Neuroticism</td>
<td>Anxious, irritable, temperamental, and moody.</td>
</tr>
</tbody>
</table>
Private traits and attributes are predictable from digital records of human behavior

Michał Kosinski1,2, David Stillwell3, and Thore Graepel3

1School of Data, The Psychometrics Centre, University of Cambridge, Cambridge CB2 8QZ United Kingdom; and 2Microsoft Research, Cambridge CB2 3RF, United Kingdom.

Computer-based personality judgments are more accurate than those made by humans

Wu Youyou1,2, Michael Kosinski1,2, and David Stillwell3

1Department of Psychology, University of Cambridge, Cambridge CB2 3RF, United Kingdom; and 2Department of Computer Science, Stanford University, Stanford, CA 94305

Edited by David Funder, University of California, Riverside, CA, and accepted by the Editorial Board December 2, 2014 (received for review September 28, 2014)

Judging others’ personalities is an essential skill in successful social living, as personality is a key driver behind people’s interactions, behaviors, and emotions. Although accurate personality judgments stem from social-cognitive skills, developments in machine learning show that computer models can also make valid judgments. This study compares the accuracy of human and computer-based personality judgments, using a sample of 84,220 volunteers who completed a 100-item personality questionnaire. We show that (i) computer predictions based on a generic digital footprint (Facebook Likes) are more accurate (r = 0.56) than those made by the participants’ Facebook friends using a personality questionnaire (r = 0.49); (ii) computer models show higher inter-judge agreement; and (iii) computer personality judgments have higher external validity when predicting life outcomes such as substance use, political attitudes, and physical health; for some outcomes, they even outperform the self-rated personality scores. Computers outgrowing humans in personality judgment presents significant opportunities and challenges in the areas of psychological assessment, marketing, and privacy.

Fig. 2. Computer-based personality judgment accuracy (y axis), plotted against the number of Likes available for prediction (x axis). The red line represents the average accuracy (correlation) of computer’s judgment across the five personality traits. The five-trait average accuracy of human judgments is positioned onto the computer accuracy curve. For example, the accuracy of an average human individual (r = 0.49) is matched by that of the computer models based on around 90–100 Likes. The computer accuracy curves are smoothed using a LOWESS approach. The gray ribbon represents the 95% CI. Accuracy was averaged using Fisher’s r-to-z transformation.

@podehaye
Ethical lapses at Cambridge University

This information is not held to any extent for the University’s own purposes under the Act.

If you are unhappy with the service you have received in relation to your request and wish to make a
Dec 2016: Das Magazin article
Cambridge Analytica methodology

Registration
Voting history
Turnout probability
Partisanship
Ideology
Issues

Openness
Conscientiousness
Extraversion
Agreeableness
Neuroticism (>< Emotional Stability)

Need for Cognition
Need for Affect (?)
Locus of Control (?)
Reciprocity
Scarcity
Authority
Fear
Social proof
Evidence: use of “Likes” (Nov 2015)

http://tinyurl.com/ca-fb-likes
Evidence: Need for Cognition (Oct 22 2016)

http://tinyurl.com/ca-cognition

“Method Effects and the Need for Cognition Scale”
Robert Mercer: the big data billionaire waging war on mainstream media

With links to Donald Trump, Steve Bannon and Nigel Farage, the rightwing US computer scientist is at the heart of a multimillion-dollar propaganda network
Facebook-based data and money flows
Investigations launched

Electoral Commission urged to investigate Farage's Brexit campaign

Stephen Kinnock suggests Leave.EU broke the law by not declaring the role of a group used by Donald Trump

Information Commissioner opens a formal investigation into the use of data analytics for political purposes:
iconewsblog.wordpress.com/2017/05/17/in...
But...
Press denials = complete ethical failure
FACEBOOK FAILED TO PROTECT 30 MILLION USERS FROM HAVING THEIR DATA HARVESTED BY TRUMP CAMPAIGN AFFILIATE

Mattathias Schwartz
March 30 2017, 8:01 p.m.
Link with MyData?
Citizens/individuals have power
Subject access requests

Quick guide to asking Cambridge Analytica for your data

Cambridge Analytica has finally responded (past deadline, after some threatening emails) to requests by individuals all over Europe and the United States for a copy of their data. I give here some advice on how to go further, and offer a template for responding at the bottom.

Voter

- voter_id
- age
- gender
- turnout probability
- county_id
- family_id

Vote Event

- voter_id
- election_id

Election

- election_id
- election type
- election date
- ...

http://tinyurl.com/ca-sar-medium
http://tinyurl.com/ca-requests
Response to Subject Access Request (Mar 2017)

Cambridge Analytics

23rd March 2017

USA

Modelled data
These are data that represent predictions we have made about you as an individual using models that we have developed as part of our general business offering. Our predictions are based on proprietary algorithms and methodologies that leverage the previous 2 categories of data at our disposal.

THE RECIPIENTS OR CLASSES OF RECIPIENTS OF PERSONAL DATA TO WHOM THE DATA WAS OR MAY HAVE BEEN DISCLOSED:

<table>
<thead>
<tr>
<th>Clients</th>
<th>Service providers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Political campaigns</td>
<td>Digital marketing platforms</td>
</tr>
<tr>
<td>Independent expenditure groups</td>
<td>Mail vendors</td>
</tr>
<tr>
<td>Non-profit organizations</td>
<td>Call centers</td>
</tr>
<tr>
<td>Commercial entities</td>
<td>Research partners</td>
</tr>
<tr>
<td></td>
<td>Affiliated data processors</td>
</tr>
<tr>
<td></td>
<td>Legal counsel</td>
</tr>
</tbody>
</table>

Some names and identifying particulars are not being disclosed to protect the identity of third parties.

Yours sincerely,

Julian Wheatland, Group COO

For and on behalf of Cambridge Analytics

<table>
<thead>
<tr>
<th>Importance Rank [1-10]</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>National Debt</td>
<td>10</td>
</tr>
<tr>
<td>Gun Rights</td>
<td>3</td>
</tr>
<tr>
<td>Traditional Social</td>
<td>9</td>
</tr>
<tr>
<td>Environment</td>
<td>5</td>
</tr>
<tr>
<td>Education</td>
<td>4</td>
</tr>
<tr>
<td>Environment Importance</td>
<td>7</td>
</tr>
<tr>
<td>Immigration</td>
<td>8</td>
</tr>
<tr>
<td>Socially Progressive</td>
<td>6</td>
</tr>
<tr>
<td>Jobs and Economy</td>
<td>1</td>
</tr>
<tr>
<td>Healthcare</td>
<td>2</td>
</tr>
<tr>
<td>Registered Partisanship</td>
<td></td>
</tr>
<tr>
<td>Unregistered Partisanship</td>
<td></td>
</tr>
<tr>
<td>2016 General Election</td>
<td></td>
</tr>
</tbody>
</table>

Very Unlikely Republican

Very Unlikely Republican

Very High
Not just “access”

- “logic of the processing”
- source of the data
- destination of the data
- legal basis for processing and transfers
Facebook forced to disclose more information about its ad targeting

Facebook now tells each of its users which advertisers are tracking them individually through so-called Custom Audiences. This change was most likely brought by my legal actions against Facebook, conducted through PersonalData.IO. This change opens up exciting new possibilities for investigative journalism and the #MyData movement, explained below.

Original request
On December 15th 2016, I asked Facebook, through PersonalData.IO, the following:

http://tinyurl.com/fb-pivot
Also in UK

Brexit - Country First, Party Second, Unfortunately!

Leave.EU info@leave.eu via registrar-servers.com
to

With the Conservatives back in government we can start the process of electoral pressure all over again.

Kind regards,

The Leave.EU Team

-10% off GoSkippy insurance

Facebook - Leave.EU
Twitter - Leave.EU
YouTube - Leave.EU
Who Targets Me? (May 2017)

TECHNOLOGY

Facebook’s Role in European Elections Under Scrutiny

By MARK SCOTT  JUNE 7, 2017
“Follow the data” from electoral rolls

Twitter: I am looking for kind volunteer(s) to help me with an FOI project related to Brexit. Sending emails & collating responses. 1/2

1:45 PM - 16 Aug 2017

Carole Cadwalladr @carolecadwalla

If you're able to help, can you email me? carole.cadwalladr@theguardian.com with 'FOI help' in the subject line. Thanks!!

Carole Cadwalladr @carolecadwalla
08 Aug 16

Omg!! Thank you all so much...deftinitely enough people now. THANK YOU!!
Core lessons

- there is a collective interest in protecting personal data, as it translates into political power
- new investigative journalism techniques can shine here, but present their own difficulties (self-implication, proxying, crowdsourcing, etc)
- deep challenge to merely address Cambridge Analytica on moral, ethical or even legal grounds around their data operations