

REGULATING PRIVACY AND THE WAY WE USE (BIG) DATA

**4th German-Brazilian
Dialogue of Science,
Research and Innovation**



disclaimers

Not a data scientist;

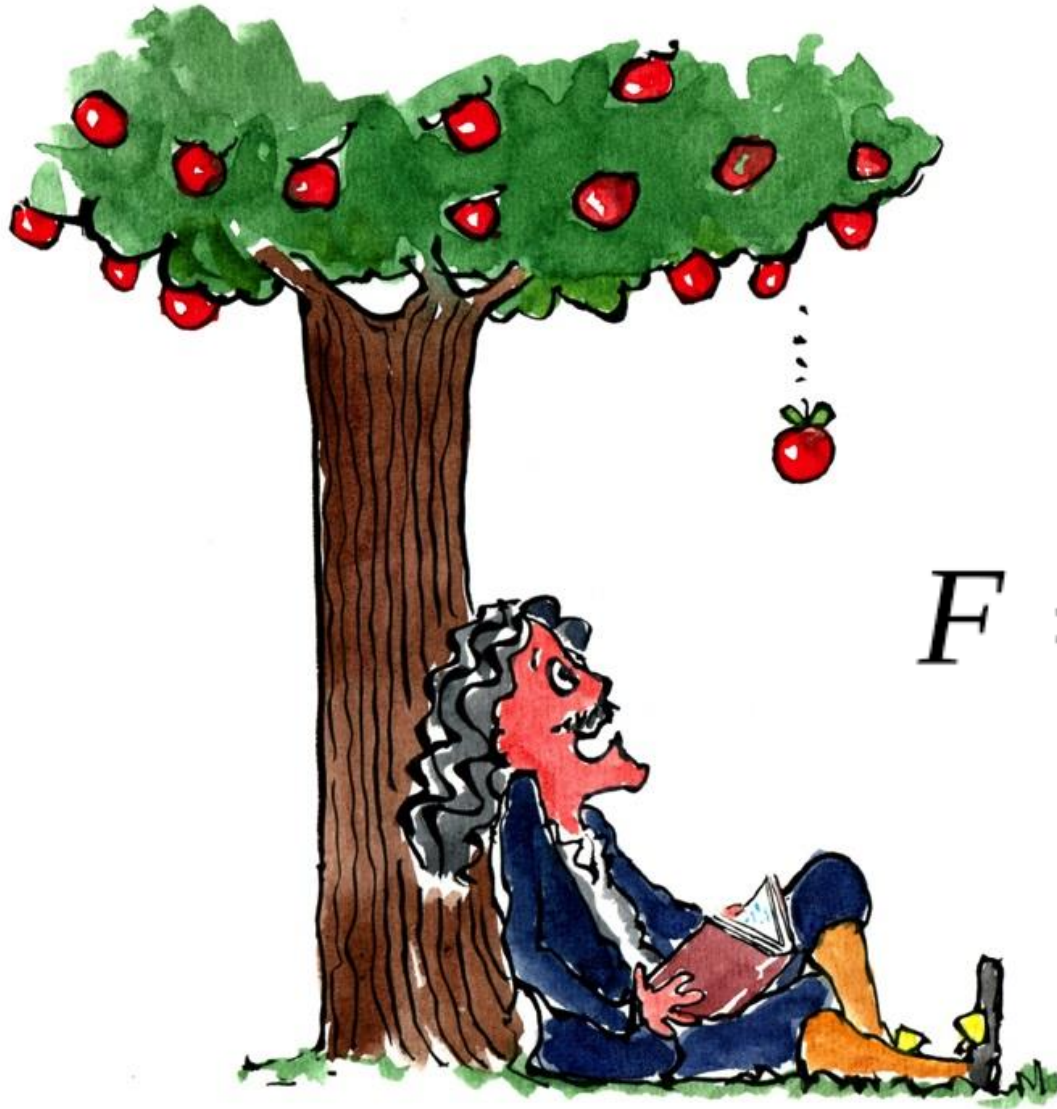
Not a developer;

Not an entrepreneur;

I am writing my PhD **in law**;

Building on Kate Crawford's work.

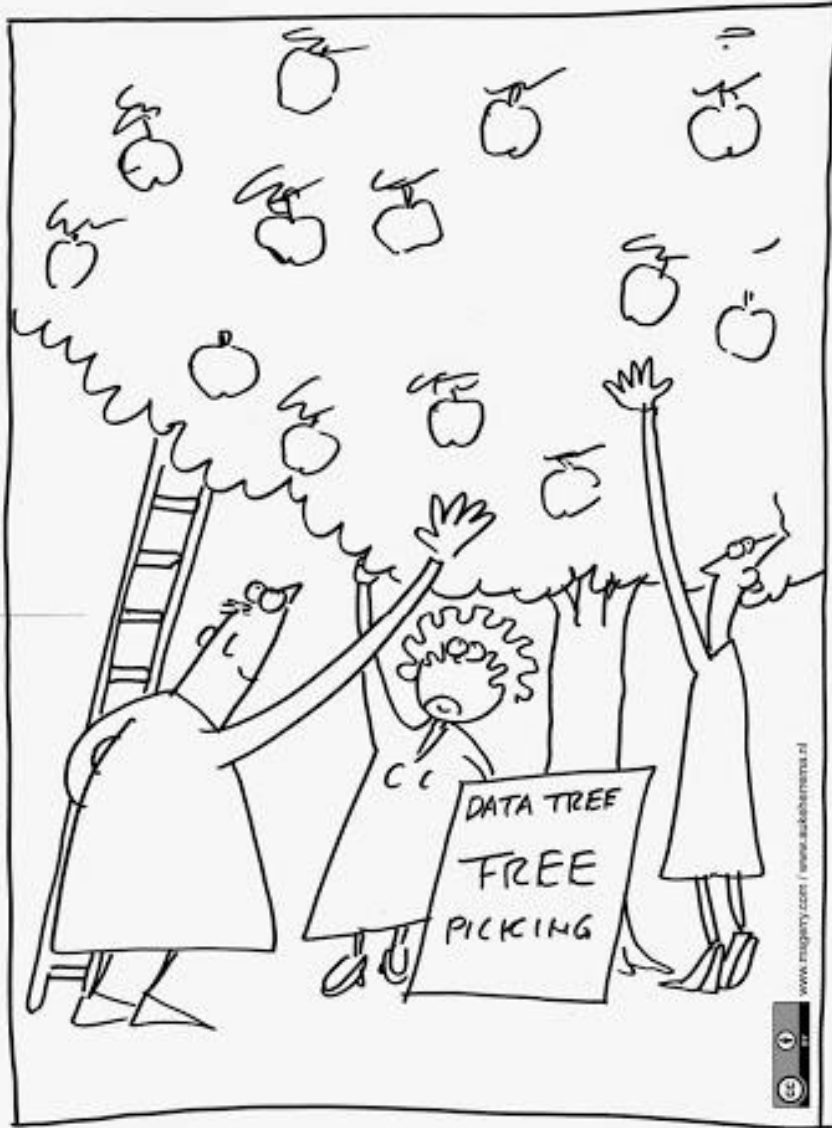
data & science



$$F = G \cdot \frac{m_{g_1} \cdot m_{g_2}}{d^2}$$

by Frits Ahlefeldt

proliferation of data



THE FUTURE?

data science

what's (big) data
science?

(big) data science is the use/manipulation of large and complex datasets that are usually in an unstructured form to...

...produce
meaningful results

“Today, all 20 Premier League football stadiums in the UK are equipped with a set of **8–10 digital cameras that track every player** on the pitch. **Ten data points** are collected **every second** for each of the 22 players on the pitch, generating **1.4 million data points per game**. Prozone analysts will then code the data to identify every tackle, shot or pass in order to enable managers and performance analysts gain insights of what exactly happened in each game, on and off the ball.” (Bernard Marr)



what is the problem?

most of the data are about
you.

most of the data are
personal data.

most of the data generate
more data about you.

most of the data **predict**
things about you.

$$\begin{aligned}
& \max_{x_s^{BS}, x_r^{RN}, \dots} \lambda \left(\sum_{\substack{s \in \mathcal{S}, \\ t \in \mathcal{T}}} b_{st}^{BS} e_{st}^{BS} + \sum_{\substack{r \in \mathcal{R}, \\ t \in \mathcal{T}}} b_{rt}^{RN} e_{rt}^{RN} \right) \quad : \text{Throughput} \\
& + \mu \sum_{\substack{s \in \mathcal{S}, \\ r \in \mathcal{R}}} z_{sr}^{BS2RN} e_{sr}^{BS2RN} \quad : \text{BS} \rightarrow \text{RN link quality} \\
& - \nu \left(\sum_{s \in \mathcal{S}} c_s^{BS} x_s^{BS} - \sum_{r \in \mathcal{R}} c_r^{RN} x_r^{RN} \right) \quad : \text{Cost}
\end{aligned}$$

subject to ... (BW limitations, minimum QoS, etc.)



these predictions may be
unpredictable

Apply Magic Sauce

PredictionAPI

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

for research

for business

for you



Please login with Facebook to submit your Facebook Likes to our prediction engine and view the output of the API. You are asked to log in so we can collect your LikeIDs but we will not post to your wall, store your Facebook profile information or any results of this predictions.

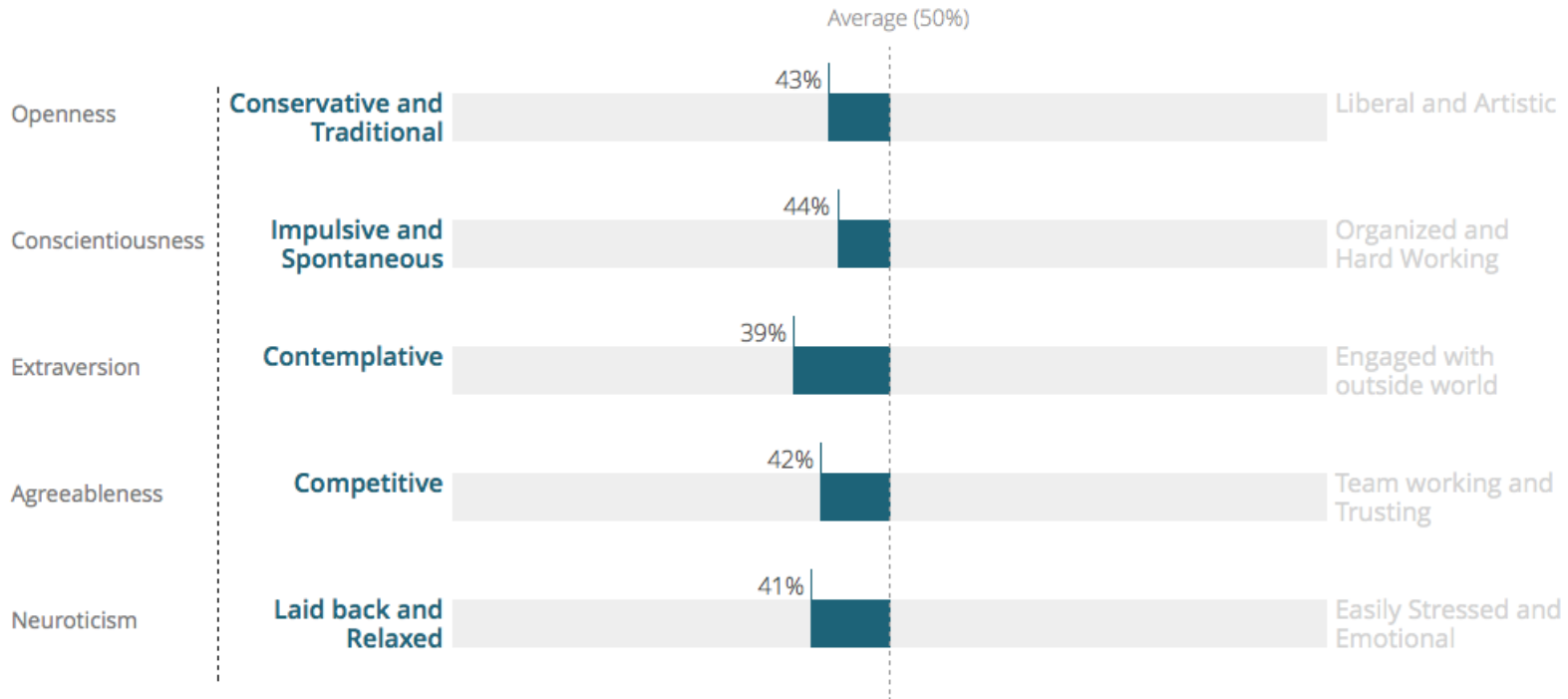


Logout from Facebook

This is a prediction of your psycho-demographic profile based on your Facebook Likes. It uses a snapshot of your digital footprint to visualise how others perceive you online and therefore may not be an entirely accurate picture of who you really are. You could [take more psychometric tests](#) as well and compare the results!

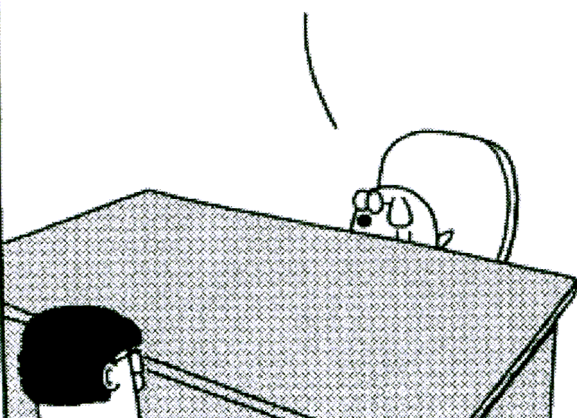
Big 5 Personality (Predictions are expressed as percentiles) ?

[Take personality test](#)



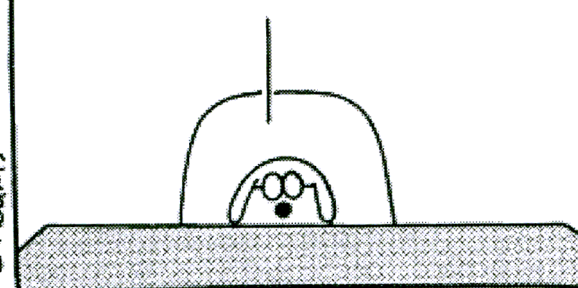
these predictions may be
part of decision-making
processes

I'VE BEEN ASKED TO
REDUCE HEADCOUNT.

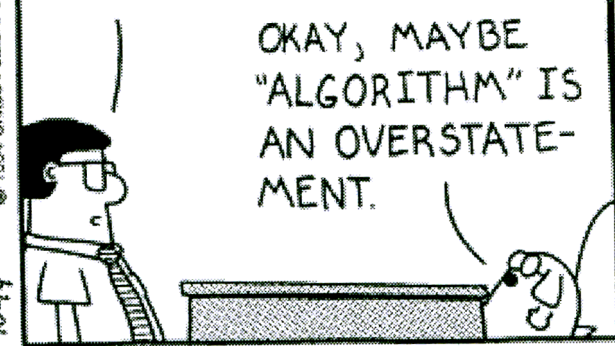


S. Adams
E-Mail: SCOTTADAMS@AOL.COM

TO BE FAIR ABOUT
IT I CREATED A
SCIENTIFIC
ALGORITHM TO
DECIDE WHO GOES.



I THOUGHT YOU WERE
FIRING THE PEOPLE
WITH THE HIGHEST
SALARIES.



© 1994 United Feature Syndicate, Inc.
10-19

they are invisible

DENIED

they might generate
discrimination

Fair Housing Act of 1968

they might be **biased**

Predictive policing



how can we control what
we don't see and don't
understand?



Data protection approach

Limits should be imposed on collection and processing of personal data;

Individuals should be entitled to make choices (informed consent);

Exceptions may include a waiver for scientific research;

Use of **anonymized data**

Data Protection draft bill (Brazilian Ministry of Justice)

Art. 11. Consent is exempt in the case of unrestricted public access data, or whenever processing is necessary to:

I – The fulfilment of a legal obligation by the controller;

II – The shared use of data pertaining to the exercise of rights or duties established by law or regulation by the public authorities;

III – The execution of pre-contractual procedures or obligations related to an contract to which the data subject is party, pursuant to section § 1 of art. 6;

IV – The performance of historical, scientific, or statistical research, ensuring, whenever possible, the dissociation of the personal data;

V – The regular exercise of rights in legal or administrative proceedings;

VI – The protection of the data subject's or a third party's life or physical safety;

VII – Health protection through a procedure performed by healthcare professional or the healthcare authorities.

do these limits restrain
science and stifle
innovation?



“We have to remember we live in a data-driven information age. And **what happens when you follow the European privacy model** and take information out of the information economy? Those are the questions that we are going to be asking because I think it is a pretty **simple answer**, and you can look at Europe and see, **revenues fall, innovation stalls, and you lose out to innovators who chose to work elsewhere.**”



Marsha Blackburn
(TN)



“That is why I am concerned about treating privacy solely as a consumer protection issue and I believe that privacy like most other issues under FTC jurisdiction must also be viewed through a competition lens.”



Maureen Ohlhausen
(FTC)



So, before we do any possible harm to the Internet, we need to understand what harm is actually being done to consumers, and where is the public outcry for legislation? Today I am simply not hearing it. I haven't gotten a single letter from anyone back home urging me to pass a privacy bill. They want data protection, but no one is beating down my door about the broader privacy issues.



Mary Bono Mack
(CA)



"Rules can force companies into innovating. [...] Data protection rules will give the EU a competitive advantage".



Sophie In 't Veld
(NL)



"Strong EU Data Protection Rules will act as a Competitive Advantage for Europe"



Paul Nemitz (DE)



"European Commission proposals over the strict handling of data will give European companies a competitive advantage in the world. Personal data is the currency of today's digital market. And like any currency, it needs stability and trust. Only if consumers can 'trust' that their data is well protected, will they continue to entrust businesses and authorities with it, buy online, and accept new services."



Viviane Reding (EU)

QUESTIONS

Is there something as the “ethics of algorithms”?

Should there be regulatory mechanisms to provide oversight of what algorithms are doing? From whom? How?

How can we ensure algorithms are not promoting discrimination? Should there be a “big data due process”?

How can we strike a balance between data protection, science, research and innovation?

OBRIGADO!

VIELEN DANK!

Dennys Antonialli

dennys@internetlab.org.br

www.internetlab.org.br