RUDE AWAKENINGS FROM BEHAVIOURIST DREAMS

METHODOLOGICAL INTEGRITY AND THE GDPR

PROF. MIREILLE HILDEBRANDT

RESEARCH PROFESSOR 'INTERFACING LAW AND TECHNOLOGY'

VRIJE UNIVERSITEIT BRUSSEL (FACULTY OF **LAW AND CRIMINOLOGY**)

RADBOUD UNIVERSITY (**SCIENCE** FACULTY)

PROF. MIREILLE HILDEBRANDT

PRINCIPLE INVESTIGATOR
ERC ADVANCED GRANT

COUNTING AS A HUMAN BEING IN THE ERA OF COMPUTATIONAL LAW

COHUBICOL
(SAY CUBICLE)

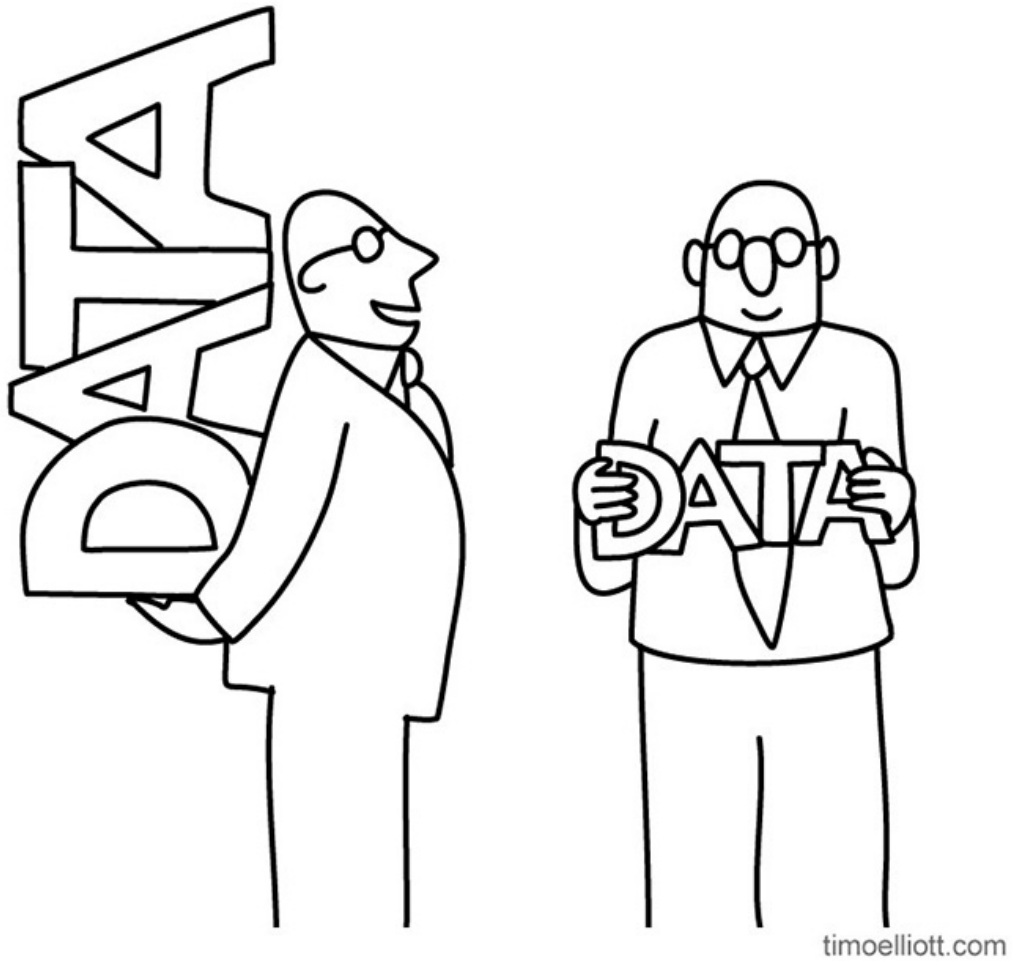
What If Google And Facebook Admitted That All This Ad Targeting Really Doesn't Work That Well?

Fri, Mar 29th 2019 10:44am — Mike Masnick

from the *the-data-obsession* dept

The little secret behind all of this that very few people want to admit is that, in **most cases** super-targeted ads are crap. They don't perform well. That's because even if you're putting the ad in front of the right demographic, most of the time they don't care or don't want to see whatever it is that you're pushing. Or, it shows an ad for something you already have (or the ever popular laugher: something you just bought and don't need to buy again).

And, if anyone should know this, it should be Google. For much of Google's existence, its big secret sauce was not deep knowledge about the people seeing the ads: it was just matching them against their search terms. That is, just a bit of simple contextual information, rather than tying it to a giant portfolio of data about you. It's really just over the last decade that Google really focused hard on building data profiles on everyone and "customizing" everything. There may be some advantages to some of those customizations -- and there are certain useful things that come with the data -- but better targeted advertisements... don't really seem to be among them.



"I think you'll find that mine is bigger..."

The case against behavioral advertising is stacking up

10:00 pm CET • January 20, 2019

But what if creepy ads don’t work as claimed? What if all the filthy lucre that’s currently being sunk into the coffers of ad tech giants — and far less visible but no less privacy-trampling data brokers — is literally being sunk, and could both be more honestly and far better spent?

MOBILE

When Procter & Gamble Cut $200 Million in Digital Ad Spend, It Increased Its Reach 10%

Unilever is also reevaluating its budget

By Lauren Johnson | March 1, 2018 PREMIUM

DIGIDAY UK

THE GDPR IMPACT

After GDPR, The New York Times cut off ad exchanges in Europe — and kept growing ad revenue

JANUARY 16, 2019 *by* Jessica Davies

POLICYBLOG

COMMENTARY

Behavioral Advertising's Benefits To Publishers Are Overstated, New Study Suggests

by **Wendy Davis** , Staff Writer **@wendyndavis**, May 30, 2019

For years, the ad industry has argued that free content online is fueled by online behavioral advertising, or tracking users across the web in order to deduce their interests and serve them with targeted ads.

The argument turns on the assumption that advertisers will pay more for targeted ads than generic ones, and that publishers will therefore garner more money from behaviorally targeted ads.

The claims -- which make some intuitive sense -- appear to have been widely accepted, even making their way into official policy documents. Last year, the Federal Trade Commission suggested in a **staff report** that publishers would be harmed by privacy rules that limited online tracking.

Online Tracking and Publishers' Revenues: An Empirical Analysis

Veronica Marotta[1], Vibhanshu Abhishek[2], and Alessandro Acquisti[3]

[1]Carlson School of Management, University of Minnesota, vmarotta@umn.edu
[2]Paul Merage School of Business, University California Irvine, vibs@uci.edu
[3]Heinz College, Carnegie Mellon University, acquisti@andrew.cmu.edu

PRELIMINARY DRAFT - MAY 2019*

Abstract

While the impact of targeted advertising on advertisers' campaign effectiveness has been vastly documented, much less is known about the value generated by online tracking and targeting technologies for publishers – the websites that sell ad spaces. In

Rude awakening?

- Advertisers moving out of RTB at YouTube?
  - Is this about RDB giving little insight to advertisers at what content their ads are displayed?
  - Or, about ineffectiveness of targeted advertising? Or both?
- Publishers moving out of behavioural advertising?
  - Is this about compliance with the GDPR? Or, about ineffectiveness?
- What about recommender systems?

What's next?

- Part I: Computer Science and the Problem of Behaviourism
- Part II: The GDPR and Methodological Integrity of ML in Recsys

Recommender systems

- Filtering is a must in era of information overload
- Search engines prove the point
- Content filtering based on behavioural profiling (locked up in the past)
- Collaborative filtering based on nearest neighbours (locked in filter bubble)
- Hybrids combined with randomiser (opening the mind, serendipity)

- Using ML for either and both assumes optimisation for a machine readable task
  - **Behavioural search, advertising & recsys = opening Pandora's box**

WHY?

Political economy of recsys

- Who is paying for recommender systems? Who is profiting from them?
  - Developers?
  - Big tech? Platforms? Publishers? Advertisers?
  - Users?

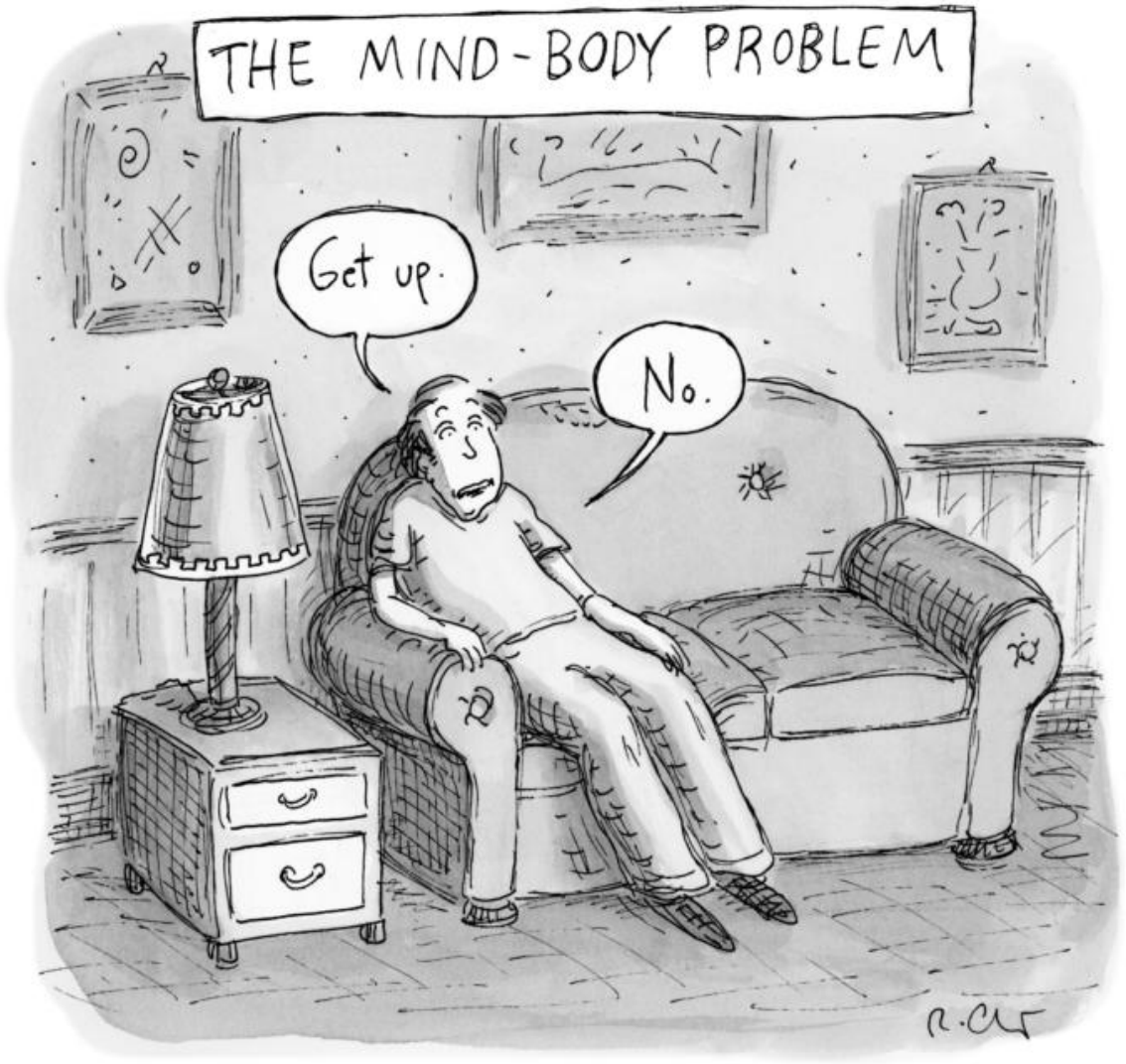
- What is the goal:
  - Mining preferences? Catering to end-users for their own sake?
  - Increasing ad revenue or sales or influencing voting behaviours?
  - **Attention grabbing, holding, hooking** → **promoting addiction**

Political economy of recsys

- Recommender systems are not just there to **'help' users** find what they don't know they want, though researchers may have this as their ultimate goal
- Providers of recommender systems operate in a 'given' **economic incentive structure**
- Economic and market forces, however, are not really 'given', so
  - What economic incentive structure is put in place? (markets are created by private & public law)
  - How does this affect the **choice architecture** of companies, developers citizens?
  - How will the GDPR change this incentive structure?

Political economy of recsys

- Mining preferences (assumes they are given, quod non)
- Inferring preferences (assumes they can be explained or predicted)
- Creating preferences (acting on inferences, creating choice architecture)
  - 'If machines define a situation as real it is real in its consequences'


THE MIND-BODY PROBLEM
Get up.
No.

Behaviourism and micro-targeting

- Data-driven recsys have trained their algorithm (learner) on behavioural data:
  - Online impression, click and conversion behaviours (RTB, AB testing)
  - Purchasing, watching, location (mobility) behaviours (LBS, Amazon, Netflix)
  - Image and voice recognition (Echo, Alexa)
  - All and any (sentiment analysis)

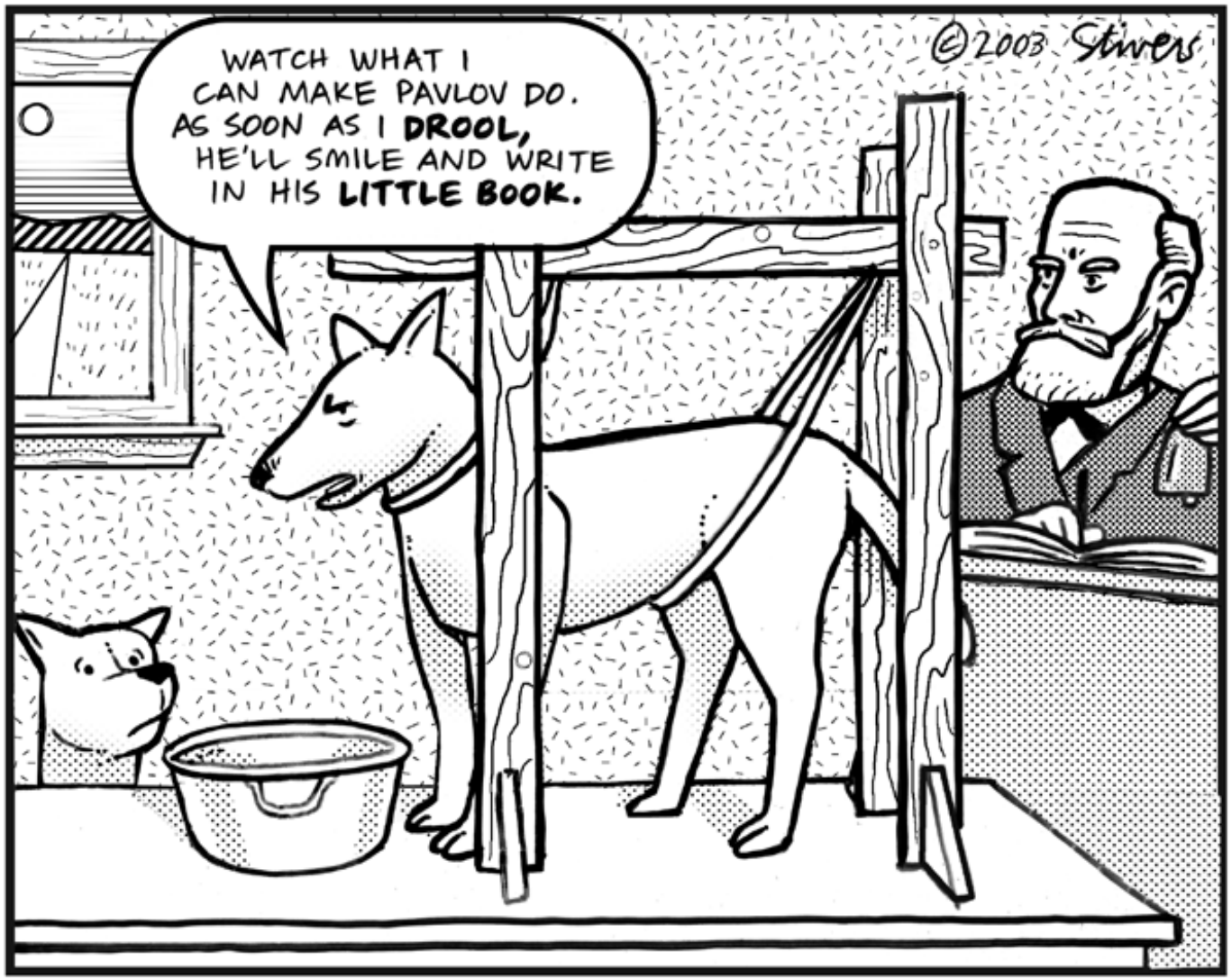
- Behavioural data =
  - Machine readable data (surf behaviour, sensor data, mobility data, etc.)
  - Data is a trace, representation, or imprint of something else
  - Based on the methodological individualism of behaviourism
  - Shares assumptions with rational choice theory and behavioural economics

Behaviourism and micro-targeting

- Behaviourism:
  - Pavlov's (stimulus-response theory of learning, acquired reflexes)
  - Watson (explicit emphasis on control and manipulation)
  - Skinner (operant conditioning, reinforcement learning)
  - Turing (testing linguistic behaviour rather than mental processes)
  - Pentland (data-driven 'social physics')
  - Helbing (computational social science)
  - Thaler & Sunstein (nudge theory)
- Anticipating machine behaviour:
  - Dennett (intentional stance based on behaviourist assumptions)
  - Rahwan et al (studying machine behaviour)

Behaviourism and micro-targeting

- Behaviourism:
  - Pavlov's (stimulus-response theory of learning, acquired reflexes)
  - **Watson (explicit emphasis on control and manipulation)**
  - Skinner (operant conditioning, reinforcement learning)
  - Turing (testing linguistic behaviour rather than mental processes)
  - Pentland (data-driven 'social physics')
  - Helbing (computational social science)
  - Thaler & Sunstein (nudge theory)
- Anticipating machine behaviour:
  - Dennett (intentional stance based on behaviourist assumptions)
  - Rahwan et al (studying machine behaviour)


WATCH WHAT I CAN MAKE PAVLOV DO. AS SOON AS I DROOL, HE'LL SMILE AND WRITE IN HIS LITTLE BOOK.
©2003 Stivers

Behaviourism and micro-targeting

- Behaviourism is based on the assumption that:
  - What can be counted matters
  - What matters can be counted

- In combination with a specific strand of cybernetics the assumption is that:
  - What can be controlled matters
  - What matters can be controlled

Behaviourism and micro-targeting

- Both assumptions cannot be proven (nor disproven)
- Best is to remain agnostic but to take them into account

Behaviourism and micro-targeting

- What can be proven mathematically is:
  - Godel's theorem
  - Wolpert's NFL theorem

- Fundamental **incomputability** (undecidability)
- Next to a prevailing **intractability** (related to computing power & complexity)

NOISE TO SIGNAL RobCottingham.com @robcottingham

Campbell, Goodhart effects Lucas critique

- **Campbell effect:** if you want to influence the future based on specified measurements, the measurement will become the target and corrupt the measurement
  - **The perverse effects of clickbait**
- **Goodhart effect:** 'when a measure becomes a target, is ceases to be a good measure'
  - **Gaming the system**
- **Lucas critique:** policy advice based on data-driven inferences is not valid, unless based on micro-foundations
  - **Correlation, causality, theory**

MacNamara Fallacy:

- The first step is to measure whatever can be easily measured.
  This is OK as far as it goes.
- The second step is to disregard that which can't be easily measured or to give it an arbitrary quantitative value. This is artificial and misleading.
- The third step is to presume that what can't be measured easily really isn't important.
  This is blindness.
- The fourth step is to say that what can't be easily measured really doesn't exist.
  This is suicide.


COUNTING AS A HUMAN BEING
IN THE ERA OF COMPUTATIONAL LAW
COHUBICOL
SAY CUBICLE • THINK WITTGENSTEIN'S CUBE
NEWS
ON THE PROJECT
RESEARCH BLOG
COMPUTATIONAL LAW
LEGAL PROTECTION
PRESS
RESEARCH OUTCOME
INNOVATION OF LEGAL METHOD
'It would be nice if all of the data which sociologists require could be enumerated because then we could run them through IBM machines and draw charts as the economists do. However, not everything that can be counted counts, and not everything that counts can be counted'.

the human condition

- **Plessner's constitutional laws:**
  1. **Artificial by nature** (we create a shared world)
  2. **Mediated immediacy** (access to reality is mediated by language and meaning)
  3. **Utopian position** (ex-centric positionality)

the human condition

- **G.H. Mead's** mind, self and society (I, me, generalised other)
- **Paul Ricoeur's** 'oneself as another' (constitution of self)
- Talcott **Parsons** and Niklas **Luhmann**'s double contingency (how to stabilise meaning)

the human condition

- The use of behavioural data results in **fundamentally unreliable results**
  - This cannot be 'fixed' by further computation
  - The implications can be addressed by
    - Awareness
    - Targeting falsification instead of verification
      - Testability (is not equivalent with reproducibility)
      - Contestability (open science, open society)

What's next?

- Part I: Computer Science and the Problem of Behaviourism
- Part II: The GDPR and Methodological Integrity of ML in Recsys


New Pub
Manage
MH
Law for Computer Scientists
Mireille Hildebrandt

GDPR Primer

- **Personal data processing:**
  - Any data relating to an identifiable person (very broad concept)
  - Data subject (natural person)
  - Data processing (from collection to storage to whatever)
  - Behavioural data that can single out = personal data

  - Goal of the GDPR is to enable processing within the Union
    - Based on equivalent protection throughout the Union
    - Mix of public enforcement and private law liability

GDPR Primer

- Whoever determines the purpose and the means of processing
  - Is liable
  - Called ‘the controller’

  - Publisher? Yes, e.g. if using analytics
  - Advertiser? Yes, e.g. if targeting
  - Webshop? Yes, e.g. if using like buttons
  - Google (analytics) yes (jointly with the website)
  - Facebook (like buttons) yes (as well as the website)

GDPR Primer

- Any processing (also of publicly available data) requires a legal ground
  1. Unambiguous informed consent for a specified purpose
  2. Contract
  3. Vital interest data subject
  4. Legal obligation
  5. Public task or public authority
  6. Legitimate interest of the controller (balancing test)

GDPR Primer

- Any processing (also of publicly available data) requires a legal ground
  1. Unambiguous informed consent for a specified purpose
  2. Contract
  3. Vital interest data subject
  4. Legal obligation
  5. Public task or public authority
  6. Legitimate interest of the controller (balancing test)

GDPR Primer

- Any processing (also of publicly available data) requires a legal ground
  - Processing must be **necessary** in relation to each ground
  - Not just appropriate or 'not excessive'

  - If not effective, processing can't be necessary!
  - This is now assumed, but that might change

GDPR Primer

- **Any processing must comply with the principles:**
  - Purpose limitation (further processing must not be incompatible)
  - Accuracy (relevance, completeness)
  - Transparency (a range of information obligations)
  - Accountability (for data controller)
  - Storage limitation (only allowed if necessary for purpose)
  - Fairness (proportionality, balancing test)
  - Lawfulness (rule of law requirements in case of eg privacy infringements)

GDPR Primer

- **Methodological integrity would benefit from compliance:**
  - Purpose limitation (further processing must not be incompatible)
  - Accuracy (relevance, completeness)
  - Transparency (a range of information obligations)
  - Accountability (for data controller)
  - Storage limitation (research exception)
  - Fairness (check for bias)
  - Lawfulness (code is not above the law)

GDPR Primer

- GDPR Consent:
  - Heavy requirements (informed, non-ambiguous)
  - As easy to withdraw as to give
  - No forcing of hand
- ePrivacy Directive:
  - Consent for cookies if not necessary for technical or functional reasons
  - Must be necessary for purpose

GDPR Primer

- **Special category data: default prohibition, limited set of exceptions**
  - Enhanced consent requirements
  - Health data is a broad concept
  - Exception for research (public interest)

GDPR Primer

- Core GDPR principle = PURPOSE
  - It determines who is responsible
  - It determines lawfulness of processing
  - Closely connected with
    - Legality in public law
    - Autonomy and trust in private law

Methodological integrity in science and society

GDPR:

- Purpose of processing of personal data by recommender systems:
  - Providing a specified type of recommendations (based on consent)
  - Serving a specific business model (based on legitimate interest provider)

Methodological integrity in science and society

ML:

- **Purpose of processing of personal data by recommender systems:**
  - What machine readable task has been used to train the system?
  - How does the system optimise for that task?
  - How effective is the system (in the real world)?

Default prohibition of micro-targeting

***Article 22* Automated individual decision-making, including profiling**

1. The data subject shall have the right not to be subject to:

- a decision
- based solely on automated processing, including profiling,
- which produces legal effects concerning him or her **or**
- similarly significantly affects him or her.

Default prohibition of micro-targeting

Recital 71

The data subject should have the right not to be subject to a **decision**, which may include

- a measure, evaluating personal aspects relating to him or her (…), such as
  - automatic refusal of an online credit application or
  - e-recruiting practices without any human intervention.

Default prohibition of micro-targeting

Recital 71

Such **processing** includes 'profiling'

- that consists of any form of automated processing of personal data evaluating the personal aspects relating to a natural person,
- in particular to analyse or predict aspects concerning the data subject's performance at work, economic situation, health, personal preferences or interests, reliability or behaviour, location or movements,
- where it produces legal effects concerning him or her or similarly significantly affects him or her.

Default prohibition of micro-targeting

GDPR Q: do recommendation decisions quality as art. 22 decisions?

"The European Data Protection Board considered that **online targeted advertisement** could have in some circumstances the capability to **sufficiently significantly affect the individuals** when, for instance, it is **intrusive or uses knowledge of vulnerabilities of the individuals**.

Given the significance of the exercise of the democratic right to vote, personalised messages which have for instance the possible effect **to stop individuals from voting** or **to make them vote in a specific way** could have the potential of meeting the criterion of significant effect."

Brussels, 12.9.2018 COM(2018) 638 final
Free and Fair elections GUIDANCE DOCUMENT

Default prohibition of micro-targeting

***Article 22* Automated individual decision-making, including profiling**

2. Paragraph 1 shall not apply if the decision:

a. is necessary for entering into, or performance of, a **contract** between the data subject and a data controller;

b. is authorised by **Union or Member State law** to which the controller is subject and which also lays down suitable measures to safeguard the data subject's rights and freedoms and legitimate interests; or

c. is based on the data subject's **explicit consent**.

Default prohibition of micro-targeting

Recital 71

However, decision-making based on such processing, including profiling, should be allowed where expressly **authorised by Union or Member State law** to which the controller is subject, including for

- fraud and tax-evasion monitoring and prevention purposes
- conducted in accordance with the regulations, standards and recommendations of Union institutions or national oversight bodies and (…)

Default prohibition of micro-targeting

Recital 71

In any case, such processing should be subject to **suitable safeguards**,

- which should include specific information to the data subject and
- the right to obtain human intervention,
- to express his or her point of view,
- **to obtain an explanation of the decision reached after such assessment** and
- **to challenge the decision**.

(…).

Default prohibition of micro-targeting

***Article 22* Automated individual decision-making, including profiling**

3. In the cases referred to in points (a) and (c) of paragraph 2, the data controller shall

- implement **suitable measures to safeguard** the data subject's rights and freedoms and legitimate interests,
- at least the **right to obtain human intervention** on the part of the controller,
- to express his or her point of view and to **contest** the decision.

Default prohibition of micro-targeting

***Article 22* Automated individual decision-making, including profiling**

- to express his or her point of view and to **contest** the decision.

Recital 71

In any case, such processing should be subject to suitable safeguards,

- to obtain an **explanation** of the decision reached after such assessment and
- to **challenge** the decision

Default prohibition of micro-targeting

Art. 13.2.f, 14.2.g, 15.1.h: Information obligations and rights concerning:

- the **existence of automated decision-making**, including profiling, referred to in Article 22(1) and (4) and, at least in those cases,
- **meaningful information about the logic involved**,
- as well as the **significance** and the **envisaged consequences** of such processing for the data subject.

Default prohibition of micro-targeting

Note that an explanation is not a justification:

- *Meaningful explanation why the system took the decision is not enough*
- *Legal justification:*
  - *For public administration: legality principle (22.2.b)*
  - *Private enterprise: freedom to conduct a business (NOT unlimited)*

Default prohibition of micro-targeting

***Article 22* Automated individual decision-making, including profiling**

4. Decisions referred to in paragraph 2 shall **not be based on special categories** of personal data referred to in Article 9(1), unless

- point (a) or (g) of Article 9(2) applies **and**
- **suitable measures to safeguard** the data subject's rights and freedoms and legitimate interests are in place.

Default prohibition of micro-targeting

Recital 71

In order to ensure **fair and transparent processing** in respect of the data subject, taking into account the specific circumstances and context in which the personal data are processed,

- the controller should use **appropriate mathematical or statistical procedures** for the profiling, implement technical and organisational measures appropriate to ensure, in particular,
  - that factors which result in inaccuracies in personal data are corrected and
  - the risk of errors is minimised,
- secure personal data in a manner that takes account of the **potential risks** involved for the interests and rights of the data subject and
- that prevents, inter alia, **discriminatory effects on natural persons** on the basis of racial or ethnic origin, political opinion, religion or beliefs, trade union membership, genetic or health status or sexual orientation, or **that result in measures having such an effect.**

Methodological integrity in science and society

GDPR:

- **Automated decisions:**
  - That target users, reconfiguring their choice architecture, making a difference
  - Contract or explicit consent
  - Information: existence, underlying logic, consequences
  - Human intervention, explanation, contestation

Methodological integrity in science and society

ML:

- **Micro targeting by recommendation systems:**
  - How does filtering, nudging affect the data subject?
  - Contract with the data subject (recom. at request)? Explicit consent?
  - Information about automation, logic, consequences?
    - Can data subject play around with counterfactuals?
  - Human intervention, explanation, contestation?
    - Is there is human customer service that can explain, reset of turn-off the system?

Transparency in ML

Hofman, Sharma, Watts (2017):

- **In exploratory analyses,**
  - researchers are free to study different tasks, fit multiple models, try various exclusion rules, and test on multiple performancemetrics.
  - When reporting their findings, however, they should transparently declare
    - their full sequence of design choices to **avoid creating a false impression** of having confirmed a hypothesis rather than simply having generated one.
  - Relatedly, they should report performance in terms of **multiple metrics**
    - to avoid creating a false appearance of accuracy.
  - In cases where data are abundant, moreover, researchers can increase the validity of exploratory research by
    - using a three-way split of their data into a **training set** used to fit models, a **validation set** used to select any free parameters that control model capacity and to compare different models, and a **test set** that is used only once to quote final performance.

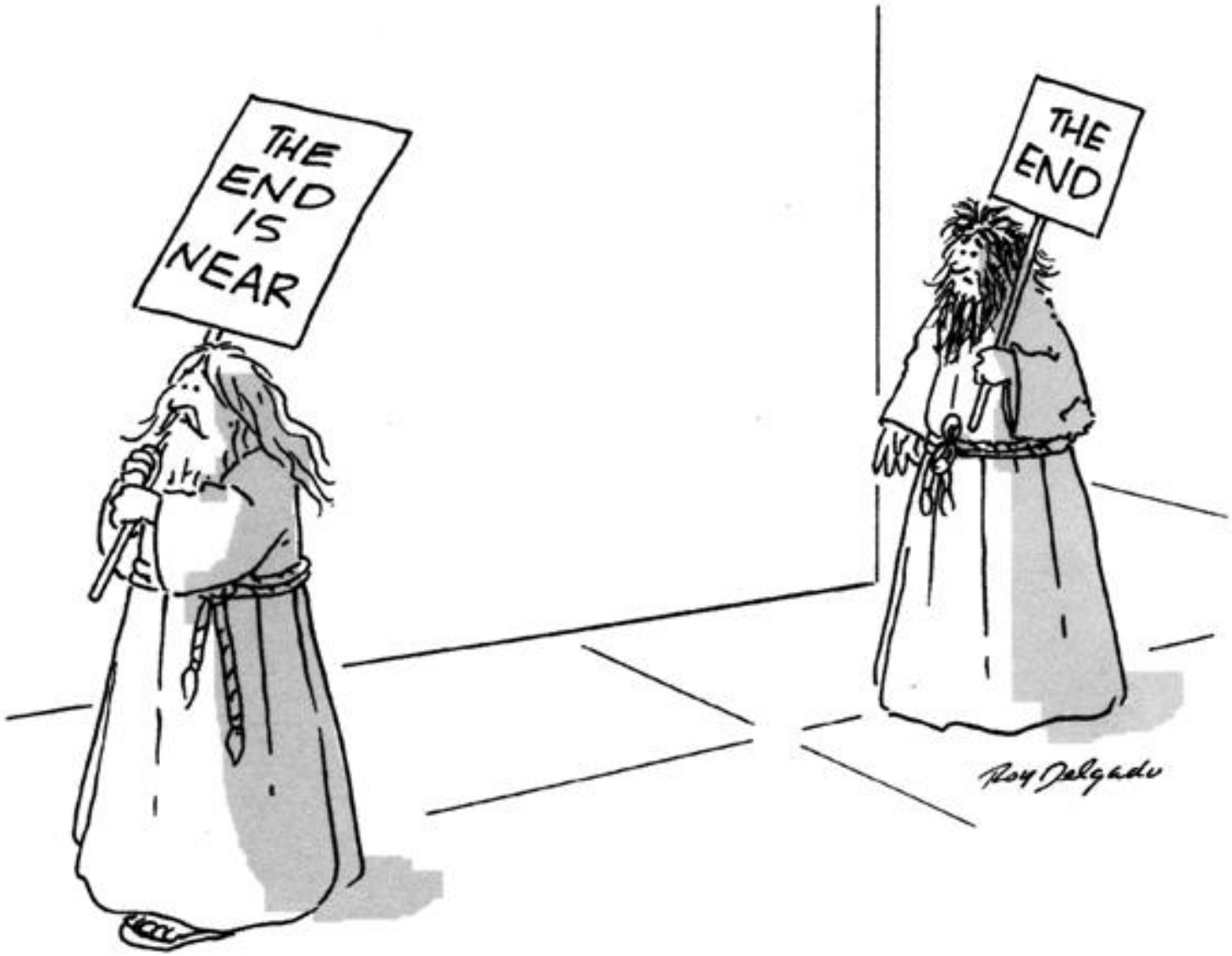
Transparency in ML

Hofman, Sharma, Watts (2017):

- To qualify research as **confirmatory**, however,
  - researchers should be required to **preregister their research designs**,
  - including data preprocessing choices, model specifications, evaluation metrics, and out-of-sample predictions,
  - in a public forum such as the **Open Science Framework (https://osf.io).**
  - Although strict adherence to these guidelines may not always be possible,
    - following them would **dramatically improve the reliability and robustness** of results,
    - as well as **facilitating comparisons across studies.**

GDPR changing the political economy of recsys

- **The GDPR changes the incentive structure for:**
  - Those who provide recsys (controllers)
  - In turn changing the incentives for developers:
    - More transparency (about purpose, about consequences)
    - Taking those whose data are processed seriously
    - Taking those who are targeted seriously
    - This will create incentives for better methodology, more integrity, better science


THE END IS NEAR
THE END
Roy Delgado