
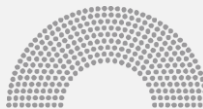



Focusing the microscope: how we can use data to understand behaviour

Joana Gonçalves de Sá
September 17th 2020



	QUESTIONS	DATA	TOOLS
HEALTH	 <p>Online vs. Offline Patterns Emergency Now-casting Antibiotic Over-prescription</p>	<p>Google Trends SNS24 Twitter ER acceptance /times SPMS e-prescriptions</p>	<p>Math Modelling ML Epidemiology</p>
POLICY	 <p>Political Decisions Gender Differences Agenda Setting Voting vs. Discourse</p>	<p>Media records Twitter Facebook Parliament data</p>	<p>NLP Networks Math Modelling Complex Systems</p>
BEHAVIOUR	 <p>Cognitive Biases Attitudes Towards Science Tracking Anxiety</p>	<p>Large scale surveys Behavioral experiments Twitter Facebook</p>	<p>Networks Math Modelling Psychology Information</p>

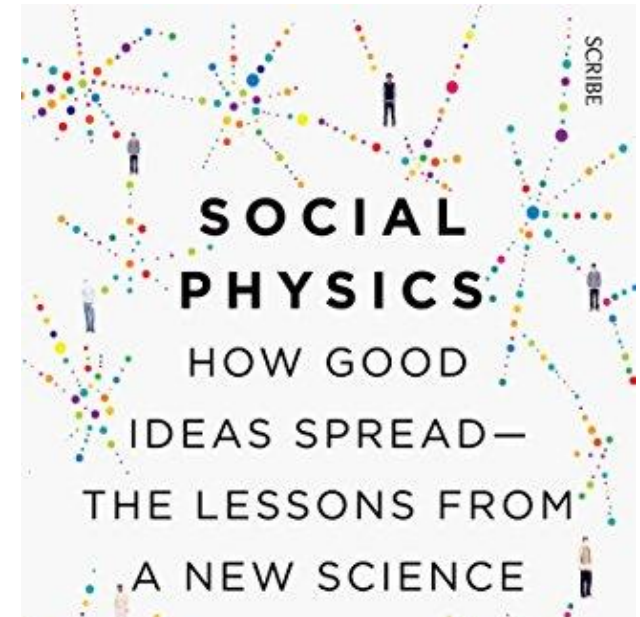


"WHO STOLE THE PEOPLE'S MONEY?" — DO TELL .N.Y.TIMES.

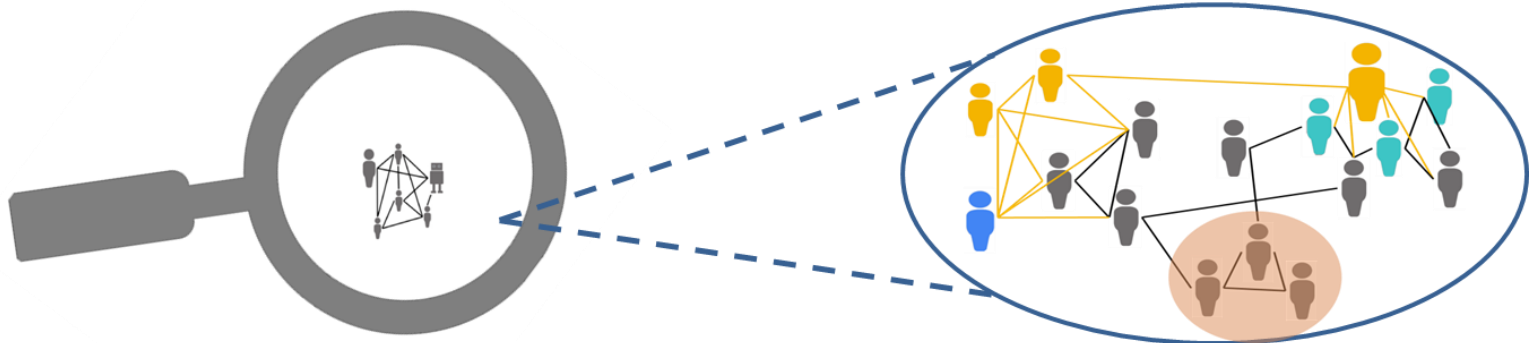
These problems—and a wide range of similar problems in the biological, medical, psychological, economic, and political sciences—are just too complicated to yield to the old nineteenth-century techniques which were so dramatically successful on two-, three-, or four-variable problems of simplicity. These new problems, moreover, cannot be handled with the statistical techniques so effective in describing average behavior in problems of disorganized complexity.

These new problems, and the future of the world depends on many of them, requires science to make a third great advance, an advance that must be even greater than the nineteenth-century conquest of problems of simplicity or the twentieth-century victory over problems of disorganized complexity. Science must, over the next 50 years, learn to deal with these problems of organized complexity.

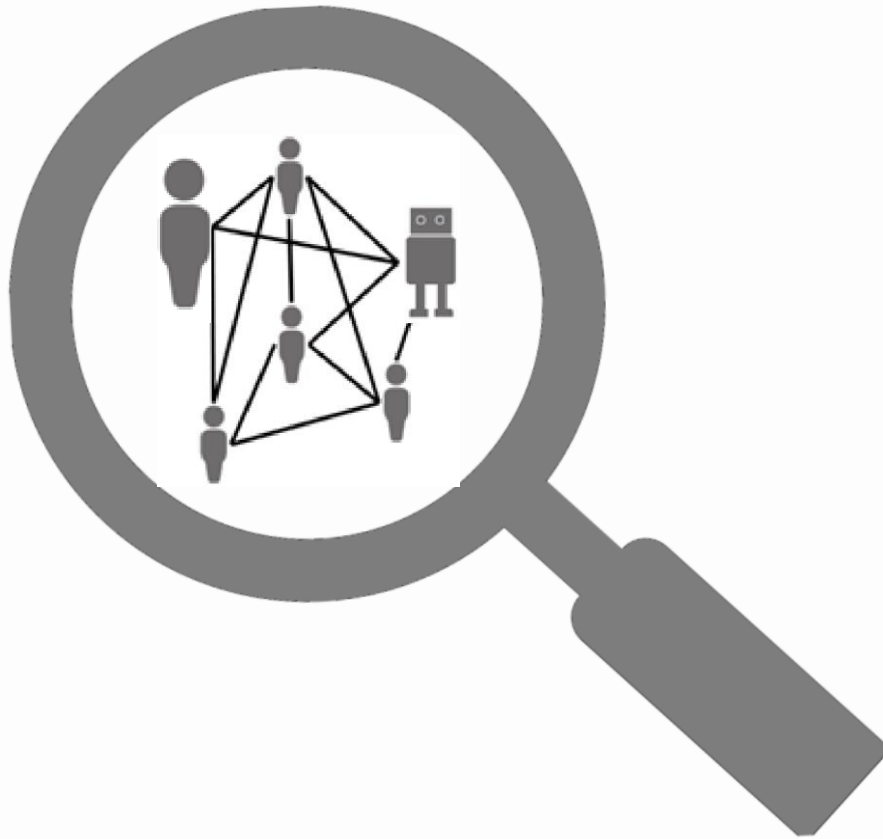
Warren Weaver, 1947



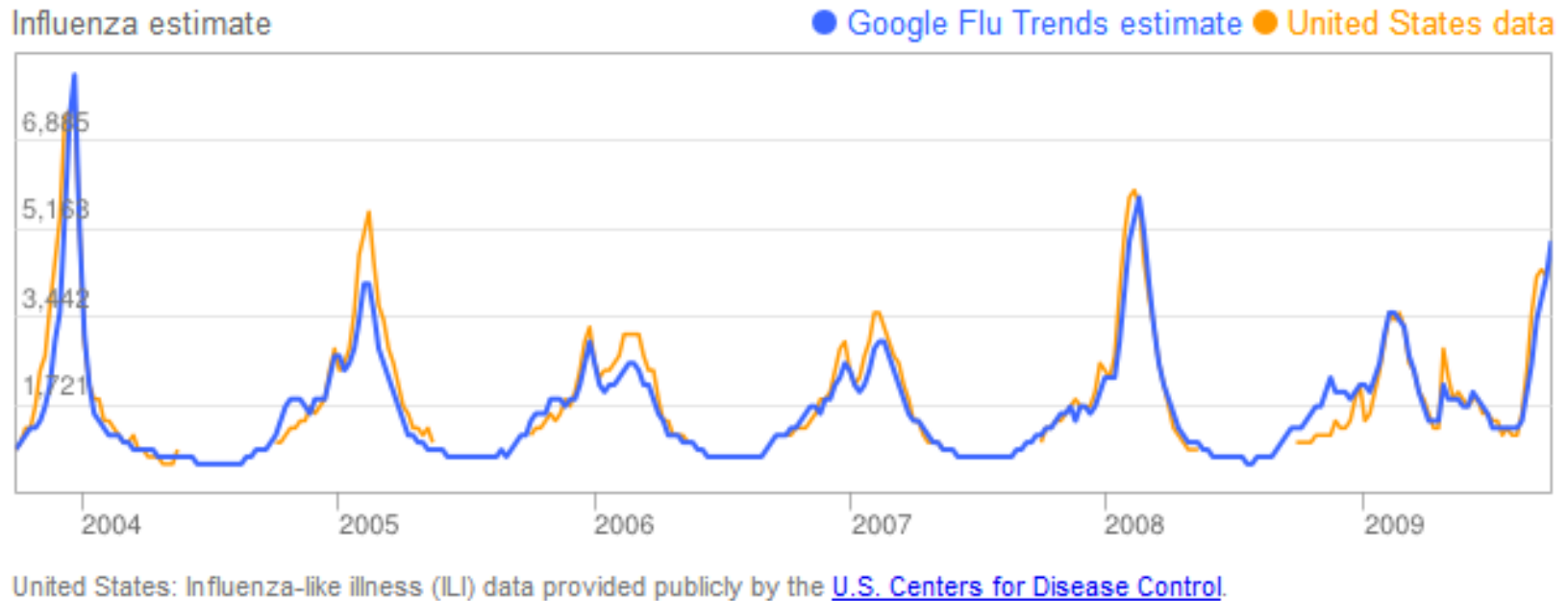
Alex Pentland, 2014



MACROSCOPE



- Individual behaviour
- Large scale
- In context
- Fast / real-time
- Not self-reported

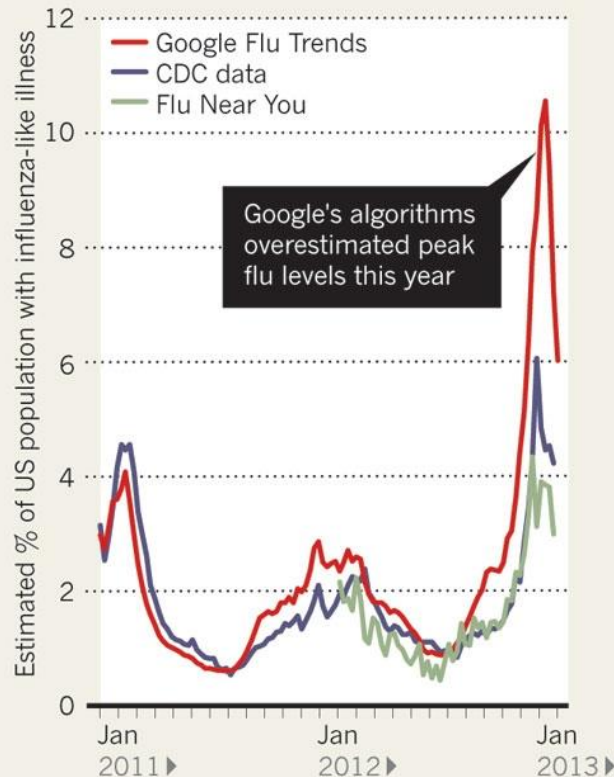


Ginsberg, Jeremy, et al. "Detecting influenza epidemics using search engine query data"
Nature 457.7232 (2009)

Google Flu Trends (never available in Portugal)

FEVER PEAKS

A comparison of three different methods of measuring the proportion of the US population with an influenza-like illness.



The New York Times

How Data Failed Us in Calling an Election

By Steve Lohr and Natasha Singer, 2016

BIG DATA
HUBRIS

Declan Butler, "When Google got it wrong"
Nature 494, 155–156 (2013)

MACROSCOPE - MAGNIFY



- Individual behaviour
- Large scale
- In context
- Fast / real-time
- Not self-reported
- Human Biases



- Profiling / Targeting
- Monetize / Incentives
- Proxies / Large impact
- Faster / not adapted
- Illusion of knowledge
- Amplification / Manipulation

FAKE NEWS: fabricated information, with the intent to mislead

FALSE NEWS: information that was not deliberately created, often being lingering misconceptions, stemming from poor reporting, misinterpretations, or even satirical pieces



FAKE NEWS

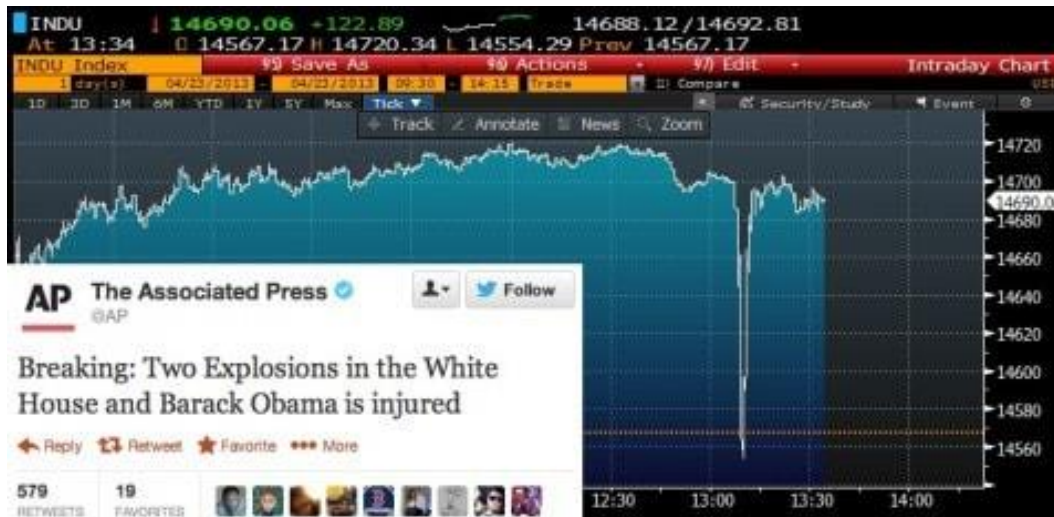
Fact-checking sites to decide

1. NOT NEW



2. SERIOUS CONSEQUENCES – PERFECT STORM

BBC NEWS India lynchings: WhatsApp sets new rules after mob killings 🕒 20 July 2018



- Fast
- Long- reaching
- Magnify
- Create new incentives
- Monetize biases
- See / alter behaviour

3. CURRENT APPROACHES



Tackling online disinformation

- MULTIDISCIPLINARITY
- TECHNOLOGY
- PLATFORM REGULATION
- EDUCATION

HUMAN BEHAVIOUR



“Falsehood diffused significantly farther, faster, deeper, and more broadly than the truth (...)

Robots accelerated the spread of true and false news at the same rate, implying that false news spreads more than the truth because **humans, not robots, are more likely to spread it.”**

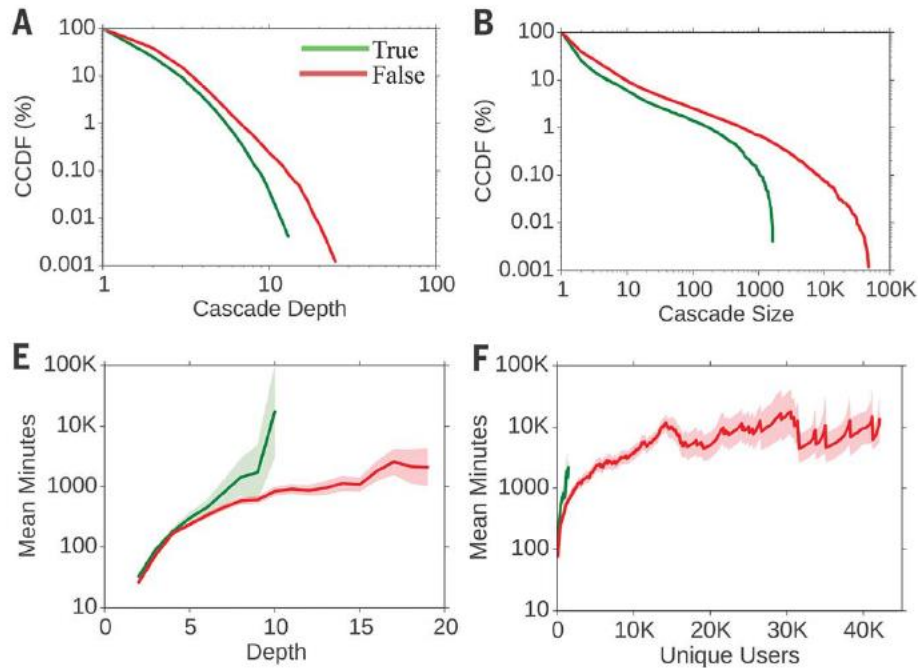


Fig. 2. Complementary cumulative distribution functions (CCDFs) of true and false rumor cascades. (A) Depth. (B) Size. (C) Maximum breadth. (D) Structural virality. (E and F) The number of minutes it takes for true and false rumor cascades to reach any (E) depth and (F) number of unique Twitter users. (G) The number of unique Twitter

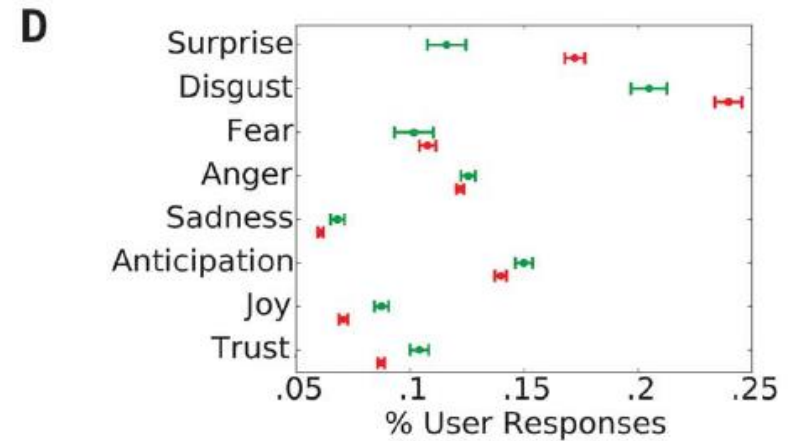
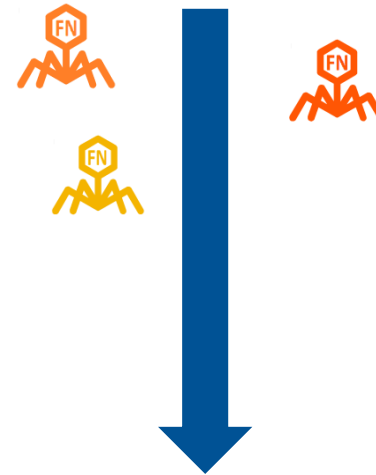
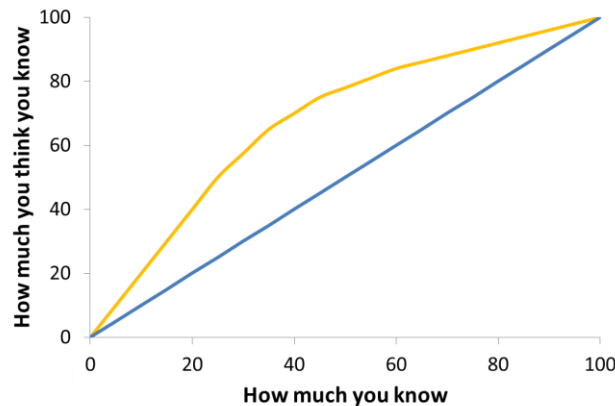
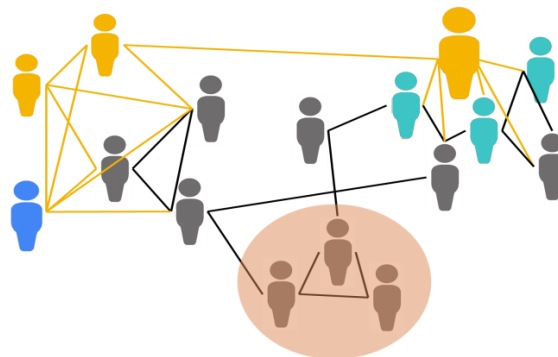


Fig. 4. (D) The emotional content of replies to true (green) and false (red) rumor tweets across seven dimensions categorized by the NRC.

Spread of Fake News is (implies) cognitive biases



We can use them as a model system to study human behavior in context



Humans have a preference for sharing fake news

- | | |
|---|---|
| 1. We cannot fully tackle the problem unless we understand the human side | 1. The social networks amplified the problem |
| 2. The problem will allow us to understand the human side | 2. The social networks will offer a way to study it |



FAKE NEWS: AN ELEGANT RESOURCE FOR SOCIAL SCIENCES



Fake News



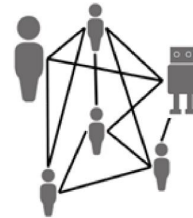
Pathogens



Humans



Hosts



Networks



Environment



DATASET1

FAKE AND REAL NEWS

RECOGNIZED FACT-CHECKING
SITES



DATASET2 INDIVIDUAL SUSCEPTIBILITIES



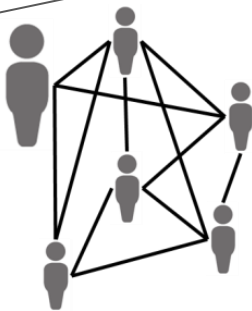
PSYCHOLOGY OF HUMAN BEHAVIOUR
COGNITIVE BIASES
VALIDATED TESTS

EXTENSIVE SURVEY

DISEASE DYNAMICS MODELS

DATASET3

ENVIRONMENT AND HISTORY



TWITTER

CONNECTIONS
PROFILES
SHARING HISTORY

1. Cognitive biases & confidence to knowledge ratios, good predictors of FN sharing
2. Position on networks should be good predictors of FN sharing
3. Past history should offer good indicators of future FN spreading

FAKE AND REAL NEWS



4 – FN that spread **slower** might activate more **complex** cognitive processes

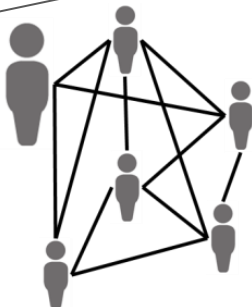


INDIVIDUAL SUSCEPTIBILITIES

1 – **Overconfident (+ unknowledgeable)** individuals more likely to share FN

DISEASE DYNAMICS MODELS

**ANONYMOUS, DISTRIBUTED
ENCRYPTED**



ENVIRONMENT AND HISTORY

- 2** – **Less diverse social networks** more likely to share FN
- 3** – Sharing in the **past** should lead to sharing in the **future**



Fake News



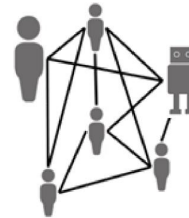
Pathogens



Humans



Hosts



Networks



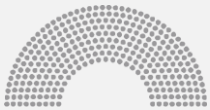
Environment





40 years of PTPARL data

POLICY

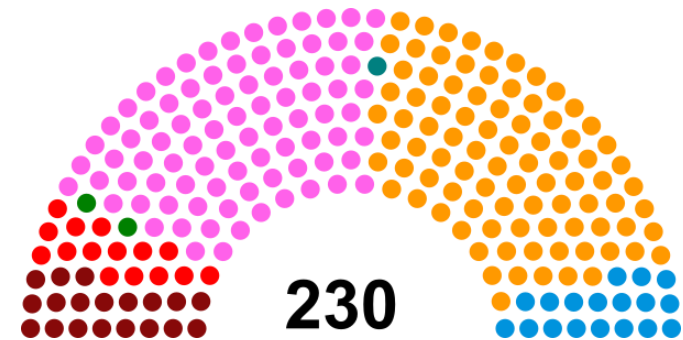


Political Decisions
Gender Differences
Agenda Setting
Voting vs. Discourse

Media records
Twitter
Facebook
Parliament data

NLP
Networks
Math Modelling
Complex Systems

- Text documents
 - Debates
- HTML documents / Parliament Open Data
 - MP biographical data
 - **Initiatives (votes)**
- CSV files
 - MP data (first five legislatures)



Manuel Pita @ U Lusófona
Nuno Mamede@ IST



40 years of PTPARL data

- Strip HTML
- Clean headers
 - Get first line
 - Fuzzy matching / Regular expressions
 - Check logs, adjust, repeat
- Detect session end
- Tag utterances
 - Iterate instances of 'Speaker : - Utterance'
 - Add <utterance> tags
 - Track utterances spanning multiple pages
- Identify President utterances
- Assign Orador
 - Only up to 10th legislature
 - For each Orador, go back to when president granted the floor and score putative Orador
- Assign speaker
 - Fuzzy matching against database
 - Skip president utterances

I SÉRIE — NÚMERO 65

partido, re'erindo-se também ao PRD e tentando pô-lo a ridículo, que nem a noção do ridículo lhe fazemos, Sr. Ministro de Estado, nem usamos da sua linguagem trauliteira.

Vozes do PSD: — Não apoiado!

O Orador: — Compreendemos que a vossa queda eminente vos coloca perante esse desespero, mas espanta-nos a sua má-criação, Sr. Ministro.

Vozes do PSD: — Não apoiado!

O Orador: — Queremos apenas que isso fique registado e mais nada.

Aplausos do MDP/CDE, do PS, do PRD e do PCP.

O Sr. Presidente: — O Sr. Ministro de Estado pretende dar explicação?

Vozes do PSD: — Não lhes dê confiança!

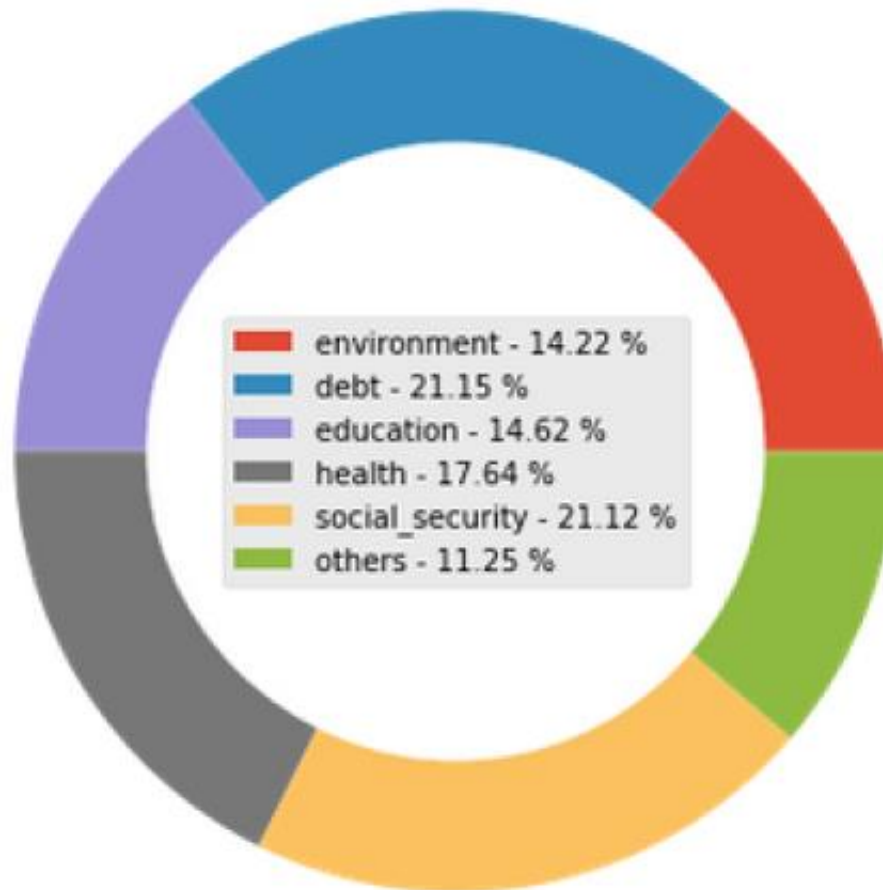
O Sr. Ministro de Estado: — Não, Sr. Presidente.

O Sr. Presidente: — O Sr. Deputado Magalhães Mota pediu a palavra para defesa da honra da sua bancada?

O Sr. Magalhães Mota (PRD): — Não, Sr. Presidente. Não vou invocar a figura regimental de defesa

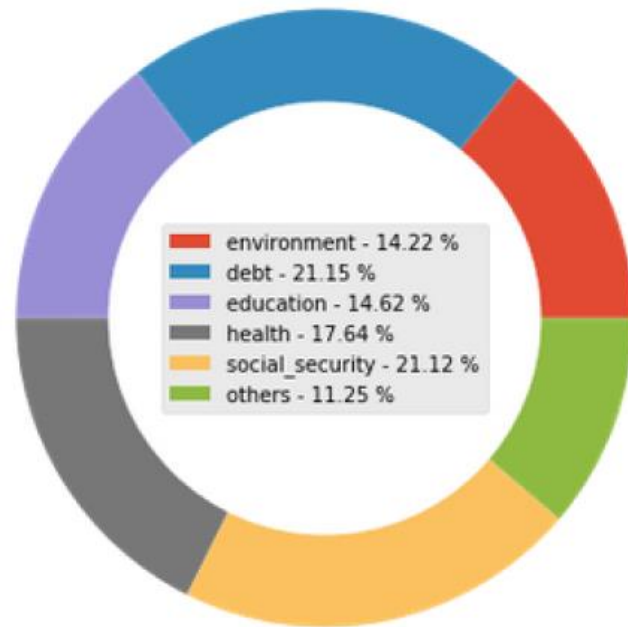


Using Intergenerational Justice (IJ) as a case-study

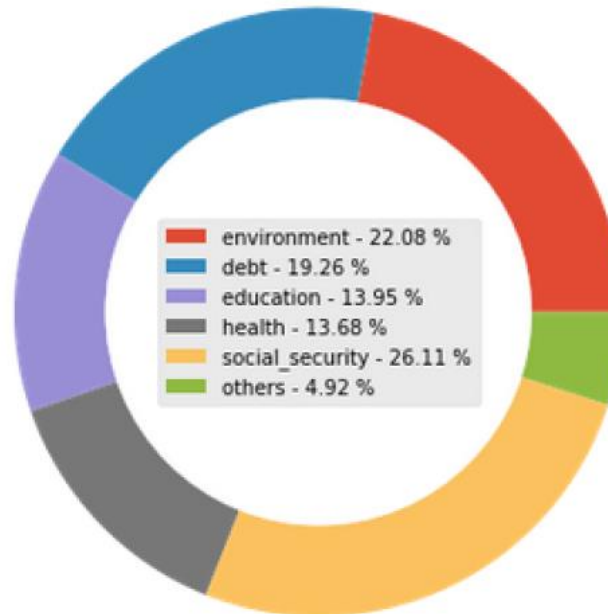




Using Intergenerational Justice (IJ) as a case-study

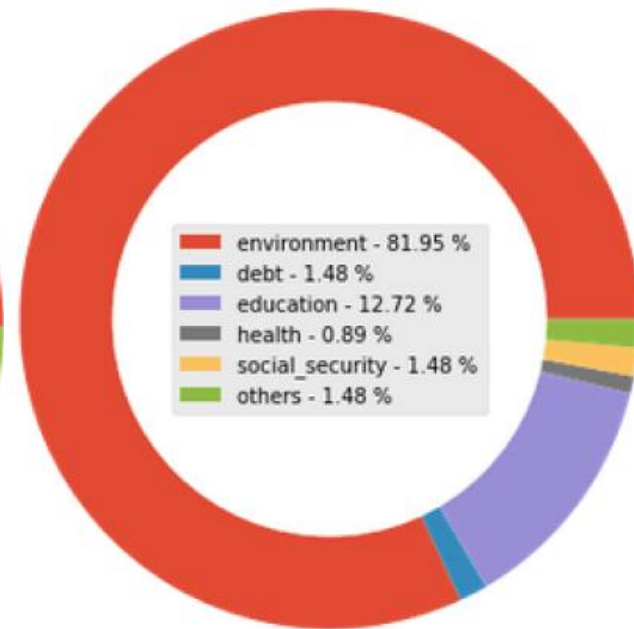


**Parliamentary
debates**



Traditional Media

2008 - date

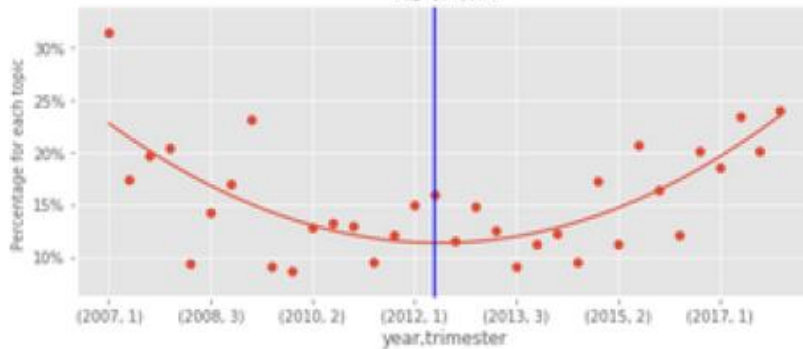


Twitter

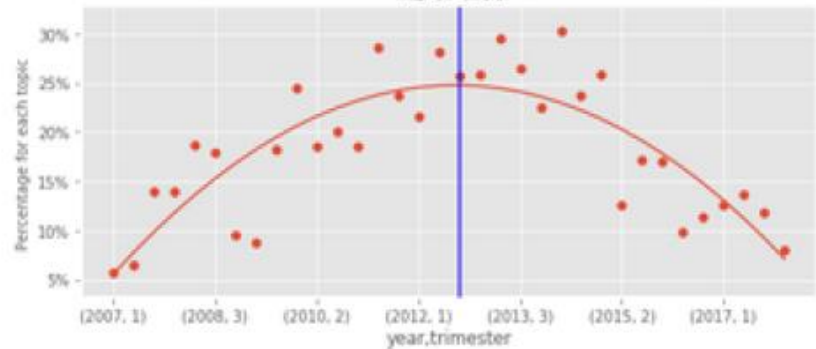


Parliament

ENVIRONMENT
With minimum at(2012, 2)
R2 of 0.47

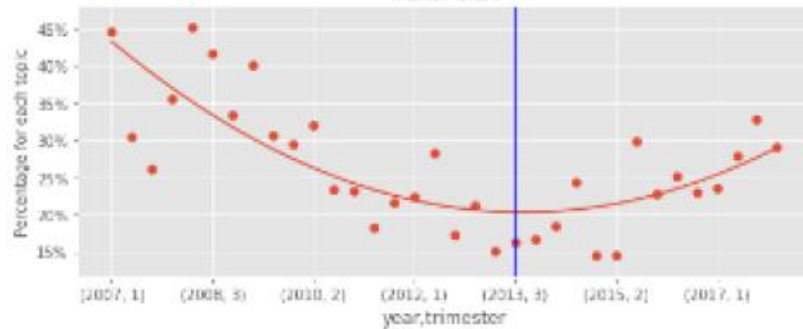


DEBT
With minimum at(2012, 3)
R2 of 0.66

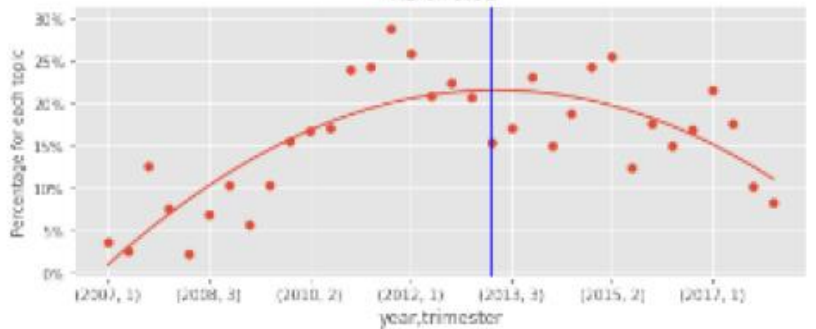


Traditional Media

ENVIRONMENT
With minimum at(2013, 3)
R2 of 0.56



DEBT
With minimum at(2013, 2)
R2 of 0.61





Fake News



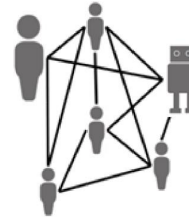
Pathogens



Humans



Hosts



Networks



Environment



Political Decisions
Gender Differences
Agenda Setting
Voting vs. Discourse

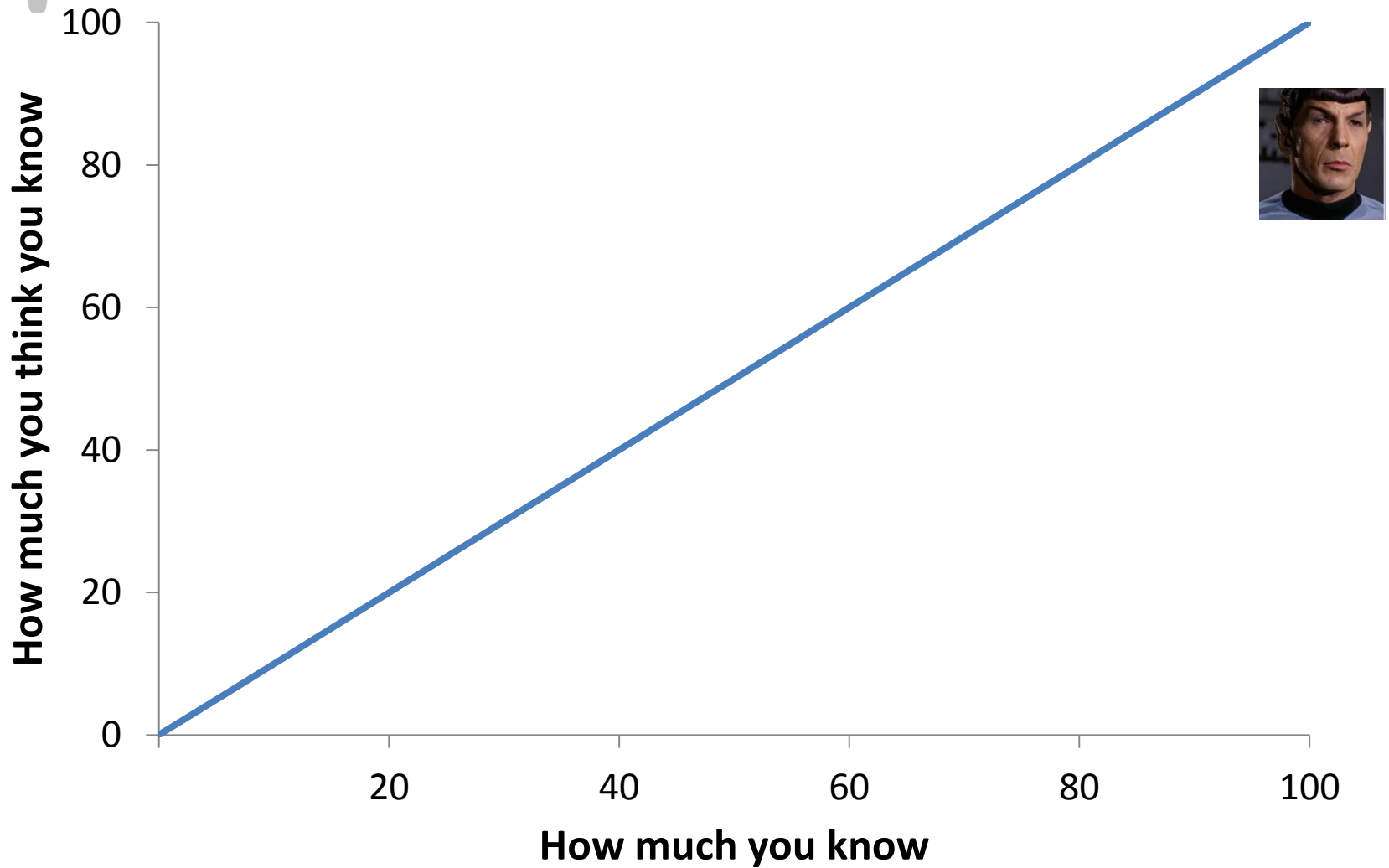
Media records
Twitter
Facebook
Parliament data

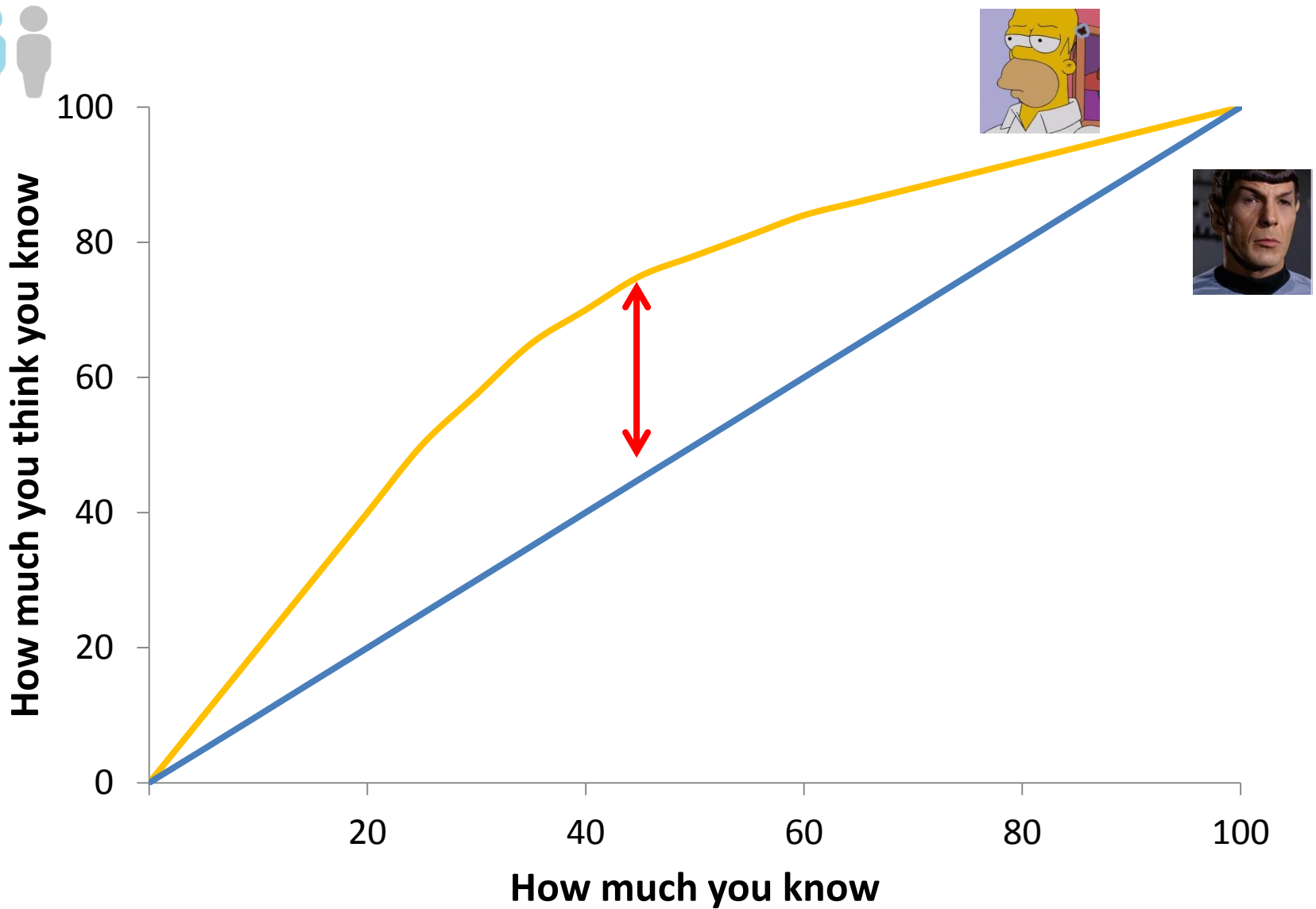
NLP
Networks
Math Modelling
Complex Systems



Fake news spreading as a deviation from “rationality”

1. Over confident
2. Confirmatory Tendencies
3. Echo chambers
4. Environment

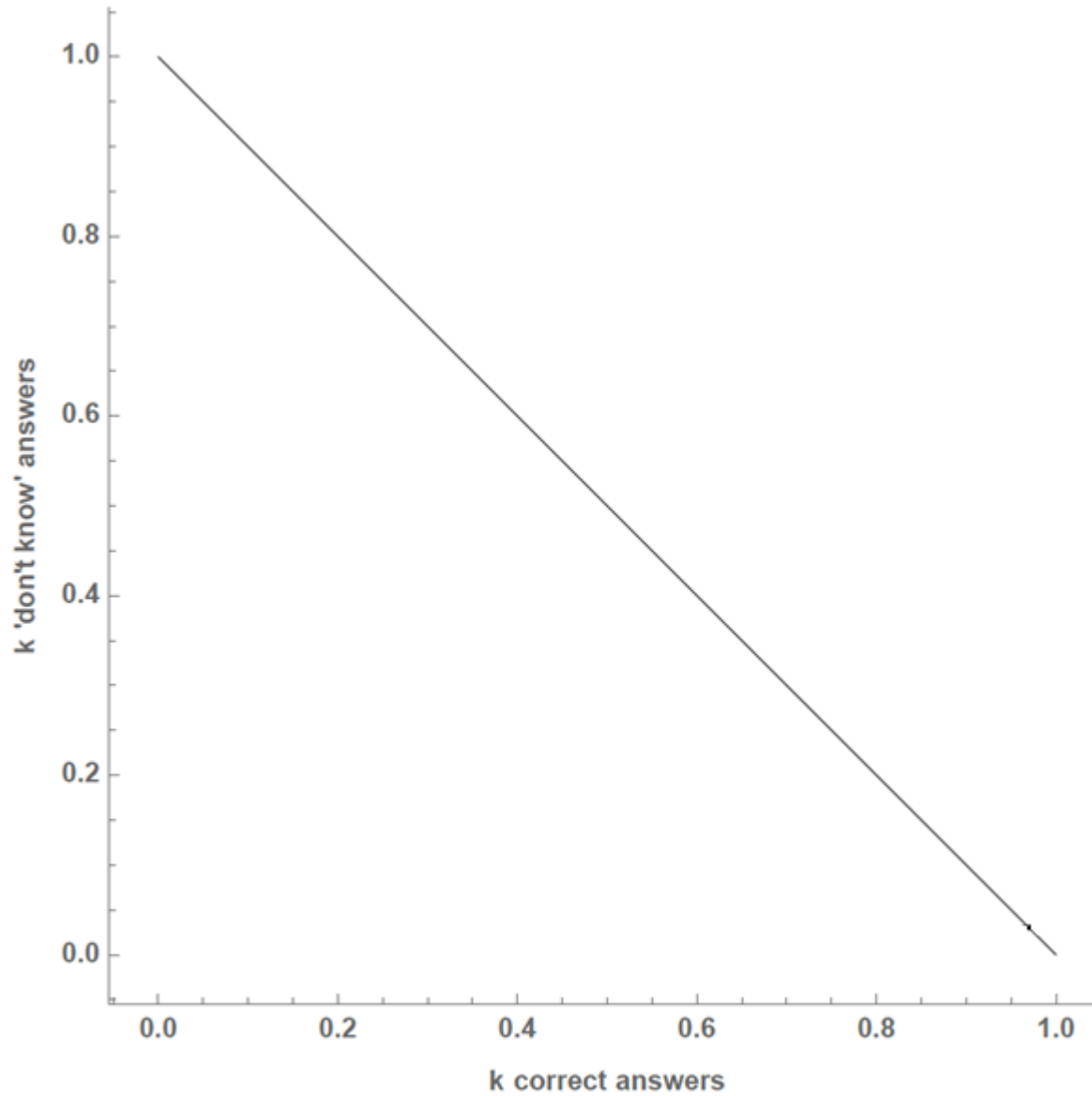


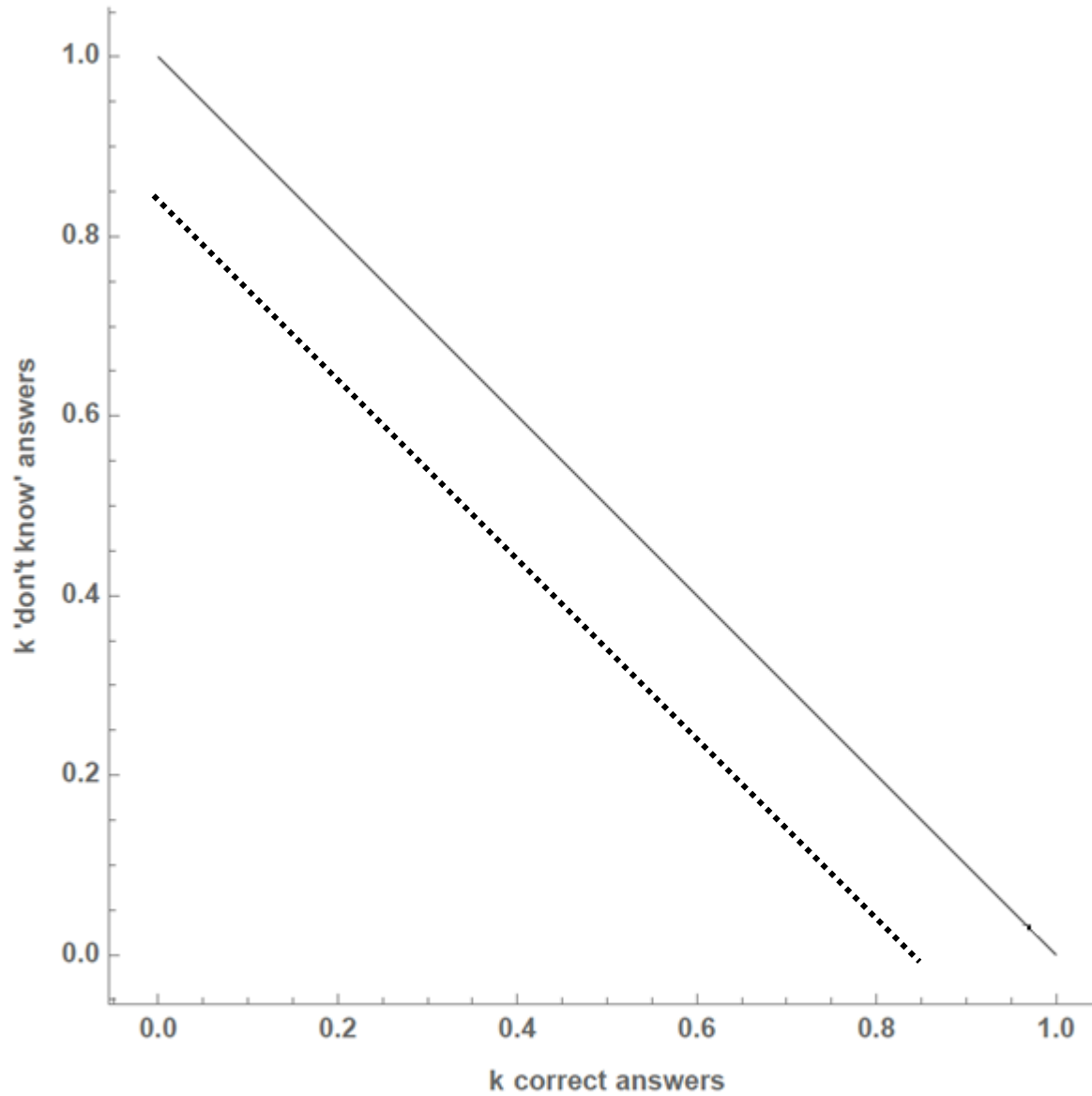


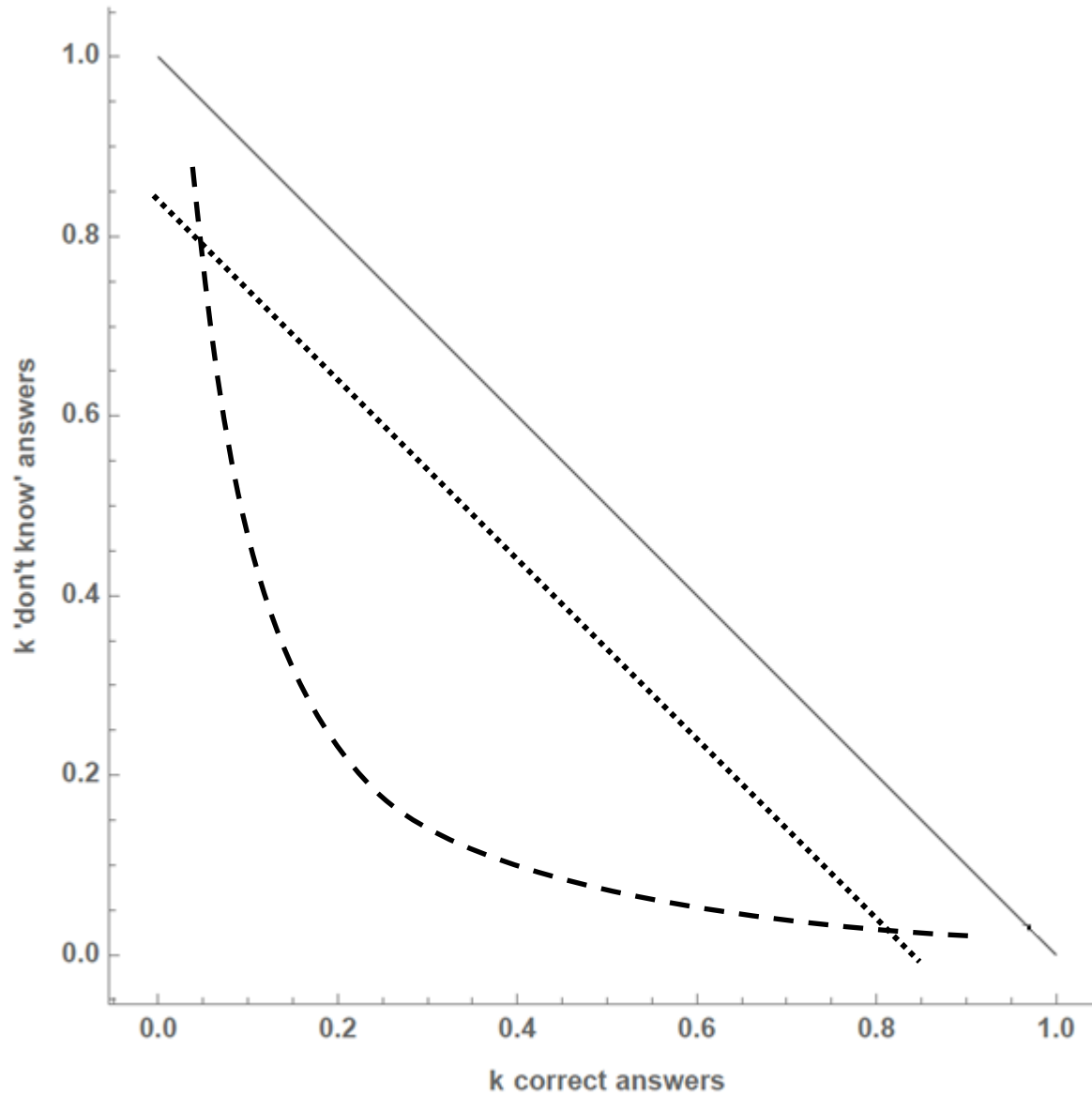
Dunning- Kruger Effect Unskilled and unaware of it



“...and will to the best of my ability, which is terrific ability, by the way. Everyone agrees, I have fantastic ability. So there’s no problem with my ability, believe me....”







Dataset from “*Europeans, Science and Technology*” Eurobarometer

Years: 1989, 1992, 2001, 2002, 2005

34 European Countries

1989, 1992: EU12

2001: EU15

2002: CY, CZ, H, ES, LV, LT, MA, PO, SL, SV, BL, RO, TK

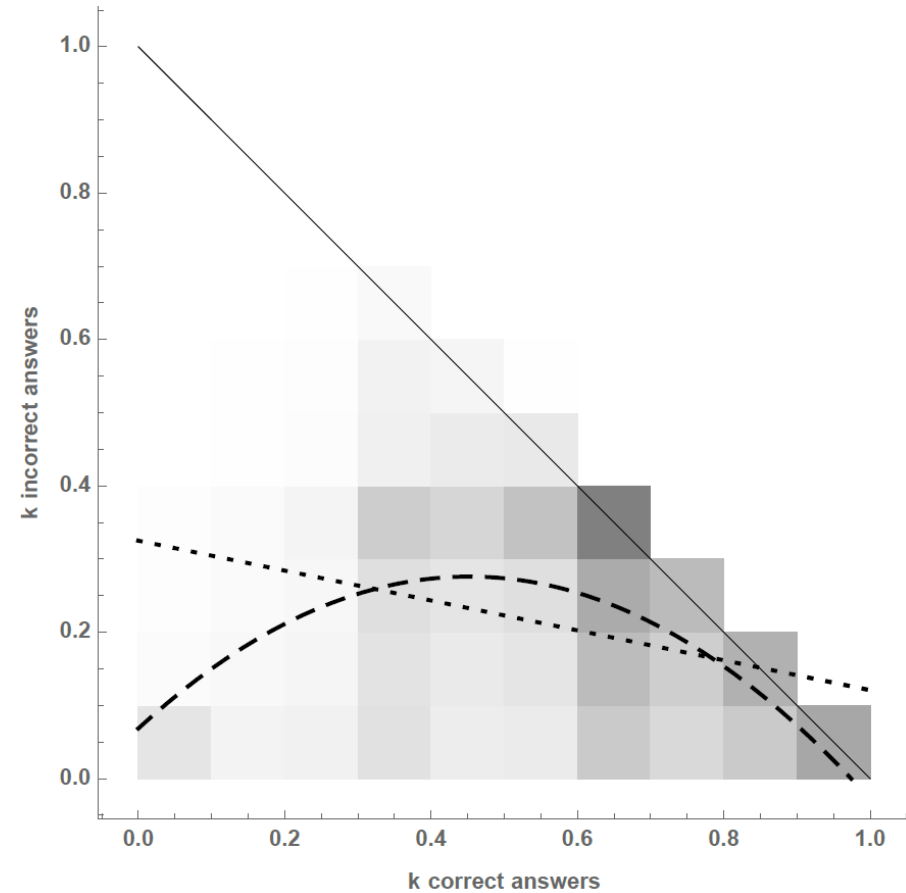
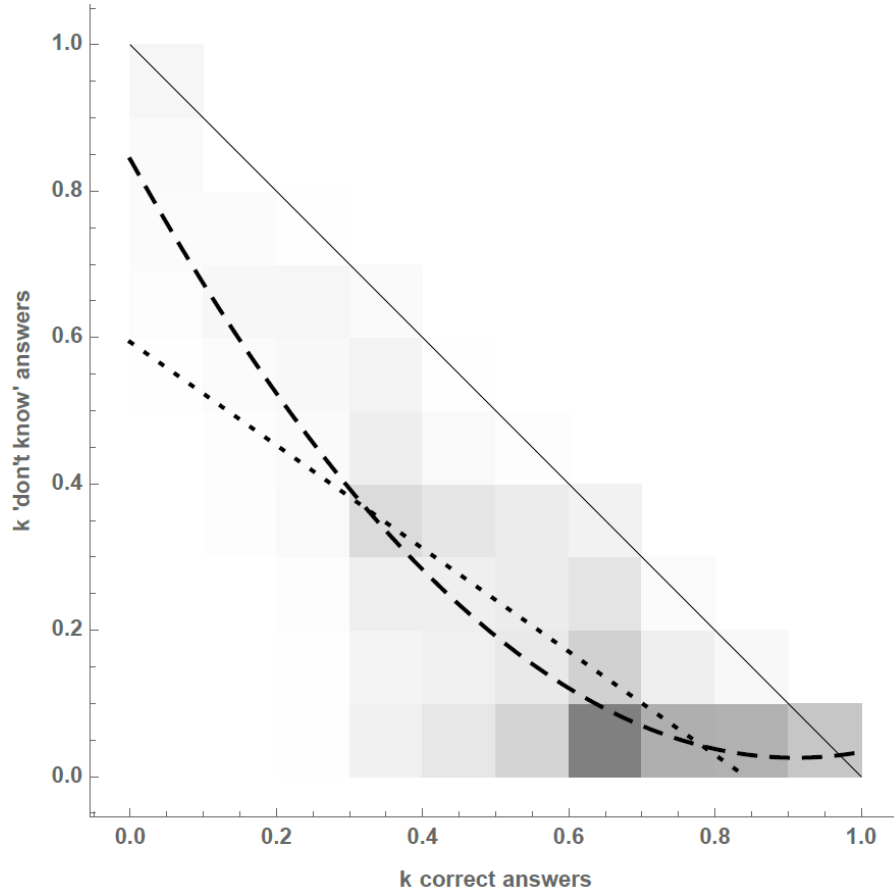
2005: all above plus IS, CR, CH, NO

N=84469

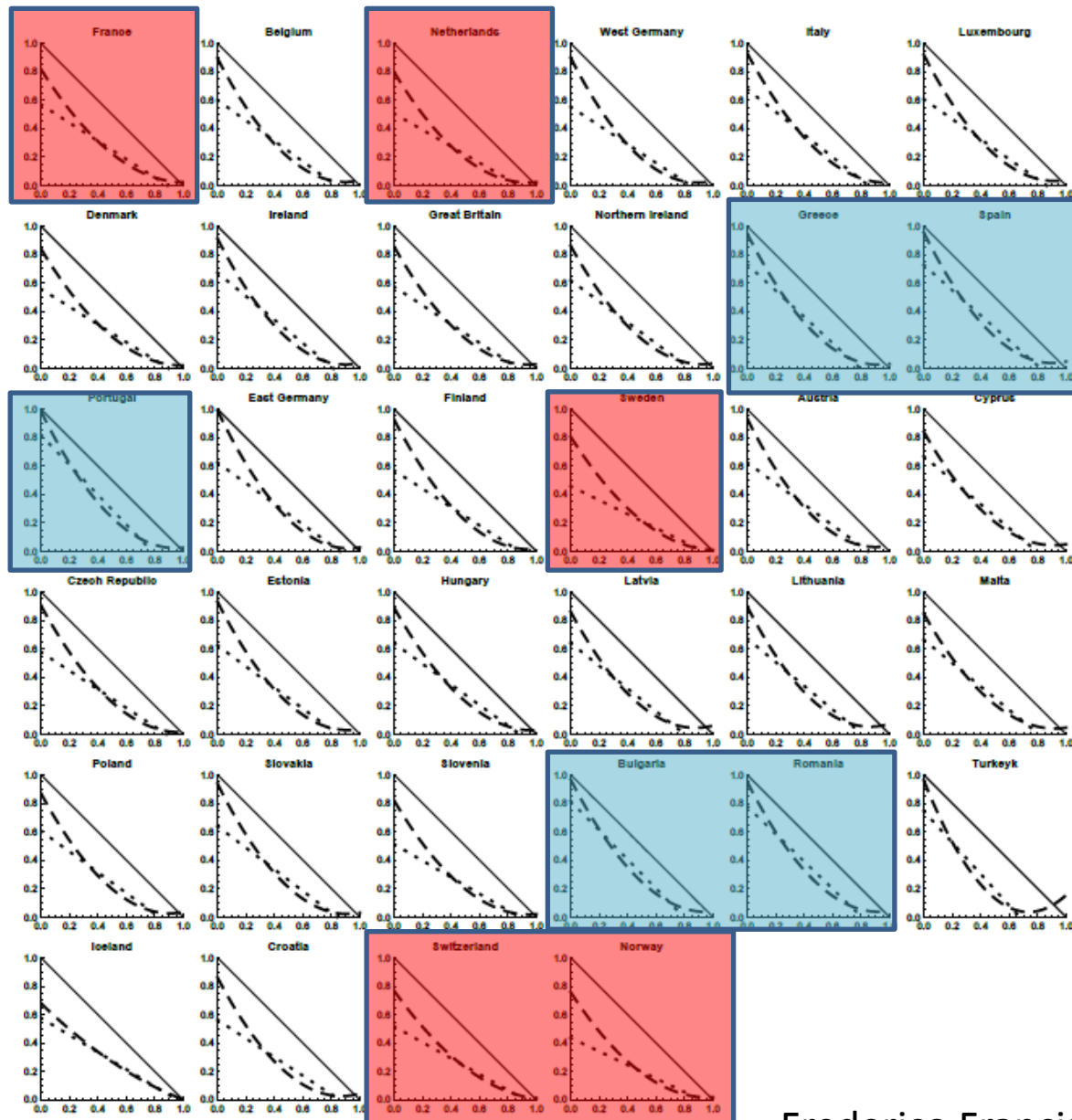
Knowledge

Attitudes

Do wrong answers scale linearly?



More wrong answers at intermediate k levels

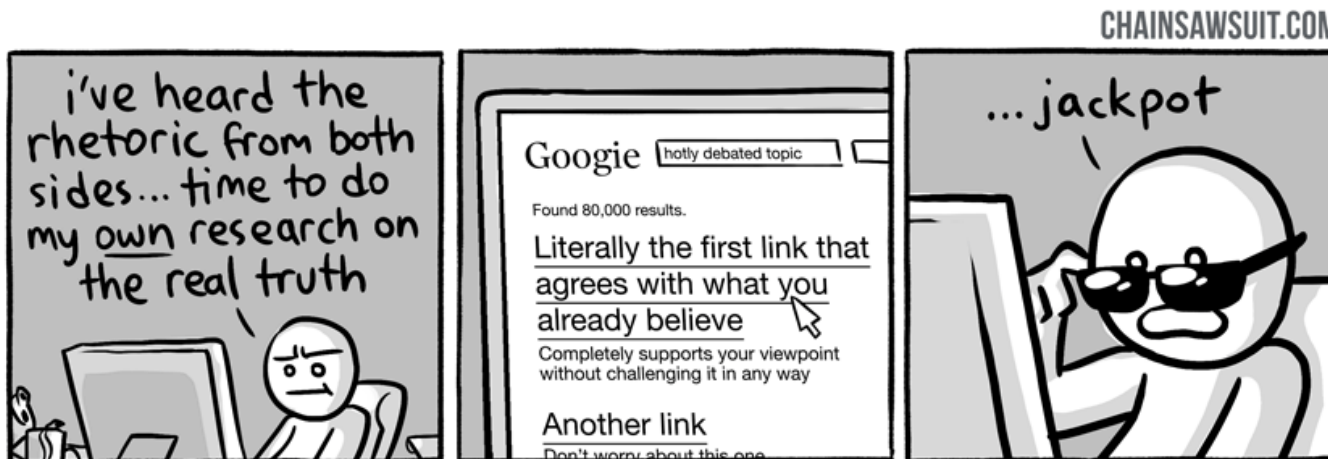




Biases in Decision-Making

The human understanding when it has once adopted an opinion ... draws all things else to support and agree with it. And though there be a greater number and weight of instances to be found on the other side, yet these it either neglects and despises ... in order that by this great and pernicious predetermination the authority of its former conclusions may remain inviolate.

Francis Bacon, Novum Organum, 1620





Fake News



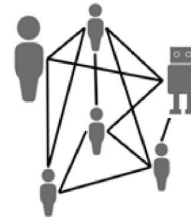
Pathogens



Humans



Hosts



Networks



Environment



Cognitive Biases
Attitudes Towards Science
Tracking Anxiety

Large scale surveys
Behavioral experiments
Twitter
Facebook

Networks
Math Modelling
Psychology
Information



Fake News



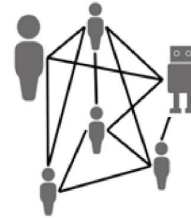
Pathogens



Humans



Hosts

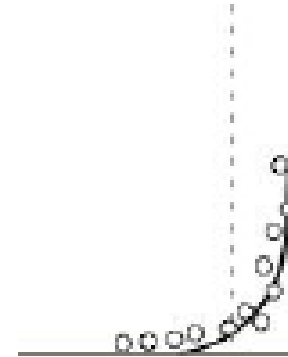
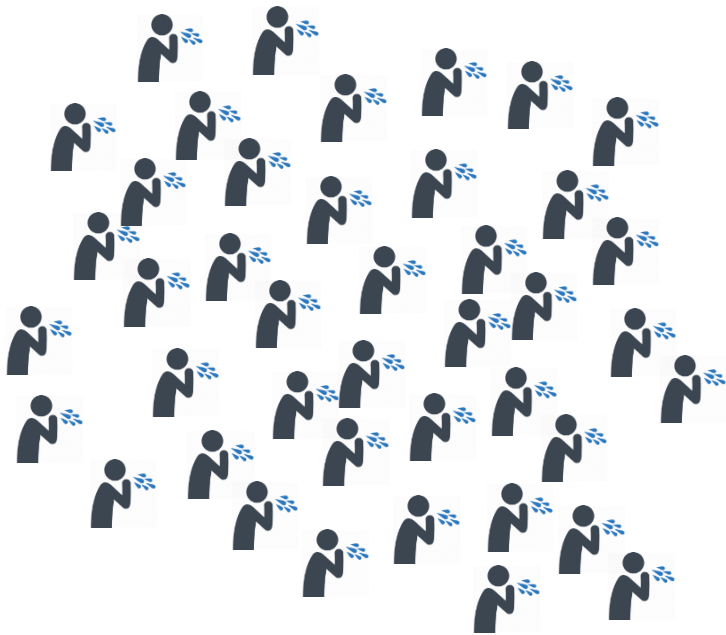


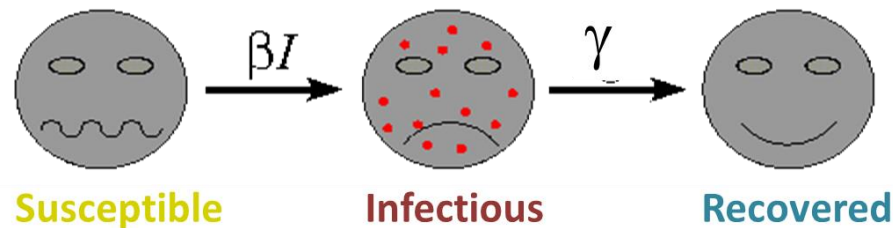
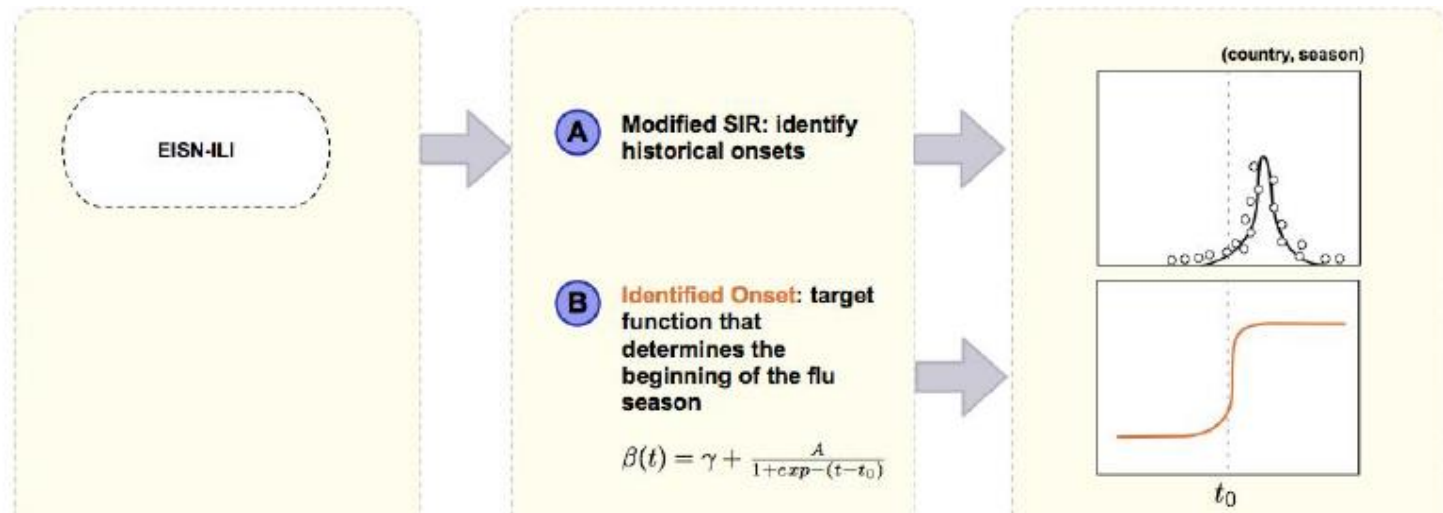
Networks



Environment



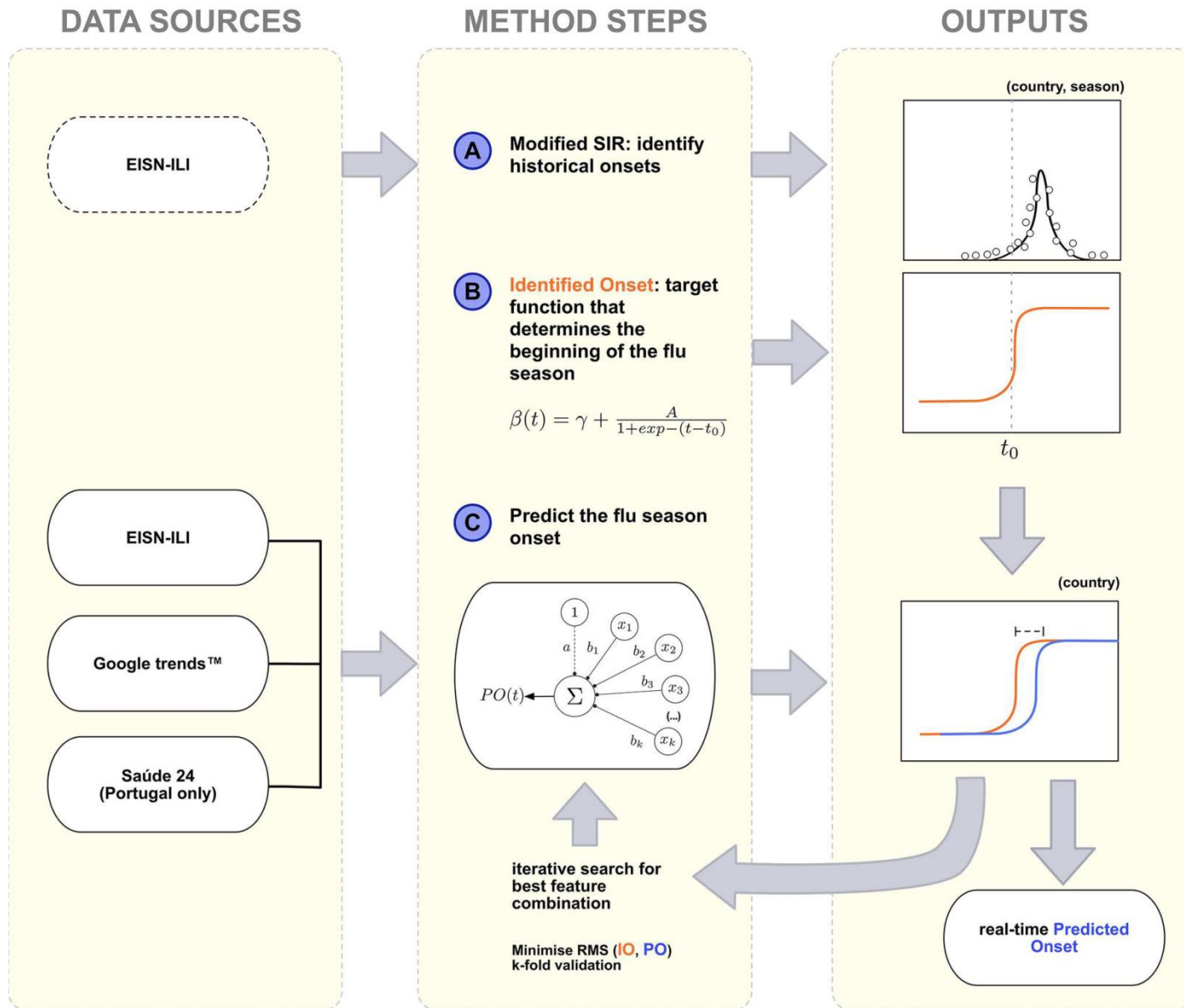


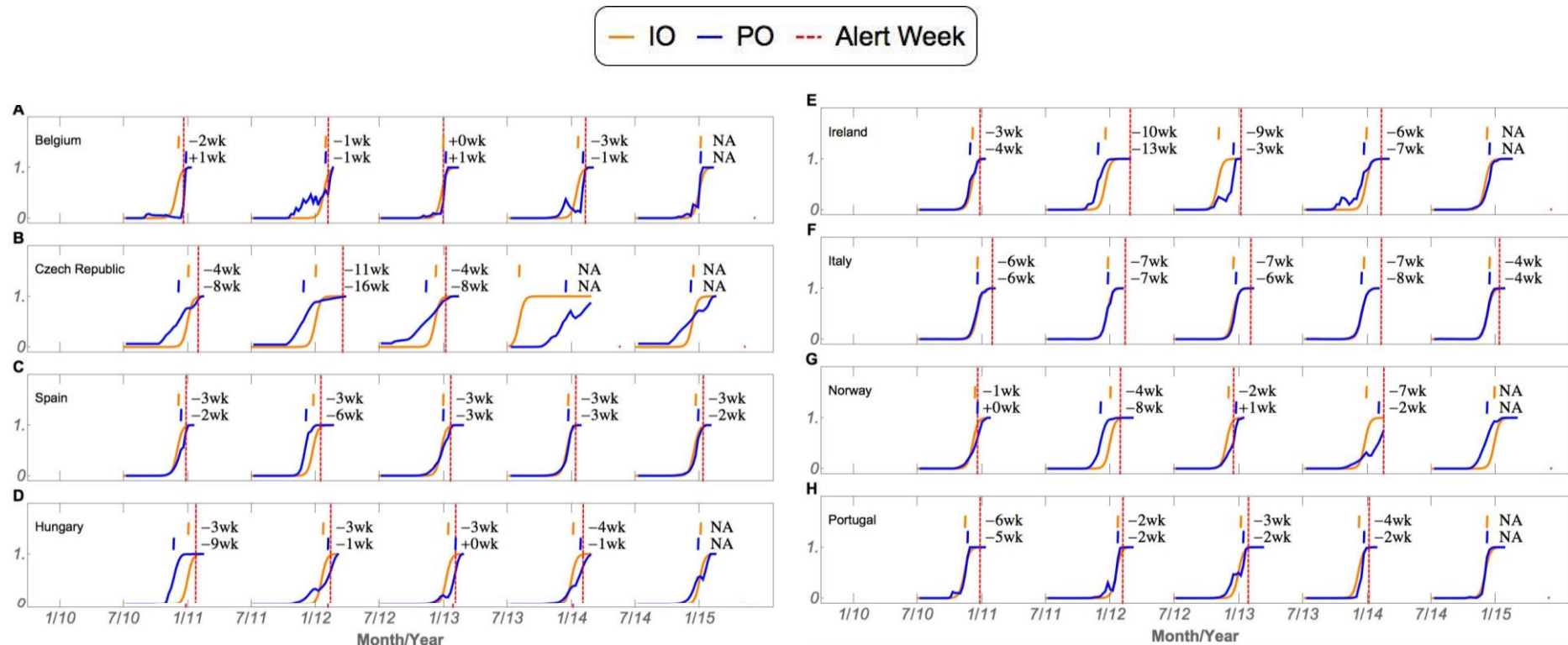


$$\begin{aligned} \frac{dS}{dt} &= -\frac{\beta SI}{N} \\ \frac{dI}{dt} &= \frac{\beta SI}{N} - \gamma I \\ \frac{dR}{dt} &= \gamma I \end{aligned}$$

I - Infected population
S - Susceptible population
R - Recovered population
 β - contact rate
 γ - recovery rate
 N - population size

$$\beta(t) = \gamma + \frac{A}{1 + \exp(-(t - t_0))}$$



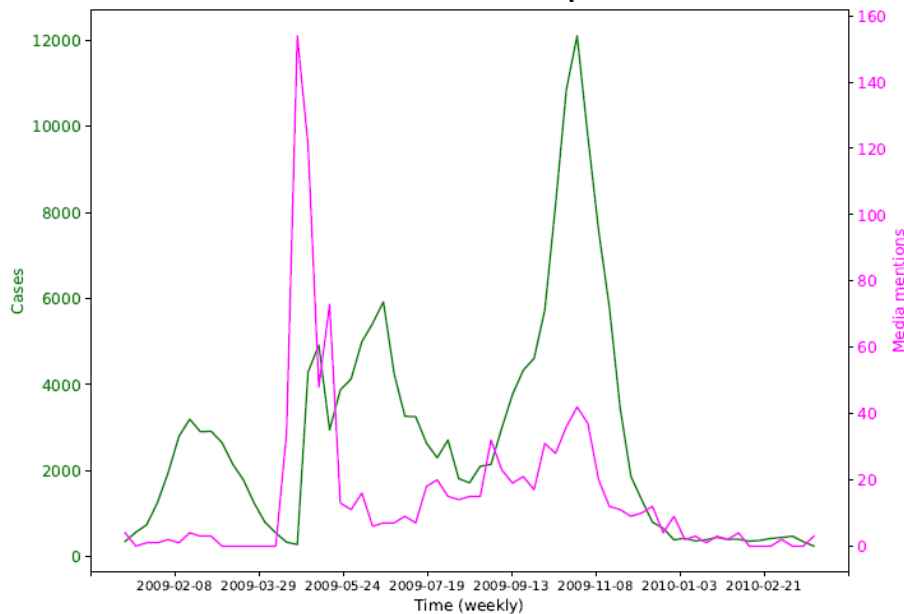


IO matches or anticipates current alerts in **all** cases studied
(by > 2 weeks in **90%** of the cases)

PO matches or anticipates current alerts in **all but three** cases
(by > 2 weeks in **70%** of the cases)

What about pandemic settings?

2009 H1N1 influenza pandemic



2020 SARS-CoV-2 pandemic



Can we distinguish between different drivers of online behaviour?

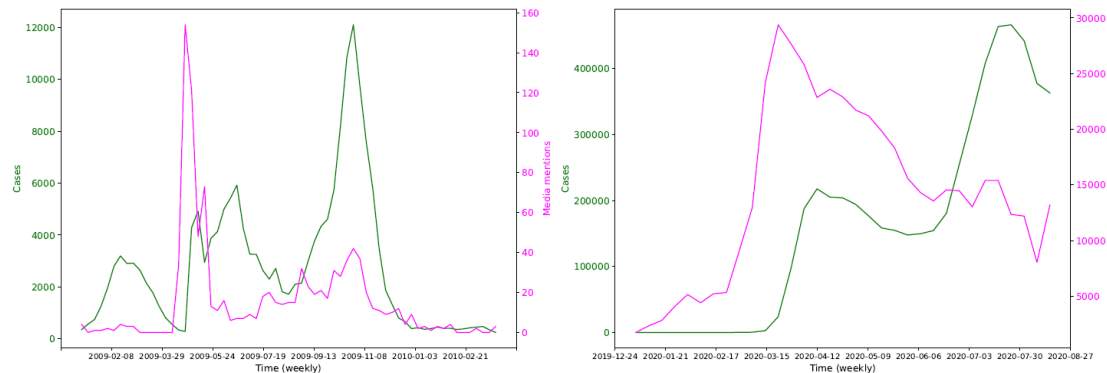
ONLINE

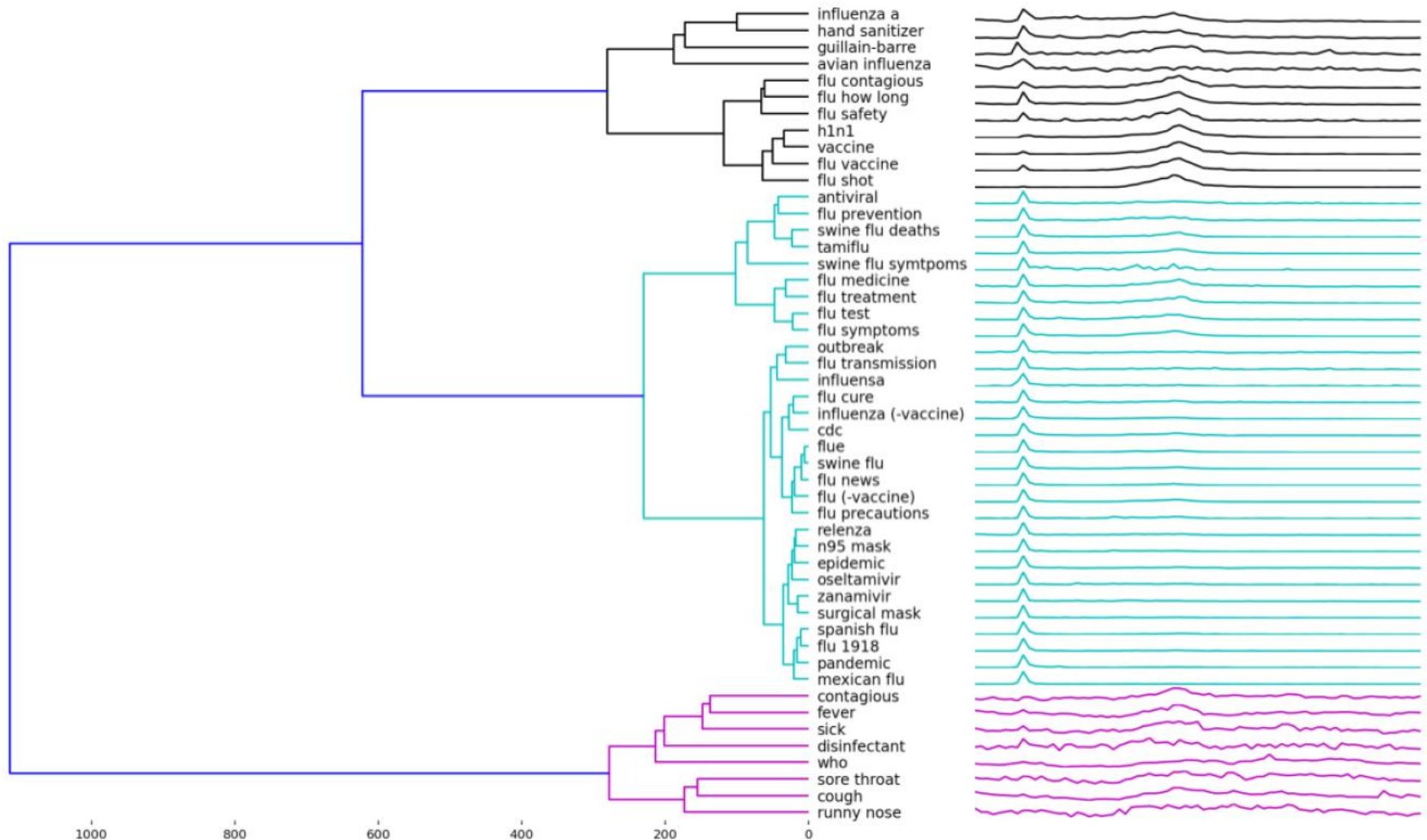
- Searches on Google (GT)
- Searches on Wikipedia (Wiki)
- Twitter posts

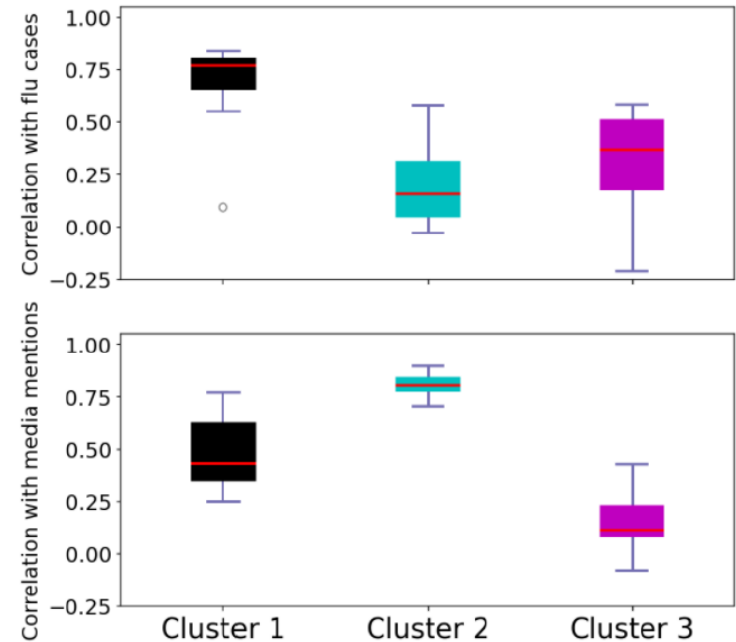
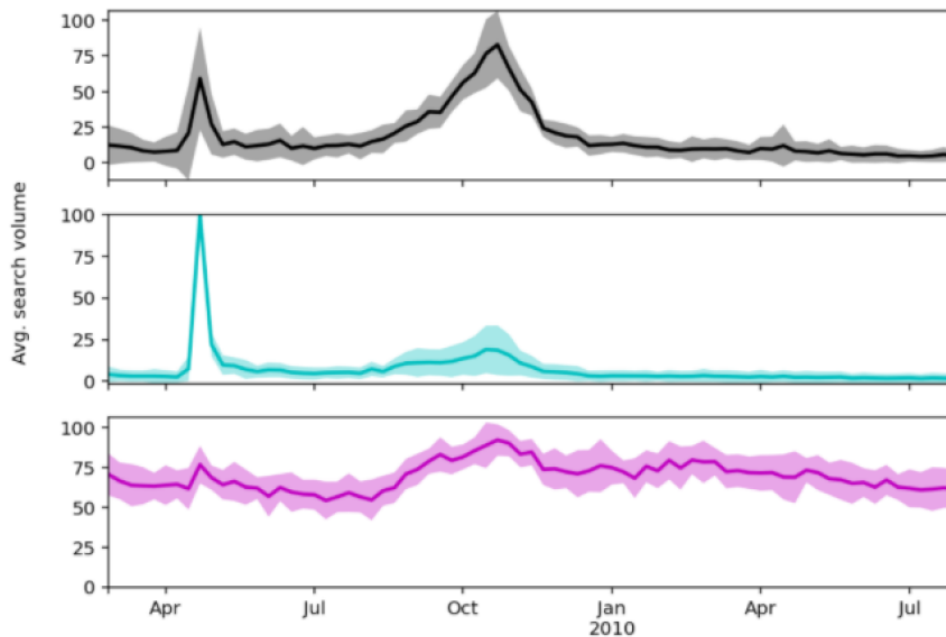
OFFLINE

- # of flu cases (WHO)
- Media reports (NYT)
- Surveys (Anxiety)

Do they have different profiles?

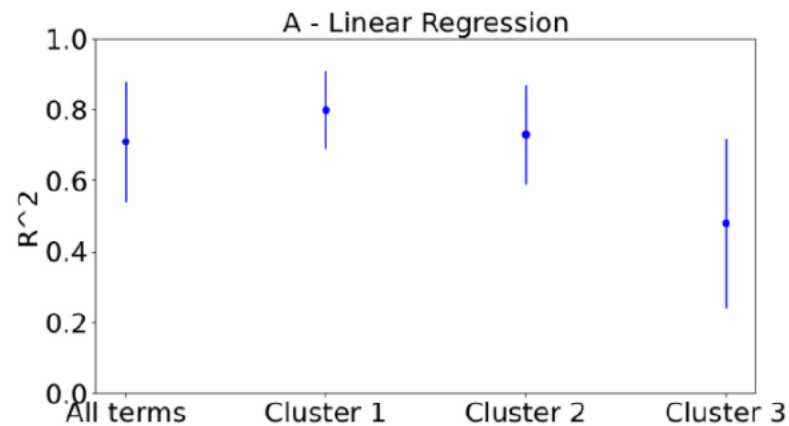
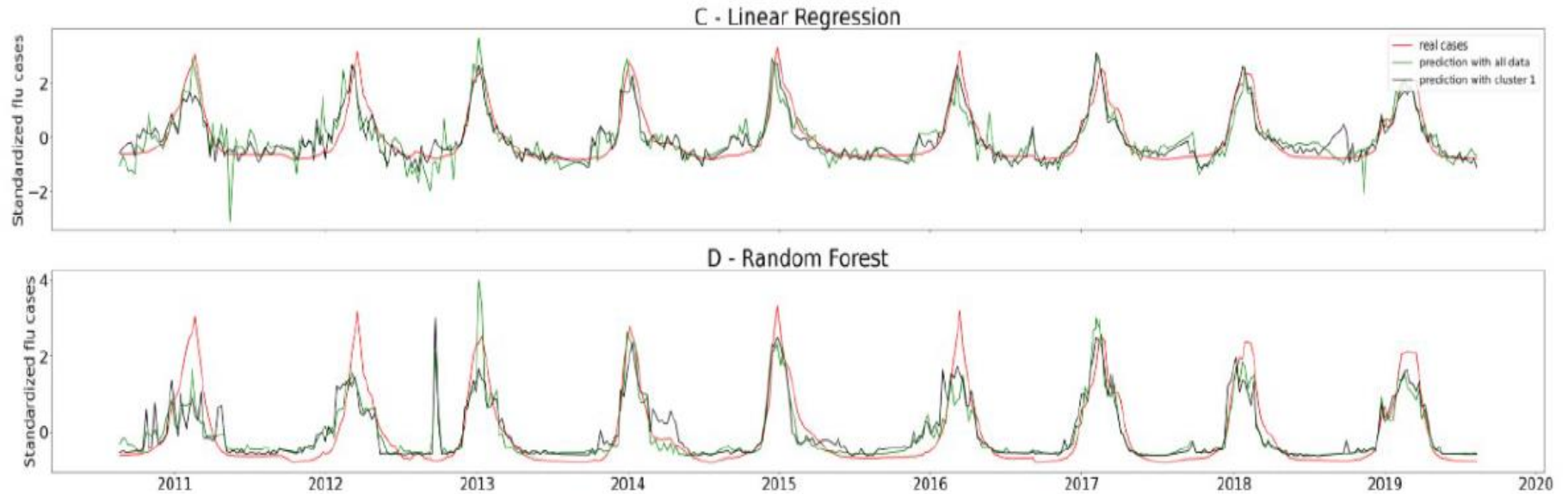






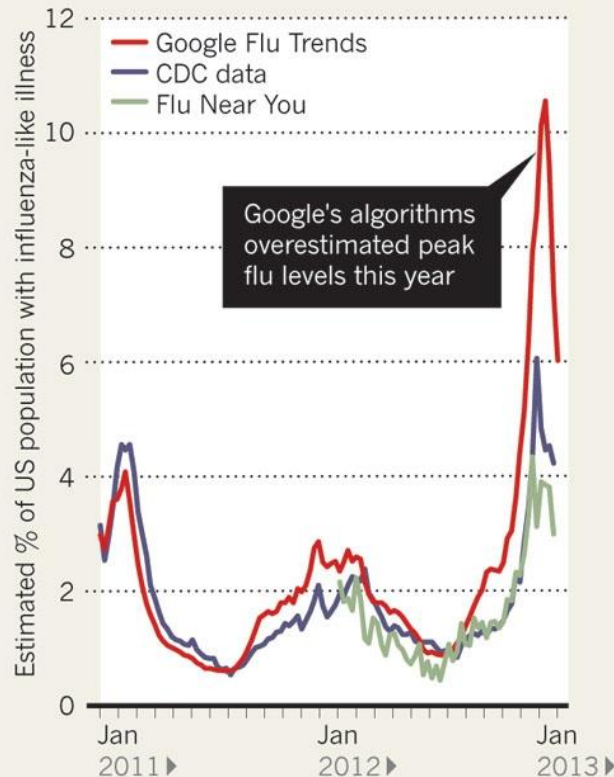
→ Cluster 1 search trends are more correlated with **flu infections**

→ Cluster 2 search trends are more correlated with **flu-related news**



FEVER PEAKS

A comparison of three different methods of measuring the proportion of the US population with an influenza-like illness.



The New York Times

How Data Failed Us in Calling an Election

By Steve Lohr and Natasha Singer, 2016

BIG DATA
HUBRIS

Declan Butler, "When Google got it wrong"
Nature 494, 155–156 (2013)



Fake News



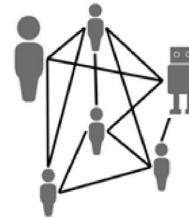
Pathogens



Humans



Hosts



Networks



Environment



HEALTH



Online vs. Offline Patterns
Emergency Now-casting
Antibiotic Over-prescription

Google Trends
SNS24
Twitter
ER acceptance /times
SPMS e-prescriptions

Math Modelling
ML
Epidemiology

FAKE AND REAL NEWS

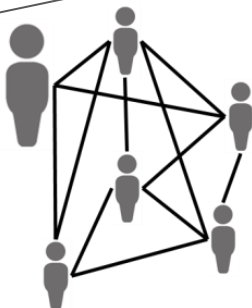


INDIVIDUAL
SUSCEPTIBILITIES

DISEASE
DYNAMICS
MODELS

ANONYMOUS, DISTRIBUTED
ENCRYPTED

ENVIRONMENT AND HISTORY



1. Cognitive biases & confidence to knowledge ratios, good predictors of FN sharing
2. Position on networks should be good predictors of FN sharing
3. Past history should offer good indicators of future FN spreading

FAKE AND REAL NEWS



4 – FN that spread **slower** might activate more **complex** cognitive processes

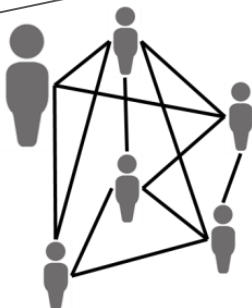


INDIVIDUAL SUSCEPTIBILITIES

1 – **Overconfident (+ unknowledgeable)** individuals more likely to share FN

DISEASE DYNAMICS MODELS

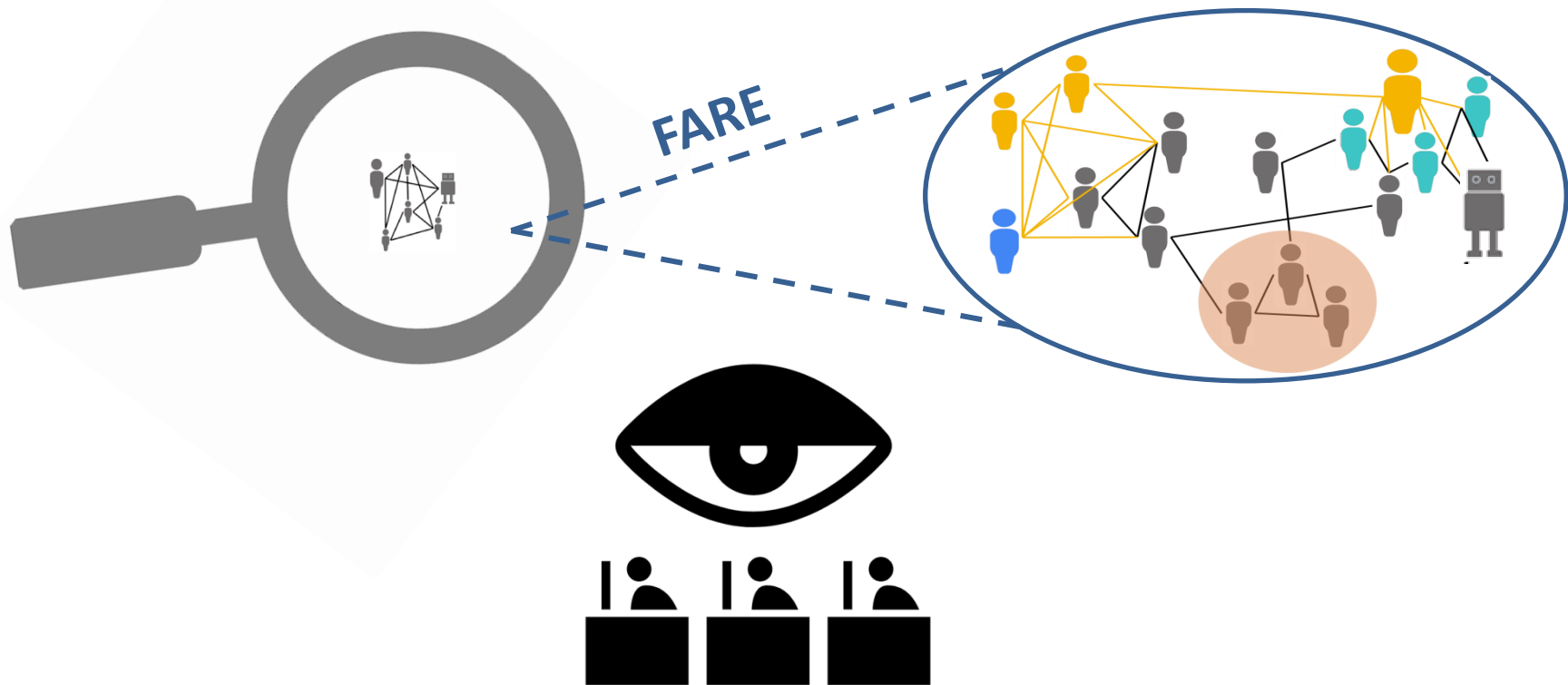
ANONYMOUS, DISTRIBUTED
ENCRYPTED



ENVIRONMENT AND HISTORY

- 2 – **Less diverse social networks** more likely to share FN
- 3 – Sharing in the **past** should lead to sharing in the **future**

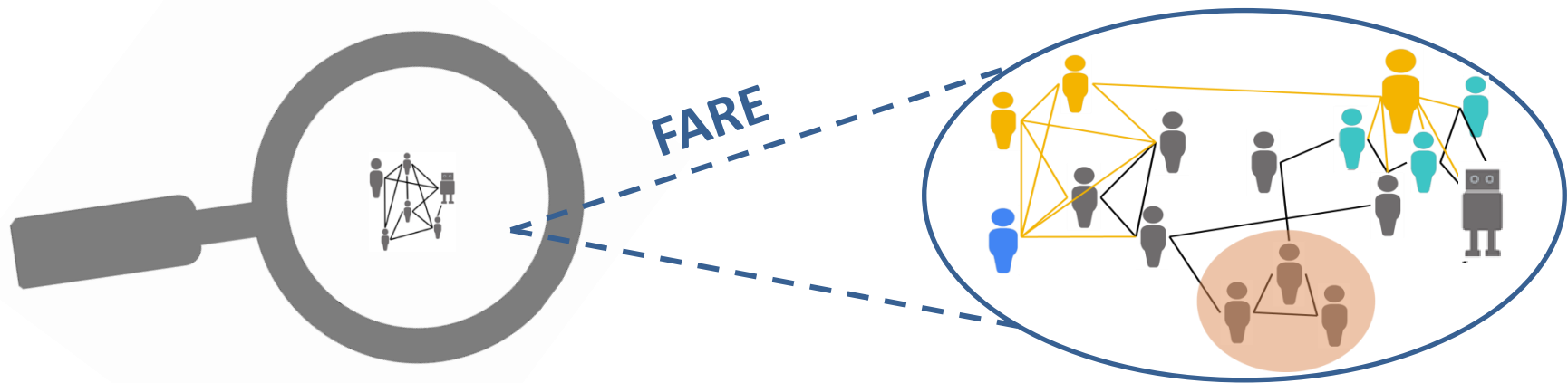
URGENT INSIGHTS TO RECLAIM THE DIGITAL REVOLUTION



EUROPE
DRIVEN

BIG DATA CAN HELP BETTER UNDERSTAND HUMAN BEHAVIOUR
A SOUND AND ETHICAL MACROSCOPE

URGENT INSIGHTS TO RECLAIM THE DIGITAL REVOLUTION



Data limits/ controls

What to do with the results

Improve privacy protection

Raise awareness

Change the narrative

These problems—and a wide range of similar problems in the biological, medical, psychological, economic, and political sciences—are just too complicated to yield to the old nineteenth-century techniques which were so dramatically successful on two-, three-, or four-variable problems of simplicity. These new problems, moreover, cannot be handled with the statistical techniques so effective in describing average behavior in problems of disorganized complexity.

These new problems, and the future of the world depends on many of them, requires science to make a third great advance, an advance that must be even greater than the nineteenth-century conquest of problems of simplicity or the twentieth-century victory over problems of disorganized complexity. Science must, over the next 50 years, learn to deal with these problems of organized complexity.

Warren Weaver, 1947

Is there any promise on the horizon that this new advance can really be accomplished? There is much general evidence, and there are two recent instances of especially promising evidence

The first piece of evidence is the wartime development of new types of electronic computing devices. These devices are, in flexibility and capacity, more like a human brain than like the traditional mechanical computing device of the past.

The second of the wartime advances is the “mixed-team” approach of operations analysis.

Thank you

