



# Hopes and challenges in information science for cosmology



IAP Colloquium

Florent Leclercq

[www.florent-leclercq.eu](http://www.florent-leclercq.eu)

Institut d'Astrophysique de Paris  
CNRS & Sorbonne Université

In collaboration with the Aquila Consortium

[www.aquila-consortium.org](http://www.aquila-consortium.org)

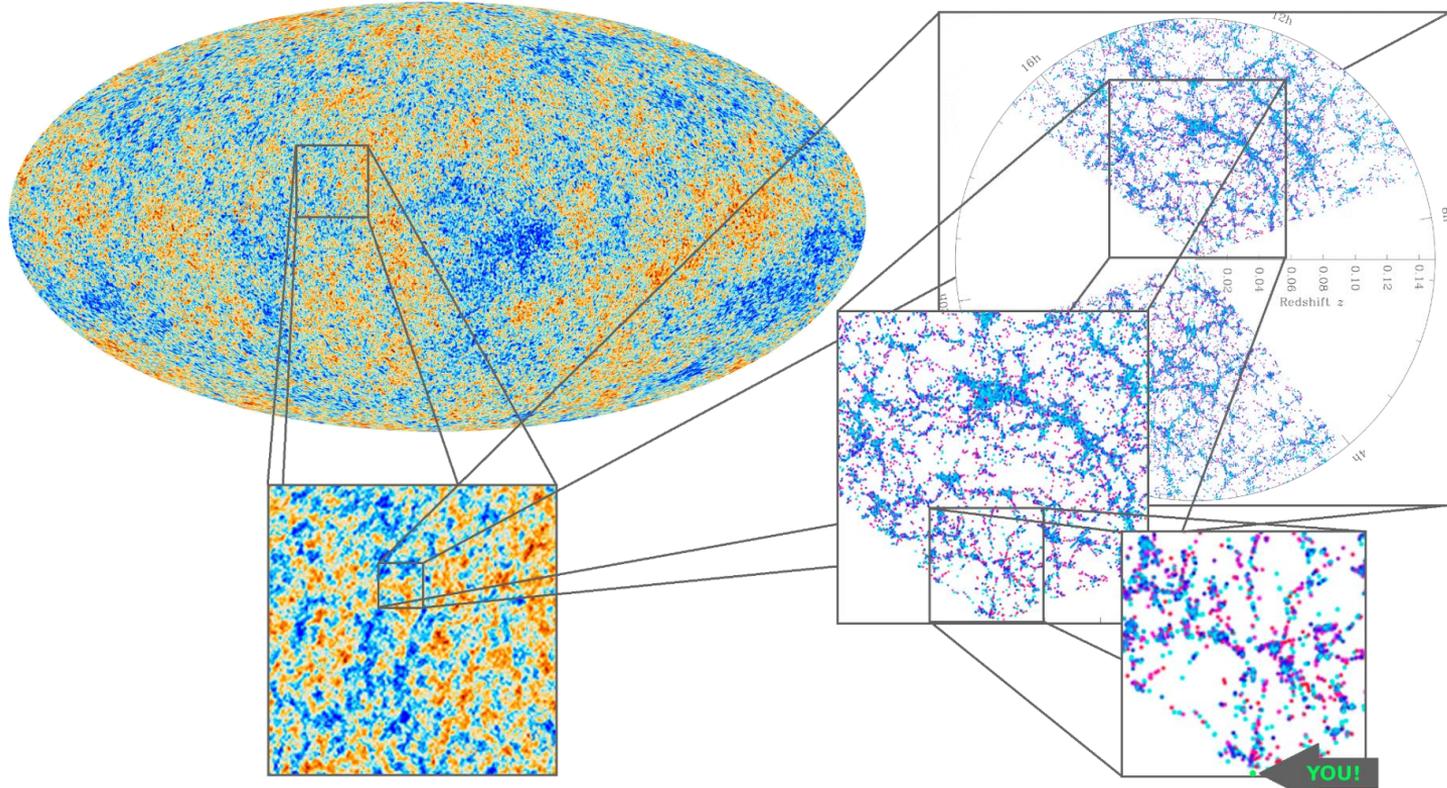
**21 October 2022**

Information science, but what for?



The big picture: the Universe is highly structured

*You are here. Make the best of it...*



Planck collaboration (2013-2015)

M. Blanton and the Sloan Digital Sky Survey (2010-2013)



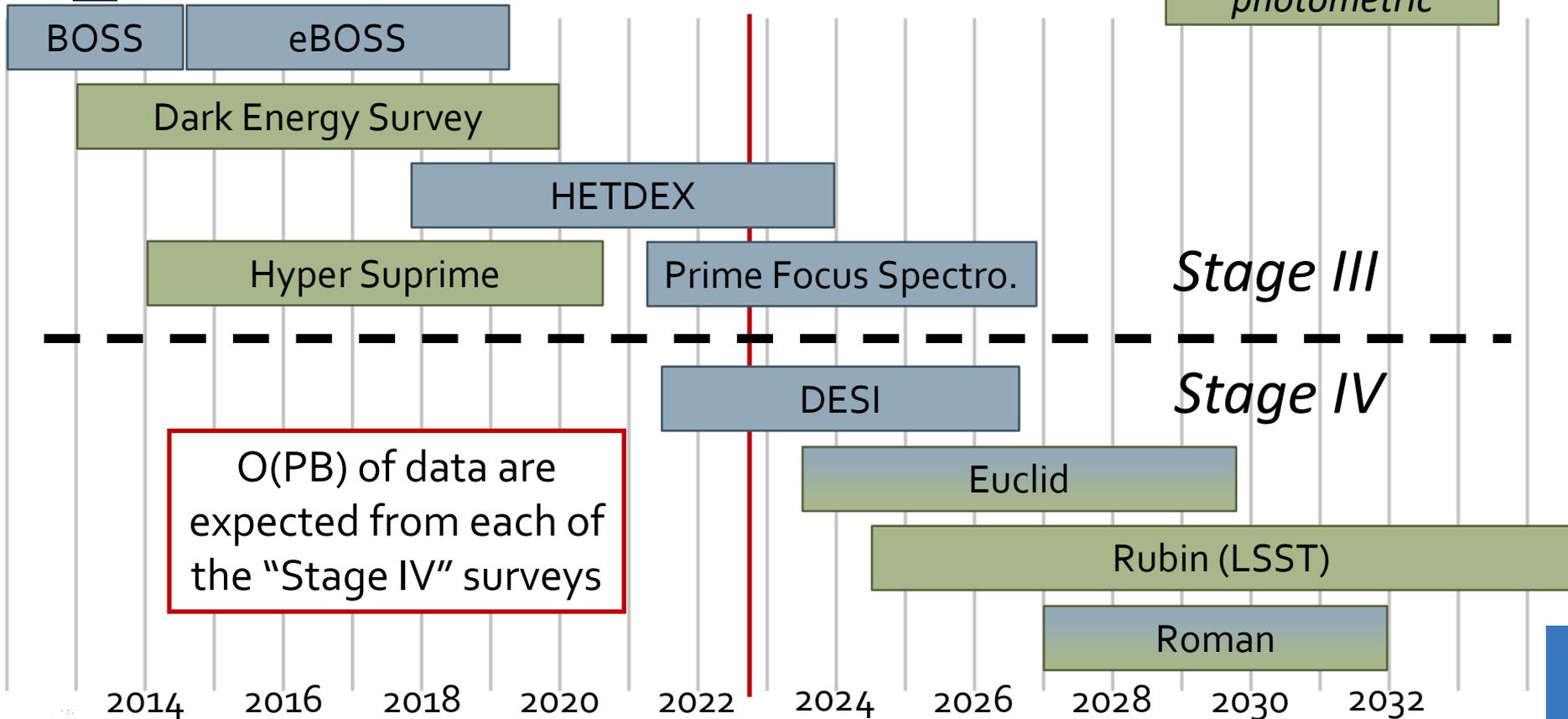
## What we want to know from the large-scale structure

The LSS is a vast source of knowledge:

- **Cosmology:**
  - $\Lambda$ CDM: cosmological parameters and tests against alternatives,
  - Physical nature of the dark components,
  - Neutrinos: number and masses,
  - Geometry of the Universe,
  - Tests of General Relativity,
  - Initial conditions and link to high energy physics
- **Astrophysics:** galaxy formation and evolution as a function of their environment
  - Galaxy properties (colours, chemical composition, shapes),
  - Intrinsic alignments, intrinsic size-magnitude correlations



# Large-scale structure surveys roadmap

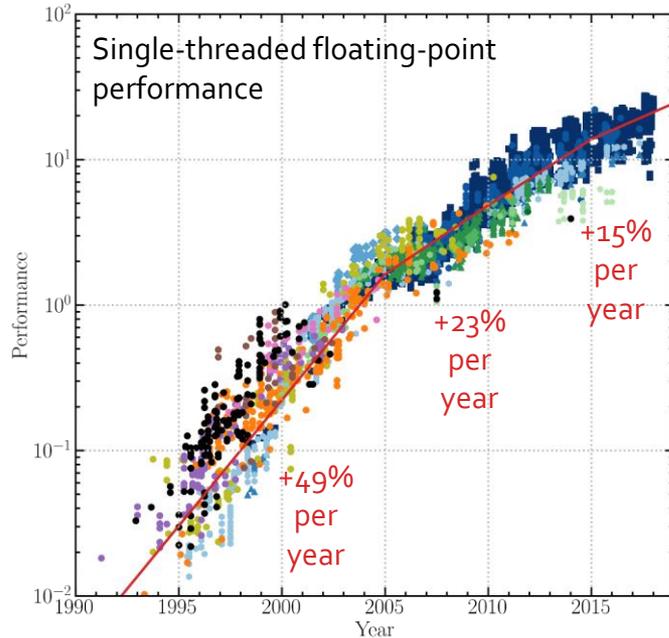


The forward problem:  
From theory to data



# Numerical data models in the exascale world

- Traditional hardware architectures are reaching their physical limit: per-core compute performance is slowing down.



Based on adjusted SPECfp® results, <http://spec.org>

- Current hardware development focuses on:
  - Packing a **larger number of cores** into each CPU: currently  $\mathcal{O}(10^5)$ , soon  $\mathcal{O}(10^6-7)$  in systems that are currently being built.
  - Developing hybrid architectures with cores + **accelerators**: GPUs, reconfigurable or dedicated chips (FPGAs/ASICs).



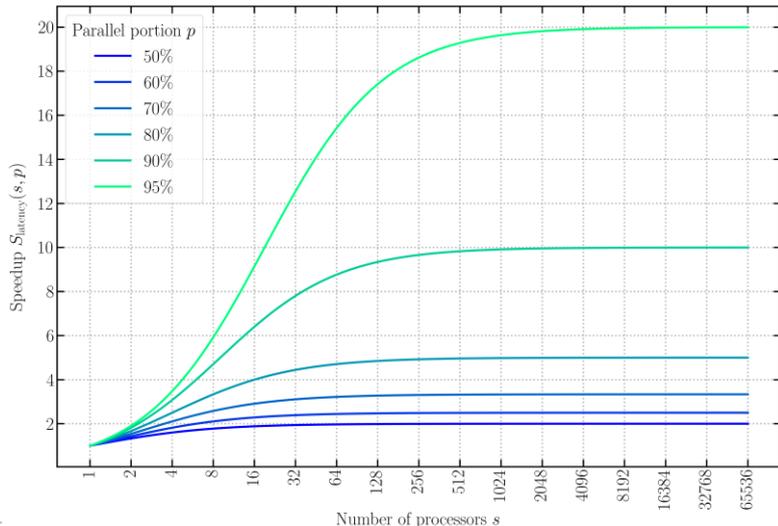
- Numerical data models cannot merely rely on computers becoming faster to reduce the computational time.



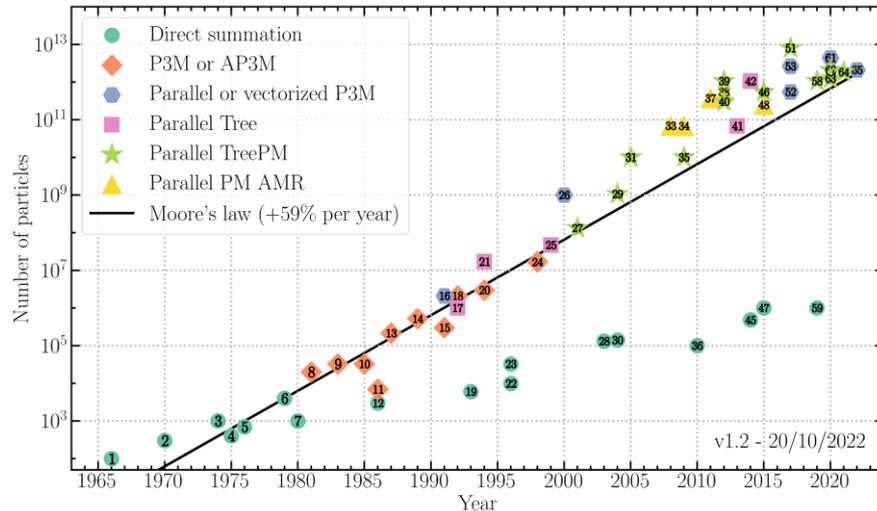
## Parallelisation of $N$ -body codes: the challenge

- Compute cycles are no longer the scarce resource. The cost is driven by **interconnections**.
- Amdahl's law: **latency kills the gains of parallelisation**.

Amdahl 1967, doi:10.1145/1465482.1465560



- Most of the work on numerical cosmology so far has focused on algorithms (such as tree, multipole, and mesh methods) that **reduce the need for communications** across the full computational volume



All references at [Github:florent-leclercq/Moore\\_law\\_cosmosims](https://github.com/florent-leclercq/Moore_law_cosmosims)

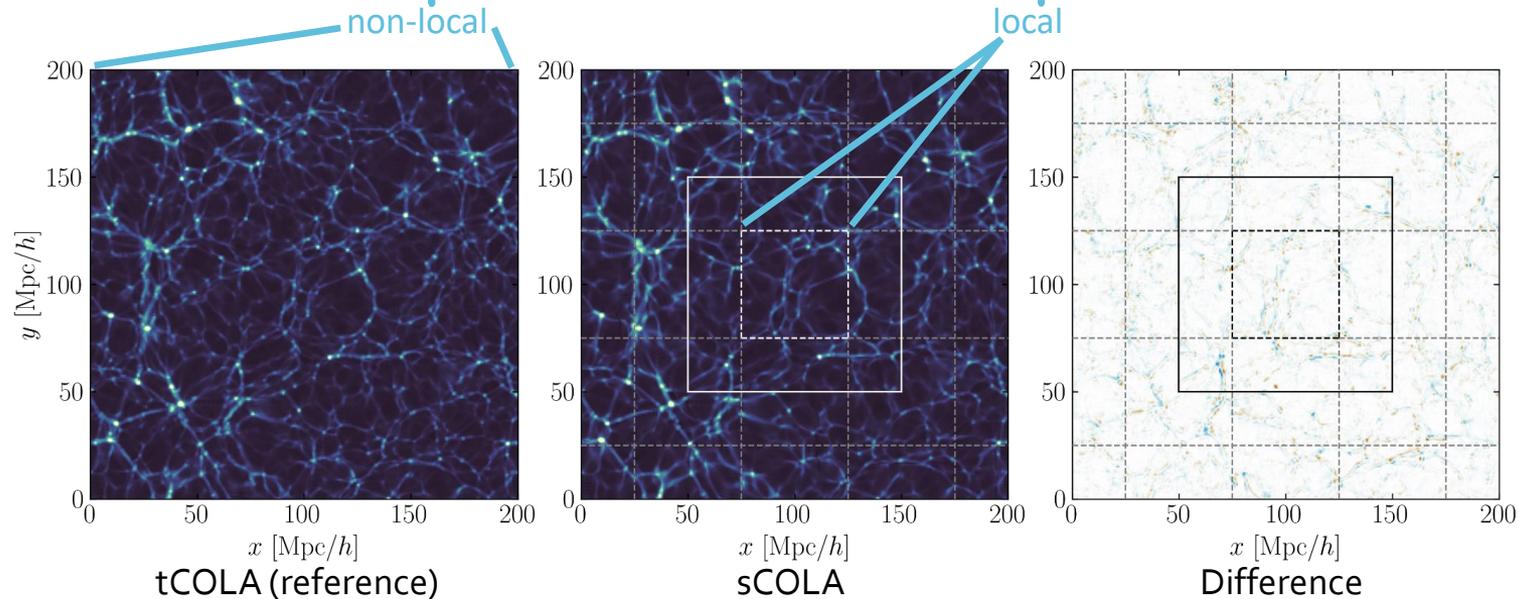


# Perfectly parallel cosmological simulations using spatial comoving Lagrangian acceleration (sCOLA)

- Can we decouple sub-volumes by using the large-scale analytical solution?

$$\frac{\partial^2 \mathbf{x}}{\partial t^2} = -\nabla \left[ \underbrace{\Delta^{-1} \delta}_{\text{non-local}} \right]$$

$$\frac{\partial^2 (\mathbf{x} - \mathbf{x}_{1.s.})}{\partial t^2} = -\nabla \left[ \underbrace{\Delta^{-1} (\delta - \delta_{1.s.})}_{\text{local}} \right]$$



Publicly available implementation:  
Bitbucket:[florent-leclercq/simbelmyne/](https://bitbucket.org/florent-leclercq/simbelmyne/)

FL, Faure, Lavaux, Wandelt, Jaffe, Heavens, Percival & Noûs 2020, 2003.04925

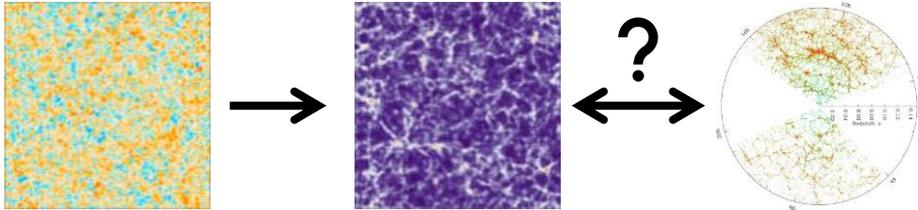


The inverse problem:  
From data to theory



# Why Bayesian inference?

- Inference of signals: an **ill-posed problem**
  - Incomplete observations: finite resolution, survey geometry, selection effects
  - Noise, biases, systematic effects
  - Cosmic variance



➔ No unique recovery is possible!

- A **natural progression** in cosmology:
  - Observations of the homogeneous and isotropic expansion (supernovæ)
  - Anisotropies of linear perturbations (CMB)
  - Non-linear cosmic structure at small scales and late times (galaxy surveys)
- Additional challenges for next-generation data:
  - Difficult data analysis questions and/or hints for new physics will first show up as **tensions** between measurements
  - **Non-linearity**: 80% of the total signal will come from non-linear structures  
e.g. LSST Science Book, 0912.0201
  - **Model misspecification**: Next-generation surveys will be dominated by (unknown) systematics

## A simple statement about building knowledge

- Bayes' theorem (1763): a statement about how we **analyse evidence** and change our minds at we get new information:

$$p(s|d) = \frac{p(d|s)p(s)}{p(d)}$$

Demonstration:

$$p(s, d) = p(s|d)p(d) = p(d, s) = p(d|s)p(s). \quad \square$$

- But why should we use it?
  - Bayes' theorem is trivial and outdated.
  - It measures **belief**. It says we can learn even from missing or incomplete data, from approximations, from ignorance. It runs counter to the conviction that science requires objectivity and precision.
  - After Laplace's death, it was pronounced dead and buried.



Thomas Bayes  
(1701-1761)



Richard Price  
(1723-1791)



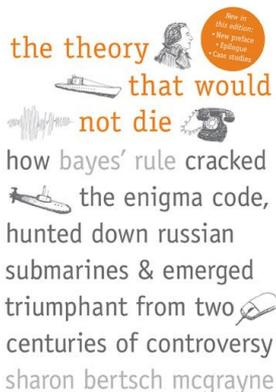
Pierre-Simon de Laplace  
(1749-1827)



Picture taken at Bunhill Fields Burial Ground, City of London, 2021

# The theory that would not die

- And yet, Bayes' theorem helped in many practical situations:
  - Exonerate Alfred Dreyfus from miscarriage of justice (1899-1906),
  - Save the Bell Telephone system from financial panic (1907),
  - Predict earthquakes and tsunamis (Harold Jeffreys, 1930-1940),
  - Direct Allied artillery fire and locate German submarines (1939-1945),
  - Break the German navy's Enigma cipher (Alan Turing, 1940-1944),
  - Prove that smoking causes lung cancer (Jerome Cornfield, 1951)...



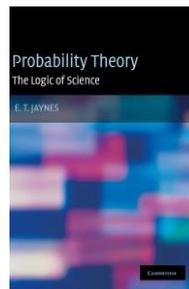
"If you're not thinking like a Bayesian, perhaps you should be."  
—John Allen Paulos, *New York Times* Book Review

Sharon Bertsch McGrayne (2012)

- The scientific battle lasted for 150 years, until computers arrived.

The superiority of Bayesian methods is now a thoroughly demonstrated fact in a hundred different areas. One can argue with a philosophy; it is not so easy to argue with a computer printout, which says to us: "Independently of all your philosophy, here are the facts of actual performance." Jaynes 2002, *Probability Theory – The logic of science*

- [Cox-Jaynes theorem](#) (1946): Any system to manipulate "plausibilities", consistent with Cox's desiderata, is isomorphic to Bayesian probability theory.



Jaynes (2002)



Richard Threlkeld Cox  
(1898-1991)



Edwin Thompson Jaynes (1922-1998)

## A bit of history: the Dreyfus affair

- 1894: Captain Alfred Dreyfus, of Jewish descent, is convicted of spying for Germany on the basis of handwriting analysis. He is sentenced to public cashiering (*dégradation*) and a life sentence in the penal colony of Devil's Island (*Île du Diable, bagne de Cayenne*, French Guiana).
- 1896-1898: Evidence comes to light which identifies the real culprit, who is acquitted.
- 1898: Novelist Émile Zola publishes an open letter to the President of the Republic Félix Faure: *J'accuse...!*
- 1899: Dreyfus is tried and convicted again, sentenced to ten years' imprisonment. He is pardoned and released, but not cleared.
- 1906: *Réhabilitation*: Dreyfus is exonerated from all charges and reinstated as *Chef d'Escadron* (commandant) in the French Army.



“The traitor”: Degradation of Alfred Dreyfus in the court of the military school in Paris, 5 January 1895



“J'accuse...!": Open letter by Émile Zola published in L'Aurore, 13 January 1898

*J'accuse [l'armée] d'avoir fait une enquête scélérate, j'entends par là une enquête de la plus monstrueuse partialité (...).  
J'accuse les trois experts en écriture (...) d'avoir fait des rapports mensongers et frauduleux (...).*

I accuse [the army] of having carried out a villainous investigation, by which I mean an investigation of the most monstrous partiality (...).  
I accuse the three handwriting experts (...) of having made deceitful and fraudulent reports (...).

# Mathematicians and probability theory in the Dreyfus affair

- Central to the incrimination was the “*système Bertillon*”, based on graphology and probability calculus, attempting to prove that Dreyfus had disguised his handwriting by imitating his own handwriting (“auto-forgery”).
- During the 1899 trial in Rennes, mathematicians Jacques Hadamard and Paul Painlevé were called as witnesses.
- Painlevé read a letter by Henri Poincaré to the Court.



Alphonse Bertillon  
(1853-1914)



Jacques Hadamard  
(1865-1963)



Paul Painlevé  
(1863-1933)

Le procès Dreyfus devant le conseil de guerre de Rennes, 7 août-9 septembre 1899 : compte-rendu sténographique in-extensa, Tome troisième

M. PAINLEVÉ, lisant :

Mon cher ami,

Vous me demandez mon opinion sur le système Bertillon. Sur le fond de l'affaire, bien entendu, je me récusé. Je n'ai pas de lumières et je ne puis que m'en rapporter à ceux qui en ont plus que moi. Je ne suis pas pas non plus graphologue, et je n'ai pas le temps de vérifier les mesures.

Maintenant, si vous voulez seulement savoir si, dans les raisonnements où M. Bertillon applique le calcul des probabilités, cette application est correcte, je puis vous donner mon avis.

Prenez le premier de ces raisonnements, le plus compréhensible de tous. (*Figaro* du 25 août, page 5, colonne 1, lignes 57 à 112.)

Sur 13 mots redoublés correspondant à 26 coïncidences possibles, l'auteur constate 4 coïncidences réalisées. Evaluant à 0,2 la probabilité d'une coïncidence isolée, il conclut que celle de la réunion de 4 coïncidences est de 0,0016.

C'est faux.

0,0016, c'est la probabilité pour qu'il y ait 4 coïncidences sur 4. Celle pour qu'il y en ait 4 sur 26 est 400 fois plus grande, soit 0,7.

Cette erreur colossale rend suspect tout ce qui suit.

[Out of 26 possible coincidences, 4 are realised. Assuming the probability of a coincidence to be 0.2, Bertillon concludes that the probability of the event is  $(0.2)^4 = 0.0016$ .]

This is wrong. (...)

This colossal mistake makes everything that follows suspect.

Binomial distribution:

$$p(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

```
binom(26,0.2).pmf(4)
```

```
0.17649844729725295
```

```
np.sum([binom(26,0.2).pmf(k) for k in range(4,27)])
```

```
0.7931603480503098
```

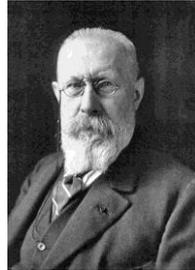
- **The prosecutor's fallacy:** "What would be the probability of [this unlikely event] if the accused were innocent? They have to be guilty."  
→ "Given that [this unlikely event] has happened, what is the probability that they are guilty?"
- Three mathematicians were mandated to write a report that shattered the "système Bertillon".



Henri Poincaré  
(1854-1912)



Gaston Darboux  
(1842-1917)



Paul Appell  
(1855-1930)

- ... Much more in this report (prior choice in "moral sciences", confirmation bias, look elsewhere effect).

Examen critique des divers systèmes ou études graphologiques auxquels a donné lieu le bordereau, Rapport de MM. Darboux, Appell & Poincaré, 1904

### NOTIONS SUR LA PROBABILITÉ DES CAUSES

Le problème qui nous occupe ici est de même nature : l'effet est connu, ce sont les coïncidences signalées sur le bordereau, et c'est la cause (forgerie ou écriture naturelle) qu'il s'agit de déterminer.

Ce sont donc les formules dites de probabilité des causes qu'il convient d'appliquer. Mais l'application de ces formules exige quelques précautions.

The problem that concerns us here is of the same nature: the effect is known, it is the coincidences noticed on the *bordereau*, and it is the cause (forgery or natural writing) that must be determined. It is therefore the so-called formulas for the probability of causes that should be applied. But the application of these formulas requires some care.

Pour pouvoir calculer, d'après un événement constaté, la probabilité d'une cause, il nous faut donc plusieurs données :

- 1° Il faut savoir quelle était *a priori*, avant l'événement, la probabilité de cette cause.

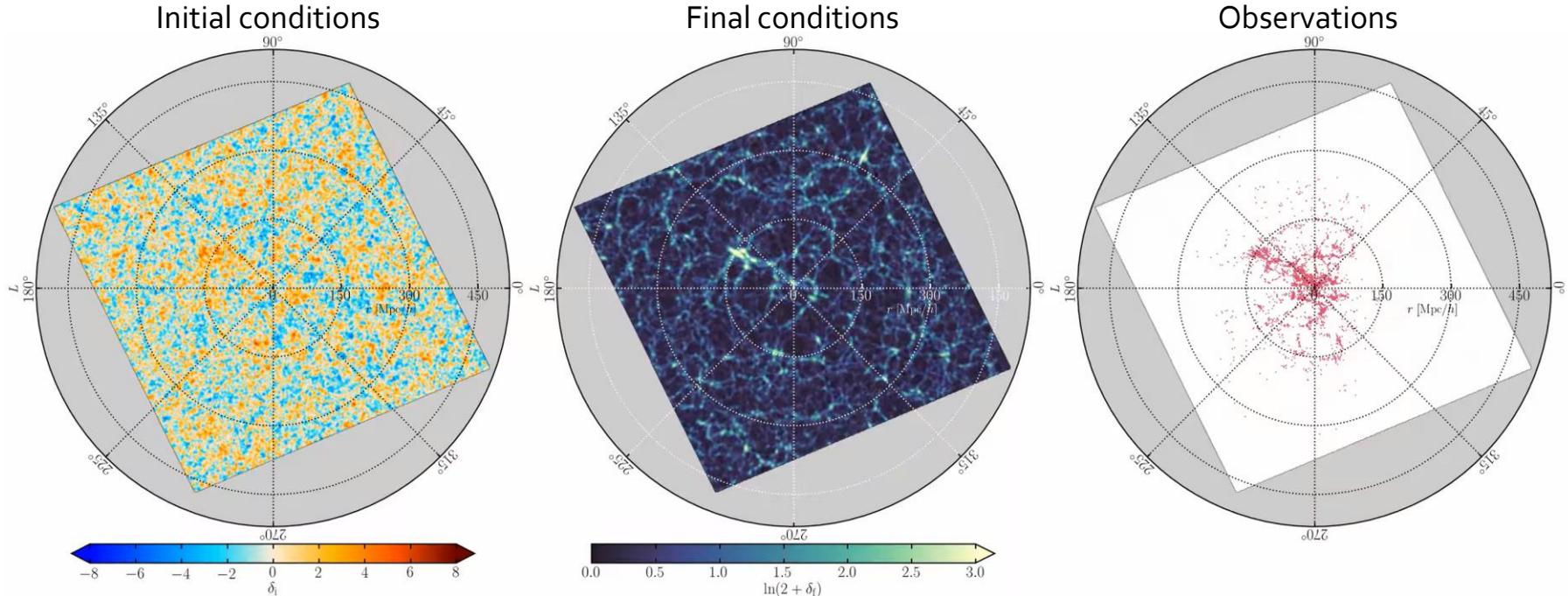
- 2° Il faut savoir ensuite quelle serait, pour chacune des causes possibles, la probabilité de l'événement constaté.

To be able to calculate, following an observed event, the probability of a cause, we need several data:

- 1- It is necessary to know what was *a priori*, before the event, the probability of this cause,

- 2- It is necessary to know what would be, for each of the possible causes, the probability of the observed event.

# Bayes at work: Bayesian Origin Reconstruction from Galaxies (BORG)



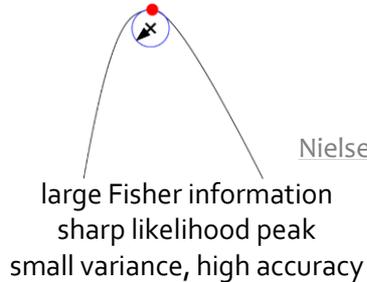
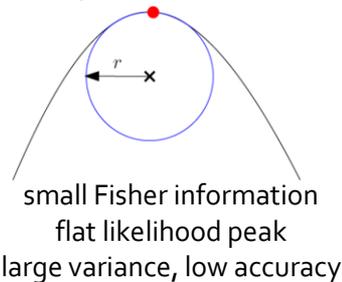
67,224 galaxies,  $\approx 17$  million parameters, 5 TB of primary data products, 10,000 samples,  $\approx 500,000$  forward and adjoint gradient data model evaluations, 1.5 million CPU-hours

Jasche & Wandelt, 1203.3639; Jasche, FL & Wandelt, 1409.6308; Jasche & Lavaux, 1806.1117; Lavaux, Jasche & FL, 1909.06396

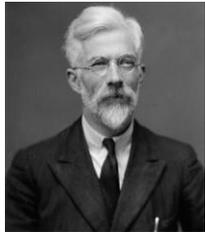


# What there is to learn: Notion of information in probability theory

- With Bayes' theorem we know *how* to learn. But how much *can* we learn? The **Fisher information** (1922) measures the amount of information that a random variable contains about an unknown parameter.



Nielsen, 1808.08271



Ronald Aylmer Fisher  
(1890-1962)



Prasanta Chandra  
Mahalanobis (1893-1972)



Calyampudi Radhakrishna  
Rao (1920-)

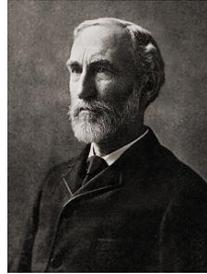
- Generalisations yield the field of **information geometry**:
  - The **Mahalanobis distance** (1927) measures the distance between a point and a distribution.
  - For a multi-dimensional problem, the Fisher information generalises to a matrix, and defines a metric: the **Fisher-Rao metric** (1945).

(...) I suggested the **differential geometric approach** in my 1945 paper by considering the space of probability distributions. I used Fisher information matrix in defining the metric, so it was called Fisher-Rao metric. Differential geometry was not well known at that time, and in order to compute the geodesic distance from the metric, I had to learn the mathematics from papers on **relativity describing Einstein metric**. It was only 30 years later, my work received attention (...).

Calyampudi Radhakrishna Rao, [Scholarpedia](#)

# Why proper statistics matter: Notion of entropy in probability theory

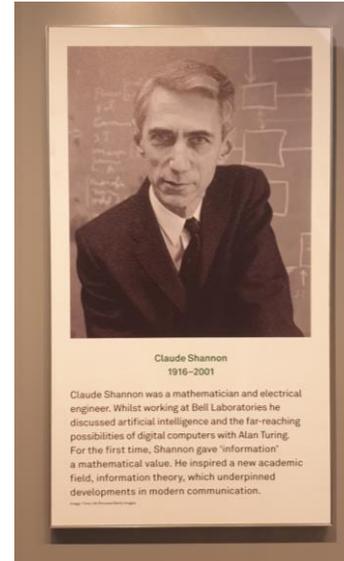
- 1884-1902: Gibbs' canonical and grand canonical ensembles, derived from the maximum entropy principle, fail to correctly predict thermodynamic properties of real physical systems.



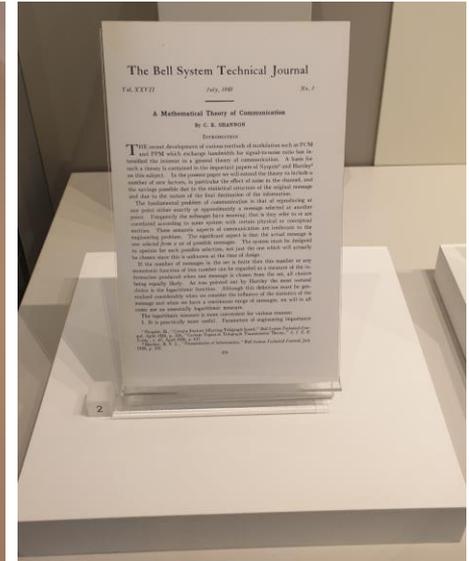
J. Willard Gibbs (1839-1903)

- The predicted entropies are always larger than the observed ones... there must exist additional microphysical constraints:
  - Discreteness of energy levels: radiation: Planck (1900), solids: Einstein (1907), Debye (1912), Ising (1925), individual atoms : Bohr (1913)...
  - ...Quantum mechanics: Heisenberg, Schrödinger (1927)
- The first clues indicating the need for quantum physics were uncovered by seemingly “unsuccessful” application of statistics!

- Information entropy (Shannon 1948)



Claude Shannon  
1916-2001  
Claude Shannon was a mathematician and electrical engineer. Whilst working at Bell Laboratories he discussed artificial intelligence and the far-reaching possibilities of digital computers with Alan Turing. For the first time, Shannon gave "information" a mathematical value. He inspired a new academic field, information theory, which underpinned developments in modern communication.



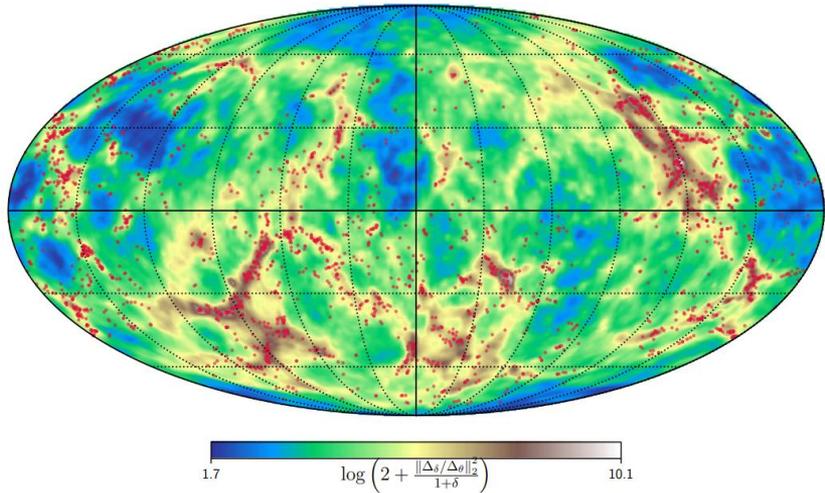
Pictures taken at the Science Museum, South Kensington, 2021

Why don't you call it entropy? In the first place, a mathematical development very much like yours already exists in Boltzmann's statistical mechanics, and in the second place, no one understands entropy very well, so in any discussion you will be in a position of advantage.  
von Neumann to Shannon, about a name for “missing information”



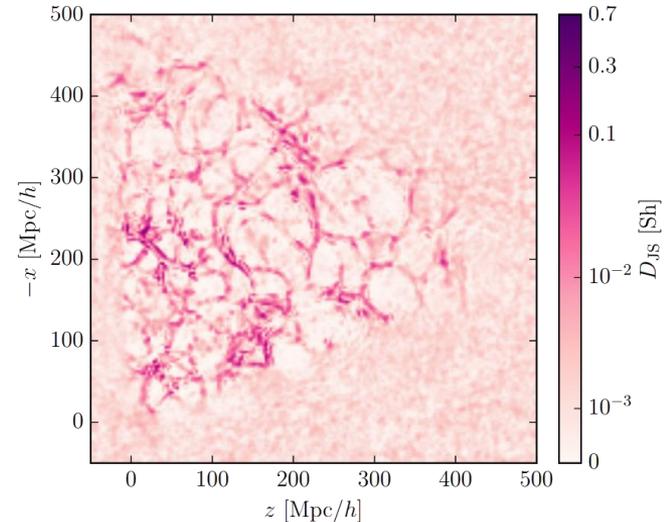
# Bayesian experimental design: Information-optimal or entropy-maximal acquisition of future cosmological data

- This is where to look if we want to measure cosmological parameters of  $\Lambda$ CDM...



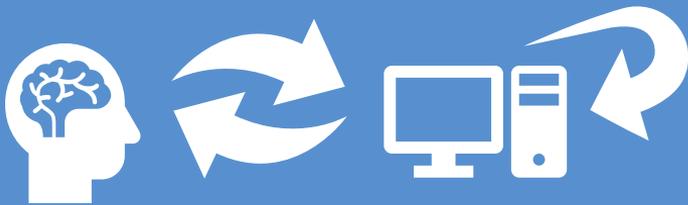
(Fisher information map for perturbative changes in the cosmological model)

- And this is where to look if we want to learn about dark energy...



(Jensen-Shannon divergence between cosmic web-type posteriors for different values of the dark energy equation of state)

The imitation problem:  
Algorithms beyond blind oracles?



## Some examples of AI in 2022...

- A binary political classifier:

C'est de gauche ou de droite ?

Quelque chose

Go

**La sécurité de l'emploi pour les jeunes chercheurs, c'est de gauche**

**Les postdocs courts et précaires, c'est de droite**

**La cosmologie, c'est de gauche**

**L'intelligence artificielle, c'est de droite**

**Le théorème de Bayes, c'est de gauche et de droite**

**La science des données, c'est de gauche**

**La science de l'information, c'est de droite**

D'accord



Pas d'accord



<https://degaucheoudedroite.delemazure.fr/>

- Text-conditional image generation:



"a cosmological simulation"



"an exascale supercomputer"



"the painting of a scientist among astronomical instruments"



Dall-E 2, <https://openai.com/dall-e-2/>, 2204.06125

## The birth of AI

- The Dartmouth Summer Research Project on Artificial Intelligence (1956).



Marvin Minsky, Claude Shannon, Ray Solomonoff and other scientists at the Dartmouth Summer Research Project on Artificial Intelligence (Photo: Margaret Minsky)

- The proposal (31 August 1955) states:

We propose that a 2-month, 10-man study of artificial intelligence be carried out during the summer of 1956 at Dartmouth College in Hanover, New Hampshire. The study is to proceed on the basis of the conjecture that ***every aspect of learning or any other feature of intelligence*** can in principle be so precisely described that ***a machine can be made to simulate it***. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer.

J. McCarthy, M. L. Minsky, N. Rochester, C. E. Shannon

Symbolic AI, explainable but costly  
Numerical AI/ML, automatic but “black-box”

- Humanity: classical theories of learning
  - Rule-based models, case-based reasoning (Aamodt & Plaza 1994)
  - Learning by practice, “chunking” (Newell & Rosenbloom 1981)
  - Reinforcement learning (Samuel 1959)
  - Non-supervised learning (Feigenbaum 1963), e.g. auto-encoders (Kramer 1991)
- Physiology: the brain
  - Artificial neuron (McCulloch & Pitts 1943), perceptron (Rosenblatt 1958)
  - Multi-layer perceptrons (Rumelhart et al. 1986, Rumelhart & McClelland 1987), gradient back-propagation (Rumelhart et al. 1986)
  - Deep learning & convolutional neural networks (LeCun et al. 2015, Goodfellow et al. 2016)
- Nature: evolution
  - Genetic algorithms (Holland 1975)
- Culture: epistemology
  - Scientific discovery (Langley et al. 1987)
  - Ontologies (Powers & Turk 1989), semantic web
- Physics: statistical mechanics, thermodynamics, quantum physics
  - Decision trees (Quinlan 1975), Bayesian networks, graphs
  - Hamiltonian Monte Carlo (Duane et al. 1987)
  - Information theory, distributed AI (Demazeau & Müller 1989)
  - Hidden Markov Models (Baum 1966)

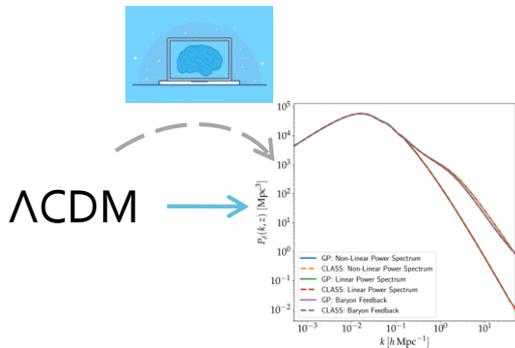
# Why machine learning for cosmology?



Last IAP conference organised by K. Benabed, G. Lavaux, H. J. McCracken

Speed up & go beyond approximations

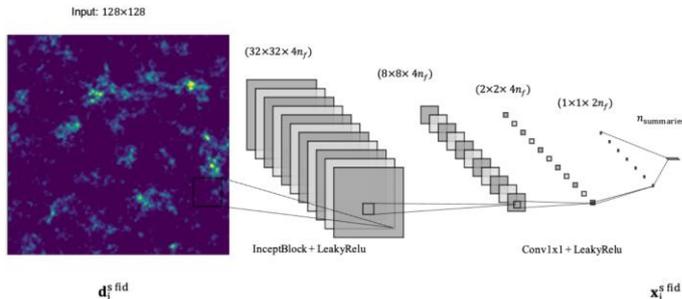
Emulators



emuPK: Mootoovaloo, Jaffe, Heavens & FL, 2105.02256

Find the information content

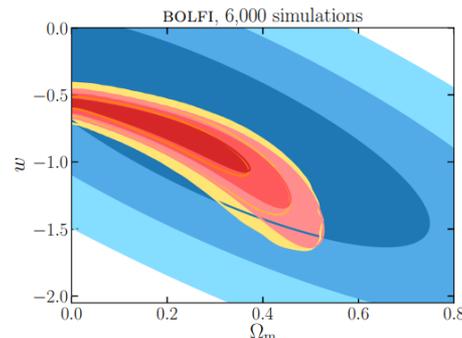
Automatic data compression



Information Maximising Neural Networks (IMNN): Charnock, Lavaux & Wandelt, 1802.03537; Makinen *et al.*, 2107.07405

Build a posterior/evidence approximator

Implicit likelihood inference



Bayesian Optimisation for Likelihood-Free Inference (BOLFI): FL, 1805.07152

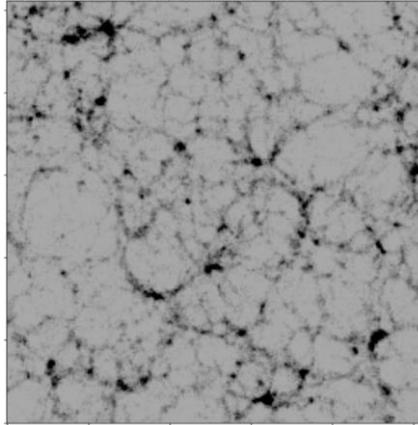


# What makes a good AI/ML model?

## Associative versus causal reasoning in scientific research

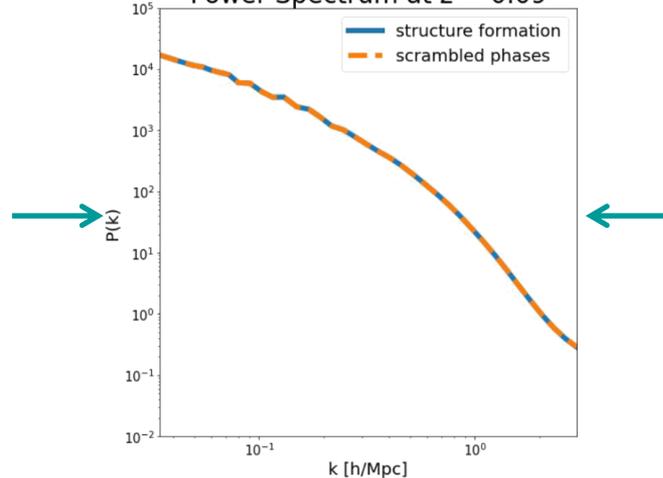
- With traditional machine learning, we obtain **associative links** between a latent space and data. But this doesn't mean we understand how nature works!

Structure formation at  $z = 0.09$

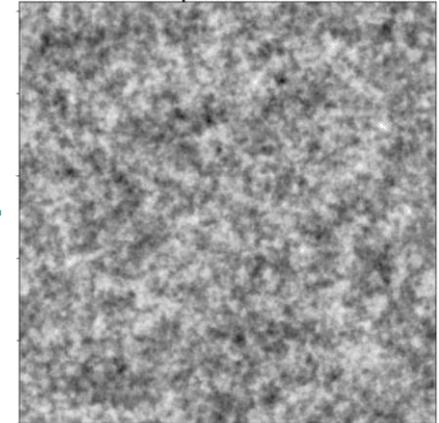


Process 1

Power Spectrum at  $z = 0.09$



Scrambled phases at  $z = 0.09$



Process 2

Credit: J. Jasche

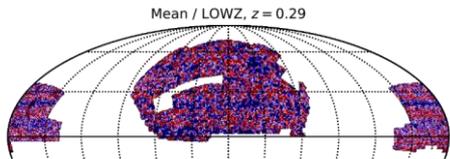
- Purely data-driven machine learning may not be sufficient for research!
  - Causally consistent generative AI/ML models produce **explanations**.
  - Predictive AI/ML models are assistance systems for **hypothesis generation**.



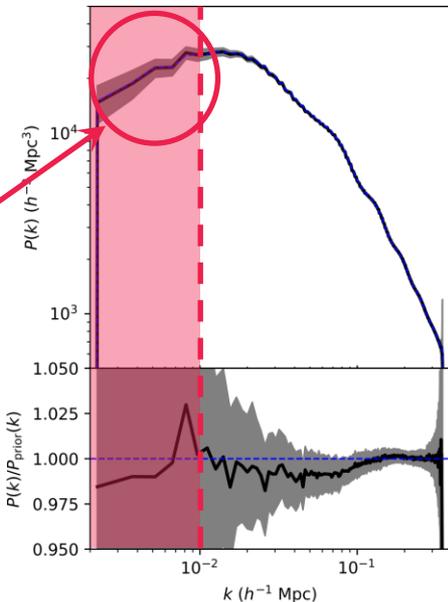
# Machine-aided report of unknown data contaminations

## Application to SDSS-III/BOSS (LOWZ+CMASS)

Map of unknown foreground contaminant



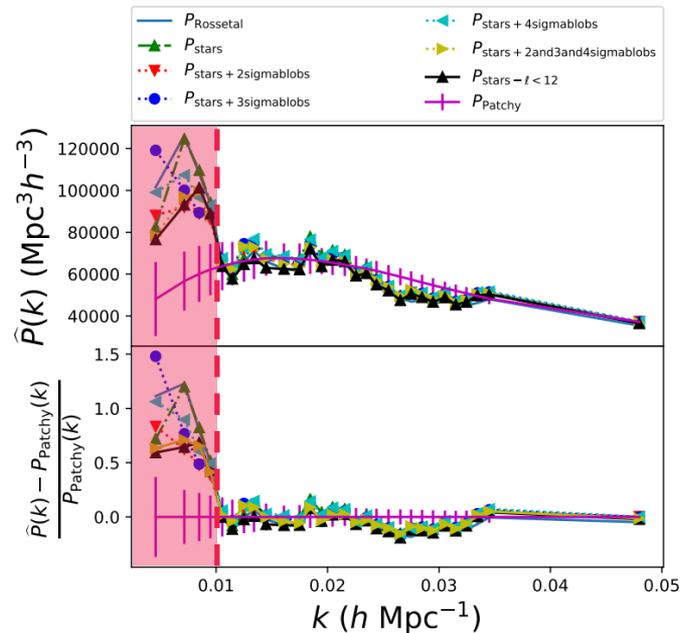
BORG *a posteriori* power spectrum



No apparent contamination, even well beyond the turn-over

Porqueres, Ramanah, Jasche & Lavaux, 1812.05113  
Lavaux, Jasche & FL, 1909.06396

State-of-the-art with backward-modelling technique (mode subtraction)



Kalus, Percival et al., 1806.02789



Conclusion:  
Hopes and challenges in information science for cosmology



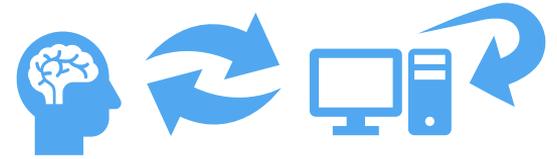
### The forward problem

- Hopes: Numerical models are the new way to formulate theory in data analysis.
- Challenges: Scalability (and energy cost!)



### The inverse problem

- Hopes: Bayesian data analysis is established as a fundamental theory of learning.
- Challenges: Control of external components in modern Bayesian models (in addition to likelihood and prior) : training data, posterior approximator...



### The imitation problem

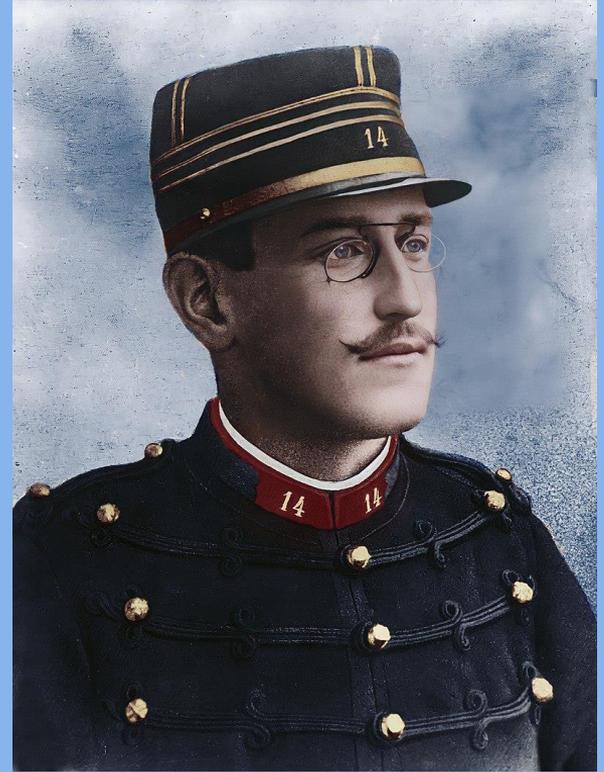
- Hopes: Machine-driven scientific discovery becomes conceivable.
- Challenges: Interpretability & explainability, proof & certifiability...

Some questions remain open...

Where does this elephant come from?!



Additional slides on the  
Dreyfus affair



## Mathematicians and probability theory in the Dreyfus affair: On prior choice in “moral sciences”

En résumé, les calculs de M. Bernard sont exacts ; ceux de M. Bertillon ne le sont pas. (*Mouvement.*)

Le seraient-ils qu'aucune conclusion ne serait pour cela légitime, parce que l'application du calcul des probabilités aux sciences morales est, comme l'a dit je ne sais plus qui, le scandale des mathématiques, parce que Laplace et Condorcet, qui calculaient bien, eux, sont arrivés à des résultats dénués de sens commun !

Or, cette probabilité *a priori*, dans des questions comme celle qui nous occupe, est uniquement formée d'éléments moraux qui échappent absolument au calcul, et si, comme nous venons de le voir, nous ne pouvons rien calculer sans la connaître, tout calcul devient impossible.

Aussi Auguste Comte a-t-il dit avec juste raison que l'application du calcul des probabilités aux sciences morales était le scandale des mathématiques.

Vouloir éliminer les éléments moraux et y substituer des chiffres, cela est aussi dangereux que vain.

En un mot, le calcul des probabilités n'est pas, comme on paraît le croire, une science merveilleuse qui dispense le savant d'avoir du bon sens.

C'est pourquoi *il faudrait s'abstenir absolument d'appliquer le calcul aux choses morales* ; si nous le faisons ici, c'est que nous y sommes contraints.

C'est des éléments moraux que doit dépendre le jugement, nous n'avons pas à en parler ici ; mais il est évident

Le procès Dreyfus devant le conseil de guerre de Rennes, 7 août-9 septembre 1899 : compte-rendu sténographique in-extensa, Tome troisième

Examen critique des divers systèmes ou études graphologiques auxquels a donné lieu le bordereau, Rapport de MM. Darboux, Appell & Poincaré, 1904



## Mathematicians and probability theory in the Dreyfus affair: On the look elsewhere effect

Outre les quatre coïncidences précitées, on en signale un grand nombre de nature différente, mettons dix mille ; mais il faudrait comparer ce nombre à celui des coïncidences *possibles*, c'est-à-dire de celles que l'auteur aurait comptées à son actif s'il les avait constatées. S'il y a 1,000 lettres dans le bordereau, cela fait 999,000 nombres, en comptant les différences des abscisses et celles des ordonnées. La probabilité pour que sur 999,000 nombres il y en ait 10,000 qui aient pu paraître « remarquables » à un chercheur aussi attentif que M. Bertillon, c'est presque la certitude.

Si l'on met en évidence certaines coïncidences, et qu'on montre qu'il y avait *a priori* peu de chances pour qu'elles se produisissent, avons-nous le droit d'en conclure qu'elles ne peuvent être l'effet du hasard ?

Si le n° 25 sort à la loterie, ce sera un événement dont la probabilité *a priori* était très faible, puisque les billets étaient fort nombreux ; mais cela ne veut pas dire que le tirage n'a pas été loyal, car il fallait bien qu'un numéro sortît ou un autre.

Ce n'est donc pas ainsi qu'il faut raisonner ; il ne s'agit pas de calculer la probabilité de telle ou telle coïncidence que vous choisissez précisément parce que vous l'avez constatée ; ce qu'il faut introduire, c'est la probabilité d'une coïncidence *quelconque* parmi celles que vous compléteriez à votre actif si elles se produisaient.

Supposons qu'il y ait 1.000 lettres dans le bordereau, avec les différences des abscisses et des ordonnées, cela fait 999.000 nombres ; qu'on trouve ensuite 10.000 coïncidences, y aura-t-il lieu de s'étonner ? La probabilité qu'il faudrait chercher, ce serait celle pour que sur 999.000 nombres, il y en eût 10.000 qui, après 10 ans de recherches, paraissent remarquables à un esprit aussi attentif que M. Bertillon ; c'est presque la certitude.

Nous avons montré, par l'application des règles du calcul des probabilités, que les coïncidences signalées, en ce qui concerne le repérage horizontal des polysyllabes, peuvent très bien s'expliquer par le hasard et ne prouvent nullement que le document ait été forgé. Ces coïncidences en effet ne sont qu'approchées.

Le procès Dreyfus devant le conseil de guerre de Rennes, 7 août-9 septembre 1899 : compte-rendu sténographique in-extensa, Tome troisième

Examen critique des divers systèmes ou études graphologiques auxquels a donné lieu le bordereau, Rapport de MM. Darboux, Appell & Poincaré, 1904



## Mathematicians and probability theory in the Dreyfus affair: On confirmation bias

Quant à l'application sur les mots des pièces de comparaison, notre examen prouve que M. Bertillon a fait un choix tout à fait arbitraire entre les mots très nombreux qu'il avait à sa disposition. C'est donc, ou qu'il avait le désir de prouver la culpabilité, ou, ce que nous croyons de préférence, qu'il a commis une grave erreur de méthode.

Ce que nous venons de dire suffit pour faire comprendre l'esprit de la « méthode » de M. Bertillon. Il l'a lui-même résumé d'un mot : « Quand on cherche, on trouve toujours. »

Quand une coïncidence est constatée, c'est une preuve accablante ; si elle fait défaut, c'est une preuve plus accablante encore, car cela prouve que le scripteur a cherché à détourner les soupçons.

On ne s'étonnera pas des résultats qu'il a obtenus par cette méthode. La naïveté avec laquelle il en a dévoilé les secrets porterait à croire à sa bonne foi.

Examen critique des divers systèmes ou études graphologiques auxquels a donné lieu le bordereau, Rapport de MM. Darboux, Appell & Poincaré, 1904



# Mathematicians and probability theory in the Dreyfus affair: On pseudoscientific statements

LE PRÉSIDENT. — Veuillez nous faire connaître le résultat de vos travaux sur le système de M. Bertillon.

M. PAINLEVÉ. — En tant que mathématicien, je me suis intéressé au système de M. Bertillon. Dès la première lecture, j'ai été frappé naturellement des erreurs de toutes sortes qui faussent ce système de fond en comble.

J'ai été frappé aussi du ton d'assurance absolue de M. Bertillon et de sa prétention d'introduire la certitude mathématique dans des questions qui ne sauraient la comporter à aucun degré.

En voyant cela, j'ai eu quelque peu d'inquiétude à la pensée que ce système, grâce à sa complication pseudo-scientifique, grâce à son ingéniosité apparente, grâce aussi au ton d'affirmation absolue, imperturbable de M. Bertillon, que ce système, dis-je, pourrait, quoique tout à fait erroné, influencer d'une façon quelconque sur l'esprit du Conseil.

Je crois donc utile de montrer très brièvement, mais d'une façon élatante, quelques-unes des erreurs essentielles de M. Bertillon.

Mais au lieu de faire moi-même cette démonstration, je crois préférable de communiquer au Conseil l'opinion de M. Poincaré, devant lequel je n'ai qu'à m'effacer.

Rien de tout cela n'a de caractère scientifique, et je ne puis comprendre vos inquiétudes. Je ne sais si l'accusé sera condamné, mais s'il l'est, ce sera sur d'autres preuves. Il est impossible qu'une pareille argumentation fasse quelque impression sur des hommes sans parti pris et qui ont reçu une éducation scientifique solide. (*Mouvement prolongé.*)

Votre bien dévoué,

H. POINCARÉ.

Le procès Dreyfus devant le conseil de guerre de Rennes, 7 août-9 septembre 1899 : compte-rendu sténographique in-extensa, Tome troisième

Mais à un certain moment, des hommes habiles comprirent quel parti on pouvait tirer de cette mine précieuse et inépuisable d'équivoques. Ils savaient que les rieurs se lassent et que les croyants ne se lassent pas ; ils savaient que le public ne fait pas attention à la valeur des arguments, mais au ton des argumentateurs ; et ils commencèrent à soutenir M. Bertillon de leurs affirmations tranchantes et réitérées.

Or, celui-ci avait un grand avantage ; l'obscurité de son système le défendait contre la critique, de même que la seiche s'entoure d'un nuage d'encre pour échapper à ses ennemis.

Son système variait constamment, et par là il pouvait prolonger la discussion, et cela pouvait faire illusion au public qui avait depuis longtemps renoncé à comprendre.

Il n'y a pas d'inventeur de la quadrature du cercle qui ne soit prêt à prolonger la résistance indéfiniment, du moment qu'on accepte de discuter avec lui.

Les arguments, que M. Bertillon a tirés de l'application des mots du bordereau les uns sur les autres ou sur ceux de la lettre du buvard, ont été réduits à leur valeur réelle, qui est nulle.

Examen critique des divers systèmes ou études graphologiques auxquels a donné lieu le bordereau, Rapport de MM. Darboux, Appell & Poincaré, 1904

# Mathematicians and probability theory in the Dreyfus affair: On pseudoscientific statements

S'il s'agissait d'un travail scientifique, nous nous arrêtons là ; nous jugerions inutile d'examiner les détails d'un système dont le principe même ne peut soutenir l'examen ; mais la Cour nous a confié une mission que nous devons accomplir jusqu'au bout.

## CONCLUSION

L'absurdité du système Bertillon est si évidente qu'on s'expliquera difficilement la longueur de cette discussion. On risquerait de n'en pas comprendre la nécessité, si on ne se rappelait l'histoire de l'affaire.

En résumé, tous ces systèmes sont absolument dépourvus de toute valeur scientifique :

- 1° Parce que l'application du calcul des probabilités à ces matières n'est pas légitime ;
- 2° Parce que la reconstitution du bordereau est fautive ;
- 3° Parce que les règles du calcul des probabilités n'ont pas été correctement appliquées.

En un mot, parce que les auteurs ont raisonné mal sur des documents faux.

Signé : POINCARÉ, DARBOUX, APPELL.

Examen critique des divers systèmes ou études graphologiques auxquels a donné lieu le bordereau, Rapport de MM. Darboux, Appell & Poincaré, 1904

Du 12 Juillet 1906.

LA COUR, CHAMBRES RÉUNIES,

Attendu, toutefois, qu'à l'appui de l'accusation les principaux témoins à charge se sont fondés sur un travail de l'expert Bertillon, prétendant démontrer géométriquement et à l'aide du calcul des probabilités que le bordereau était un document truqué, forgé par Dreyfus ; que celui-ci, usant d'un gabarit placé sous le papier pelure,

Attendu que, les études graphologiques de Bertillon et autres devant, par suite, être éliminées du débat, il reste acquis que le bordereau a été écrit par Esterhazy et non par Dreyfus ;

Attendu qu'ainsi, au point de vue soit de l'écriture, soit du texte, l'accusation, dont le bordereau était la base légale, est entièrement injustifiée ;

Bulletin des arrêts de la Cour de Cassation rendus en matière criminelle, tome CXI, N°1, janvier 1906



# Mathematicians and probability theory in the Dreyfus affair: On the opinion of experts

M. PAINLEVÉ. — Comme le nom de M. Poincaré peut prêter à confusion, je spécifie : il est question de M. Henri Poincaré, membre de l'Académie des sciences.

Tous les membres du Conseil connaissent le nom de M. Poincaré, une des gloires de l'École polytechnique. M. Poincaré est le plus illustre des mathématiciens contemporains. Il n'y a pas un pays civilisé où les savants ne soient prêts à s'incliner avec admiration et respect devant l'autorité de M. Poincaré.

J'ajoute que M. Poincaré, dont l'activité et la compétence ont embrassé tout le champ des sciences rationnelles, a été pendant plus de dix ans professeur de calcul des probabilités à la Sorbonne.

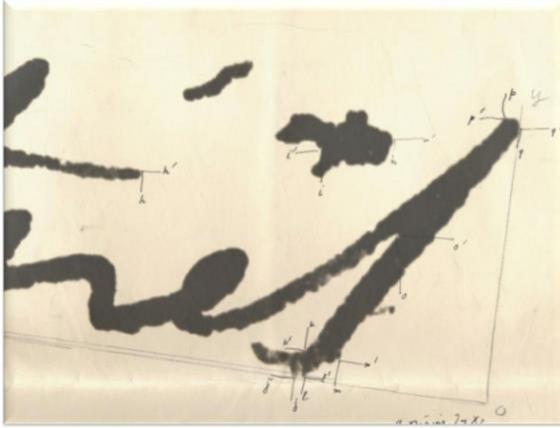
Attendu que, par ordonnance du 18 avril 1904, le président de la chambre criminelle a commis les membres de l'institut Darboux, secrétaire perpétuel de l'Académie des sciences, Appell, doyen de la Faculté des sciences de Paris, et Poincaré, professeur à la même faculté, pour examiner, en provoquant toutes précisions et explications de la part de leurs auteurs, les études graphologiques de Bertillon,

Le procès Dreyfus devant le conseil de guerre de Rennes, 7 août-9 septembre 1899 : compte-rendu sténographique in-extensa, Tome troisième

Bulletin des arrêts de la Cour de Cassation rendus en matière criminelle, tome CXI, N°1, janvier 1906



## Astronomers in the Dreyfus affair: The macro-micromètre of the Paris Observatory



**AN UNUSUAL USE OF AN ASTRONOMICAL INSTRUMENT:  
THE DREYFUS AFFAIR AND THE PARIS 'MACRO-MICROMÈTRE'**

SUZANNE DÉBARBAT, Observatoire de Paris

The year 1994 marked the centenary of what was called in France “L’Affaire Dreyfus”<sup>1</sup> and of what was, around the beginning of the present century, a decade of difficulties for the French Government and its army. The year 1904 saw the beginning of the end of these troubles. These circumstances led the Library of the Paris Observatory to arrange a display of documents discovered in the notebooks of astronomers, documents that show that a scientific instrument designed to measure astronomical plates was — surprisingly — used at that time for non-scientific purposes.

Débarbat 1996, *Journal for the History of Astronomy*, 27, 1, 45

La reconstitution du bordereau est fausse, les erreurs sont généralement d'un demi-millimètre et peuvent aller jusqu'à un ou deux millimètres ; c'est de cette reconstitution fautive que M. Bertillon s'est uniquement servi.

Ses planches sont le résultat d'un traitement compliqué infligé au document primitif et d'où celui-ci est sorti altéré. Il a subi une série d'agrandissements et de réductions photographiques et même de calquages, décalquages, recalquages, découpages, collages, gouachages, badigeonnages et retouches.

Examen critique des divers systèmes ou études graphologiques auxquels a donné lieu le bordereau, Rapport de MM. Darboux, Appell & Poincaré, 1904

Attendu que Bertillon a édité son système, non d'après l'original du bordereau, mais d'après un document artificiel, le bordereau reconstitué par lui ;

Attendu que les trois experts ont dressé, à l'unanimité, un rapport dans lequel ils établissent que la reconstitution du bordereau, effectuée par Bertillon est fautive, que « ses planches sont le résultat d'un traitement compliqué, infligé au document primitif, et d'où celui-ci est sorti altéré, après avoir subi une série d'agrandissements et de réductions photographiques et même de calquages, recalquages, découpages, collages, gouachages, badigeonnages et retouches » ;

Bulletin des arrêts de la Cour de Cassation rendus en matière criminelle, tome CXI, N°1, janvier 1906

# A bit of history: the Dreyfus affair

## Epilogue

Du 12 Juillet 1906.

LA COUR, CHAMBRES RÉUNIES,

Où aux audiences publiques des 18, 19, 20, 21, 22, 25, 26, 27, 28 et 30 juin, 2, 3, 5, 6 et 7 juillet, M. le conseiller Moras, en son rapport, M. le procureur général Baudouin, en ses réquisitions, et M<sup>e</sup> Mornard, avocat d'Alfred Dreyfus, intervenant, dans ses conclusions;

Vu la lettre du 25 décembre 1903 par laquelle le Garde des sceaux, Ministre de la justice, a, en vertu des articles 443, § 4, et 444 du Code d'instruction criminelle, chargé le procureur général près la Cour de déléger à la chambre criminelle le jugement du conseil de guerre de Rennes qui, le 9 septembre 1899, a condamné Alfred Dreyfus à dix ans de détention et à la dégradation militaire, par application des articles 76 et 463 du Code pénal et 1<sup>er</sup> de la loi du 8 juin 1850;

Attendu que, les études graphologiques de Bertillon et autres devant, par suite, être éliminées du débat, il reste acquis que le bordereau a été écrit par Esterhazy et non par Dreyfus;

Attendu qu'ainsi, au point de vue soit de l'écriture, soit du texte, l'accusation, dont le bordereau était la base légale, est entièrement injustifiée;

Attendu, en dernière analyse, que de l'accusation portée contre Dreyfus rien ne reste debout;

Et que l'annulation du jugement du conseil de guerre ne laisse rien subsister qui puisse à sa charge être qualifié crime ou délit;

Attendu dès lors, que, par application du paragraphe final de l'article 445, aucun renvoi ne doit être prononcé;

Par ces motifs,

ANNULE le jugement du conseil de guerre de Rennes qui, le 9 septembre 1899, a condamné Dreyfus à dix ans de détention et à la dégradation militaire, par application des articles 76 et 463 du Code pénal et 1<sup>er</sup> de la loi du 8 juin 1850;

DIT que c'est par erreur et à tort que cette condamnation a été prononcée;

DONNE ACTE à Dreyfus de ce qu'il déclare renoncer à demander l'indemnité pécuniaire que l'article 446 du Code d'instruction criminelle permettait de lui allouer;

ORDONNE qu'en conformité de cet article le présent arrêt sera affiché à Paris et à Rennes et sera inséré au *Journal officiel*, ainsi que dans cinq journaux, au choix de Dreyfus;

AUTORISE Dreyfus à le faire publier aux frais du Trésor et au taux des insertions légales dans cinquante journaux de Paris et de province, à son choix;

ORDONNE que l'arrêt sera transcrit sur les registres du conseil de guerre de Rennes et que mention en sera faite en marge de la décision annulée.

Ainsi jugé et prononcé, etc. — Chambre criminelle.