

Recherche Reproductible

Etat des lieux

8 & 9 Mars 2023, Institut Pasteur

Un nombre croissant de chercheurs s'intéresse à la problématique de la reproductibilité, dont la définition même peut grandement varier d'une discipline à une autre (observationnelle, expérimentale, statistique, computationnelle, etc). Or, nous avons rarement l'occasion de poser un regard inter-disciplinaire sur nos approches et définitions respectives. Ce workshop se veut donc un lieu d'échanges et d'information pour dresser un premier état des lieux de la reproductibilité en France.

S. Cohen Boulakia, A. Legrand, F. Lemoine, N. Rougier



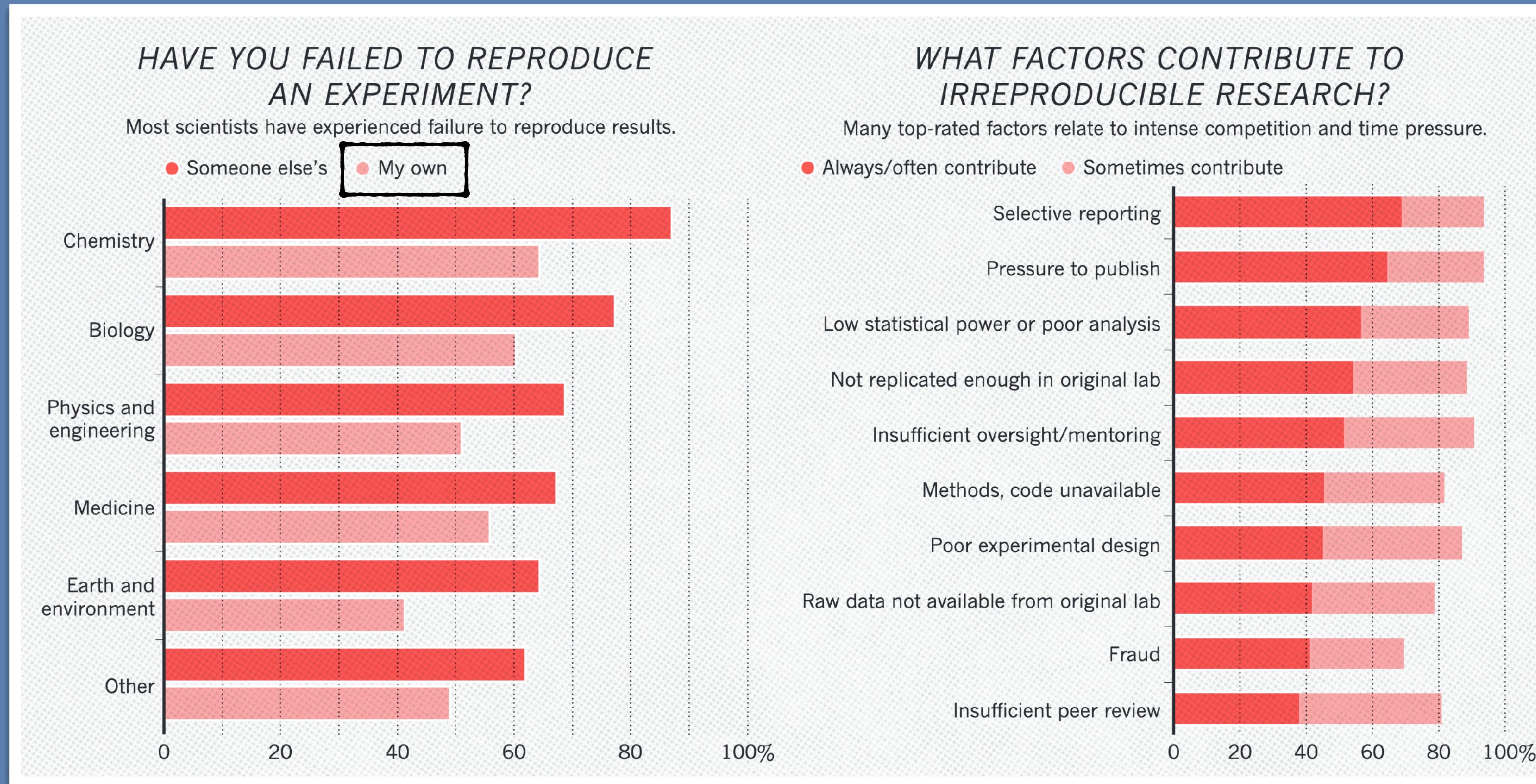
Réseau de la recherche reproductible
<http://www.recherche-reproducible.fr>



"Is there a reproducibility crisis?

A Nature survey

(1576 researchers, 2016)



COMMENT

AVIAN INFLUENZA Shift expertise to track mutations where they emerge p.534

EARTH SYSTEMS Past climates give valuable clues to future warming p.537

HISTORY OF SCIENCE Descartes' lost letter tracked using Google p.540

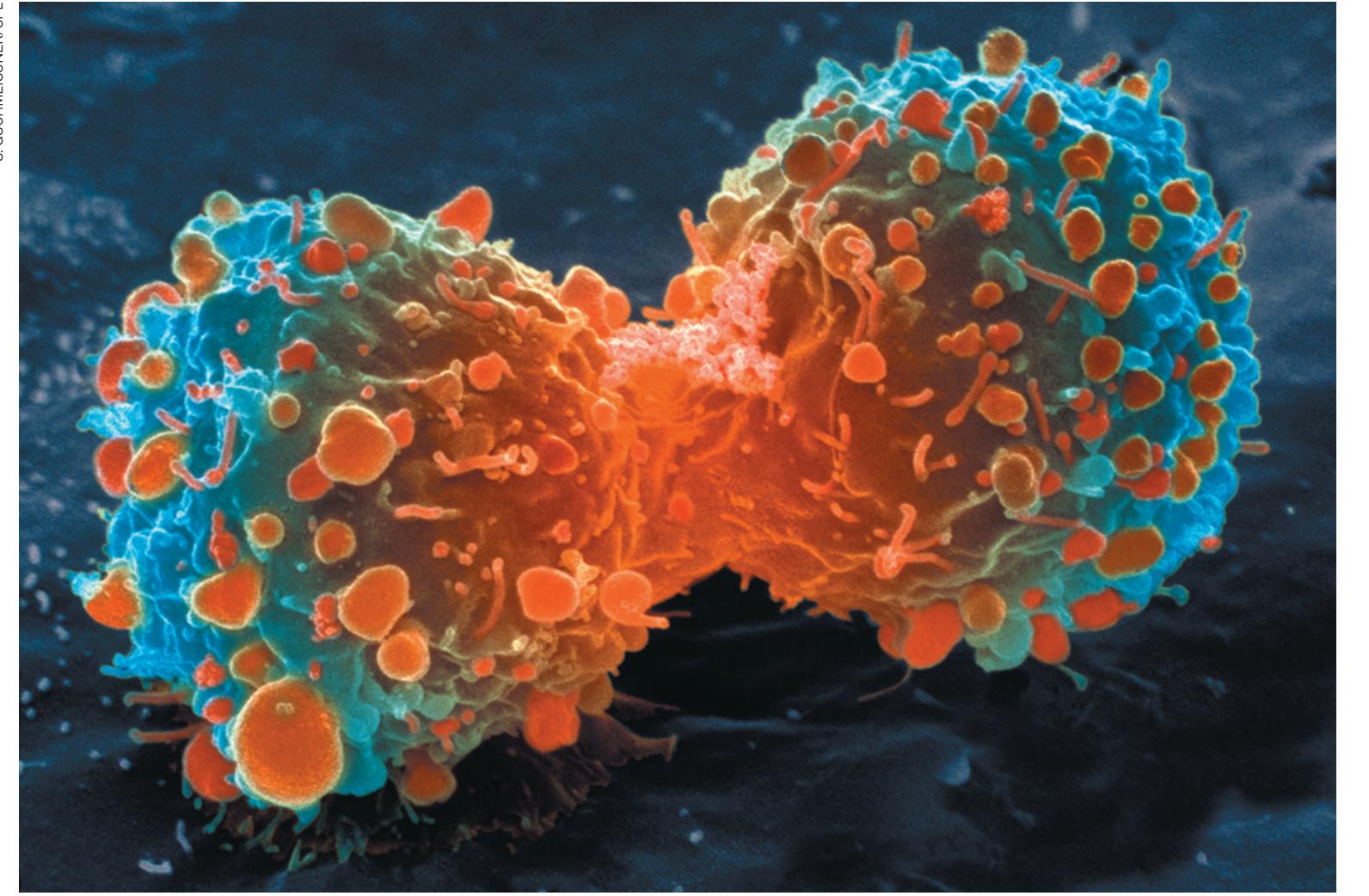
OBITUARY Wylie Vale and an elusive stress hormone p.542



If a job is worth doing, it's worth doing twice

Researchers and funding agency need to put a premium on ensuring that results are reproducible argues **Jonathan F. Russel**

The case for open computer programs



Many landmark findings in preclinical oncology research are not reproducible, in part because of inadequate cell lines and animal models.

Raise standards for preclinical cancer research

C. Glenn Begley and Lee M. Ellis propose how methods, publications and incentives must change if patients are to benefit.

Efforts over the past decade to characterize the genetic alterations in human cancers have led to a better understanding of molecular drivers of this complex set of diseases. Although we in the cancer field hoped that this would lead to more effective drugs, historically, our ability to translate cancer research to clinical success has been remarkably low¹. Sadly, clinical

trials in oncology have the highest failure rate compared with other therapeutic areas. Given the high unmet need in oncology, it is understandable that barriers to clinical development may be lower than for other disease areas, and a larger number of drugs with suboptimal preclinical validation will enter oncology trials. However, this low success rate is not sustainable or acceptable, and

investigators must reassess their approach to translating discovery research into greater clinical success and impact.

Many factors are responsible for the high failure rate, notwithstanding the inherently difficult nature of this disease. Certainly, the limitations of preclinical tools such as inadequate cancer-cell-line and mouse models² make it difficult for even ▶

Must try harder

Too many sloppy mistakes are creeping into scientific papers at the data — and at themselves.

Six red flags for suspect work

C. Glenn Begley explains how to recognize the preclinical papers in which the data won't stand up.

Error prone

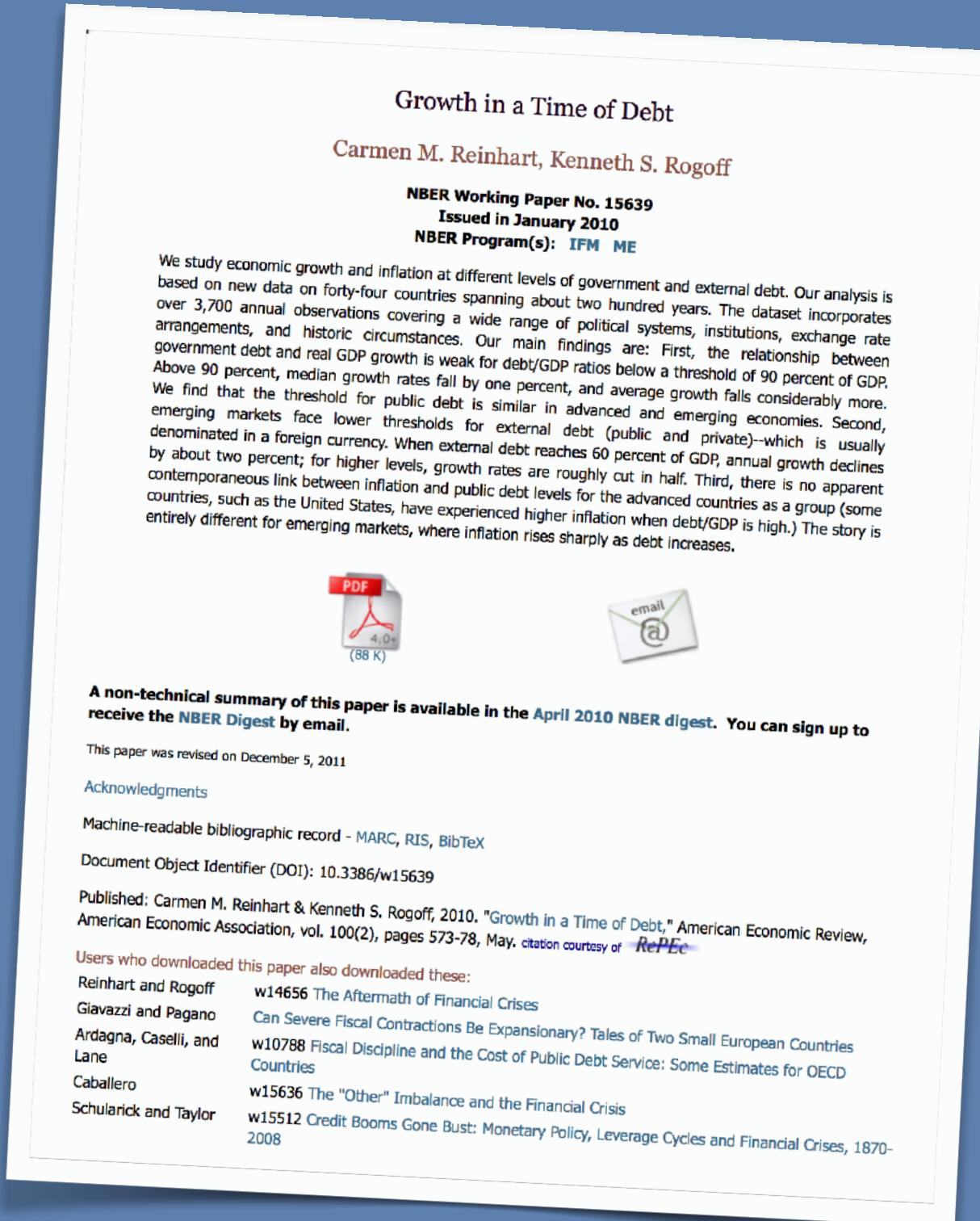
Biologists must realize the pitfalls of massive amount of data

Know when your numbers are significant

- Nekrutenko & Taylor, *Nature Genetics* (2012)
- Alsheikh-Ali et al. *PLoS ONE* (2011)
- Begley & Ellis *Nature* (2012)

"The Excel error that changed history

Coy, Bloomberg, 2013



Harvard University economists C. Reinhart and K. Rogoff have acknowledged making a spreadsheet calculation mistake in a 2010 research paper which has been widely cited to justify budget-cutting. But the authors stand by their conclusion that higher government debt is associated with slower economic growth.

Reinhart and Rogoff's work showed average real economic growth slows (a 0.1% decline) when a country's debt rises to more than 90% of gross domestic product (GDP) – and this 90% figure was employed repeatedly in political arguments over high-profile austerity measures. When that error was corrected, the 0.1% decline data became a 2.2% average increase in economic growth.

Questions de Vocabulaire

• Empirical reproducibility

- Detailed information about non-computational empirical scientific experiments and observations
- In practice this is enabled by making data freely available, as well as details of how the data was collected.

Rerun

Redo

Reproduce

Repeat

Replicate

Reuse

Register

Report

• Statistical reproducibility

- Detailed information about the choice of statistical tests, model parameters, threshold values, etc.
- This relates to pre-registration of study design to prevent p-value hacking and other manipulations.

Repeat
Same experiment
Same laboratory

Replicate
Same experiment
Different laboratory

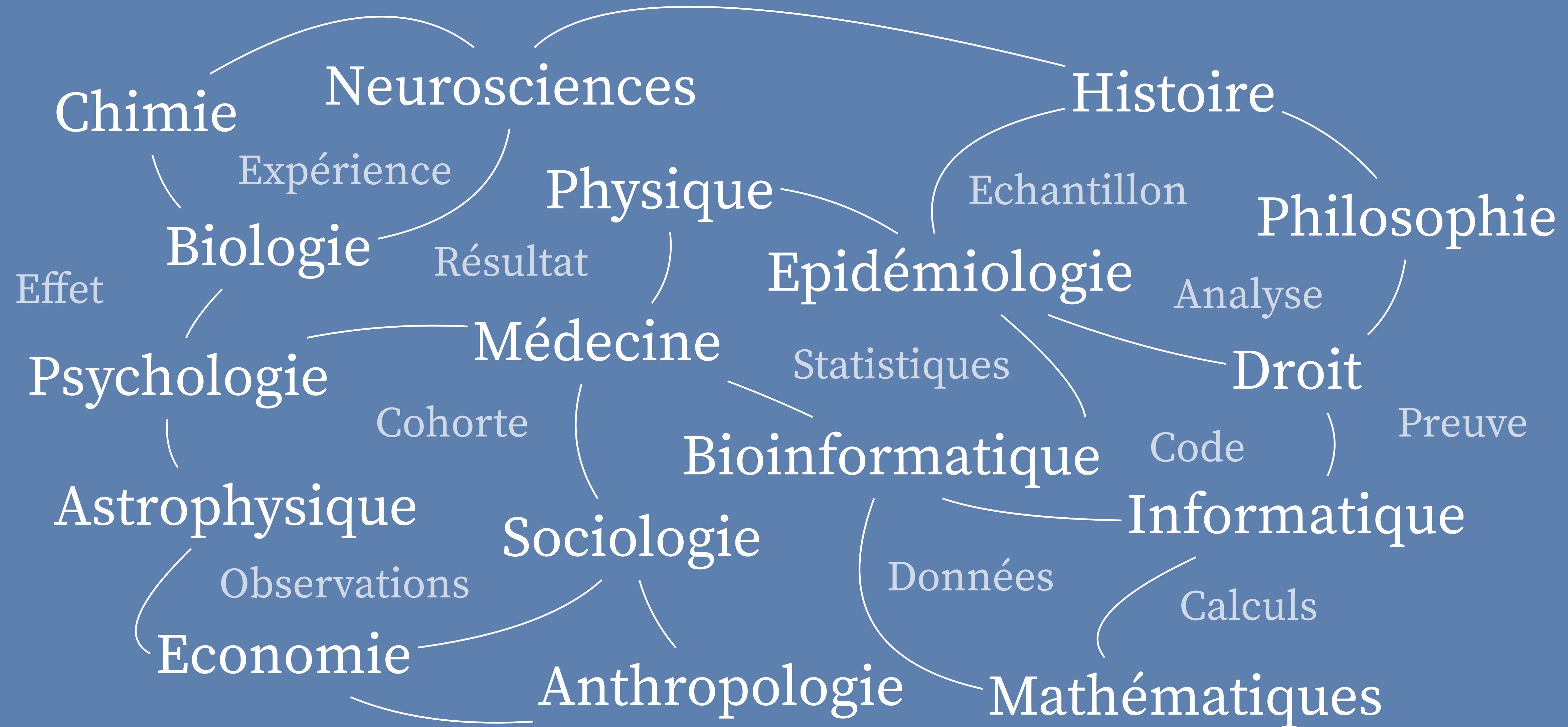
• Computational reproducibility

- Detailed information about code, software, hardware and implementation detail
→ Goal: document how data has been produced

Reproduce
Same experiment
Different setup

Reuse
Same experiment
Different use

Questions de Pratiques



Questions d'Education

MOOC

*Recherche reproductible : principes méthodologiques
pour une science transparente*

ReproducibiliTea

Journal club for Open Science

ReproHackathon

Série de hackathons pour la reproductibilité
Une action MaDICS ReproVirtuFlow

Vers une recherche reproductible

Faire évoluer ses pratiques

Benchopt: Reproducible, efficient and
collaborative optimization benchmarks

Ten Years Reproducibility Challenge

Would you dare to run the code from your past self?

Programme des journées

Mercredi 8 mars 2023

14h30 - 15h30 **Reproductibilité observationnelle**

Alexandre Hocquet, Timothée Giraud, Thomas Vuillaume

15h30 - 16h30 **Reproductibilité computationnelle**

Konrad Hinsen, Roberto Di Cosmo, Camille Maumet,
Thomas Moreau

17h00 - 18h00 **Reproductibilité statistique**

Nelle Varoquaux, Christophe Hurlin, Isabelle Boutron

18h00 - 19h00 **Discussion générale**

19h00 - 21h00 **Cocktail à l’Institut**

Jeudi 9 mars 2023

9h30 - 10h30 **Reproductibilité expérimentale**

François-Xavier Coudert, Florian Naudet,
François Ric, Yoann Lafon

10h30 - 11h30 **Formation et enseignement**

Frédéric Lemoine, Eduarda Centeno, Fjola Hysseni,
Laurent Oudre, Christophe Pouzat

11h30 - 12h00 **Discussion générale**

12h00 - 13h00 **Déjeuner**

13h00 - 14h00 **Europe et international**

Etienne Roesch, Enrico Glerean, Isabelle Blanc

14h00- 15h00 **Organisation, finances**

15h00 - 15h30 **Discussion générale**