université de BORDEAUX

Laboratoire de Psychologie, UR 4139

La reproductibilité: Qu'est-ce que c'est? Quelles implications?

François Ric (LabPsy, UR 4139, U. Bordeaux)

Reproductibilité: Peut-on reproduire les résultats d'une recherche (d'une analyse: i.e. reproductibilité statistique/computationnelle) en suivant la même procédure?

Elément central de la démarche scientifique:

- Avancement théorique
- Applications (e.g., traitements)
- Investissement (temps et argent: estimation coût 28 Md$/an aux USA; Freedman, Cockburn & Simcoe, 2015)

Reproductibilité = fiabilité

De nombreux résultats difficiles à reproduire

Exposition à l'argent (Caruso, Vohs, Baxter, and Waytz, 2013)

Effet Mozart (Rauscher, Shaw, & Ky, 1993)

Ego-depletion (Baumeister, Bratslavsky, Muraven, & Tice, 1998)

Pré-cognition (Bem, 2011)

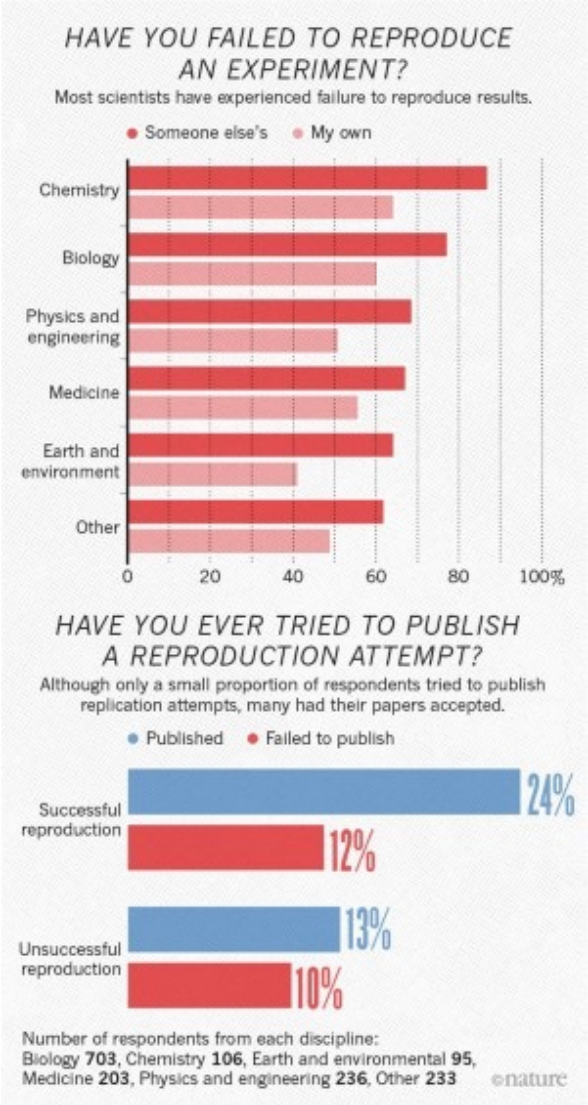
…

Les données de la science sont-elles fiables?

Peut-on les reproduire?

- Psychologie: 100 études, 36% Open Science Collaboration project (2015)
- Sciences sociales, 21 études publiées dans *Nature* et *Science* entre 2010 et 2015, 62% (Camerer et al., 2018)
- Economie: 18 études, 66% (Camerer et al. 2016)
- Essais cliniques cancer: 53 études, 11% (Begley & Ellis, 2012)
- Biologie du cancer: 50 études, 46% (Errington et al., 2021)

(…)


HAVE YOU FAILED TO REPRODUCE AN EXPERIMENT?
Most scientists have experienced failure to reproduce results.
Someone else's
My own
Chemistry
Biology
Physics and engineering
Medicine
Earth and environment
Other
0
20
40
60
80
100%
HAVE YOU EVER TRIED TO PUBLISH A REPRODUCTION ATTEMPT?
Although only a small proportion of respondents tried to publish replication attempts, many had their papers accepted.
Published
Failed to publish
Successful reproduction
24%
12%
Unsuccessful reproduction
13%
10%
Number of respondents from each discipline:
Biology 703, Chemistry 106, Earth and environmental 95, Medicine 203, Physics and engineering 236, Other 233
©nature

Nature, 2016

Pourquoi?

Fraude ?

Elle existe dans de nombreux domaines…

mais semble marginale…on l'espère.

NEWS

SCIENTIFIC INTEGRITY

Study reveals industrial-scale publishing fraud

Sophisticated global networks are infiltrating journals to publish fake papers CATHLEEN O'GRADY

Science, 2025

Pourquoi?

Open access, freely available online

Essay

Why Most Published Research Findings Are False

John P. A. Ioannidis

False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant

Joseph P. Simmons[1], Leif D. Nelson[2], and Uri Simonsohn[1]
[1]The Wharton School, University of Pennsylvania, and [2]Haas School of Business, University of California, Berkeley

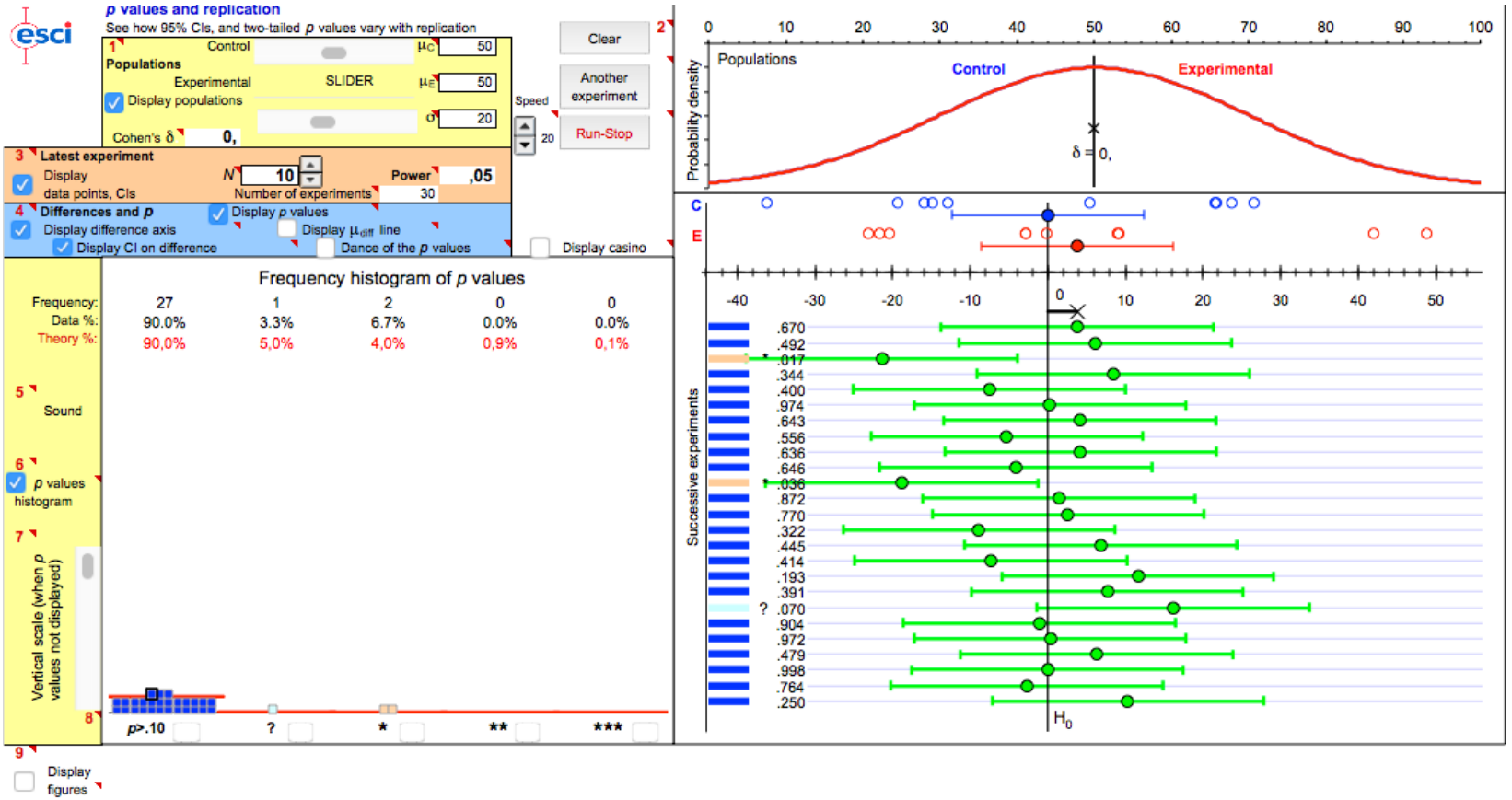
Journal of Abnormal and Social Psychology
1962, Vol. 65, No. 3, 145–153

THE STATISTICAL POWER OF ABNORMAL–SOCIAL PSYCHOLOGICAL RESEARCH: A REVIEW [1]

JACOB COHEN
New York University

Facteurs de non-réplicabilité

- Faible taille d’échantillons (instabilité; file drawer)


esci
p values and replication
See how 95% CIs, and two-tailed p values vary with replication
1
Populations
Control
Experimental
Display populations
SLIDER
μC 50
μE 50
σ 20
Cohen's δ 0,
Clear
Another experiment
Speed
20
Run-Stop
2
3 Latest experiment
Display data points, CIs
N 10
Number of experiments 30
Power ,05
4 Differences and p
Display difference axis
Display CI on difference
Display p values
Display μdiff line
Dance of the p values
Display casino
Frequency histogram of p values
Frequency: 27 1 2 0 0
Data %: 90.0% 3.3% 6.7% 0.0% 0.0%
Theory %: 90,0% 5,0% 4,0% 0,9% 0,1%
5 Sound
6 p values histogram
7 Vertical scale (when p values not displayed)
8
p>.10 ? * ** ***
9 Display figures
Probability density
Populations
Control
Experimental
δ = 0,
C
E
Successive experiments
.670
.492
.017
.344
.400
.974
.643
.556
.636
.646
.036
.872
.770
.322
.445
.414
.193
.391
.070
.904
.972
.479
.998
.764
.250
H0

Facteurs de non-réplicabilité

- Faible taille d'échantillons (instabilité; file drawer)
- Qualité méthode faible (peu de contrôle)
- Biais de confirmation
- Compétences (méthodologiques, statistiques, …)

➔ Réduire les « degrés de liberté » du chercheur

Study 2: musical contrast and chronological rejuvenation

Using the same method as in Study 1, we asked 20 University of Pennsylvania undergraduates to listen to either "When I'm Sixty-Four" by The Beatles or "Kalimba." Then, in an ostensibly unrelated task, they indicated their birth date (mm/dd/yyyy) and their father's age. We used father's age to control for variation in baseline age across participants.

An ANCOVA revealed the predicted effect: According to their birth dates, people were nearly a year-and-a-half younger after listening to "When I'm Sixty-Four" (adjusted $M = 20.1$ years) rather than to "Kalimba" (adjusted $M = 21.5$ years), $F(1, 17) = 4.92$, $p = .040$.

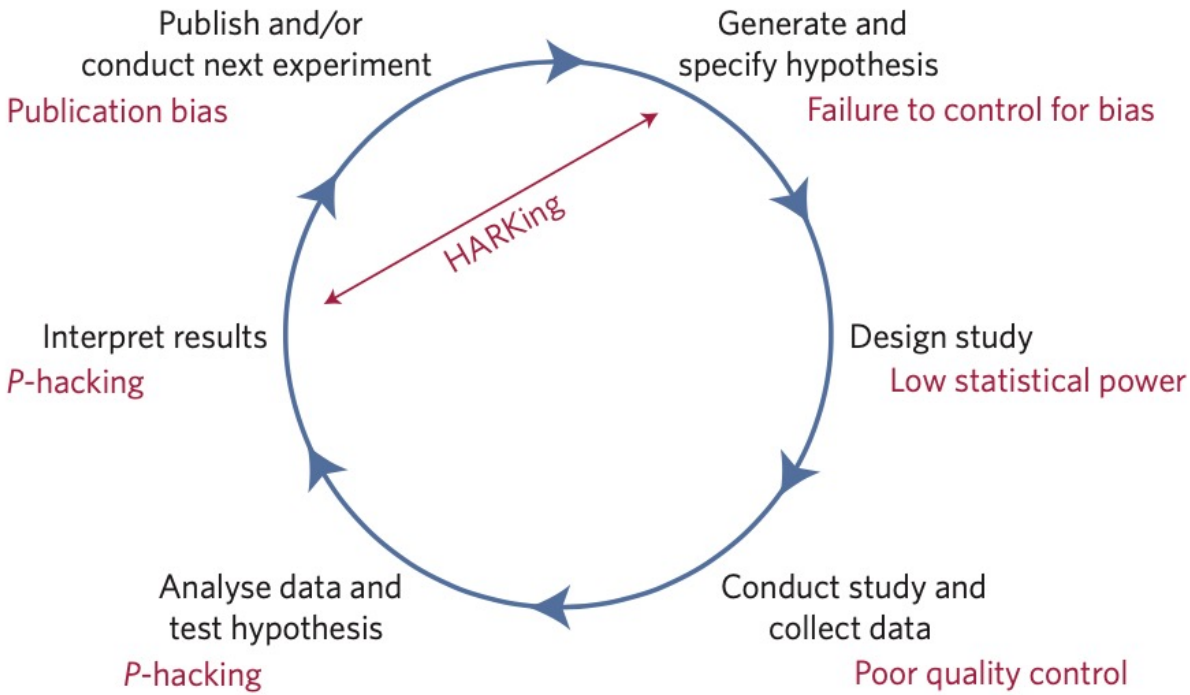
Simmons, Nelson & Simonsohn (2011)

Table 3. Study 2: Original Report (in Bolded Text) and the Requirement-Compliant Report (With Addition of Gray Text)

Using the same method as in Study 1, we asked ~~20~~ 34 **University of Pennsylvania undergraduates to listen** only **to either "When I'm Sixty-Four" by The Beatles or "Kalimba"** or "Hot Potato" by the Wiggles. We conducted our analyses after every session of approximately 10 participants; we did not decide in advance when to terminate data collection. **Then, in an ostensibly unrelated task, they indicated** only **their birth date (mm/dd/yyyy) and** how old they felt, how much they would enjoy eating at a diner, the square root of 100, their agreement with "computers are complicated machines," **their father's age**, their mother's age, whether they would take advantage of an early-bird special, their political orientation, which of four Canadian quarterbacks they believed won an award, how often they refer to the past as "the good old days," and their gender. **We used father's age to control for variation in baseline age across participants**.

An ANCOVA revealed the predicted effect: According to their birth dates, people were nearly a year-and-a-half younger after listening to "When I'm Sixty-Four" (adjusted *M* = 20.1 years) rather than to "Kalimba" (adjusted *M* = 21.5 years), *F*(1, 17) = 4.92, *p* = .040. Without controlling for father's age, the age difference was smaller and did not reach significance (*M*s = 20.3 and 21.2, respectively), $F(1, 18) = 1.01$, $p = .33$.

Pourquoi?



Munafo et al. (2017)

Des solutions

Transparence

- Pré-enregistrement des études
- Détermination des tailles d'échantillons a priori
- Description précise de la procédure (e.g., exclusions)
- Partage du matériel, des données et des scripts d'analyse
- Certification des résultats (CASCAD)

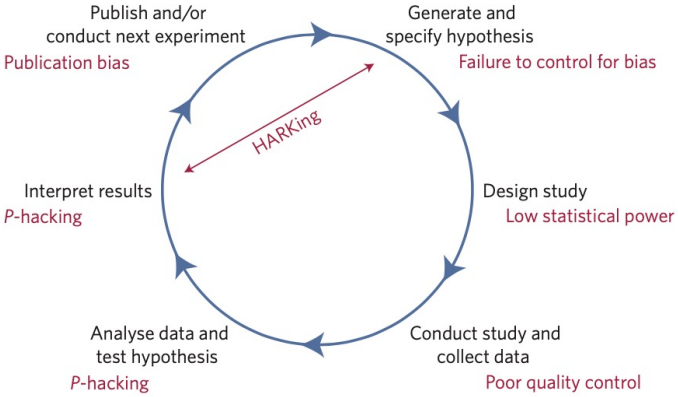
Psychological Science
Volume 35, Issue 7, July 2024, Pages 708-711

https://doi.org/10.1177/09567976231221573

Sage Journals

Editorials

Transparency Is Now the Default at *Psychological Science*



Est-ce que cela a un impact?

Résultats semblent plus stables (Bogdan, 2025)

Tailles d'échantillons sont plus importantes (e.g. Sassenberg et al., 2019)

Mais: Information parfois incomplète (Crüwell et al., 2023; Hardwicke et al., 2021)

Protzko et al. (2023), 16 études socio-comportementales répliquées par 4 laboratoires, 86% de réplication, taille d'effets similaires (97%).

Mais… article rétracté en raison d'un défaut de pré-enregistrement.