



Reproducible Research: A Retrospective

Annual Review of Public Health

Vol. 42:79-93 (Volume publication date April 2021)

First published as a Review in Advance on January 19, 2021

<https://doi.org/10.1146/annurev-publhealth-012420-105110>

Roger D. Peng and Stephanie C. Hicks

Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland 21205, USA; email: rdpeng@jhu.edu, shicks19@jhu.edu

[https://github.com/richelbilderbeek/
journal_club_20220602](https://github.com/richelbilderbeek/journal_club_20220602)



Science?

I have discovered a marvelous
proof to this theorem,
that this margin is
too narrow to contain



Why this article

**I care about
reproducible science
Article is recent**

Reproducible
Science

Code

Data

The authors

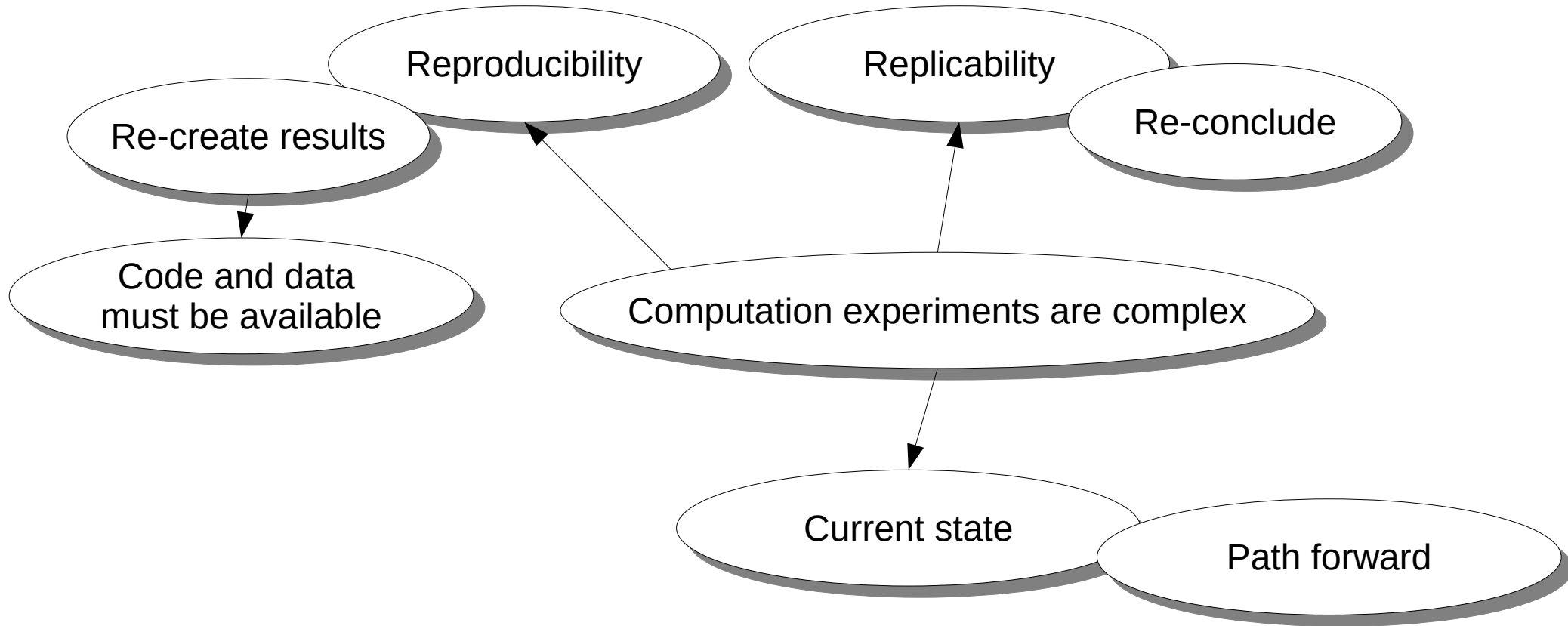


Roger D Peng



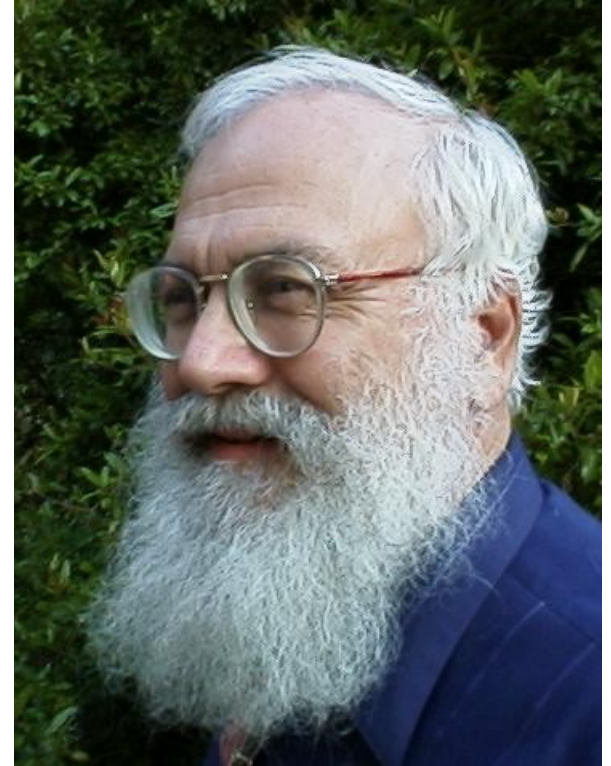
Stephanie Hicks

One-slide overview



Introduction 1/3

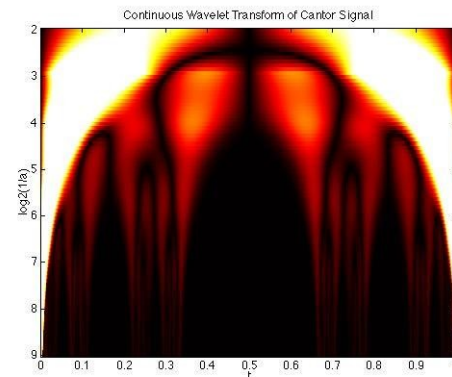
[...] it often seems that the greatest beneficiary of preparing the work in a reproducible form is the original author!



Introduction 2/3

An article [...] is merely advertising of [...] scholarship.

The **actual** scholarship is the complete software development environment and the complete set of instructions which generated the figures.



Introduction 3/3

Our reviewers were not able to conduct an independent and private peer review and therefore notified us of their withdrawal from the peer-review process.

THE LANCET

Volume 391 Number 11115 Pages 947-958 March 6-12, 2020

www.thelancet.com

TRUMP IS TAKING
HYDROXYCHLOROQUINE:
WHY EXPERTS SAY YOU
SHOULDN'T



Definition of reproducible research

A published data analysis is reproducible
if the analytic **data** sets
and the computer **code** used to create the data analysis
are made available to others for independent study and analysis

Subset of data set

Analysis and visualization

Tuning parameters

?

Definition of reproducible research

A published data analysis is reproducible
if the analytic **data** sets
and the computer **code** used to create the data analysis
are made available to others for independent study and analysis

Subset of data set

Analysis and visualization

Tuning parameters

Sensitive data set?

How to make available?

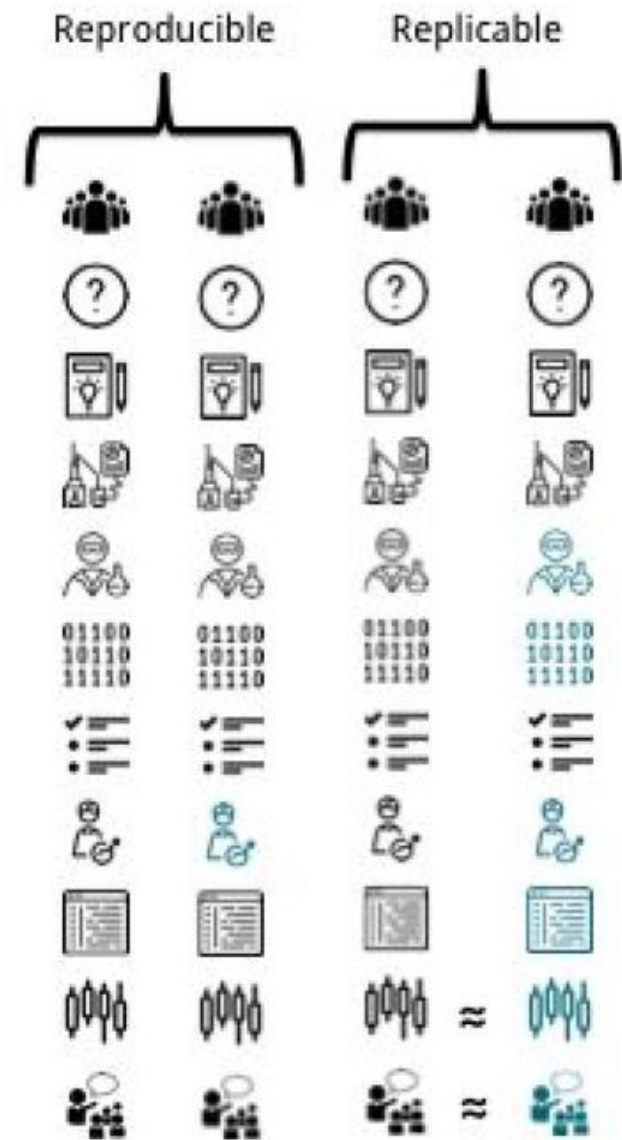
Testing data set?

Requirements for code?

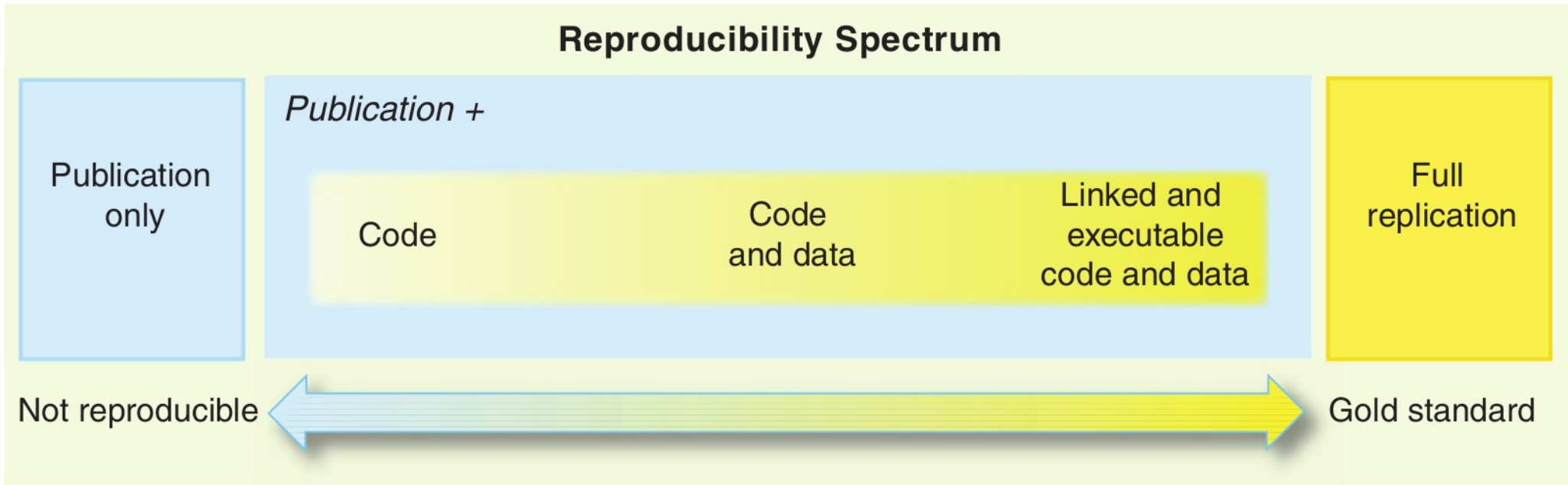
Reproduction:
re-do experiment

Replication:
re-conclude

Legend

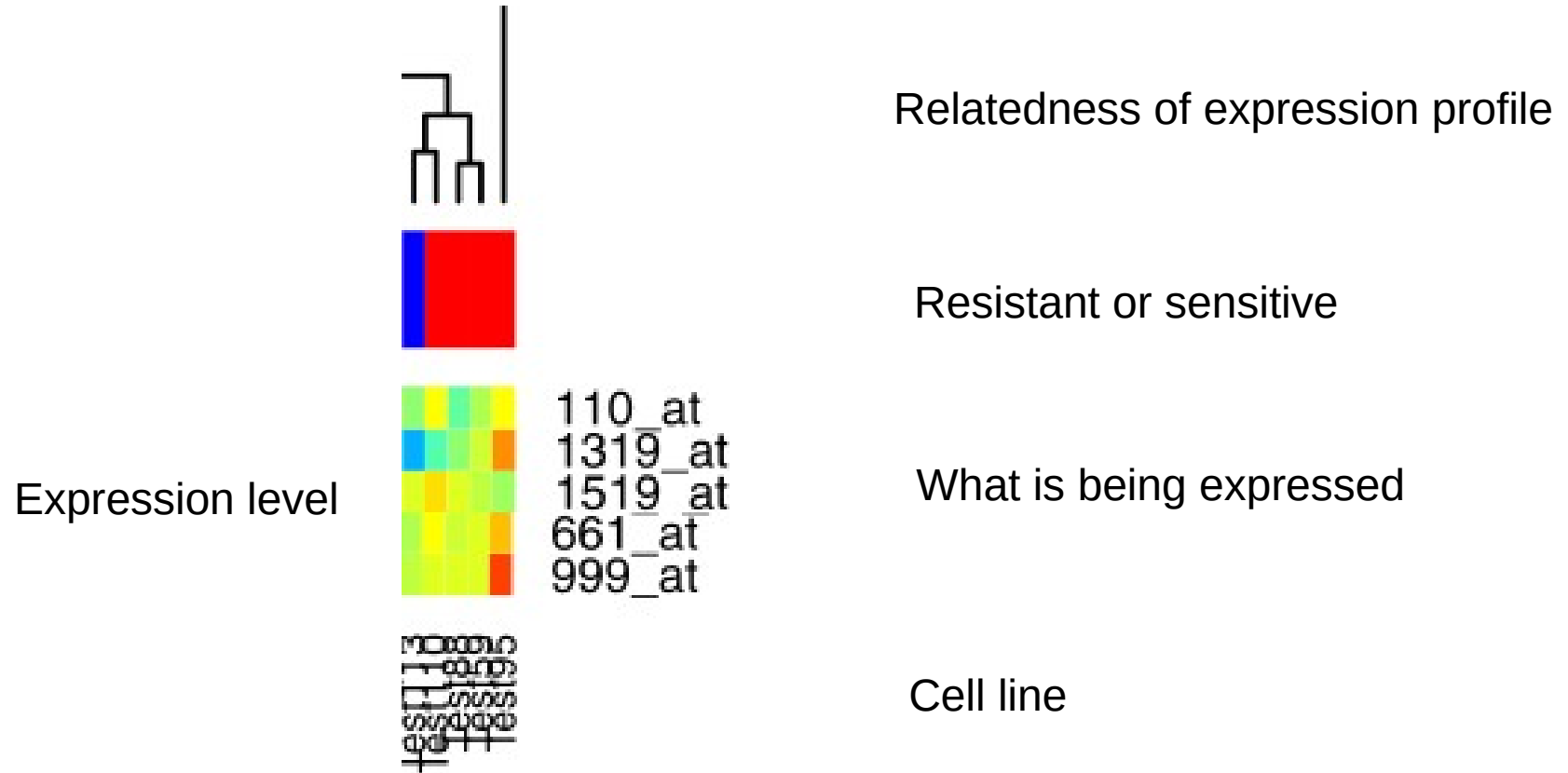


Looking back



Peng, Roger D. "Reproducible research in computational science." *Science* 334.6060 (2011): 1226-1227.

Forensic bioinformatics

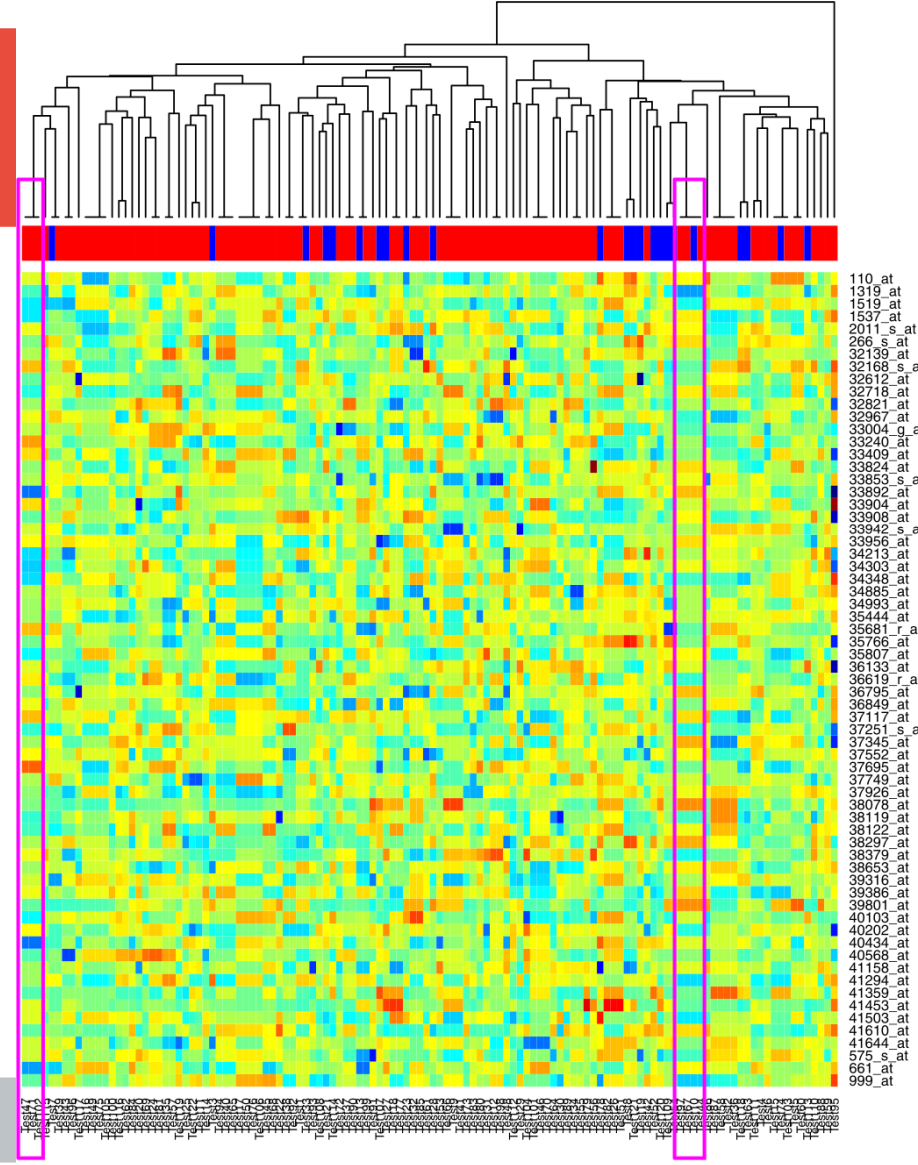


Forensic bioinformatics

An off-by-one error

Most common errors are simple

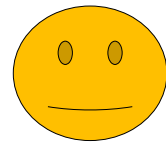
Simple errors are common



Refining reproducibility

Data

Code



Refining reproducibility

Data

More commonly shared

Big datasets are/can be uploaded

Stripping data from original context may be problematic

Best practices emerge: Tidy Data

Code

?



Refining reproducibility

Data

Code

Software development has become easier
Easier to put code in packages
Increase of testing and test-based development



Refining reproducibility

Data

Sensitive data set?

Homomorphic algorithms

Testing data set?

Send code to data

Benchmark data set?

Code

How to make available?

Supply as runnable container?

Requirements for code?

Proof of building?

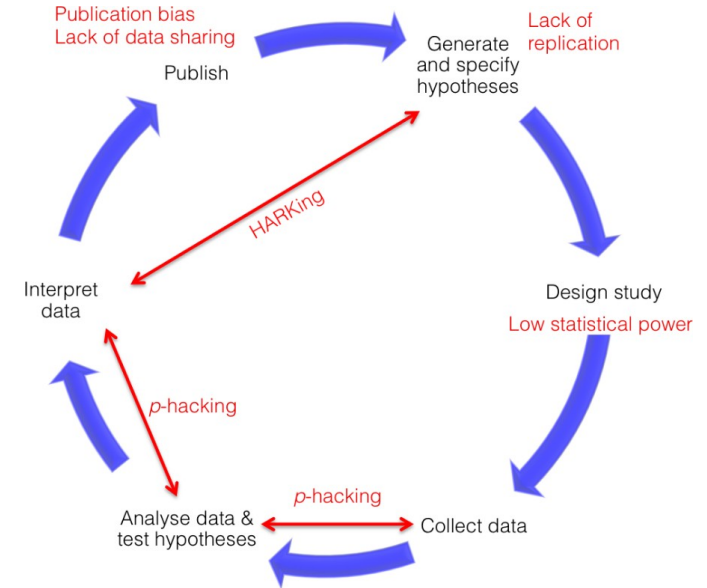
Proof of correctness?

Refining reproducibility

Paper

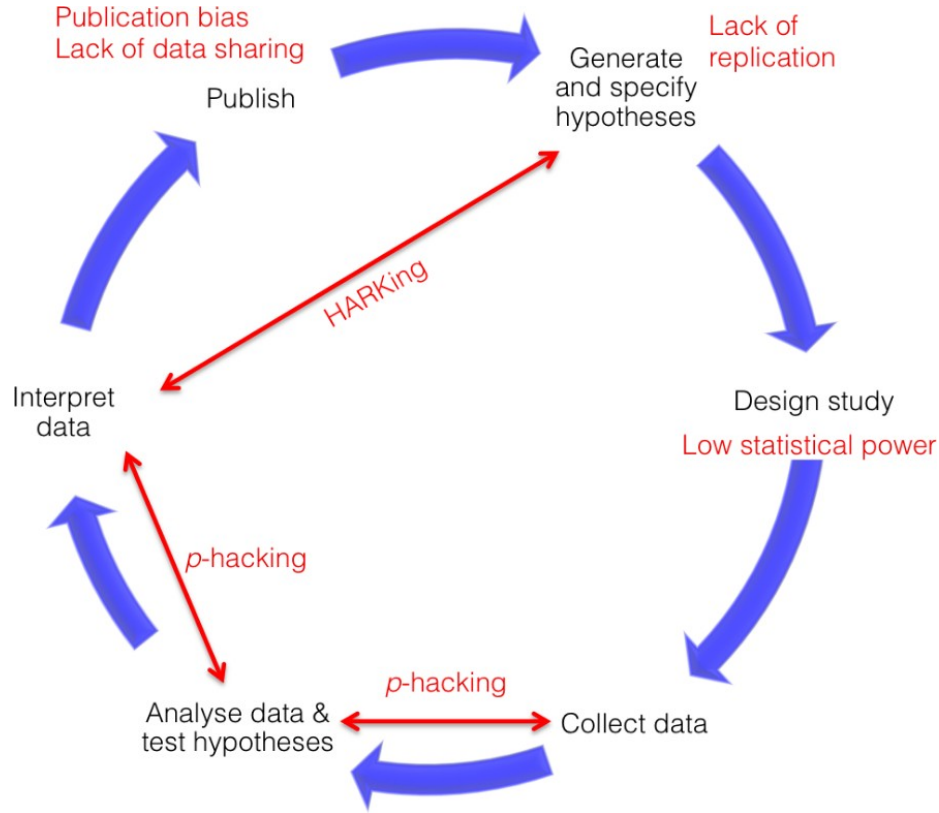
Pre-registration?

Requirements for history?



Chambers, Christopher D., et al. "Instead of" playing the game" it is time to change the rules: Registered Reports at AIMS Neuroscience and beyond." AIMS Neuroscience 1.1 (2014): 4-17.

Refining reproducibility



Chambers, Christopher D., et al. "Instead of" playing the game" it is time to change the rules: Registered Reports at AIMS Neuroscience and beyond." AIMS Neuroscience 1.1 (2014): 4-17.

Refining reproducibility

Paper

Pre-registration?

Requirements for history?

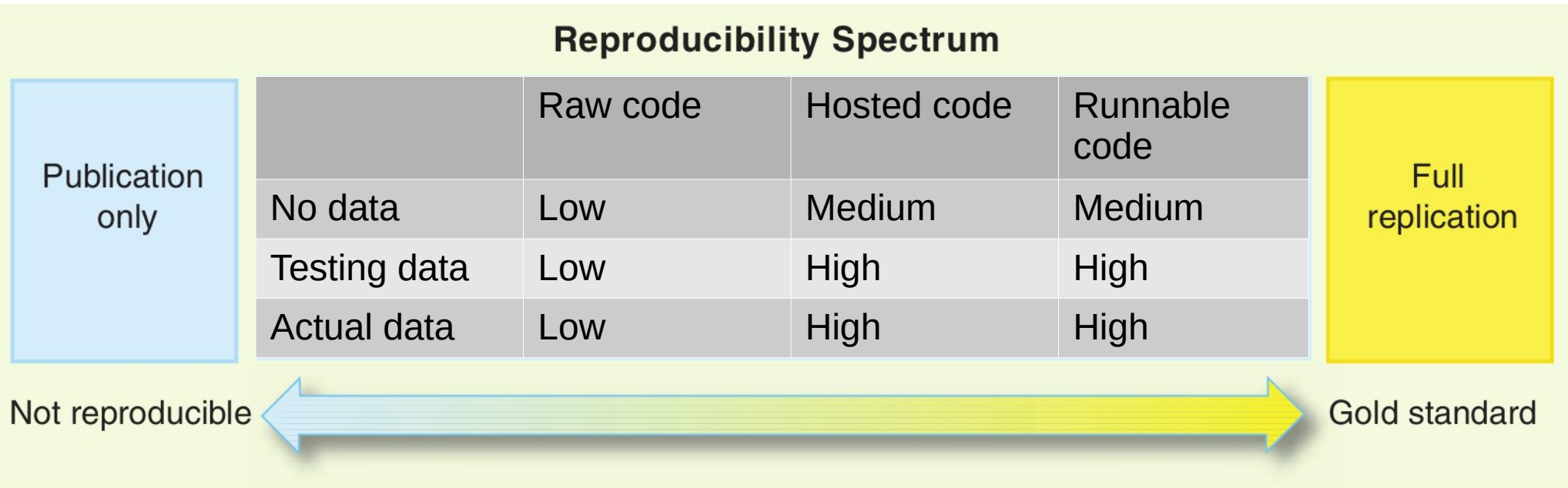


What do you think?

[https://github.com/richelbilderbeek/
journal_club_20220602](https://github.com/richelbilderbeek/journal_club_20220602)



Code and data are separate components



In school and university

Is there a solution for
 $a^n + b^n = c^n$
for every $n > 2$?

Yes!



Is there a solution for
 $a^n + b^n = c^n$
for every $n > 2$?

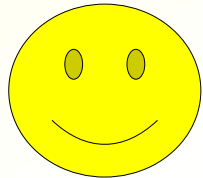
[complex proof here]



In academia

Complex
computational
pipeline

No code



Complex
computational
pipeline

Code public

