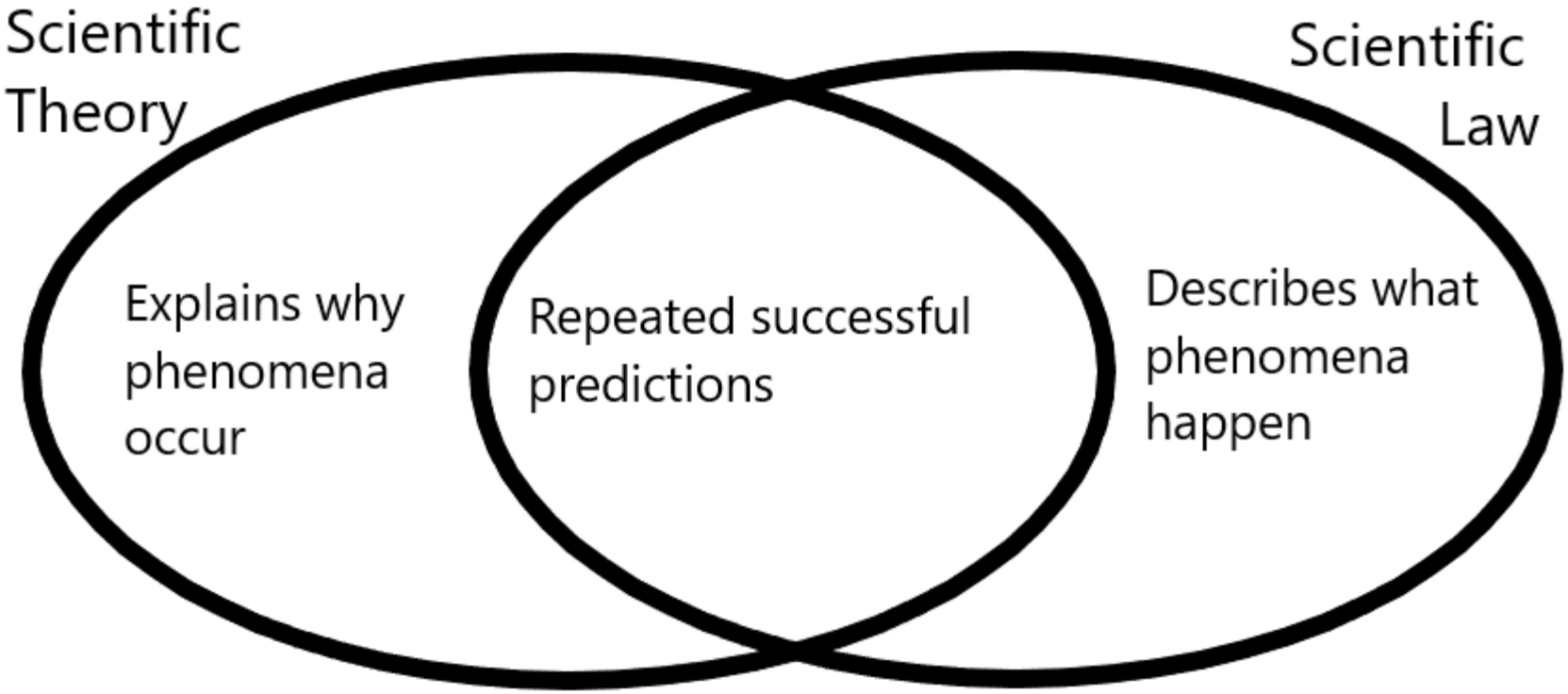
Openness and Reproducible Research Practices

Elisabeth Julie Vargo, PhD

COURSE OBJECTIVES

- Discuss the “Reproducibility Crisis” in Psychology
- Describe benefits of open and reproducible research practices
- Understand emerging incentives for those practices

What is the Replication Crisis?


Scientific Theory
Scientific Law
Explains why phenomena occur
Repeated successful predictions
Describes what phenomena happen

CORRESPONDENCE

LINK TO O

Believe it or not: how much can we rely on published data on potential drug targets?

Florian Prinz, Thomas Schlange and Khusru Asadullah

A recent report by Arrowsmith noted that the success rates for new development projects in Phase II trials have fallen from 28% to 18% in

to 'feasible/marketable', and the financial costs of pursuing a full-blown drug discovery and development programme for a particular tar-

results that are publish
duce. However, there is
this apparently widespr
public recognition (for
and the surprisingly f
tions dealing with this
knowledge, so far there
in-depth, systematic a
reproduced results with
wet-lab experiments rela
tion and validation.

Early research in the
try, with a dedicated bud
mainly work on target

© 2011 Macmillan Publishers Limited https://doi.org/10.1038/nrd3439-c1

Power failure: why small sample size undermines the reliability of neuroscience

Katherine S. Button[1,2], John P. A. Ioannidis[3], Claire Mokrysz[1], Brian A. Nosek[4], Jonathan Flint[5], Emma S. J. Robinson[6] and Marcus R. Munafò[1]

Abstract | A study with low statistical power has a reduced chance of detecting a true effect, but it is less well appreciated that low power also reduces the likelihood that a statistically significant result reflects a true effect. Here, we show that the average statistical power of studies in the neurosciences is very low. The consequences of this include overestimates of effect size and low reproducibility of results. There are also ethical dimensions to this problem, as unreliable research is inefficient and wasteful. Improving reproducibility in neuroscience is a key priority and requires attention to well-established but often ignored

© 2013 Macmillan Publishers https://doi.org/10.1038/nrn3475

Open access, freely available online

Essay

Why Most Published Research Findings Are False

John P.A. Ioannidis

Summary

There is increasing concern that most current published research findings are false. The probability that a research claim is true may depend on study power and bias, the number of other studies on the same question, and, importantly, the ratio of true to no relationships among the relationships probed in each scientific field. In this framework, a research finding is less likely to be true when the studies conducted in a field are smaller; when effect sizes are smaller; when there is a greater number and lesser preselection of tested relationships; where there is greater flexibility in designs, definitions, outcomes, and analytical modes; when there is greater financial and other interest and prejudice; and when more teams are involved in a scientific field in chase of statistical significance. Simulations show that for most study designs and settings, it is more likely for a research claim to be false than true. Moreover, for many current scientific fields, claimed research findings may often be simply accurate measures of the prevailing bias. In this essay, I discuss the implications of these problems for the conduct and interpretation of research.

Published research findings are sometimes refuted by subsequent evidence, with ensuing confusion and disappointment. Refutation and controversy is seen across the range of research designs, from clinical trials and traditional epidemiological studies [1–3] to the most modern molecular research [4,5]. There is increasing concern that in modern research, false findings may be the majority or even the vast majority of published research claims [6–8]. However, this should not be surprising. It can be proven that most claimed research findings are false. Here I will examine the key factors that influence this problem and some corollaries thereof.

The Essay section contains opinion pieces on topics of broad interest to a general medical audience.

Modeling the Framework for False Positive Findings

Several methodologists have pointed out [9–11] that the high rate of nonreplication (lack of confirmation) of research discoveries is a consequence of the convenient, yet ill-founded strategy of claiming conclusive research findings solely on the basis of a single study assessed by formal statistical significance, typically for a p-value less than 0.05. Research is not most appropriately represented and summarized by p-values, but, unfortunately, there is a widespread notion that medical research articles

It can be proven that most claimed research findings are false.

should be interpreted based only on p-values. Research findings are defined here as any relationship reaching formal statistical significance, e.g., effective interventions, informative predictors, risk factors, or associations. "Negative" research is also very useful. "Negative" is actually a misnomer, and the misinterpretation is widespread. However, here we will target relationships that investigators claim exist, rather than null findings.

As has been shown previously, the probability that a research finding is indeed true depends on the prior probability of it being true (before doing the study), the statistical power of the study, and the level of statistical significance [10,11]. Consider a 2×2 table in which research findings are compared against the gold standard of true relationships in a scientific field. In a research field both true and false hypotheses can be made about the presence of relationships. Let R be the ratio of the number of "true relationships" to "no relationships" among those tested in the field. R is characteristic of the field and can vary a lot depending on whether the field targets highly likely relationships or searches for only one or a few true relationships among thousands and millions of hypotheses that may be postulated. Let us also consider, for computational simplicity, circumscribed fields where either there is only one true relationship (among many that can be hypothesized) or the power is similar to find any of the several existing true relationships. The pre-study probability of a relationship being true is $R/(R+1)$. The probability of a study finding a true relationship reflects the power $1 - \beta$ (one minus the Type II error rate). The probability of claiming a relationship when none truly exists reflects the Type I error rate, α. Assuming that c relationships are being probed in the field, the expected values of the 2×2 table are given in Table 1. After a research finding has been claimed based on achieving formal statistical significance, the post-study probability that it is true is the positive predictive value, PPV. The PPV is also the complementary probability of what Wacholder et al. have called the false positive report probability [10]. According to the 2×2 table, one gets $PPV = (1 - \beta)R/(R - \beta R + \alpha)$. A research finding is thus

Citation: Ioannidis JPA (2005) Why most published research findings are false. PLoS Med 2(8): e124.

Copyright: © 2005 John P.A. Ioannidis. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abbreviation: PPV, positive predictive value

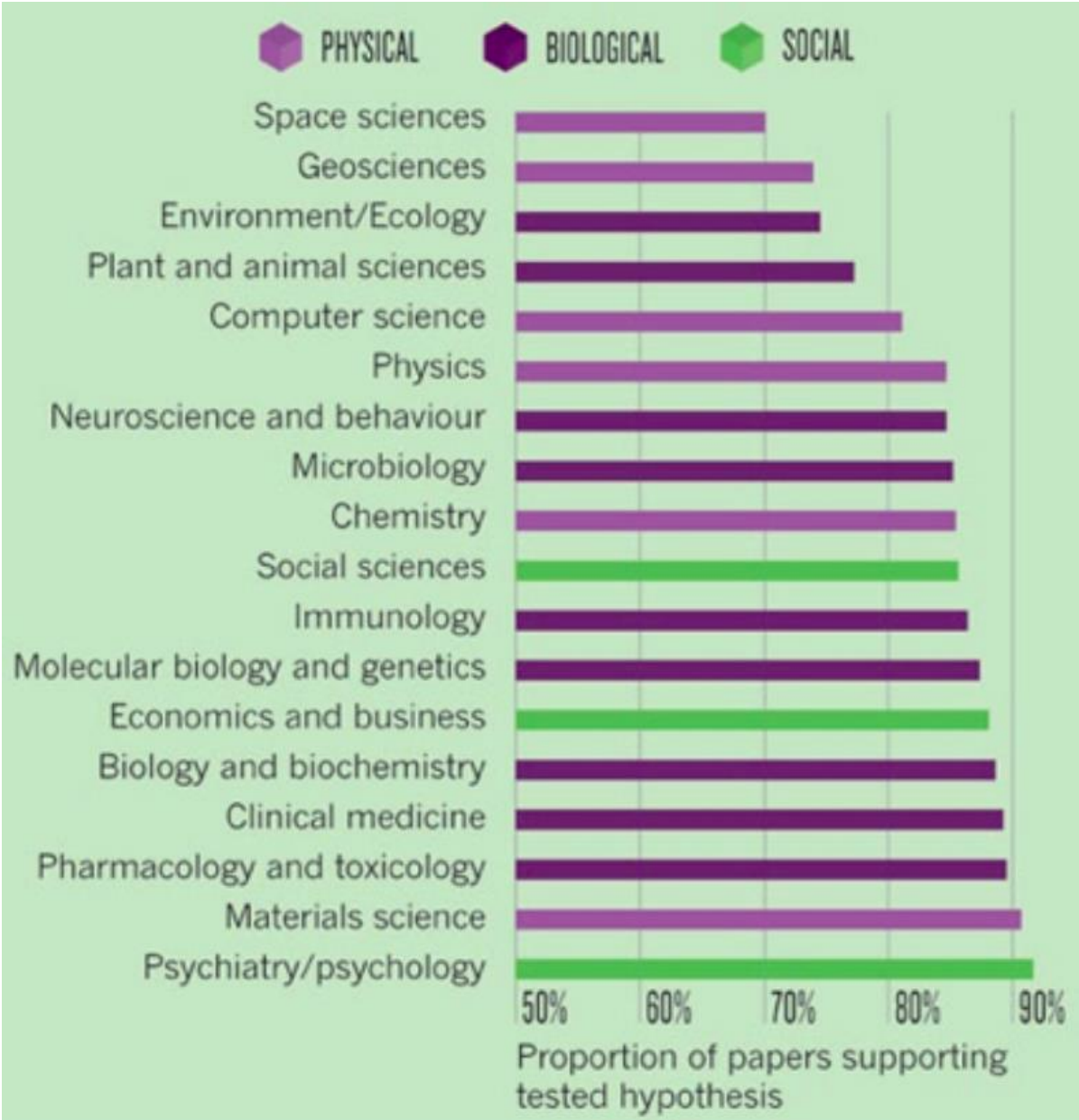
John P.A. Ioannidis is in the Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina, Greece, and Institute for Clinical Research and Health Policy Studies, Department of Medicine, Tufts-New England Medical Center, Tufts University School of Medicine, Boston, Massachusetts, United States of America. E-mail: jioannid@cc.uoi.gr

Competing Interests: The author has declared that no competing interests exist.

DOI: 10.1371/journal.pmed.0020124

CC-BY © 2005 John P. A. Ioannidis https://doi.org/10.1371/journal.pmed.0020124

Positive Results by Discipline



Fanelli D (2010) "Positive" Results Increase Down the Hierarchy of the Sciences. PLOS ONE 5(4): e10068. CC-BY
doi:10.1371/journal.pone.0010068

PLOS | ONE

Researchers have many decisions to make when conducting a study:

- Choosing a sample size - stopping data collection
- How to deal with outliers/illegitimate responses
- Which conditions/groups should be compared
- Creating variables
- Which items? Transformations?
- Which variables should be included in analyses
- IVs, DVs, controls, mediators, moderators

Making each decision increases the false positive rate

SOURCES OF ISSUES IN REPRODUCIBILITY

- Methodological, statistical, and reporting practices
- Structural and organizational practices
- Intentional scientific misconduct (p-hacking, HARK)

WHAT IS REPRODUCIBILITY?

Computational Reproducibility:

If we took your data and code/analysis scripts and reran it, we can reproduce the numbers/graphs in your paper

Methods Reproducibility:

We have enough information to rerun the experiment or survey the way it was originally conducted

Results Reproducibility/Replicability:

We use your exact methods and analyses, but collect new data, and we get the same statistical conclusion

Replicability of results is essential for science
If results do not replicate, how can we be sure that they exist at all?

Open Science Collaboration attempted to replicate 100 studies published in 3 top psychology journals in 2008.

Results: 39% of the original studies were successfully replicated

- 25% of social psychology studies replicated
- 50% of cognitive psychology studies replicated

Effect sizes overestimated in original studies

WHY SHOULD YOU CARE?

- Increases efficiency of your own work
- Can reduce false leads
- Data sharing citation advantage

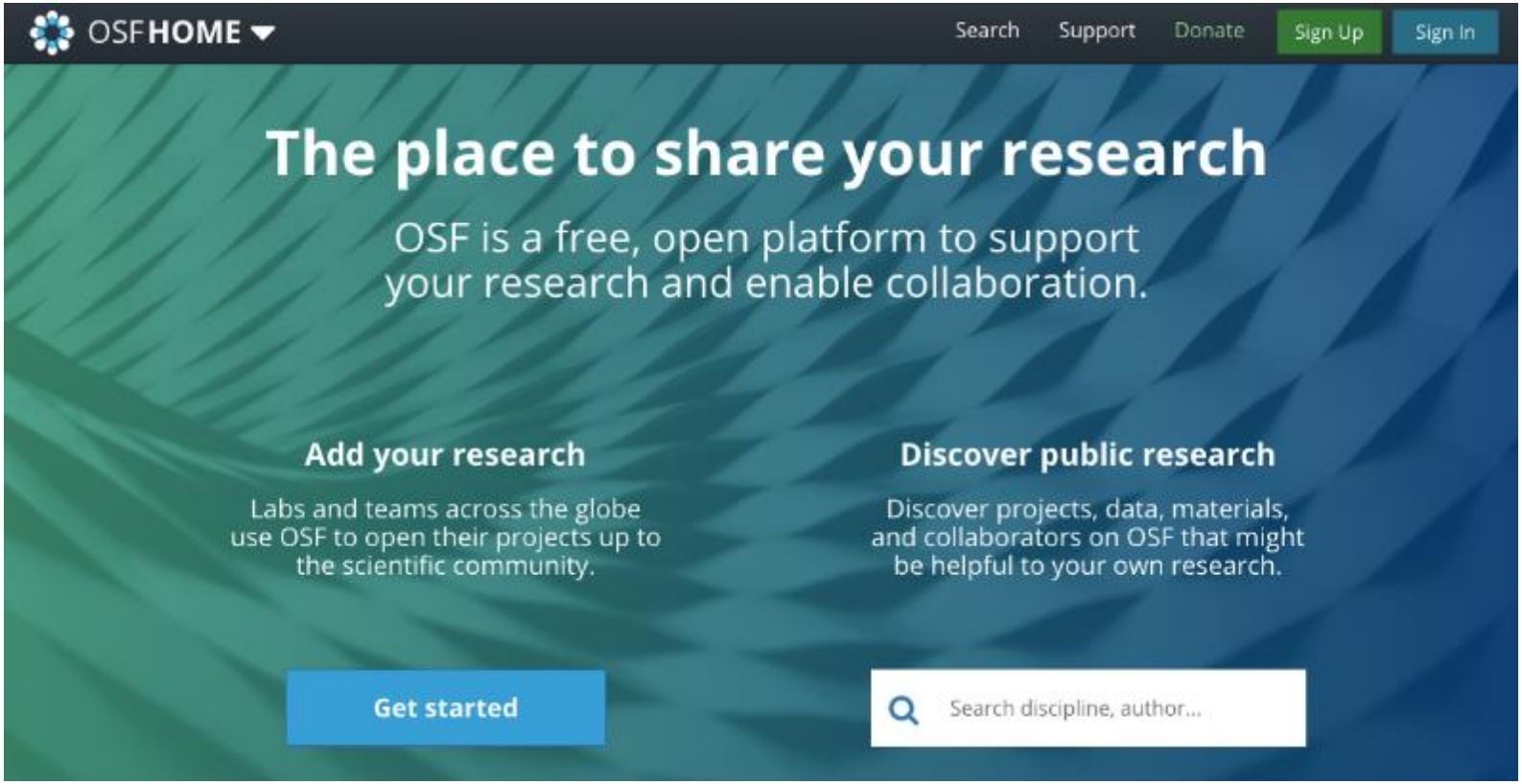
> It takes some effort to organize your research to be reproducible...the principal beneficiary is generally the author herself.

– Jon Claerbout

Making Scientific Contributions Reproducible

sepwww.stanford.edu/oldsep/matt/join/redoc/web/iris.html

osf.io



FREE, OPEN SOURCE

PREREGISTRATION

Documenting your research plan in a read-only public repository before you conduct the study.

Practice originated in clinical research and is now expanding more broadly.

PREREGISTRATION

Benefits of preregistering your study depend on how much information you include. At a minimum a preregistration should include the "what" of a study:

- Research question
- Population and sample size
- General design
- Variables to be collected, or dataset you'll be using

WHY PREREGISTER?

- Preregistration helps reduce the “file drawer effect” by increasing discoverability of unpublished studies.
- Preregistered analysis plans help improve study accuracy and replicability by guarding against unintended false positive inflation.

EXPLORATORY VS. CONFIRMATORY ANALYSES

Exploratory

- Interested in discovering possible patterns/relationships in data to develop hypotheses

Confirmatory

- Have a specific hypothesis you want to test

PRE-ANALYSIS PLAN

Details the analyses planned for hypothesis testing:

- Sample size
- Data processing and cleaning procedures
- Exclusion criteria
- Statistical analyses

Including a pre-analysis plan in your preregistration helps decrease researcher degrees of freedom.

RESEARCHER DEGREES OF FREEDOM

All data processing and analytical choices made ***after*** seeing and interacting with your data

- Should I collect more data?
- Which observations should I exclude?
- Which conditions should I compare?
- What should be my main DV?
- Should I look for an interaction effect?

“The first principle is that you must not fool yourself, and you are the easiest person to fool.

After you’ve not fooled yourself, it’s easy not to fool other scientists. You just have to be honest in a conventional way after that.”

– Richard P. Feynman

Cargo Cult Science (1974)

WHY YOU MIGHT WANT TO SHARE

- Journal/funder mandates
- Increase impact of your work
- Get recognition for your good research practices

WHAT TO SHARE

- Sharing is a continuum
  - Data underlying just results reported in a paper
  - Data underlying publication + information about other variables collected
  - Data underlying publication + embargo on full dataset
  - All data collected for that study

TRADITIONAL ACADEMIC PUBLISHING MODEL



REGISTERED REPORTS



cos.io/rr

Take-Home Message

- Replication should be just as important as innovation in science
- Registered replication reports
- Science is a continual process of updating what we know, self-correcting as we go along: innovation, doesn't replicate, figure out why → more innovation, replicate
- Psychology has changed in the last 5-6 years. Open science becoming the norm.

FEELING OVERWHELMED?

You don’t have to do all of this at once:

- One piece at a time
- Series of small steps

Even incremental changes
can have significant impact

WHERE TO GET HELP

- osf.io/support
- Center for Open Science YouTube channel

Thank you!