The Reproducibility Crisis in science and how to solve it

Zoltan Dienes
University of Sussex

1. Evidence that there is a problem

2. Solutions!

1. The problem.

Psychological Bulletin
1979, Vol. 86, No. 3, 638–641

The "File Drawer Problem" and Tolerance for Null Results

Robert Rosenthal
Harvard University

For any given research area, one cannot tell how many studies have been conducted but never reported. The extreme view of the "file drawer problem" is that journals are filled with the 5% of the studies that show Type I errors, while the file drawers are filled with the 95% of the studies that show nonsignificant results. Quantitative procedures for computing the tolerance for filed and future null results are reported and illustrated, and the implications are discussed.

General Article

Psychological Science
22(11) 1359–1366

DOI: 10.1177/0956797611417632
http://pss.sagepub.com

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

Abstract

In this article, we accomplish two things. First, we show that despite empirical psychologists' nominal endorsement of a low rate of false-positive findings (≤ .05), flexibility in data collection, analysis, and reporting dramatically increases actual false-positive rates. In many cases, a researcher is more likely to falsely find evidence that an effect exists than to correctly find evidence that it does not. We present computer simulations and a pair of actual experiments that demonstrate how unacceptably easy it is to accumulate (and report) statistically significant evidence for a false hypothesis. Second, we suggest a simple, low-cost, and straightforwardly effective disclosure-based solution to this problem. The solution involves six concrete requirements for authors and four guidelines for reviewers, all of which impose a minimal burden on the publication process.

Table 1. Likelihood of Obtaining a False-Positive Result

	Significance level		
Researcher degrees of freedom	$p < .1$	$p < .05$	$p < .01$
Situation A: two dependent variables ($r = .50$)	17.8%	9.5%	2.2%
Situation B: addition of 10 more observations per cell	14.5%	7.7%	1.6%
Situation C: controlling for gender or interaction of gender with treatment	21.6%	11.7%	2.7%
Situation D: dropping (or not dropping) one of three conditions	23.2%	12.6%	2.8%
Combine Situations A and B	26.0%	14.4%	3.3%
Combine Situations A, B, and C	50.9%	30.9%	8.4%
Combine Situations A, B, C, and D	81.5%	60.7%	21.5%

Research Article

Psychological Science
23(5) 524–532

DOI: 10.1177/0956797611430953
http://pss.sagepub.com

Measuring the Prevalence of Questionable Research Practices With Incentives for Truth Telling

Leslie K. John[1], George Loewenstein[2], and Drazen Prelec[3]
[1]Marketing Unit, Harvard Business School; [2]Department of Social & Decision Sciences, Carnegie Mellon University; and [3]Sloan School of Management and Departments of Economics and Brain & Cognitive Sciences, Massachusetts Institute of Technology

Abstract

Cases of clear scientific misconduct have received significant media attention recently, but less flagrantly questionable research practices may be more prevalent and, ultimately, more damaging to the academic enterprise. Using an anonymous elicitation format supplemented by incentives for honest reporting, we surveyed over 2,000 psychologists about their involvement in questionable research practices. The impact of truth-telling incentives on self-admissions of questionable research practices was positive, and this impact was greater for practices that respondents judged to be less defensible. Combining three different estimation methods, we found that the percentage of respondents who have engaged in questionable practices was surprisingly high. This finding suggests that some questionable practices may constitute the prevailing research norm.

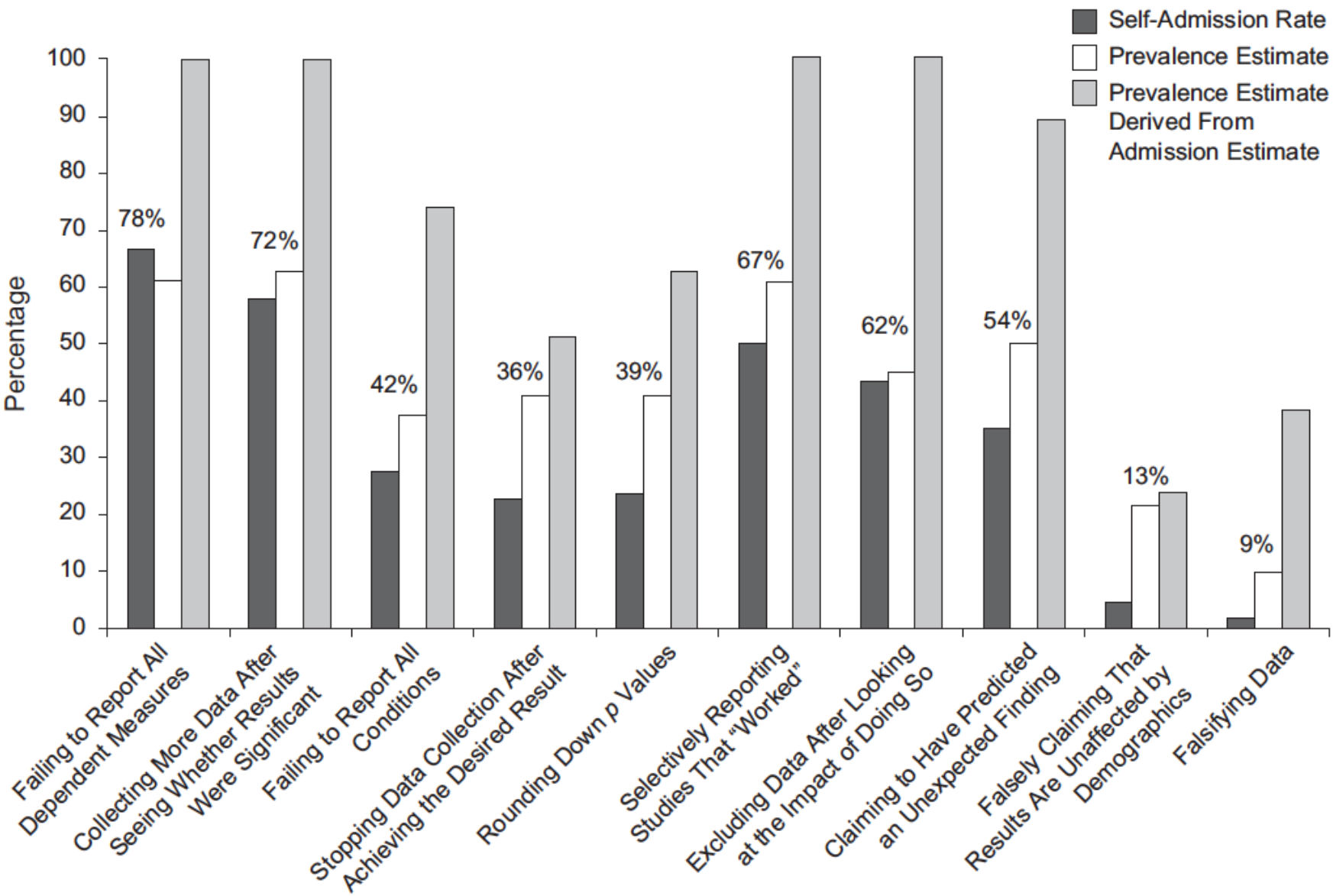


Fig. 1. Results of the Bayesian-truth-serum condition in the main study. For each of the 10 items, the graph shows the self-admission rate, prevalence estimate, prevalence estimate derived from the admission estimate (i.e., self-admission rate/admission estimate), and geometric mean of these three percentages (numbers above the bars). See Table 1 for the complete text of the items.

THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY
2012, i*First*, 1–9

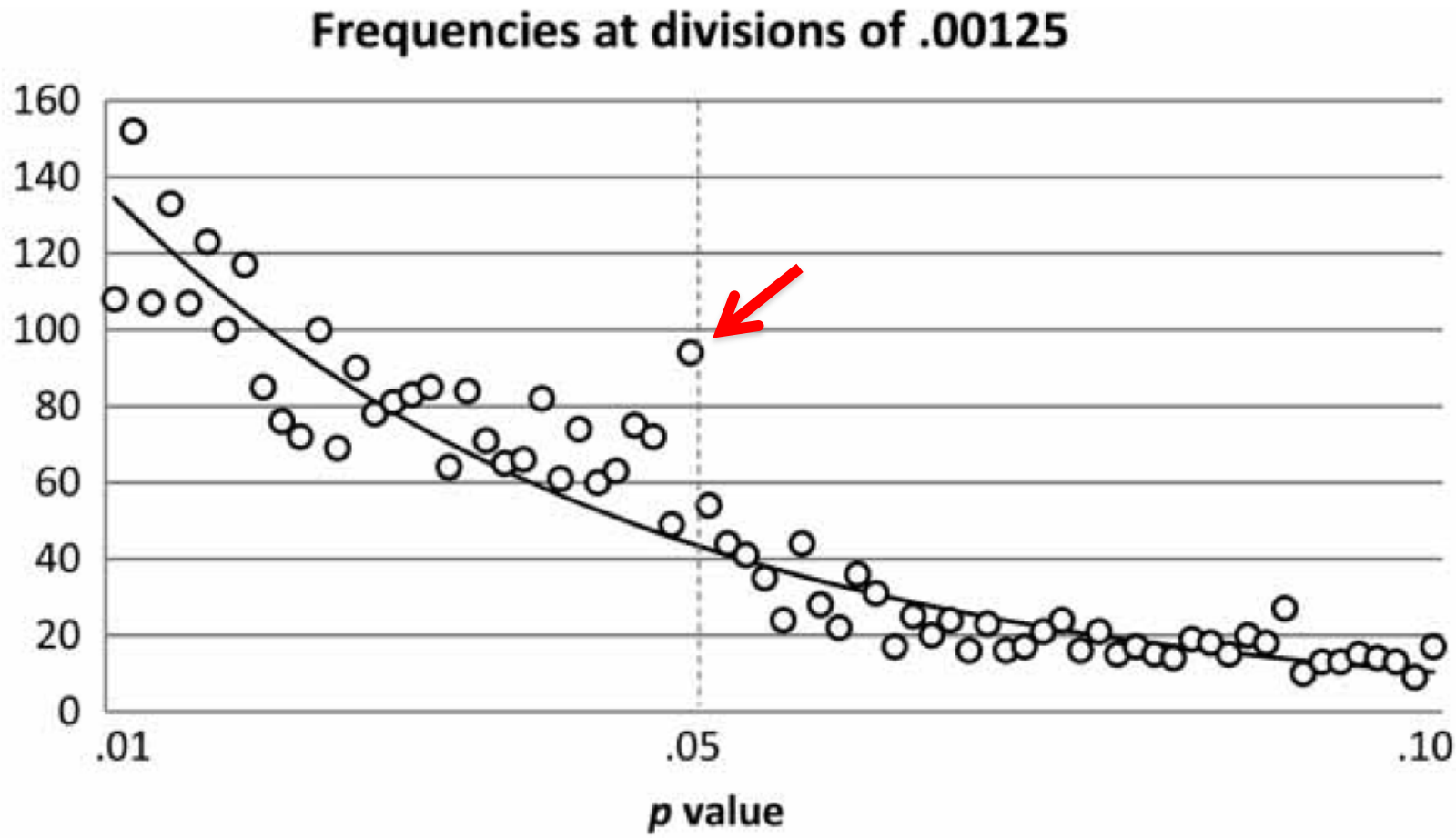
Psychology Press
Taylor & Francis Group

A peculiar prevalence of *p* values just below .05

E. J. Masicampo[1], and Daniel R. Lalande[2]

[1]Department of Psychology, Wake Forest University, Winston-Salem, NC, USA
[2]Department of Health Sciences, Université du Québec à Chicoutimi, Chicoutimi, QC, Canada

In null hypothesis significance testing (NHST), *p* values are judged relative to an arbitrary threshold for significance (.05). The present work examined whether that standard influences the distribution of *p* values reported in the psychology literature. We examined a large subset of papers from three highly regarded journals. Distributions of *p* were found to be similar across the different journals. Moreover, *p* values were much more common immediately below .05 than would be expected based on the number of *p* values occurring in other ranges. This prevalence of *p* values just below the arbitrary criterion for significance was observed in all three journals. We discuss potential sources of this pattern, including publication bias and researcher degrees of freedom.



Distribution of 3,627 p values from *JEP: General, JPSP, Psych. Sci.*

The behavioural priming saga …

PLOS | ONE | Subject Areas | For Authors | Ab

OPEN ACCESS PEER-REVIEWED
RESEARCH ARTICLE

Priming Intelligent Behavior: An Elusive Phenomenon

David R. Shanks, Ben R. Newell, Eun Hee Lee, Divya Balakrishnan, Lisa Ekelund, Zarus Cenac, Fr

Published: April 24, 2013 • DOI: 10.1371/journal.pone.0056515

Article | About the Authors | Metrics | Comments

PLOS | ONE | Subject Areas | For

OPEN ACCESS PEER-REVIEWED
RESEARCH ARTICLE

Behavioral Priming: It's All in the Mind, but

Stéphane Doyen, Olivier Klein, Cora-Lise Pichon, Axel Cleeremans

Published: January 18, 2012 • DOI: 10.1371/journal.po

Article | About the Authors | Met

Abstract

The perspective that b
widespread in social p
participants unwittingl
laboratory, was instru
aimed at replicating th
larger sample, our first
manipulating the belief
walk slower when pri
Strikingly, we o
participants wo
expectations ar
suggestive of a

Replication

Does Cleanliness Influence Moral Judgments?

A Direct Replication of Schnall, Benton, and Harvey (2008)

David J. Johnson, Felix Cheung, and M. Brent Donnellan

Department of Psychology, Michigan State University, East Lansing, MI, USA

Abstract. Schnall, Benton, and Harvey (2008) hypothesized that physical cleanliness reduces the severity of moral judgments. In support of
ess (Exp. 1) and when they
als supplied by the original
sizes that provided over .99
rom Experiment 1 (original
2 (original $d = -0.85$; 95%
population effect sizes are
aliness and morality should

OPEN ACCESS Freely available online | PLOS | ONE

Two Failures to Replicate High-Performance-Goal Priming Effects

Christine R. Harris[1*], Noriko Coburn[1], Doug Rohrer[2], Harold Pashler[1]

1 Psychology Department, University of California San Diego, La Jolla, California, United States of America, 2 Psychology Department, University of South Florida, Tampa, Florida, United States of America

Abstract

Bargh et al. (2001) reported two experiments in which people were exposed to words related to achievement (e.g.

Journal of Experimental Social Psychology 49 (2013) 959–964

Contents lists available at SciVerse ScienceDirect

Journal of Experimental Social Psychology

journal homepage: www.elsevier.com/locate/jesp

Can the goal of honesty be primed?

Harold Pashler [a,*], Doug Rohrer [b], Christine R. Harris [a]

[a] University of California, San Diego, USA
[b] University of South Florida, USA

HIGHLIGHTS

- We attempted to replicate the "Honesty Priming" effect of Rasinski et al. (2005).
- Two direct replications failed to reproduce the effect.
- An additional conceptual replication using a broader outcome measure also failed.

ARTICLE INFO

Article history:
Received 3 May 2013
Revised 5 May 2013
Available online 4 June 2013

Keywords:
Honesty
Candor
Social desirability
Priming
Social priming

ABSTRACT

In a simple study involving 64 participants, Rasinski, Visser, Zagatsky, and Rickett (2005) reported that requiring people to make semantic judgments involving four words related to honesty (embedded among other words) increased the likelihood that they would later admit to having engaged in problematic alcohol-related behaviors (e.g., drinking to the point of blackout). If valid, this honesty-priming effect would offer a powerful intervention to improve the validity of self-report data in many different contexts. To determine whether the effect is repeatable, we first attempted two replications using the same materials, tasks, and measures used by Rasinski et al. Experiment 1 repeated the study with a sample of 150 students. No priming effects were observed here, nor in a follow-up study using adults recruited on the web (Experiment 2). Experiment 3 used the same priming manipulation together with a more refined measure of response candor (derived from Paulhus, 1991). Again, the honesty-related primes had no detectable effects.

© 2013 Elsevier Inc. All rights reserved.

Replication

Replication of "Experiencing Physical Warmth Promotes Interpersonal Warmth" by Williams and Bargh (2008)

Dermot Lynott,[1] Katherine S. Corker,[2] Jessica Wortman,[3] Louise Connell,[1]
M. Brent Donnellan,[3] Richard E. Lucas,[3] and Kerry O'Brien[4,5]

OPEN ACCESS Freely available online | PLOS | ONE

Priming of Social Distance? Failure to Replicate Effects on Social and Food Judgments

Harold Pashler*, Noriko Coburn, Christine R. Harris

University of California San Diego, La Jolla, California, United States of America

Abstract

Williams and Bargh (2008) reported an experiment in which participants were simply asked to plot a single pair of points on a piece of graph paper, with the coordinates provided by the experimenter specifying a pair of points that lay at one of three different distances (close, intermediate, or far, relative to the range available on the graph paper). The participants who had graphed a more distant pair reported themselves as being significantly less close to members of their own family than did those who had plotted a more closely-situated pair. In another experiment, people's estimates of the caloric content of different foods were reportedly altered by the same type of spatial distance priming. Direct replications of both results were attempted, with precautions to ensure that the experimenter did not know what condition the participant was assigned to. The results showed no hint of the priming effects reported by Williams and Bargh (2008).

Citation: Pashler H, Coburn N, Harris CR (2012) Priming of Social Distance? Failure to Replicate Effects on Social and Food Judgments. PLoS ONE 7(8): e42510. doi:10.1371/journal.pone.0042510

Editor: Jan Lauwereyns, Kyushu University, Japan

Received May 13, 2012; **Accepted** July 6, 2012; **Published** August 29, 2012

Copyright: © 2012 Pashler et al. 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 author and source are credited.

Funding: This work was supported a collaborative activity grant from the James S. McDonnell Foundation and by the National Science Foundation (Grant SBE-0820542013 to the University of California San Diego Temporal Dynamics of Learning Center). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: hpashler@ucsd.edu

Reproducibility Project (Open Science Collaboration, 2015)

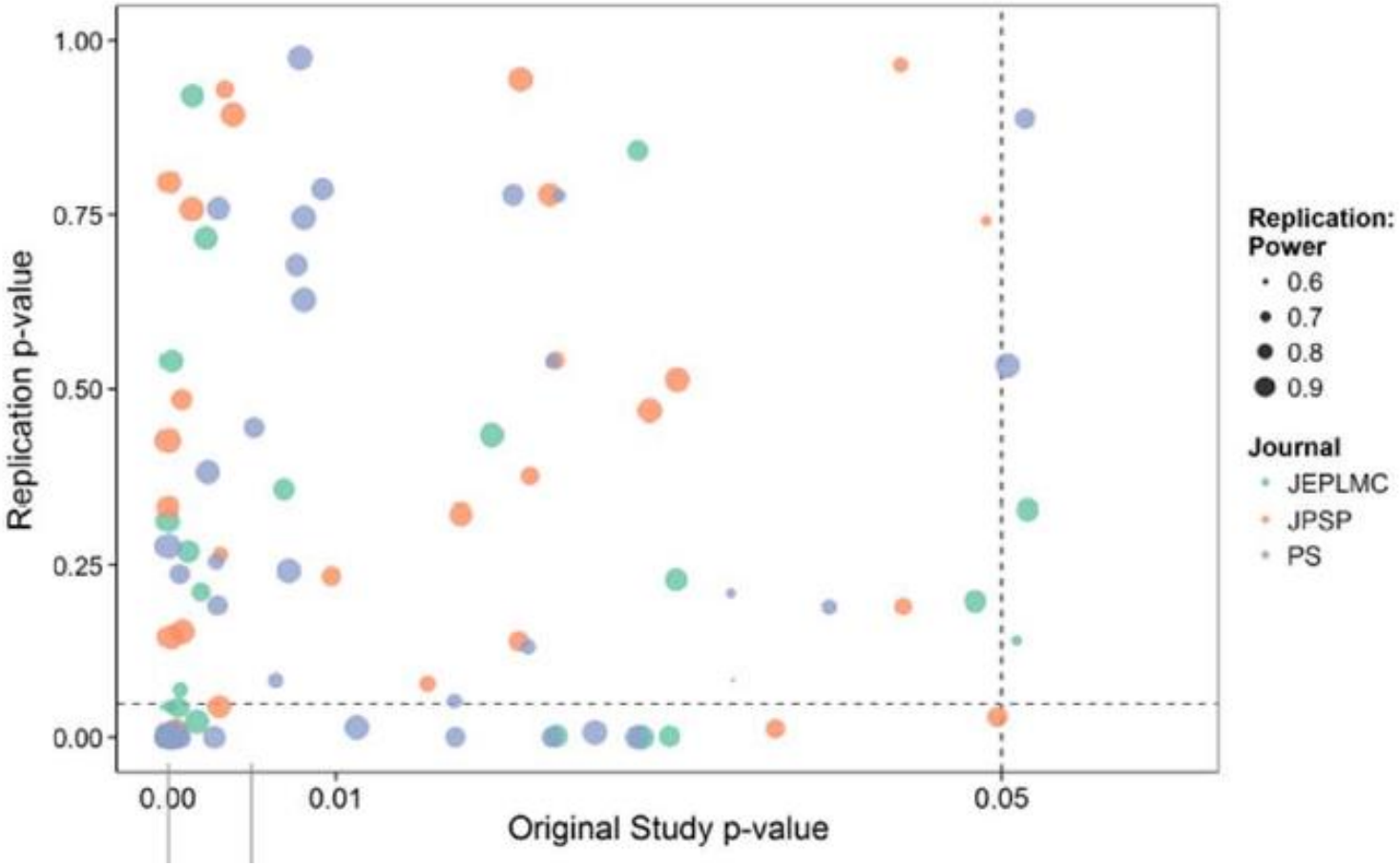
97 studies taken from 2008 volume of *Journal of Personality and Social Psychology, Psychological Science,* and *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Exact replications attempted by academic crowd sourcing.

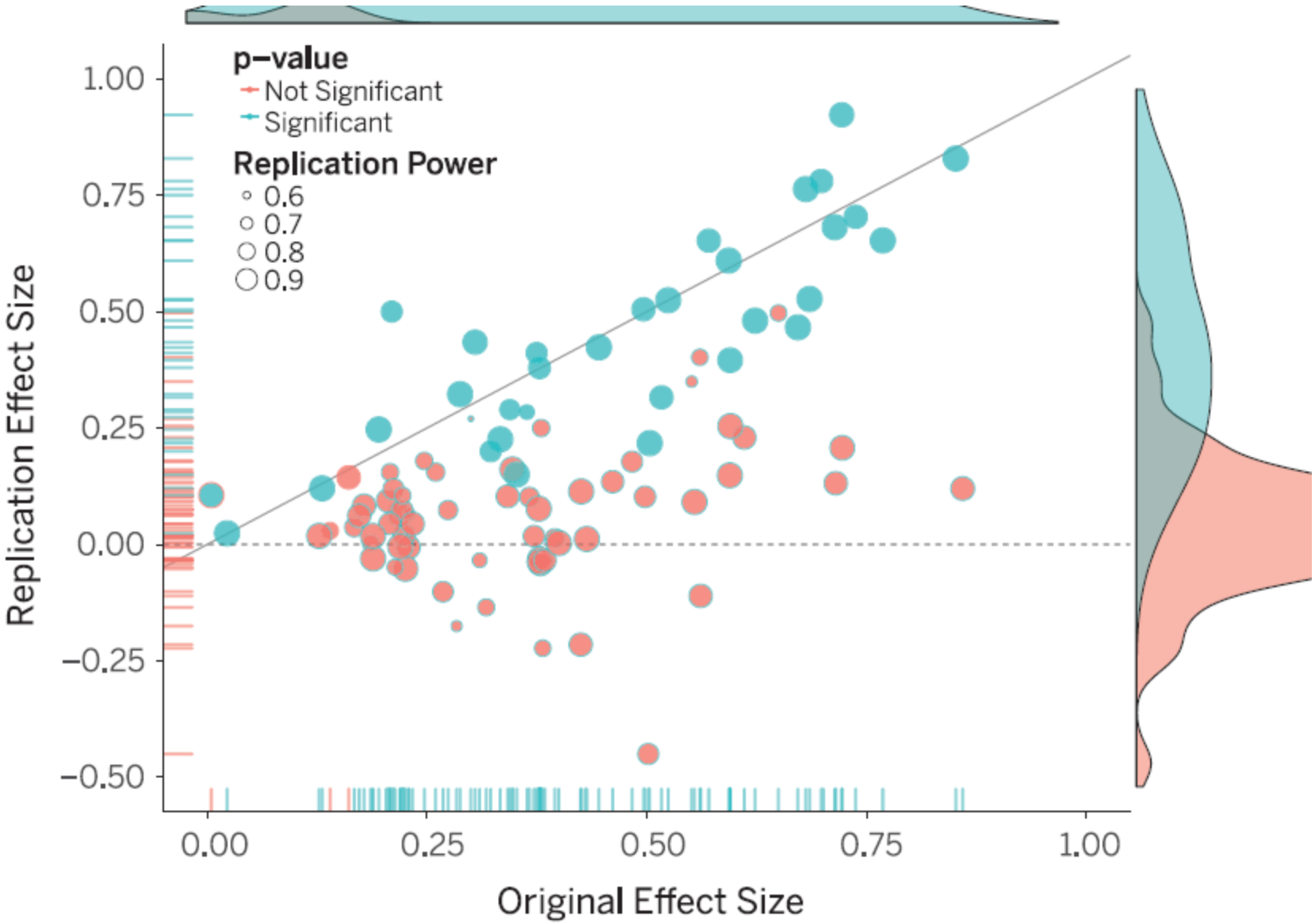
What proportion of replications will be significant, given average power 92%?

What proportion of replications will be significant, given power 92%?

35/97 replications significant, i.e. 36%



Social 14/55 (25%)
Cognitive 21/42 (50%)


p−value
Not Significant
Significant
Replication Power
0.6
0.7
0.8
0.9
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
Replication Effect Size
0.00
0.25
0.50
0.75
1.00
Original Effect Size

What can be done?

i. Pre-registration (https://osf.io/8mpji/wiki/home/)
Change incentives to get quality of science right independent of results.
(*Cortex* ….and about 20 other journals now)

ii. Publication of direct replications in prestige journals

iii. Transparency: Full online easy public availability of materials, programs, raw data, analyses scripts. (Cf. Open Science Framework: https://osf.io/)

iv.Statistical reform

v.Institutional reform

What can be done?

i.Pre-registration

Why pre-registration?

Science has an incentive problem

What's best for science

Transparent and high quality research, regardless of outcome

What's best for scientists

Producing a lot of "good results"

see Nosek, Spies & Motyl (2012). *Perspectives on Psychological Science, 7(6)*: 615–631

Because we place too much importance on the **results** of experiments and not enough on the **processes** that produce them

Results make science exciting but judging the quality of science (and scientists) according to the results condemns us to being a "soft" science

Can we fix this? Yes

Philosophy:

What gives hypothesis-testing its scientific value is:

- the QUESTION it asks
- the QUALITY of the method it uses
- not the RESULT it produces

If we accept this philosophy then editorial decisions at journals should be *blind* to results

Registered Reports

CORTEX 49 (2013) 609–610

Available online at www.sciencedirect.com

SciVerse ScienceDirect

Journal homepage: www.elsevier.com/locate/cortex

Editorial

Registered Reports: A new publishing initiative at *Cortex*

Christopher D. Chambers

Cardiff University Brain Research Imaging Centre (CUBRIC), School of Psychology, Cardiff University, United Kingdom

Four central aspects of the Registered Reports model:

- Researchers decide hypotheses, experimental procedures, and main analyses *before* data collection
- Part of the peer review process takes place before experiments are conducted
- Passing this stage of review virtually guarantees publication
- Original studies and high-value replications are welcome

Cortex

Available online 13 February 2015

In Press, Uncorrected Proof — Note to users

Registered report

The P600 as a correlate of ventral attention network reorientation

Jona Sassenhagen[a], Ina Bornkessel-Schlesewsky[b, a]

Show more

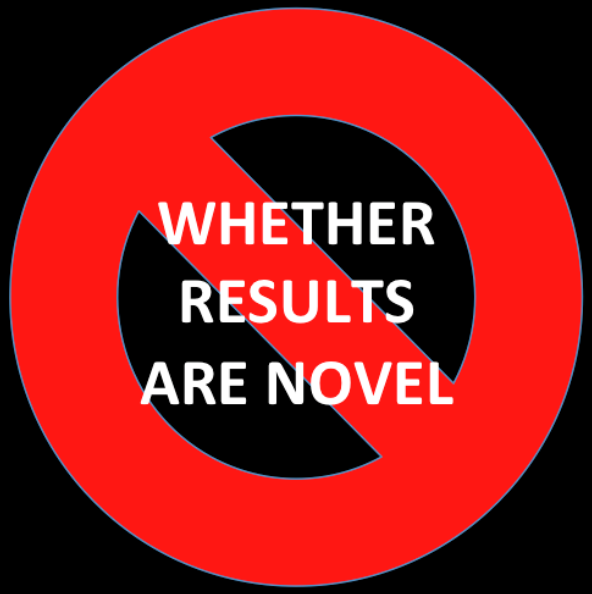
doi:10.1016/j.cortex.2014.12.019

Abstract

When, during language processing, a reader or listener is confronted with a structurally deviant phrase, this typically elicits a late positive ERP deflection (P600). The P600 is often understood as a correlate of structural analysis. This assumption has informed a number of neurocognitive models of language. However, the P600 strongly resembles the P3, likely a more general electrophysiological correlate of reorientation behaviour supported by noradrenergic input to the ventral attention network/VAN. Some researchers have proposed that the P600 is an instance of the P3, not a distinct component reflecting the analysis of structured inputs. Here, we tested the P600-as-P3 hypothesis by estimating the alignment of the P600 elicited in a visual sentence processing task to simultaneously collected behavioural measures. A similar analysis was undertaken for a P3 elicited in a separate non-linguistic (face detection) task. Since the P3 is usually aligned to reaction time/RT, the same should hold for the P600; a failure to find RT alignment of the P600 would pose a problem for the P600-as-P3 hypothesis. In contrast, RT alignment of the P600 would associate it with the well-established VAN/Locus Coeruleus – Noradrenaline – Network subserving cortical reorientation.

We failed to falsify the hypothesis of RT alignment. Secondary measures, while less unambiguous, were more in agreement with the P600-as-P3 hypothesis. We interpret our results as corroborating the hypothesis that the P600 is a P3, in that it shows that the P600 is RT-aligned. This perspective is

None of these things matter



What is the evidence the pre-registration even works? Does it reduce publication bias or improve study quality?

- At least one journal has pre-registered a trial of Registered Reports

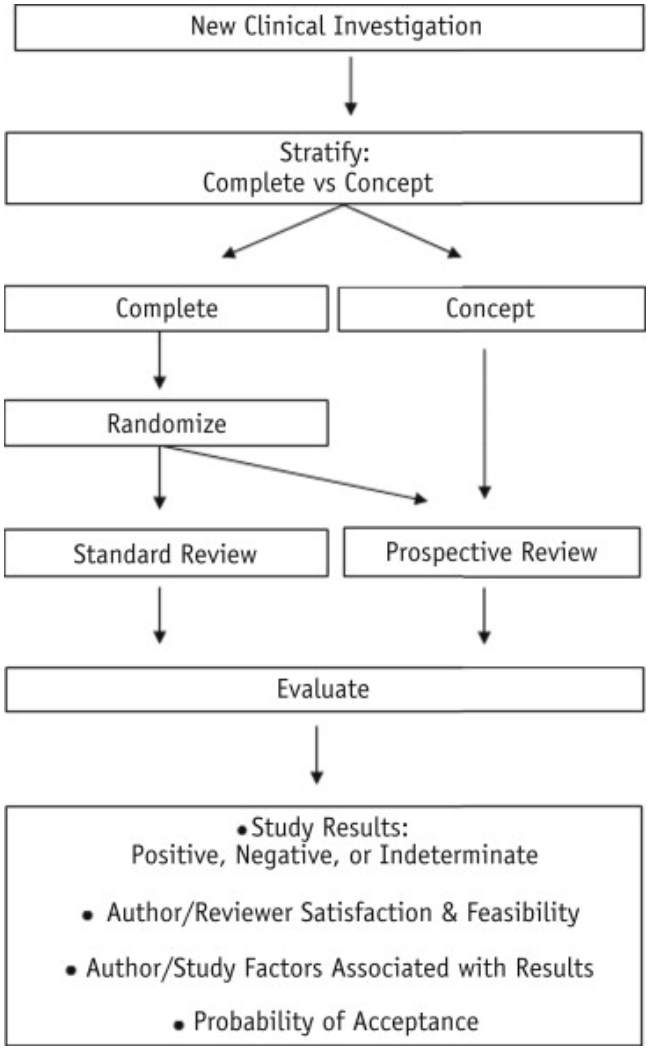
Introducing Prospective Manuscript Review to Address Publication Bias

Loren K. Mell, MD,* and Anthony L. Zietman, MD†

**Department of Radiation Medicine and Applied Sciences, University of California San Diego, La Jolla, California; and †Department of Radiation Oncology, Massachusetts General Hospital, Boston, Massachusetts*

International Journal of Radiation Oncology biology • physics

www.redjournal.org



http://www.redjournal.org/article/S0360-3016%2814%2903596-2/fulltext

Number of studies finding medical interventions effective before preregistration introduced:
17/30 (55%)

Afterwards:

PLOS | ONE

Publish | About | Browse

OPEN ACCESS PEER-REVIEWED

RESEARCH ARTICLE

Likelihood of Null Effects of Large NHLBI Clinical Trials Has Increased over Time

Robert M. Kaplan, Veronica L. Irvin

Published: August 5, 2015 • DOI: 10.1371/journal.pone.0132382

Article | Authors | Metrics | Comments | Related Content

Abstract

Background

We explore whether the number of null results in large National Heart Lung, and Blood Institute (NHLBI) funded trials has increased over time.

Methods

We identified all large NHLBI supported RCTs between 1970 and 2012 evaluating drugs or

Number of studies finding medical interventions effective before preregistration introduced:
17/30 (55%)

Afterwards:
2/25 (8%)

PLOS | ONE

Publish | About | Browse

OPEN ACCESS PEER-REVIEWED

RESEARCH ARTICLE

Likelihood of Null Effects of Large NHLBI Clinical Trials Has Increased over Time

Robert M. Kaplan, Veronica L. Irvin

Published: August 5, 2015 • DOI: 10.1371/journal.pone.0132382

Article | Authors | Metrics | Comments | Related Content

Abstract

Background

We explore whether the number of null results in large National Heart Lung, and Blood Institute (NHLBI) funded trials has increased over time.

Methods

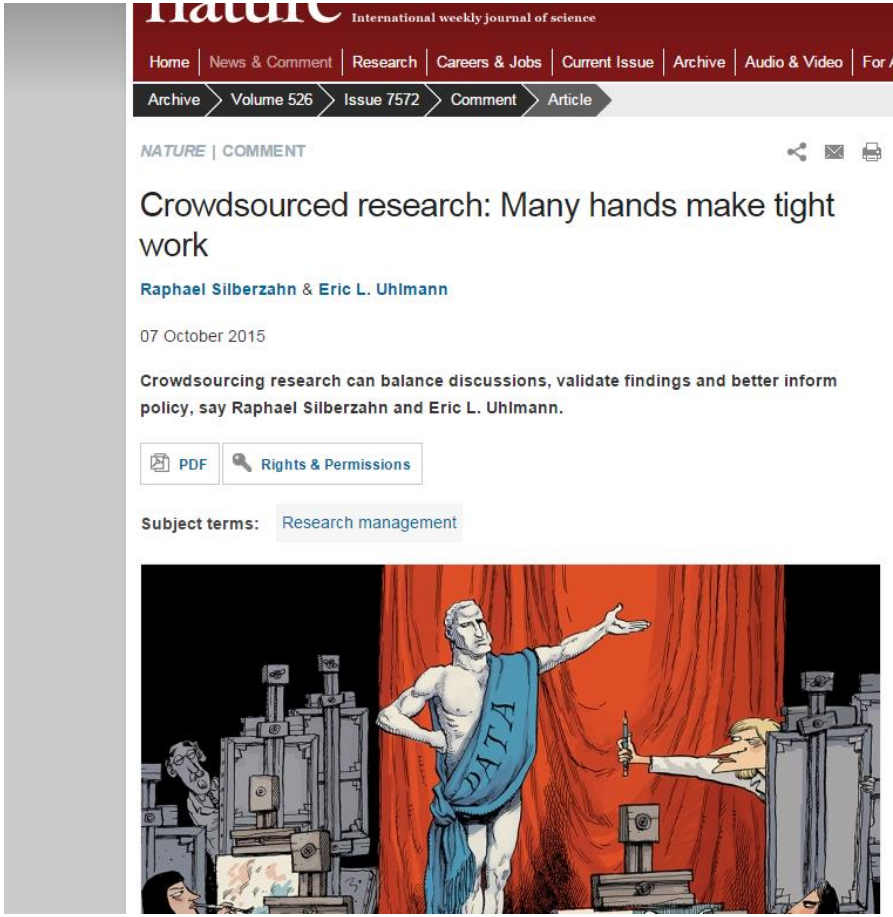
We identified all large NHLBI supported RCTs between 1970 and 2012 evaluating drugs or

Variant:
Create a number of datasets by re-arranging labels, adding a bit of noise or offset etc.
Analyst does not know which is real data set.
Work out analysis protocol so that it is solid
Break blind - and bingo unbiased analyses!

Further variant in analyses:

Have different team analyse the same data in their own way.

Submission type for a journal?



Nature International weekly journal of science

Home | News & Comment | Research | Careers & Jobs | Current Issue | Archive | Audio & Video | For A

Archive > Volume 526 > Issue 7572 > Comment > Article

NATURE | COMMENT

Crowdsourced research: Many hands make tight work

Raphael Silberzahn & Eric L. Uhlmann

07 October 2015

Crowdsourcing research can balance discussions, validate findings and better inform policy, say Raphael Silberzahn and Eric L. Uhlmann.

PDF | Rights & Permissions

Subject terms: Research management

Encourage **adversarial collaborations** (Wagenmakers):

People with opposite views run a replication/test, both collecting data and both analysing the data

What can be done?

i. Pre-registration

ii. Publication of direct replications in prestige journals

Journals should be obliged to published direct replications with high sensitivity?

Pre-registration should encourage direct replications.

What can be done?

i. Pre-registration

ii. Publication of direct replications in prestige journals

iii. Transparency

What can be done?

i. Pre-registration

ii. Publication of direct replications in prestige journals

iii. Transparency

Peer Reviewers' Openness Initiative

WELCOME | THE INITIATIVE | JOIN THE INITIATIVE | GUIDELINES FOR REVIEWERS | GUIDELINES FOR AUTH

Supporting the spread of open research practices

Search ...

RECENT COMMENTS

Zoltan Dienes on Join the Initiative

Sonali Marwaha on Join the Initiative

Richard Morey on Join the Initiative

Sonali Marwaha on Join the Initiative

Jeff Rouder on Join the Initiative

The open research revolution begins June 1, 2016. Will you join us?

We believe that openness and transparency are core values of science. For a long time, technological obstacles existed preventing transparency from being the norm. With the advent of the internet, however, these obstacles have largely disappeared. The promise of open research can finally be realized, but this will require a cultural change in science. The power to create that change lies in the peer-review process.

We suggest that beginning June 1, 2016, **reviewers make open practices a pre-condition for more comprehensive review.** This is

American Journal of Political Science

All analyses checked before publication on submitted datasets

ajps American Journal of Political Science

HOME ABOUT SUBSCRIBE/JOIN AJPS ARTICLES SUBMIT REVIEW SUBMIT MANUSCRIPT REPLICATION CONTACT

The AJPS is the flagship journal of the MPSA. Founded in 1939, the MPSA is dedicated to the advancement of scholarship in all areas of political science. Each April, the MPSA holds one of the largest academic political science conferences.

GUIDELINES FOR MANUSCRIPTS

INSTRUCTIONS FOR SUBMITTING AUTHORS

The *American Journal of Political Science* (*AJPS*) seeks manuscripts that make outstanding contributions to scholarly knowledge about notable theoretical concerns, empirical issues, or methodological strategies in any subfield of political science. The *AJPS* will not review manuscripts

Search this site...

ARCHIVES

Select Month

RECENT POSTS

- The Voters' Curses: Why we need

What can be done?

i. Pre-registration

ii. Publication of direct replications in prestige journals

iii. Transparency

iv. Statistical reform

Now is a time of ferment of exploring different statistical approaches …

Journals should be open to new approaches to help the exploration

A key debate:

Orthodoxy (significance testing) versus Bayes

Evidence for H0
(no difference)

No evidence to
speak of

Evidence for H1
(some difference)

P-values make a two-way distinction:

Evidence for H0
(no difference)

No evidence to
speak of

Evidence for H1
(some difference)

P-values make a two distinction:

Evidence for H0
(no difference)

No evidence to
speak of

Evidence for H1
(some difference)

NO MATTER WHAT THE P-VALUE, NO DISTINCTION MADE WITHIN THIS BOX

Bayes factors make the three way distinction:

0 … 1/3	1/3 … 3	3 …
Evidence for H0 (no difference)	No evidence to speak of	Evidence for H1 (some difference)

www.lifesci.sussex.ac.uk/home/Zoltan_Dienes/inference/bayes_factor.swf

Apps AGR'O Cal MyView Add Bayes post. R'der Wish Am

Calculate your Bayes factor

Is the distribution of p(population value|theory) uniform? yes no

What is the sample standard error?

What is the sample mean?

Go!

The likelihood of the obtained data given your theory is

The likelihood of the obtained data given the null is

The Bayes factor is

frontiers in Psychology | Quantitative Psychology and Measurement

‹ Archive

METHODS ARTICLE

Front. Psychol., 29 July 2014 | http://dx.doi.org/10.3389/fpsyg.2014.00781

Using Bayes to get the most out of non-significant results

Zoltan Dienes*

School of Psychology and Sackler Centre for Consciousness Science, University of Sussex, Brighton, UK

ientific conclusion follows automatically from a statistically non-significant result, yet people routinely use non-
icant results to guide conclusions about the status of theories (or the effectiveness of practices). To know whether a
ignificant result counts against a theory, or if it just indicates data insensitivity, researchers must use one of: power,
als (such as confidence or credibility intervals), or else an indicator of the relative evidence for one theory over
er, such as a Bayes factor. I argue Bayes factors allow theory to be linked to data in a way that overcomes the
nesses of the other approaches. Specifically, Bayes factors use the data themselves to determine their sensitivity in
guishi often easiest to
y (unli ess theory).
factor pothesis over a

JASP

A Fresh Way to Do Statistics

HOME ABOUT WORKSHOPS DOWNLOAD TEACHING JASP DEVELOPMENT

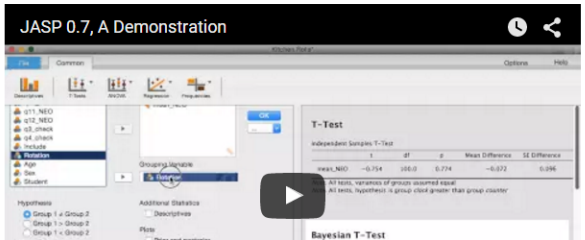
JASP

JASP, a low fat alternative to SPSS, a delicious alternative to R. Bayesian statistics made accessible.

Version 0.7.1.12 available now!

Follow us on twitter! follow us on Facebook!

Read the latest article on JASP and interview with Jonathon at JEPS



Publications ▼ Bayes Factor Calculators ▼ Code/Data/Stimuli For Lab Members

Bayes Factor Calculators

Our goal is to provide a convenient set of web-based Bayes factor calculators. Currently, we have imp
calculators for:

- Paired or one-sample t-tests
- Grouped or two-sample t-tests
- Regression
- Paired t-tests where the null is an equivalence region rather than a point
- Binomially Distributed Observation

News: Release of Bayes Factor Package

Nosek et al, *Science* 2015

	LEVEL 0	LEVEL 1	LEVEL 2	LEVEL 3
Citation standards	Journal encourages citation of data, code, and materials—or says nothing.	Journal describes citation of data in guidelines to authors with clear rules and examples.	Article provides appropriate citation for data and materials used, consistent with journal's author guidelines.	Article is not published until appropriate citation for data and materials is provided that follows journal's author guidelines.
Data transparency	Journal encourages data sharing—or says nothing.	Article states whether data are available and, if so, where to access them.	Data must be posted to a trusted repository. Exceptions must be identified at article submission.	Data must be posted to a trusted repository, and reported analyses will be reproduced independently before publication.
Analytic methods (code) transparency	Journal encourages code sharing—or says nothing.	Article states whether code is available and, if so, where to access them.	Code must be posted to a trusted repository. Exceptions must be identified at article submission.	Code must be posted to a trusted repository, and reported analyses will be reproduced independently before publication.
Research materials transparency	Journal encourages materials sharing—or says nothing	Article states whether materials are available and, if so, where to access them.	Materials must be posted to a trusted repository. Exceptions must be identified at article submission.	Materials must be posted to a trusted repository, and reported analyses will be reproduced independently before publication.
Design and analysis transparency	Journal encourages design and analysis transparency or says nothing.	Journal articulates design transparency standards.	Journal requires adherence to design transparency standards for review and publication.	Journal requires and enforces adherence to design transparency standards for review and publication.

Summary of the eight standards and three levels of the TOP guidelines

	LEVEL 0	LEVEL 1	LEVEL 2	LEVEL 3
Preregistration of studies	Journal says nothing.	Journal encourages preregistration of studies and provides link in article to preregistration if it exists.	Journal encourages preregistration of studies and provides link in article and certification of meeting preregistration badge requirements.	Journal requires preregistration of studies and provides link and badge in article to meeting requirements.
Preregistration of analysis plans	Journal says nothing.	Journal encourages preanalysis plans and provides link in article to registered analysis plan if it exists.	Journal encourages preanalysis plans and provides link in article and certification of meeting registered analysis plan badge requirements.	Journal requires preregistration of studies with analysis plans and provides link and badge in article to meeting requirements.
Replication	Journal discourages submission of replication studies—or says nothing.	Journal encourages submission of replication studies.	Journal encourages submission of replication studies and conducts blind review of results.	Journal uses Registered Reports as a submission option for replication studies with peer review before observing the study outcomes.

Levels 1 to 3 are increasingly stringent for each standard. Level 0 offers a comparison that does not meet the standard.

What can be done?

i. Pre-registration

ii. Publication of direct replications in prestige journals

iii. Transparency

iv. Statistical reform

v. Institutional reform

Why the credibility crisis now?

Maybe performance management based on simplistic criteria (Key Performance Indicators): getting papers published, getting grants …..

Change in last 10 years: A separate management class who are only incentivized to apply strong top down pressure to achieve KPIs

What if we implemented an Athenian democracy at a university so that academics – who care about research quality – made the decisions …??