

Making Replication Mainstream

Rolf A. Zwaan

Erasmus University Rotterdam

rolfzwaan@gmail.com

@rolfzwaan

<https://rolfzwaan.blogspot.nl/>



Behavioral and Brain Sciences

Article

Metrics

October 2017 , pp. 1-50

MAKING REPLICATION MAINSTREAM

Rolf A. Zwaan ^(a1), Alexander Etz ^(a1), Richard E. Lucas ^(a1) and M. Brent Donnellan ^(a1) 

<https://doi.org/10.1017/S0140525X17001972> Published online: 25 October 2017



Starting Assumptions...

- Scientific research is important. Research helps us better understand the world and often has practical implications.
- The credibility of scientific claims rests, in part, on transparency and independent verification.
- Knowledge accumulates incrementally and mistakes happen. Progress is faster when accuracy is a priority.

Threats to Accuracy

- *False Positives (Type I Errors)* – Inferring a result is legitimate when it is a fluke
- *False Negatives (Type II Errors)* - Missing (often) small but nonetheless “real” effects
- *Estimation Errors: Type M (magnitude) and Type S (sign) errors* (Gelman & Tuerlinckx, 2000; Gelman & Carlin, 2014)

HARKing

Low power

P-hacking

Publication bias



Credit: Dorothy Bishop

Progressive vs. Degenerative Research Programmes (Lakatos)



Progressive vs. Degenerative Research Programmes (Lakatos)

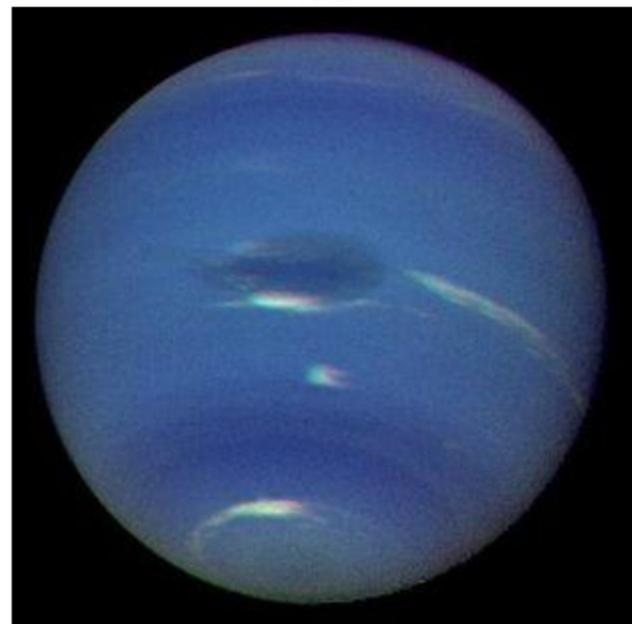
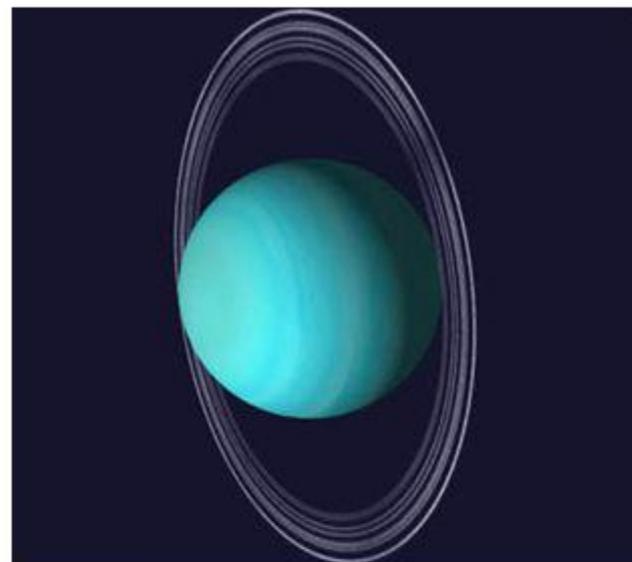
- Hard core of theoretical assumptions
- Protective belt of auxiliary hypotheses (which are actually tested)
- Can fail without destroying entire programme

Progressive vs. Degenerative Research Programmes (Lakatos)

- Progressive research programmes
 - testing auxiliary hypotheses increases explanatory power.
- Degenerative research programmes
 - Changes auxiliary hypotheses to fend off contradictory evidence.

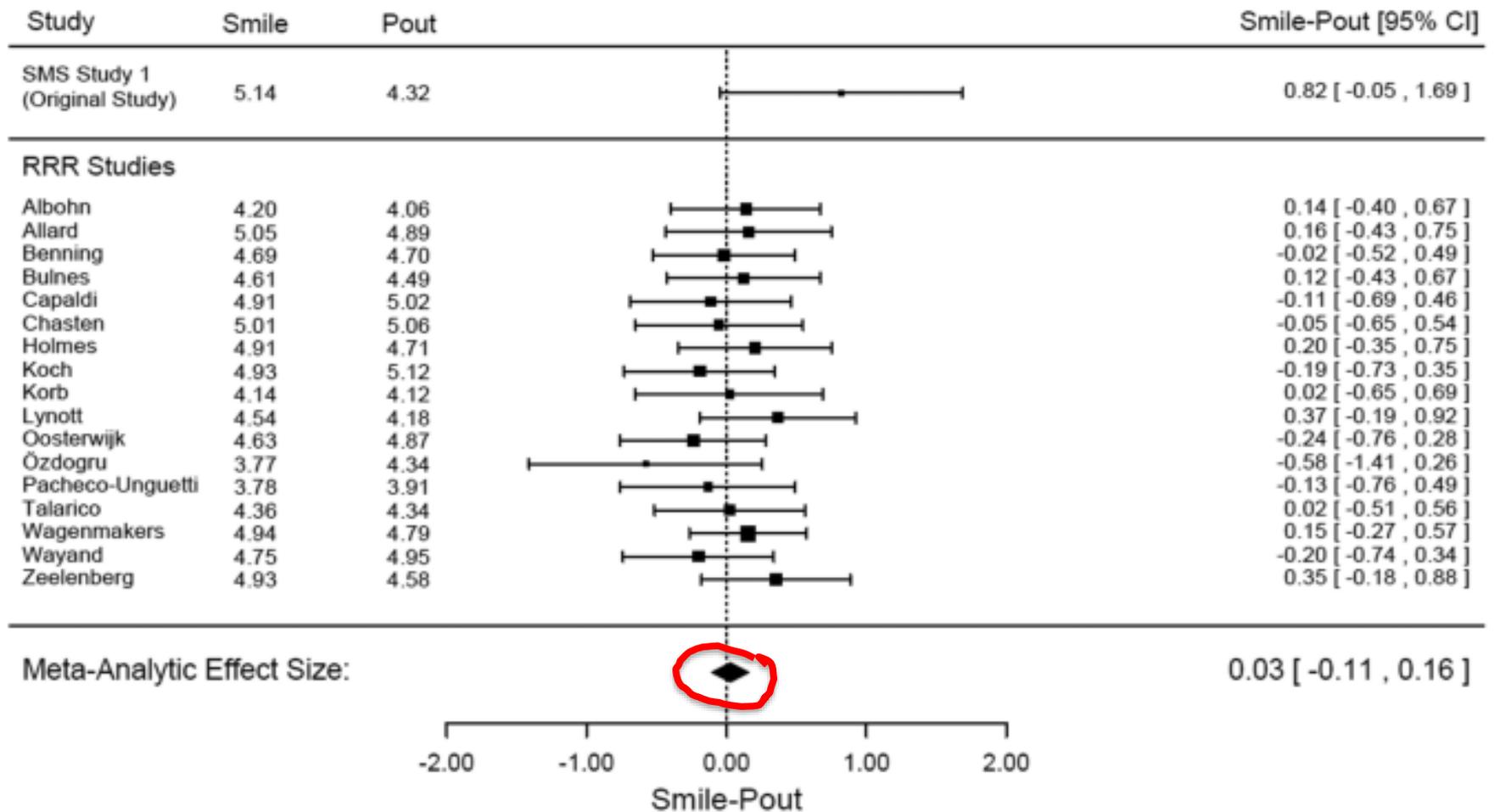
Example: Neptune

- Astronomers discovered that the orbit of Uranus didn't match Newton's predictions
- They did NOT give up Newtonian physics
- They DID add a new item to the protective belt: something else must be "perturbing" the orbit of Uranus
- This turned out to be Neptune: Progressive change to research program
- What if...no Neptune? Could hypothesize that some unobservable force acts only on Uranus → no new predictions = degenerative shift



Example: Facial Feedback





“The proof established by the test must have a specific form, namely, repeatability. The issue of the experiment must be a statement of the hypothesis, the conditions of test, and the results, in such form that another experimenter, from the description alone, may be able to repeat the experiment. Nothing is accepted as proof, in psychology or in any other science, which does not conform to this requirement.” (Dunlap, 1926).



Knight Dunlap

THE EXPERIMENTAL METHODS OF PSYCHOLOGY*

KNIGHT DUNLAP



Replication Terminology

- Design Considerations – *Did I do the same thing?*
 - Direct Replication – Same Method/Procedure
 - Exact repetition of the procedures – *Impossible?*
 - Repetition of the **Critical elements**
 - Goal: Evaluate the reliability of a specific effect
 - Conceptual Replication – Different Method/Procedures
 - Evaluate theory in a novel way
 - Goal: Test and Extend generalizability

Replication Terminology

- Outcome Considerations – *Did I get the same results?*
 - *What statistical tools will be used to evaluate this question?*

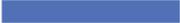
FRAUD



*How Confident Are You that the
Majority of Findings In Your Area
Will Replicate?*

1=Not at All Confident to 4 = Very Confident

How confident are you that the majority of findings in social psychology will replicate?

#	Answer	Bar	Response	%
1	Not at all confident		347	29.56%
2	Slightly confident		443	37.73%
3	Moderately confident		345	29.39%
4	Very confident		39	3.32%
	Total		1174	100.00%

Min Value	Max Value	Average Value	Variance	Standard Deviation	Total Responses	Total Respondents
1	4	2.06	0.72	0.85	1174	1174

Results as of 10:43 am on 25 Feb 2015:

https://uic.qualtrics.com/WRReport/?RPID=RP2_bjhdTM0IAkWZMk5&P=CP

Reproducibility Project: Psychology

RESEARCH ARTICLE

PSYCHOLOGY

Estimating the reproducibility of psychological science

Open Science Collaboration*†

Reproducibility is a defining feature of science, but the extent to which it characterizes current research is unknown. We conducted replications of 100 experimental and correlational studies published in three psychology journals using high-powered designs and original materials when available. Replication effects were half the magnitude of original effects, representing a substantial decline. Ninety-seven percent of original studies had statistically significant results. Thirty-six percent of replications had statistically significant results; 47% of original effect sizes were in the 95% confidence interval of the replication effect size; 39% of effects were subjectively rated to have replicated the original result; and if no bias in original results is assumed, combining original and replication results left 68% with statistically significant effects. Correlational tests suggest that replication success was better predicted by the strength of original evidence than by characteristics of the original and replication teams.

How Reliable Are Psychology Studies?

A new study shows that the field suffers from a reproducibility problem, but the extent of the issue is still hard to nail down.

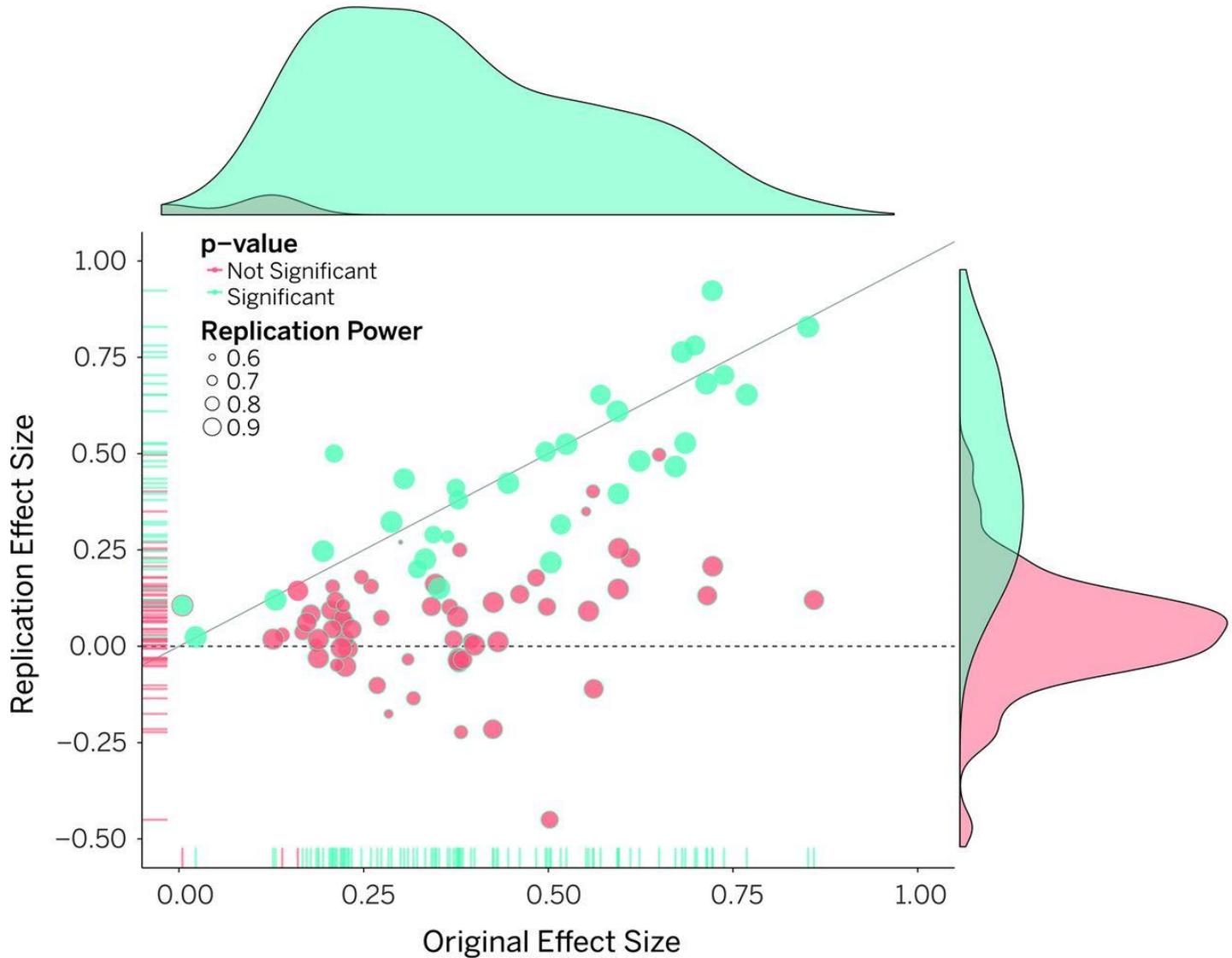


Lali Masriera / Flickr

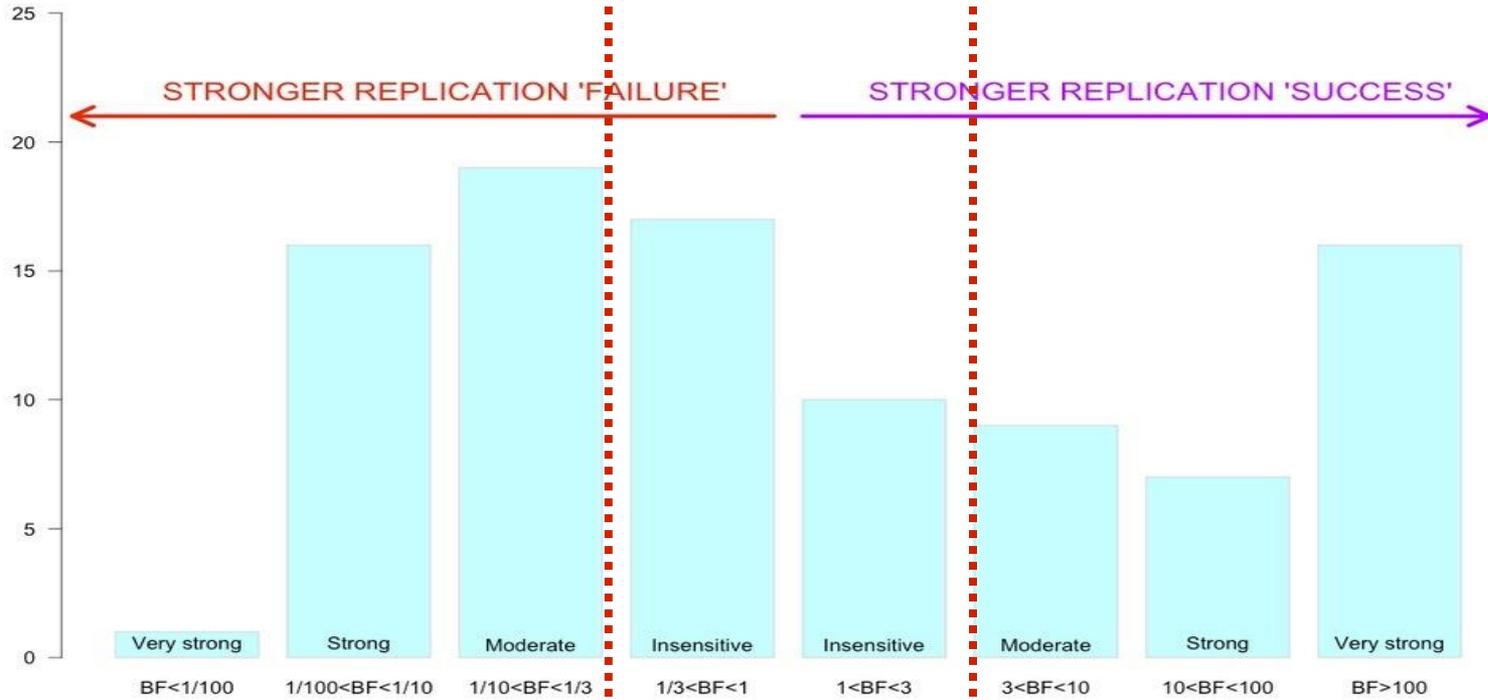
ED YONG
AUG 27, 2015

No one is entirely clear on how Brian Nosek pulled it off, including Nosek himself. Over the last three years, the psychologist from the University of Virginia persuaded some 270 of his peers to channel their free time into [repeating 100 published psychological experiments](#) to see if they could get the same results a second time around. There would be no glory, no

Original study effect size versus replication effect size (correlation coefficients).



Bayesian Replication Outcomes from the Reproducibility Project: Psychology



38%

FAIL

28%

INCONCLUSIVE

34%

SUCCESS

Concerns About Direct Replication

- 1: Experimental context is too variable
- 2: Direct replications have limited theoretical value
- 3: Replications are not feasible in certain domains
- 4: Replications are a distraction
- 5: Replications affect reputations
- 6: There is no standard way to assess replications

1: “Experimental context
is too variable”

Response

Context is important and theoretically interesting. Articulate **constraints on generality** (Simons et al., 2017).

Raise reporting standards. Preregister within lab replications. ***Be Your Own Replicator.***

Response

Strack & Stroebe commentary:

... replicators must take the effort of identifying the actual determinants of the original finding if they believe that it was not caused by the factors claimed in the original study.

We disagree

2: “Direct replications have limited theoretical value”

Response:

- direct vs. conceptual = false dichotomy
- direct necessary before moving on to conceptual .

Response

The combined effects of researcher degrees of freedom, chance findings from small sample studies, and the existence of publication bias mean that it is possible to assemble a seemingly solid set of studies that appear to support an underlying theory, even though no single study from that set could survive a direct replication attempt.



3: “Replications are not feasible in certain domains”

- such as large-scale observational and clinical epidemiological studies (Coyne, 2016).
- certain studies may capitalize on extremely rare events (e.g., occurrence of a natural disaster or an astronomical event), and replicating studies that test the effects of these events is simply impossible.

Response

Concerns about feasibility are orthogonal to the overarching value of direct replications for advancing scientific knowledge. The fact that replication studies are not always possible does not undermine their value when they can be conducted.

StudySwap: A platform for interlab replication, collaboration, and research resource exchange



Title ^ v	Author ^ v	Category ^ v	Date Created ^ v	Downloads ^ v
PPIR Needed. How powerful are focal points? Label salience and payoff asymmetry ...	Chartier	need	2017-03-10 03:06 PM	 55
[complete] Short negotiation study - data can be collected online!	Schweinsberg	need	2017-03-22 08:42 PM	 34
[in progress] 100 In-Lab Participants for 30 Minutes	Chartier	have	2017-03-01 06:01 PM	 53
[complete] McCarthy-have ~200 Intro Psych Students	McCarthy	have	2017-03-17 03:28 PM	 56
Needed: Collaborator with experience in computational modelling (connectionist m...	Schmalz	need	2017-05-30 11:39 AM	 23
[in progress] Short scenario negotiation study - data can be collected online!	Schweinsberg	need	2017-05-31 10:32 AM	 40
[in progress] Cross-cultural Study of Effects of Teacher Characteristics on Academic ...	Karunagharan	need	2017-06-20 08:37 AM	 19
East-West Data Collection Collaborations: Thai Participants	Manley	have	2017-06-24 06:50 AM	 25
German participants needed for my Bachelor Thesis: rating of journalistic excerpts (...	Kahre	need	2017-07-14 05:22 PM	 18
Call for Papers: Environmental Engagement and Cultural Value: Global Perspectives ...	Jia	need	2017-08-08 07:29 PM	 12
pipeline project 2	Schweinsberg	need	2017-08-09 11:29 PM	 24
[complete] Needed: Data collection partners for Many Labs 5 (ML5) replication of Al...	Corker	need	2017-08-14 07:24 PM	 18
The efficacy of self-guided virtual reality exposure therapy with biofeedback for soci...	Daly	have	2017-08-16 01:51 PM	 29
~100 in-lab participants for 40 minutes	Simons	have	2017-08-28 09:12 PM	 20
International research on figurative language// Native English and Spanish speakers...	Vernarská	need	2017-10-12 09:45 AM	 27
Simulating parties in the lab	Gerber	need	2017-11-27 10:25 PM	 3
StudySwap live at #SPSP2018: Fostering Collaborations to Reach Hard-to-Reach Pon	Wilson	have	2017-12-19 09:53 PM	

4: “Replications are a distraction”

“[o]ur position is that the current obsession with replication is a red herring, distracting attention from the real threats to the validity of cumulative knowledge in the behavioral sciences” (Schmidt & Oh, 2016).

Response

- Replication is but one tool in the kit. It might become less important as practices evolve. *It won't solve everything, but it might improve a few things. Hard to imagine a science without it.*

5: “Replications affect reputations”

- Reputation of the “replicatee.”
- Reputation of the replicator.
 - Roy Baumeister: *replication crisis has created “a career niche for bad experimenters.”*

Response

Replicators should go out of their way to describe their results carefully, objectively, and without exaggeration about the implications for the original work. In addition, those whose studies are the focus of replication attempts should give replicators the benefit of the doubt when considering the contribution of the replication study and the replicators' motivations.

Response

As more replications are conducted, the experience of having a study fail to replicate will become more normative and hopefully less unpleasant.

Response

Gordon Pennycook's commentary: "You are not your data!"

Response

Commentary Pennycook: “You are not your data!”

Commentary Vazire & Tullett: “Doing science is like jointly solving a jigsaw puzzle.”

6. “There is no standard way to assess replications.”

Response

Given the multiple options available, investigators should consider multiple approaches and also consider pre-registering analytic plans and committing to how evidence will be interpreted before analyzing the data. Inferences that are robust across approaches are more likely to be more scientifically defensible.

Response

- “Small telescopes” approach (Simonsohn, 2015).

Imagine an astronomer claiming to have found a new planet with a telescope. Another astronomer tries to replicate the discovery using a larger telescope and finds nothing. Although this does not prove that the planet does not exist, it does nevertheless contradict the original findings, because planets that are observable with the smaller telescope should also be observable with the larger one.

Response

- “Small telescopes” approach (Simonsohn, 2015).
- *The idea is to consider what effect size the original study would have 33% power to detect and then use this value as a benchmark for the replication study. If the 90% confidence interval from the replication study excludes this value then we say the original study could not have meaningfully examined this effect.*

Response

- “replication Bayes factor” approach (Ly, Etz, Marsman, Wagenmakers, 2017; Verhagen & Wagenmakers, 2014; Wagenmakers, Verhagen, Ly, 2016).

Closing Thoughts

- There are no serious concerns about making direct replication mainstream.
- Direct and conceptual replication both have a role to play.
- A failure to replicate does not falsify a theory.
- Direct replication is necessary but not sufficient to improve the accuracy of our science.
- Preregistration is also key.

A Few Plugs



**SOCIETY FOR THE
IMPROVEMENT OF
PSYCHOLOGICAL SCIENCE**

[Home](#)

[Mission](#)

[About SIPS](#)

[Meetings](#)

[Committees](#)

[Become a Member!](#)

What is SIPS?

The Society for the Improvement of Psychological Science (SIPS) brings together scholars working to improve methods and practices in

Upcoming SIPS Meeting

Registration is **NOW OPEN** for the next SIPS meeting, happening **June 24 to 26, 2018 in Grand Rapids,**

Become a Member!

[Click here to join SIPS!](#) Anyone interested in improving psychological research is welcome to join, regardless of experience.

A Few Plugs

<https://improvingpsych.org/>

A Few Plugs



SOCIETY FOR THE
IMPROVEMENT OF
PSYCHOLOGICAL SCIENCE

[Program](#) [Travel And Hotels](#) [Social Events](#) [Contact](#)



SIPS 2018

Grand Rapids, MI: June 24-26, 2018

A Few Plugs

Next year: Netherlands

A Few Plugs

Collabra: Psychology: <https://www.collabra.org/>

The screenshot shows the homepage of the journal Collabra: Psychology. At the top is a navigation bar with links for Home, Articles, Who We Are, How It Works, FAQ, Get Involved, and Contact. A search bar and a user greeting 'Hi, Rolf' are on the right. Below the navigation bar are logos for the University of California Press and the Society for the Improvement of Psychological Science (SIPS). Two buttons, 'Start Submission' and 'Become a Reviewer', are prominently displayed. The main content area features a large banner for SIPS with the text 'SOCIETY FOR THE IMPROVEMENT OF PSYCHOLOGICAL SCIENCE' and 'Collabra: Psychology is now the official journal of SIPS'. Below the banner are social media links for Twitter, Facebook, and a sign-up for news. A section titled 'About this Journal' provides a mission statement. On the right, a 'LATEST ARTICLES' sidebar lists three recent publications with their titles, authors, and dates.

Home Articles Who We Are How It Works FAQ Get Involved Contact Search... Hi, Rolf

UNIVERSITY of CALIFORNIA PRESS Collabra: Psychology Start Submission Become a Reviewer SIPS SOCIETY FOR THE IMPROVEMENT OF PSYCHOLOGICAL SCIENCE

SIPS SOCIETY FOR THE IMPROVEMENT OF PSYCHOLOGICAL SCIENCE

Collabra: Psychology is now the official journal of SIPS

Follow on Twitter Like on Facebook Sign up for News Collabra Blog

About this Journal

Collabra: Psychology, the official journal of the [Society for the Improvement of Psychological Science](#), is a mission-driven Open Access (OA) journal from the University of California Press that shares not only the research it publishes, but also the value created by the psychology community during the peer-review process. Collabra: Psychology has 7 sections representing the broad field of psychology, and a highlighted focus area of "Methodology and Research Practice."

LATEST ARTICLES

-  **The Influence of Direct and Indirect Speech on Source Memory**
Eerland & Zwaan — 20 Feb 2018
Share: [f](#) [t](#) [g+](#) [in](#)
-  **A Conceptual Review of Lab-Based Aggression Paradigms**
McCarthy & Elson — 26 Jan 2018
Share: [f](#) [t](#) [g+](#) [in](#)
-  **Eye Movement Evidence for Context-Sensitive Derivation of Scalar Inferences**
Politzer-Ahles & Husband — 26 Jan 2018
Share: [f](#) [t](#) [g+](#) [in](#)
-  **Children's Third-Party Understanding of Communicative Interactions in a Foreign Language**

[bra.org/articles/10.1525/collabra.123/](https://www.collabra.org/articles/10.1525/collabra.123/)

Grazie!

Increase Transparency

- *How can our intellectual life...be arranged so as to expose our beliefs, conjectures, positions, sources of ideas, traditions, and the like ... to maximum criticism?"*
 - Source Bartley (1984, p. 113)

Transparency and Openness Promotion (TOP) Guidelines

<https://cos.io/top/>

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

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

	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.
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.

Lykken (1991) – Feynman's Principle of Total Scientific Honesty

- Details that could throw doubt on your interpretation must be given if you know them...If you make a theory, for example, you must also put down all the facts that disagree with it...(Feynman, 1986, as cited in Lykken, 1991, p. 36)
- *Many of us are able to tear other people's research limb from limb while we smile upon our own like an indulgent parent*

Preregistration

Avoiding the Garden of Forking Paths.
Distinguish between confirmatory and
exploratory research

Prospective declaration of the primary outcome variable is important because it eliminates the possibility of selecting for reporting an outcome among many different measures included in the study. In order to investigate this issue, we looked at the statistical significance of other

The garden of forking paths

Andrew Gelman & Eric Loken, 2013

