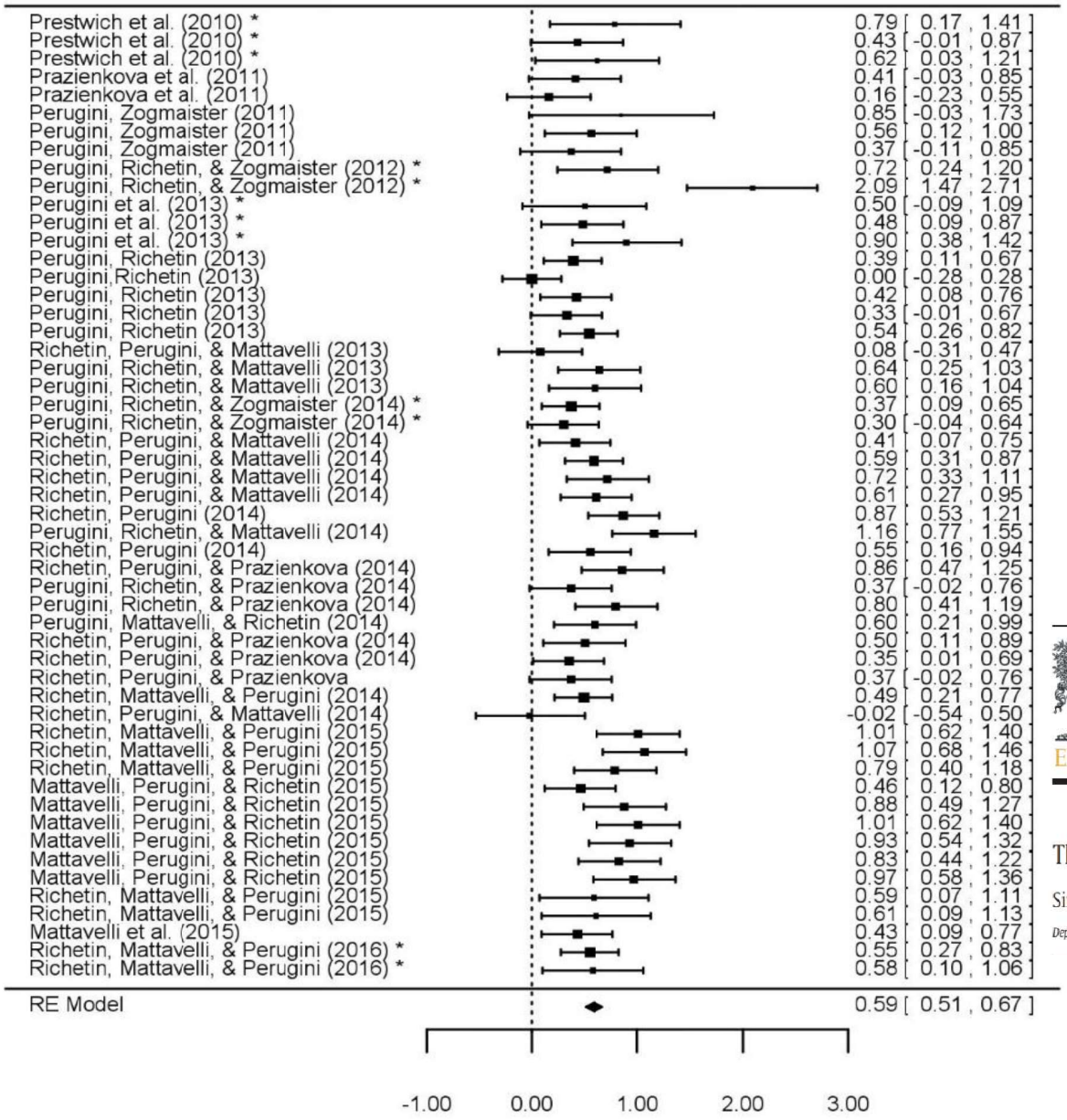
Real data are a bit messy…
Self-Referencing (Implicit; k=53)



d=0.59
[0.51, 0.67]

Journal of Experimental Social Psychology 71 (2017) 68–82

Contents lists available at ScienceDirect

Journal of Experimental Social Psychology

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

The Self-Referencing task: Theoretical overview and empirical evidence

Simone Mattavelli *, Juliette Richetin, Marcello Gallucci, Marco Perugini

Department of Psychology, University of Milan-Bicocca, Piazza dell'Ateneo Nuovo, 1, 20126 Milan, Italy

4th pointer: Meta-conditional

- Avoid dichotomous thinking (and, if you can, also dichotomous theorizing…)
- Everything happens under some circumstances
- Try to identify these circumstances and understand whether they are robust
- Meta-conditional approach: what, how, when, for whom, how much something happens

THE AMERICAN STATISTICIAN
2019, VOL. 73, NO. S1, 271–280: Statistical Inference in the 21st Century
https://doi.org/10.1080/00031305.2018.1518266

Taylor & Francis
Taylor & Francis Group

OPEN ACCESS Check for updates

The New Statistics for Better Science: Ask How Much, How Uncertain, and What Else Is Known

Robert J. Calin-Jageman[a] and Geoff Cumming[b]

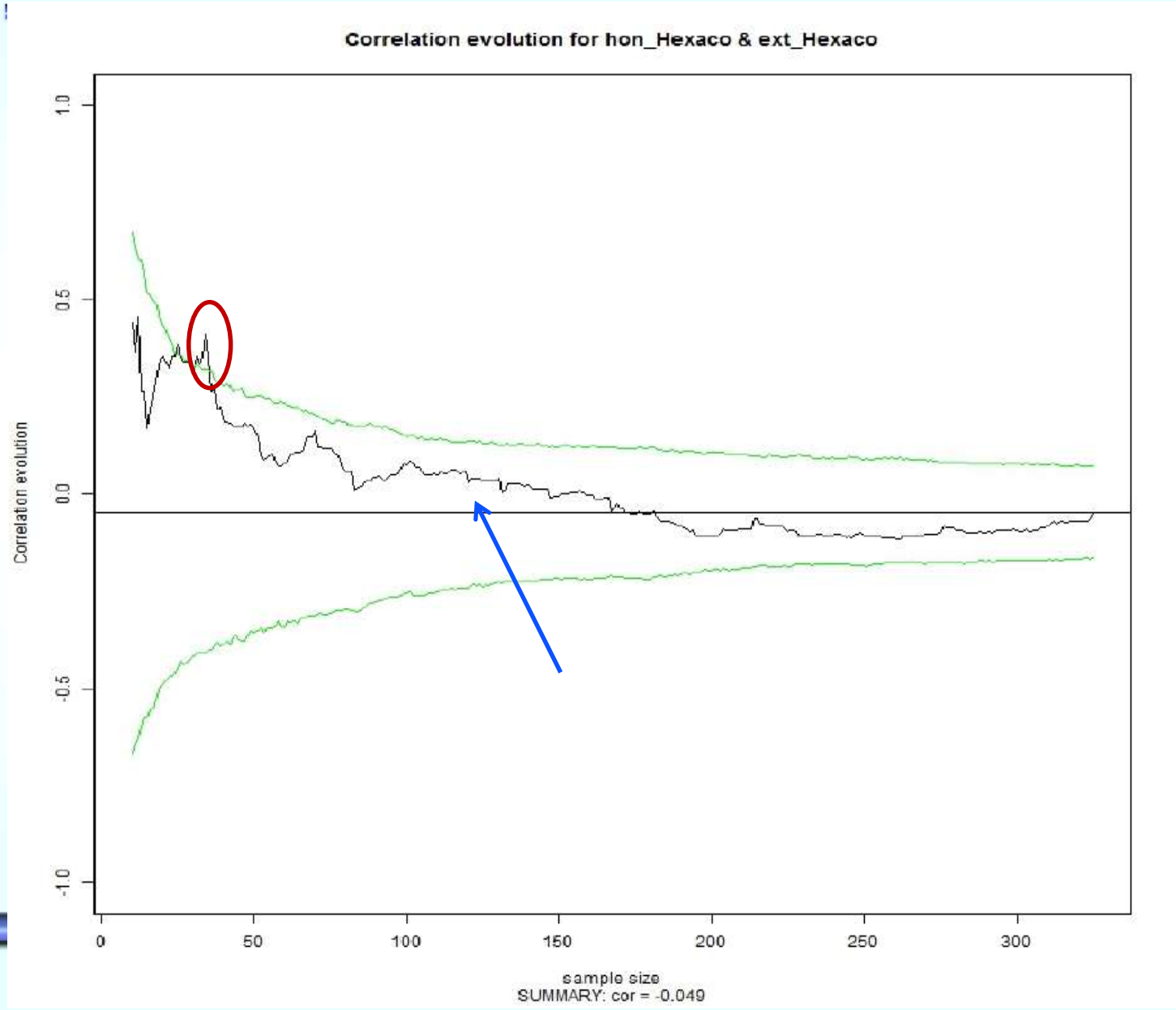
5th pointer: Stability

- Results stabilize with bigger sample sizes
- Try to have a decent sample size
- Sometimes results can be significant in opposite directions with small sample sizes
- For example, stability of correlation coefficients (cf. Schonbrodt & Perugini, 2013)

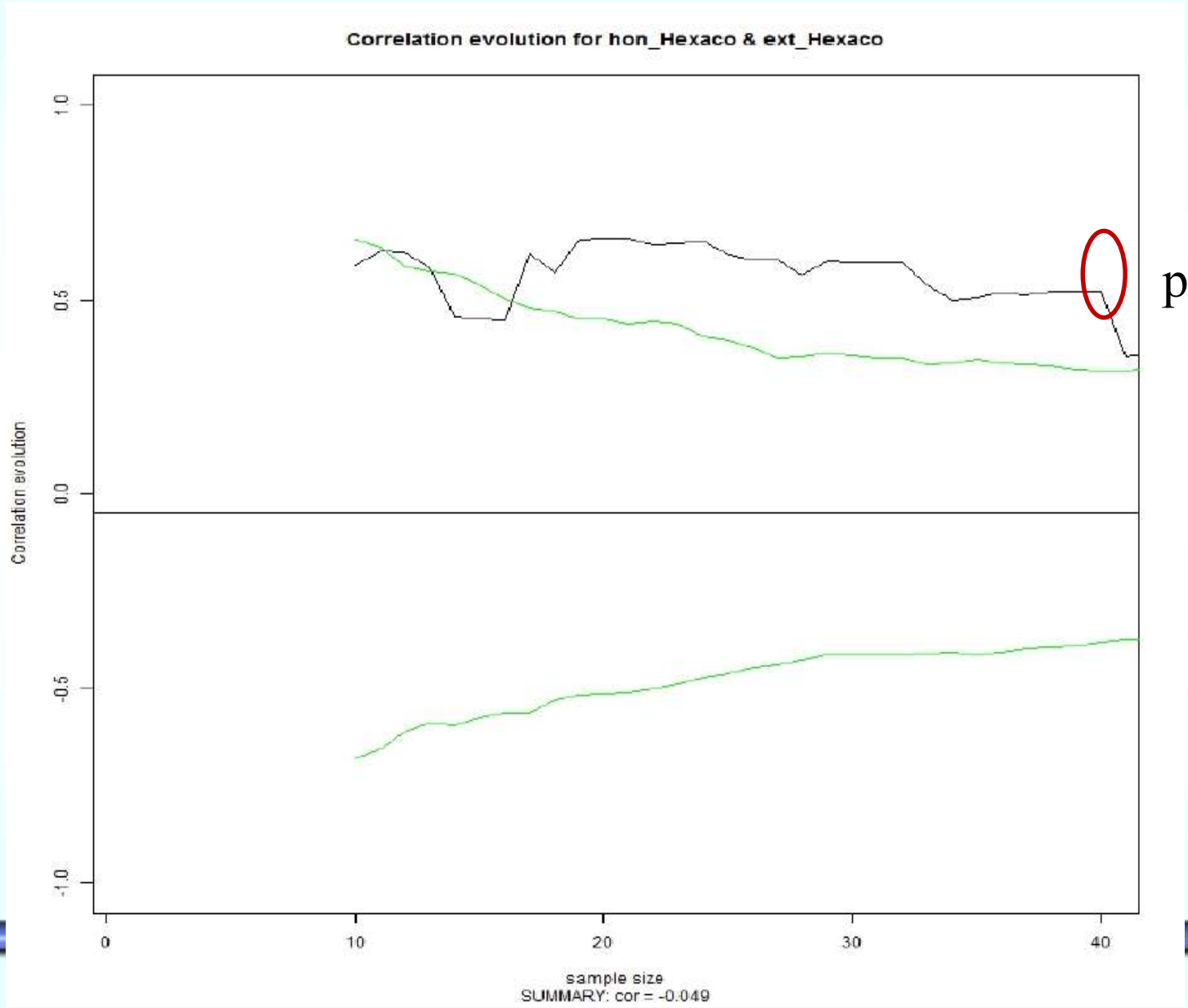
Stability of correlations

- Example from my own data
- One effect that is trivial in the full sample (r between H/H_{quest} and Ext_{quest}=-.05)
- Correlations calculated adding Ss at each step starting from N=10 to full sample (evolution of r)
- Real Ss order
- Boostrapped (s=1000) CI 95%

H vs. E


Correlation evolution for hon_Hexaco & ext_Hexaco
Correlation evolution
1.0
0.5
0.0
-0.5
-1.0
0
50
100
150
200
250
300
sample size
SUMMARY: cor = -0.049

H vs. E


Correlation evolution for hon_Hexaco & ext_Hexaco
Correlation evolution
1.0
0.5
0.0
-0.5
-1.0
0
10
20
30
40
sample size
SUMMARY: cor = -0.049

p<.001!

Contents lists available at SciVerse ScienceDirect

Journal of Research in Personality

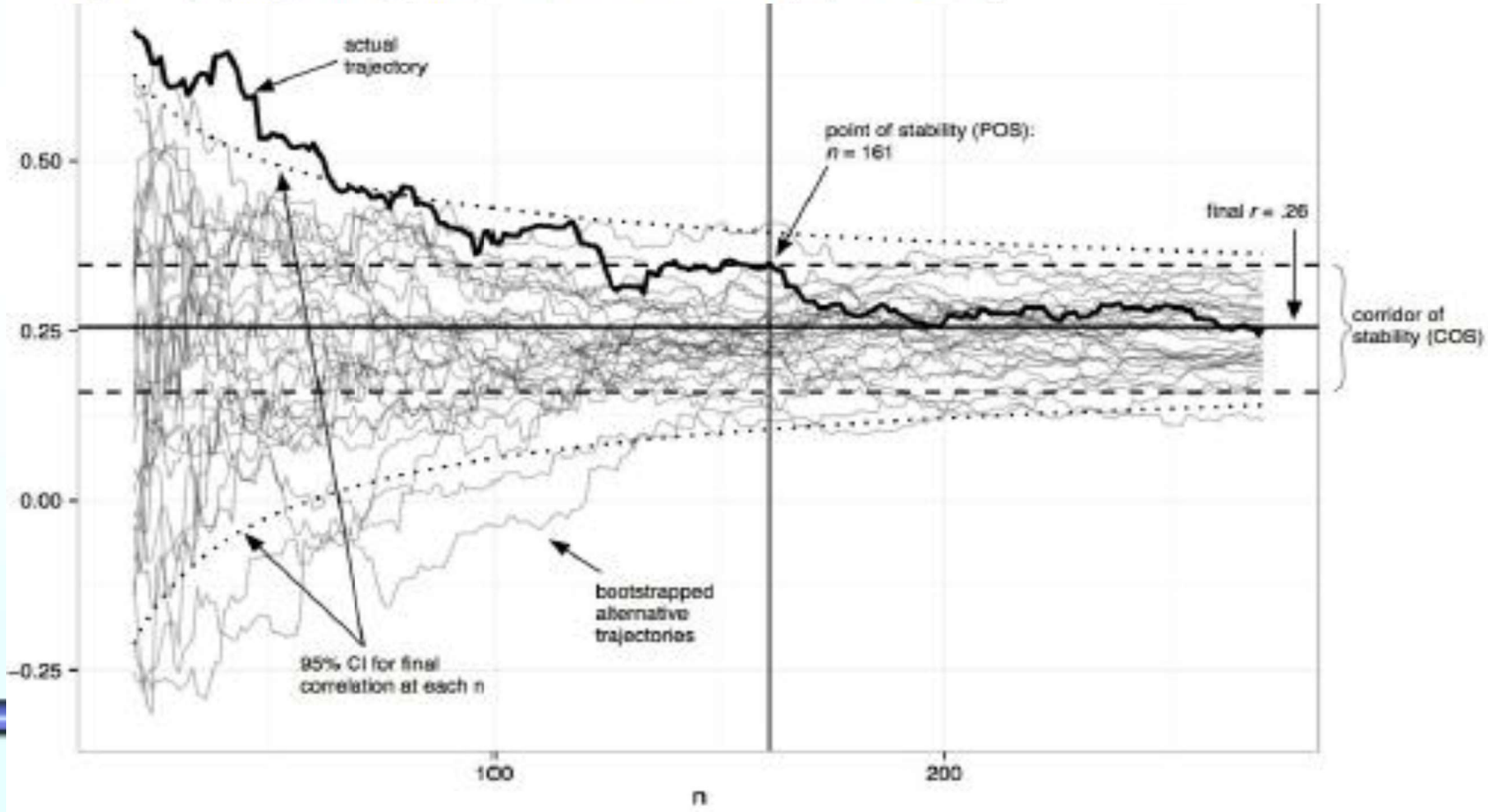
journal homepage: www.elsevier.com/locate/jrp

Brief Report

At what sample size do correlations stabilize?

Felix D. Schönbrodt [a,*], Marco Perugini [b]

[a] Department of Psychology, Ludwig-Maximilians-Universität, Leopoldstr. 13, 80802 München, Germany
[b] Department of Psychology, University of Milan, Bicocca, Piazza dell'Ateneo Nuovo 1 (U6), 20126 Milan, Italy



Implications

- Sequential effects can be devastating for small samples (e.g., N≤60)
- Estimates start to stabilize for N≥150 (but it depends on the expected correlation and desired width; e.g., with w=.1, N≈180 for r = .4 & N ≈65 for r=.7)
- Small samples (N≤60) can give many false positives/negatives, especially for small effects. But there are appropriate sequential approaches

European Journal of Social Psychology, Eur. J. Soc. Psychol. **44**, 701–710 (2014)
Published online in Wiley Online Library (wileyonlinelibrary.com) **DOI**: 10.1002/ejsp.2023

Special issue article: Methods and statistics in social psychology: Refinements and new developments

Performing high-powered studies efficiently with sequential analyses

DANIËL LAKENS*
Human Technology Interaction Group, Eindhoven University of Technology, Eindhoven, The Netherlands

hological Methods
, Vol. 22, No. 2, 322–339

1082-989X/17/$12.00 http://dx.doi.org/10.1037/met0000061

Sequential Hypothesis Testing With Bayes Factors: Efficiently Testing Mean Differences

Felix D. Schönbrodt
Ludwig-Maximilians-Universität München

Eric-Jan Wagenmakers
University of Amsterdam

Michael Zehetleitner
Ludwig-Maximilians-Universität München

Marco Perugini
University of Milan–Bicocca

6th pointer: Reduce variability

- Results stabilize with smaller standard errors
- Standard errors depend on N and SD
- Smaller SD means smaller SE
- SD can be reduced (*ceteris paribus*) with more reliable measures, more precise experimental designs, less Ss variability
- Plan your design as simple and as clean as possible

$$SE = \sqrt{\frac{S^2}{n}}$$

Distinguish conceptually between unnecessary ("added noise") and necessary ("natural") variance

Improve your design. Optimize it.

Reduce the **noise**! Increase the **signal**!

Increasing Statistical Power Without Increasing Sample Size

Gary H. McClelland
University of Colorado at Boulder

August 2000 • American Psychologist 963

Increasing the Power of Your Study by Increasing the Effect Size

TOM MEYVIS
STIJN M. J. VAN OSSELAER

Journal of Consumer Research. Feb2018, Vol. 44 Issue 5, p1157-1173.

General logic behind Effect Sizes

$$\hat{\eta}^2 = \frac{SS_{\text{Effect}}}{SS_T}, \qquad \hat{\eta}_p^2 = \frac{SS_{\text{Effect}}}{SS_{\text{Effect}} + SS_{\text{s/Cells}}}, \qquad \hat{\omega}_p^2 = \frac{SS_{\text{Effect}} - df_{\text{Effect}} MS_{\text{s/Cells}}}{SS_{\text{Effect}} + (N - df_{\text{Effect}})\, MS_{\text{s/Cells}}}$$

$$\boldsymbol{d = \frac{M_1 - M_2}{pooled\ SD}} \qquad r = \frac{Covariance\ (x,y)}{S.D.(x)S.D.(y)} \qquad Z_{\text{CL}} = \frac{|\overline{X}_1 - \overline{X}_2|}{\sqrt{S_1^2 + S_2^2}}$$

$$r = \frac{n\Sigma xy - (\Sigma x)(\Sigma y)}{\sqrt{[n\Sigma x^2 - (\Sigma x)^2][n\Sigma y^2 - (\Sigma y)^2]}}$$

- Effect sizes go up when "**signal**" (*numerator*) increases relative to "**noise**" (*denominator*)

7th pointer:
Parameter estimate vs. Statistical inference

- Parameter estimate ≠ Statistical inference
- More data points are **always** better than less data points
- Parameters are estimated more accurately, error is reduced
- It is a basic statistical principle
- NHST or Bayes (NHBT) are formal tools for statistical inference
- Do not mix them up with parameter estimate!!

Precision vs. Power

- They have different aims

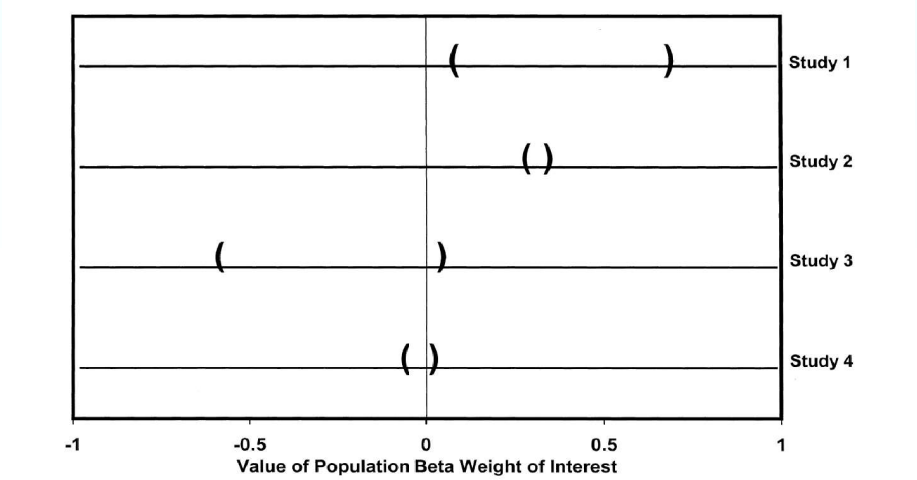
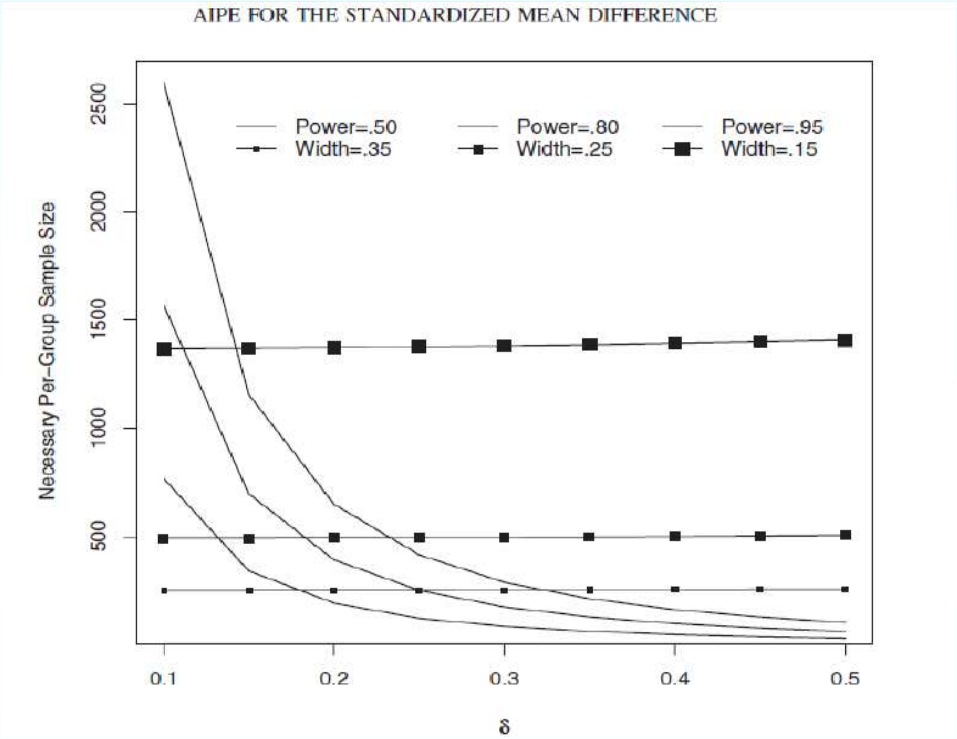


Figure 1. Illustration of possible scenarios in which planned sample size was considered a "success" or "failure" according to the accuracy in parameter estimation and the power analysis frameworks. Parentheses are used to indicate the width of the confidence interval.



- Precision is valuable no matter everything else, but…

  BIG sample sizes are needed for precise estimates no matter the effect size

8th pointer: Don't get it personal

- Get it right ≠ I am right
- Try to plan studies pitting against different hypotheses and predictions (e.g., also use of Bayes Factors)
- Try to depersonalize your preferred theoretical explanation (strong inference, Platt 1964)
- Consider adversarial collaborations
- Try to start from "truth-seeking" and add good aesthetic standards (Giner Sorolla, 2012)

9th pointer: Don't QRPs

- Avoid Questionable Research Practices (QRP)
- Do not cherry-pick DVs among many that you have, do not exclude cases as is, do not make multiple interim analyses to decide whether to collect additional Ss, correct for multiple testing (FDR) when study is exploratory
- Read Simmons et al. (2011): not every recommendation is perfect, but they do give many good ones

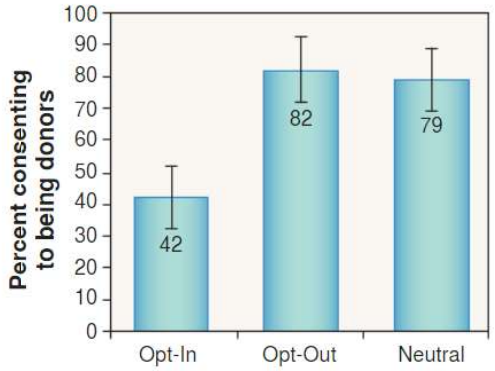
10th pointer: Be wary of "easy fixes"

- Be wary of easy quick fixes with subtle "psychological" manipulations to big problems
- Some seem to work (e.g., Default organ donor)

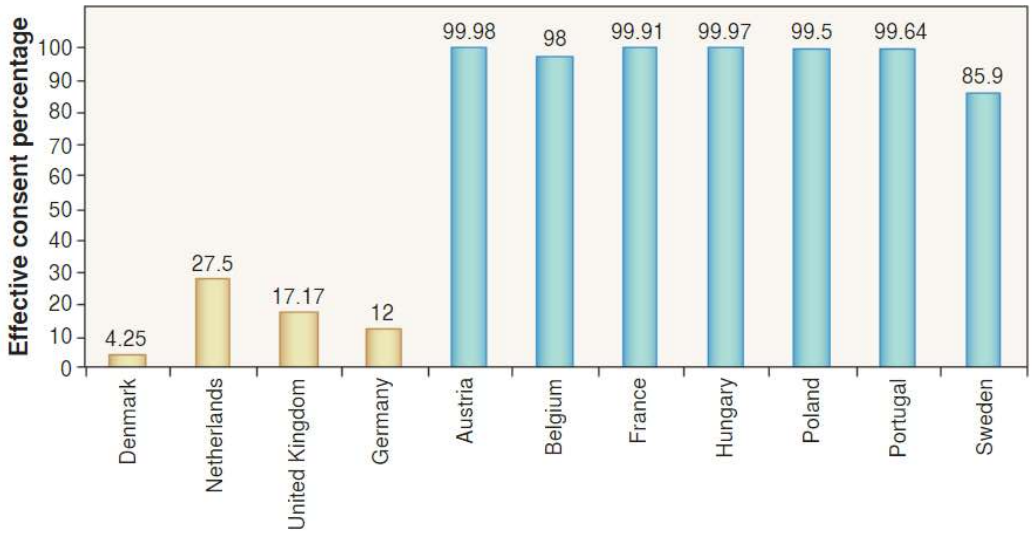
Do Defaults Save Lives?

Eric J. Johnson* and Daniel Goldstein

21 NOVEMBER 2003 VOL 302 SCIENCE



Effective consent rates, online experiment, as a function of default.



Effective consent rates, by country. Explicit consent (opt-in, gold) and presumed consent (opt-out, blue).

…but others don't

Signing at the beginning makes ethics salient and decreases dishonest self-reports in comparison to signing at the end

PNAS | September 18, 2012 | vol. 109 | no. 38 | 15197–15200

Lisa L. Shu[a], Nina Mazar[b,1], Francesca Gino[c], Dan Ariely[d], and Max H. Bazerman[c]

Experiment 3: Participants and Procedure. We conducted a field experiment with an insurance company in the southeastern United States asking some of their existing customers to report their odometer reading.

We compared the reported current odometer mileage on 13,488 completed policy forms for 20,741 cars to the latest records of each car's odometer mileage to calculate its use (number of miles driven). Customers who signed at the beginning on average revealed higher use (M = 26,098.4, SD = 12,253.4) than those who signed at the end [M = 23,670.6, SD = 12,621.4; $F(1, 13{,}485)$ = 128.63, $P < 0.001$]. The difference was 2,427.8 miles per car. That is, asking customers to sign at the beginning of the form led to a 10.25% increase in implied miles driven (based on reported odometer readings) over the current practice of asking for a signature at the end. Follow-up analyses suggested that the higher

Signing at the beginning versus at the end does not decrease dishonesty

PNAS | March 31, 2020 | vol. 117 | no. 13 | 7103–7107

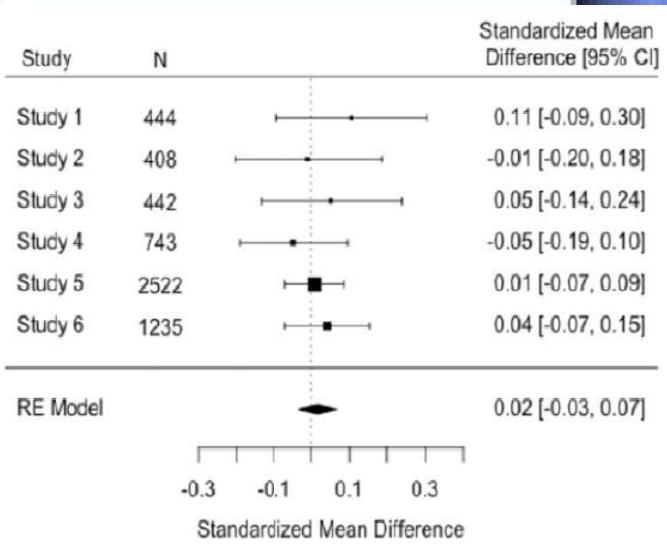
Ariella S. Kristal[a], Ashley V. Whillans[a], Max H. Bazerman[a], Francesca Gino[a,1], Lisa L. Shu[b], Nina Mazar[c], and Dan Ariely[d]

To test this hypothesis in the original PNAS paper (1), there were two laboratory experiments (n = 101 and n = 60, respectively) and one field experiment (n = 13,488). Across the two laboratory ex-

Table 2. Effect sizes of the experiments in the current and original investigation demonstrating the effect of having people sign a veracity statement attesting to their honest reporting placed before versus after reporting

Study	Sample size	Number of conditions	Cheating task	Population	Average performance reported effect size (d) [95% CI]*
This study					
Study 1	444	6	Die rolling	Community laboratory	0.11 [−0.09, 0.30]
Study 2	408	4	Anagrams	Community laboratory	−0.01 [−0.20, 0.18]
Study 3	442	2	Anagrams	MTurk	0.05 [−0.14, 0.24]
Study 4	743	3	Anagrams	MTurk	−0.05 [−0.19, 0.10]
Study 5	2,522	2	Anagrams	Naive MTurk	0.01 [−0.07, 0.09]
Study 6 (direct replication of PNAS study 1)	1,235	2	Paper matrix; self-reported travel expenses	Community laboratory	−0.04 [−0.07, 0.15]†
Shu et al. (1) study					
Study 1	101	3	Paper matrix; self-reported travel expenses	Students	−1.05 [−1.55, −0.53]†
Study 2	60	2	Paper matrix; self-reported travel expenses	Students	−0.53 [−1.04, −0.01]†
Study 3	13,488	2	Odometer reading reported on audit form	Automobile insurance clients	−0.20 [−0.16, −0.23]

*For all tasks, effect sizes are reported for the differences in total amounts reported between conditions. Negative effect size indicates reduction in cheating
†Effect sizes reported in the last column are based on the paper matrix performance only, not the claimed travel expenses.



…with a twist

SCIENTIFIC INTEGRITY

Honesty study was based on fabricated data

Made-up data set raises questions about behavioral scientist Dan Ariely

By **Cathleen O'Grady**

27 AUGUST 2021 • VOL 373 ISSUE 6558 SCIENCE

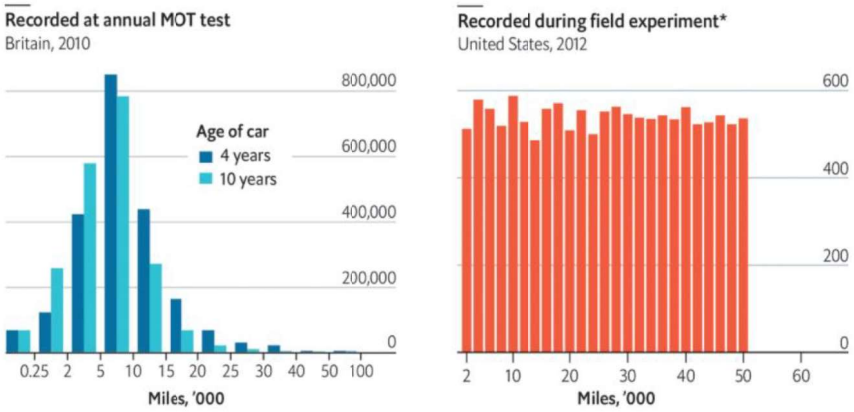
Daily chart

A study on dishonesty was based on fraudulent data

The numbers were clearly faked. No one will admit to faking them

Lies, damned lies and faked statistics

Distribution of miles driven, number of cars



Source: Data Colada

*Mileage allegedly self-reported during experiment minus mileage before experiment. Time interval unknown

[98] Evidence of Fraud in an Influential Field Experiment About Dishonesty

Posted on August 17, 2021 by Uri, Joe, & Leif

This post is co-authored with a team of researchers who have chosen to remain anonymous. They uncovered most of the evidence reported in this post. These researchers are not connected in any way to the papers described herein.

NEWS

DukeWeek: Star Duke professor Dan Ariely faces allegations of research fraud

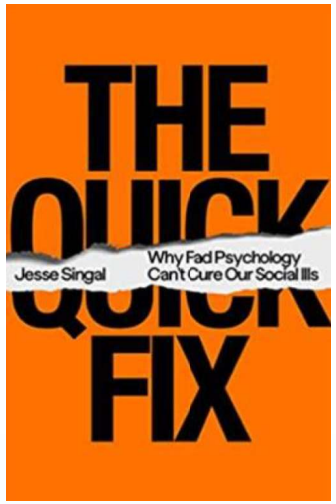
10th pointer: Be wary of "easy fixes"

NATURE HUMAN BEHAVIOUR | VOL 4 | NOVEMBER 2020 | 1092–1094 |

Use caution when applying behavioural science to policy

Social and behavioural scientists have attempted to speak to the COVID-19 crisis. But is behavioural research on COVID-19 suitable for making policy decisions? We offer a taxonomy that lets our science advance in 'evidence readiness levels' to be suitable for policy. We caution practitioners to take extreme care translating our findings to applications.

Hans IJzerman, Neil A. Lewis Jr., Andrew K. Przybylski, Netta Weinstein, Lisa DeBruine, Stuart J. Ritchie, Simine Vazire, Patrick S. Forscher, Richard D. Morey, James D. Ivory and Farid Anvari



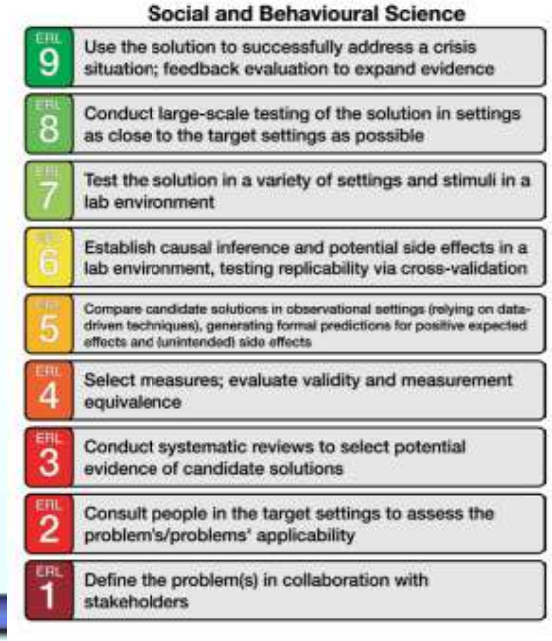


Fig. 2 | Proposed social and behavioural sciences evidence readiness levels.

11th pointer: Think about α

- Consider using p<.005 as significant (instead of p<.05, suggestive) for **NOVEL** findings

NATURE HUMAN BEHAVIOUR | www.nature.com/nathumbehav

comment

Redefine statistical significance

We propose to change the default *P*-value threshold for statistical significance from 0.05 to 0.005 for claims of new discoveries.

Daniel J. Benjamin, James O. Berger, Magnus Johannesson, Brian A. Nosek, E.-J. Wagenmakers, Richard Berk, Kenneth A. Bollen, Björn Brembs, Lawrence Brown, Colin Camerer, David Cesarini, Christopher D. Chambers, Merlise Clyde, Thomas D. Cook, Paul De Boeck, Zoltan Dienes, Anna Dreber, Kenny Easwaran, Charles Efferson, Ernst Fehr, Fiona Fidler, Andy P. Field, Malcolm Forster, Edward I. George, Richard Gonzalez, Steven Goodman, Edwin Green, Donald P. Green, Anthony Greenwald, Jarrod D. Hadfield, Larry V. Hedges, Leonhard Held, Teck Hua Ho, Herbert Hoijtink, Daniel J. Hruschka, Kosuke Imai, Guido Imbens, John P. A. Ioannidis, Minjeong Jeon, James Holland Jones, Michael Kirchler, David Laibson, John List, Roderick Little, Arthur Lupia, Edouard Machery, Scott E. Maxwell, Michael McCarthy, Don Moore, Stephen L. Morgan, Marcus Munafó, Shinichi Nakagawa, Brendan Nyhan, Timothy H. Parker, Luis Pericchi, Marco Perugini, Jeff Rouder, Judith Rousseau, Victoria Savalei, Felix D. Schönbrodt, Thomas Sellke, Betsy Sinclair, Dustin Tingley, Trisha Van Zandt, Simine Vazire, Duncan J. Watts, Christopher Winship, Robert L. Wolpert, Yu Xie, Cristobal Young, Jonathan Zinman and Valen E. Johnson

nature human behaviour

Comment | Published: 26 February 2018

Justify your alpha

Daniel Lakens, Federico G. Adolfi, Casper J. Albers, Farid Anvari, Matthew A. J. Apps, Shlomo E. Argamon, Thom Baguley, Raymond B. Becker, Stephen D. Benning, Daniel E. Bradford, Erin M. Buchanan, Aaron R. Caldwell, Ben Van Calster, Rickard Carlsson, Sau-Chin Chen, Bryan Chung, Lincoln J. Colling, Gary S. Collins, Zander Crook, Emily S. Cross, Sameera Daniels, Henrik Danielsson, Lisa DeBruine, Daniel J. Dunleavy, Brian D. Earp, Michele I. Feist, Jason D. Ferrell, James G. Field, Nicholas W. Fox, Amanda Friesen, Caio Gomes, Monica Gonzalez-Marquez, James A. Grange, Andrew P. Grieve, Robert Guggenberger, James Grist, Anne-Laura van Harmelen, Fred Hasselman, Kevin D. Hochard, Mark R. Hoffarth, Nicholas P. Holmes, Michael Ingre, Peder M. Isager, Hanna K. Isotalus, Christer Johansson, Konrad Juszczyk, David A. Kenny, Ahmed A. Khalil, Barbara Konat, Junpeng Lao, Erik Gahner Larsen, Gerine M. A. Lodder, Jiří Lukavský, Christopher R. Madan, David Manheim, Stephen R. Martin, Andrea E. Martin, Deborah G. Mayo, Randy J. McCarthy, Kevin McConway, Colin McFarland, Amanda Q. X. Nio, Gustav Nilsonne, Cilene Lino de Oliveira, Jean-Jacques Orban de Xivry, Sam Parsons, Gerit Pfuhl, Kimberly A. Quinn, John J. Sakon, S. Adil Saribay, Iris K. Schneider, Manojkumar Selvaraju, Zsuzsika Sjoerds, Samuel G. Smith, Tim Smits, Jeffrey R. Spies, Vishnu Sreekumar, Crystal N. Steltenpohl, Neil Stenhouse, Wojciech Świątkowski, Miguel A. Vadillo, Marcel A. L. M. Van Assen, Matt N. Williams, Samantha E. Williams, Donald R. Williams, Tal Yarkoni, Ignazio Ziano & Rolf A. Zwaan - Show fewer authors

In response to recommendations to redefine statistical significance to P ≤ 0.005, we propose that researchers should transparently report and justify all choices they make when designing a study, including the alpha level.

arXiv.org > stat > arXiv:1709.07588

Statistics > Methodology

Abandon Statistical Significance

Blakeley B. McShane, David Gal, Andrew Gelman, Christian Robert, Jennifer L. Tackett

(Submitted on 22 Sep 2017 (v1), last revised 10 Apr 2018 (this version, v2))

Why?

- Four main reasons:

1) **Findings with p<.005 were twice more likely to be replicated than with p<.05 (in both Psychology and Experimental Economics)**
2) Your intuition tells you something about .05 that is not true (BF vs. 95%)
3) The base rate of correct hypotheses in Psychology is low
4) **It is less draconian than it might seem at first (usually less than 70% increase in sample size)**

12th pointer: Look around you

- Some of this stuff is already implemented in top level journals

DISCLOSURE QUESTIONS:

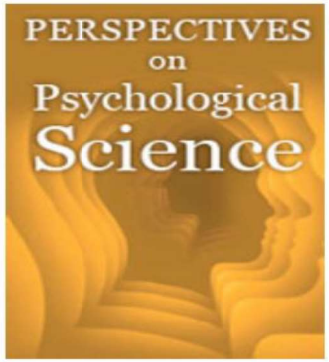
For all studies in your recently published article titled [publication title], please endorse the following statements: (please type an X to indicate your answer)

We reported the total number of observations which were excluded (if any) and the criterion for doing so. (If no observations excluded, please indicate Yes)
Yes: ___ No: ___
If no, please report this information here (e.g., data from 3 participants in Study 2 excluded due to computer malfunction; 4 participants in Study 1 excluded for not following instructions):

We reported all tested experimental conditions, including failed manipulations. Yes: ___ No: ___
If no, please provide brief explanation for not reporting this information (e.g., critical software implementation error; editorial request):

We reported all administered measures/items. Yes: ___ No: ___
If no, please provide brief explanation for not reporting this information (e.g., measures not related to research question; scores from unreported measure insufficiently reliable):

We reported (a) how we determined our sample size **and** (b) our data collection stopping rule. Yes: ___ No: ___
If no, please describe (a) the basis for the sample sizes used and (b) how you decided to stop collecting data (e.g., decided ahead of time to collect data until minimum sample size achieved and this was followed; sample size determined by power analysis but did not achieve it by the end of term):



*

*

- Some journals have a registered reports option
- Some journals have a replication study option also with registered reports
- Some journals provide badges
- Most journals accepts or ask for SM
- Increased collaborative efforts (team research)
- Before 2011 little if any of this was present

Current scenario

Persistent problems of low power (and small samples)

PLOS | BIOLOGY

META-RESEARCH ARTICLE

Empirical assessment of published effect sizes and power in the recent cognitive neuroscience and psychology literature

Denes Szucs[1]*, John P. A. Ioannidis[2]

1 Department of Psychology, University of Cambridge, Cambridge, United Kingdom, 2 Meta-Research Innovation Center at Stanford (METRICS) and Department of Medicine, Department of Health Research and Policy, and Department of Statistics, Stanford University, Stanford, California, United States of America

ELSEVIER

Contents lists available at ScienceDirect

NeuroImage

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

Cross-validation failure: Small sample sizes lead to large error bars

Gaël Varoquaux [a, b, c, *]

[a] Parietal Project-team, INRIA Saclay-Île de France, France
[b] CEA/Neurospin bât 145, 91191 Gif-Sur-Yvette, France
[c] Université Paris-Saclay, Saclay, France

Table 1. Median and mean power to detect small, medium, and large effects in the current study and in three often-cited historical power surveys.
The bottom row shows mean power computed from 25 power surveys.

		Small effect		Medium effect		Large effect	
Subfields or other surveys	**Records/Articles**	**Median**	**Mean**	**Median**	**Mean**	**Median**	**Mean**
Cognitive neuroscience	7,888/1,192	0.11	0.14	0.40	0.44	0.70	0.67
Psychology	16,887/2,261	0.16	0.23	0.60	0.60	0.81	0.78
Medical	2,066/348	0.15	0.23	0.59	0.57	0.80	0.77
All subfields	26,841/3,801	0.11	0.17	0.44	0.49	0.73	0.71
Cohen (1962)	2,088/70	0.17	0.18	0.46	0.48	0.89	0.83
Sedlmeier & Gigerenzer (1989)	54 articles	0.14	0.21	0.44	0.50	0.90	0.84
Rossi (1990)	6,155/221	0.12	0.17	0.53	0.57	0.89	0.83
Rossi (1990); means of surveys	25 surveys		0.26		0.64		0.85

Current scenario

In Social Psychology sample sizes (and power) seem to go up (but also other less desirable changes)

Research in Social Psychology Changed Between 2011 and 2016: Larger Sample Sizes, More Self-Report Measures, and More Online Studies

Kai Sassenberg[1,2] **and Lara Ditrich**[1]
[1]Leibniz-Institut für Wissensmedien (Knowledge Media Research Center), Tübingen, Germany, and
[2]School of Science, University of Tübingen

Advances in Methods and Practices in Psychological Science
2019, Vol. 2(2) 107–114

DOI: 10.1177/2515245919838781
www.psychologicalscience.org/AMPPS
SAGE

Table 3. Mean Sample Size, Mean Percentages of Studies Using Online Data Collection and Only Self-Report Measures, and Mean Number of Studies per Article, by Journal and Publication Year

Variable and year	*JESP*	*JPSP*	*PSPB*	*SPPS*
Sample size				
2009	113 (98)		122 (115)	
2011	112 (98)	102 (72)	138 (99)	130 (111)
2016	142 (106)	195 (120)	180 (131)	198 (158)
2018	203 (134)		185 (115)	
Overall	145 (117)	145 (108)	161 (120)	165 (141)
Online data collection (%)				
2009	9.0		2.6	
2011	11.6	5.0	12.2	18.7
2016	33.7	64.2	37.7	42.3
2018	49.2		50.5	
Overall	26.6	32.6	28.2	30.9
Only self-report measures (%)				
2009	43.8		48.7	
2011	33.1	44.6	40.9	36.3
2016	32.7	76.4	57.9	66.0
2018	71.7		63.8	
Overall	46.0	59.5	53.4	51.6
Studies per article				
2009	1.96 (1.17)		2.62 (1.21)	
2011	2.55 (1.21)	3.78 (1.60)	2.43 (1.03)	1.88 (0.93)
2016	3.09 (1.25)	5.94 (1.84)	3.25 (1.64)	2.13 (1.35)
2018	3.24 (1.40)		3.79 (1.62)	
Overall	2.65 (1.34)	4.50 (1.96)	2.98 (1.47)	2.00 (1.15)

Note: Standard deviations are given in parentheses. $N = 1{,}300$ coded studies. *JESP* = *Journal of Experimental Social Psychology*; *JPSP* = *Journal of Personality and Social Psychology*; *PSPB* = *Personality and Social Psychology Bulletin*; *SPPS* = *Social Psychology and Personality Science*.

Current scenario

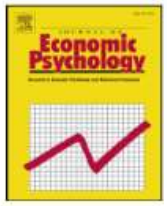
-Good practices and methodological advances

Journal of Economic Psychology 75 (2019) 102117

Contents lists available at ScienceDirect

Journal of Economic Psychology

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



Predicting replication outcomes in the Many Labs 2 study

Eskil Forsell[a,1], Domenico Viganola[b,1], Thomas Pfeiffer[c], Johan Almenberg[d], Brad Wilson[e], Yiling Chen[f], Brian A. Nosek[g,h], Magnus Johannesson[b], Anna Dreber[b,i,*]

nature human behaviour

PERSPECTIVE

PUBLISHED: 10 JANUARY 2017 | VOLUME: 1 | ARTICLE NUMBER: 0021

OPEN

A manifesto for reproducible science

Marcus R. Munafò[1,2*], Brian A. Nosek[3,4], Dorothy V. M. Bishop[5], Katherine S. Button[6], Christopher D. Chambers[7], Nathalie Percie du Sert[8], Uri Simonsohn[9], Eric-Jan Wagenmakers[10], Jennifer J. Ware[11] and John P. A. Ioannidis[12,13,14]

The Psychological Science Accelerator: Advancing Psychology Through a Distributed Collaborative Network

Advances in Methods and Practices in Psychological Science
2018, Vol. 1(4) 501–515

DOI: 10.1177/2515245918797607
www.psychologicalscience.org/AMPPS
SAGE

Hannah Moshontz[1], Lorne Campbell[2], Charles R. Ebersole[3], Hans IJzerman[4], Heather L. Urry[5], Patrick S. Forscher[6],

Current scenario

2600–2606 | PNAS | March 13, 2018 | vol. 115 | no. 11

The preregistration revolution

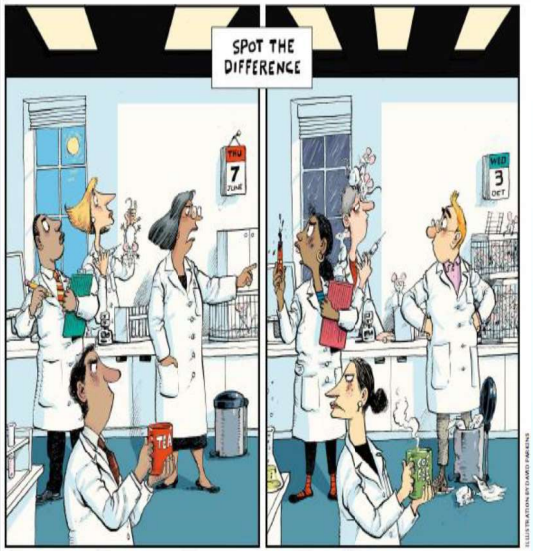
Brian A. Nosek[a,b,1], Charles R. Ebersole[b], Alexander C. DeHaven[a], and David T. Mellor[a]

[a]Center for Open Science, Charlottesville, VA 22903; and [b]Department of Psychology, University of Virginia, Charlottesville, VA 22904

Putting the Self in Self-Correction: Findings from the Loss-of-Confidence Project

Julia Rohrer,[1,2] Warren Tierney,[3] Eric L. Uhlmann,[3] Lisa M. DeBruine,[4] Tom Heyman,[5,6] Benedict Jones,[7] Stefan C. Schmukle,[2] Raphael Silberzahn,[8] Rebecca M. Willén,[9] Rickard Carlsson,[10] Richard E. Lucas,[11] Julia Strand,[12] Simine Vazire,[13] Jessica K. Witt,[14] Thomas R. Zentall,[15] Christopher F. Chabris,[16] Tal Yarkoni[17]

Comment



Argue about what a replication means before you do it

ROYAL SOCIETY OPEN SCIENCE

royalsocietypublishing.org/journal/rsos

Research

Are replication rates the same across academic fields? Community forecasts from the DARPA SCORE programme

Cite this article: Gordon M *et al.* 2020 Are replication rates the same across academic fields? Community forecasts from the DARPA SCORE programme. *R. Soc. Open Sci.* **7**: 200566. http://dx.doi.org/10.1098/rsos.200566

The Defense Advanced Research Projects Agency (DARPA) programme 'Systematizing Confidence in Open Research and Evidence' (SCORE) aims to generate confidence scores for a large number of research claims from empirical studies in the social and behavioural sciences. The confidence scores will provide a quantitative assessment of how likely a claim will hold up in an independent replication. To create the scores, we follow earlier approaches and use prediction markets and surveys to forecast replication outcomes. Based on an initial set of forecasts for the overall replication rate in SCORE

Registered reports

cos.io/rr/

CENTER FOR OPEN SCIENCE

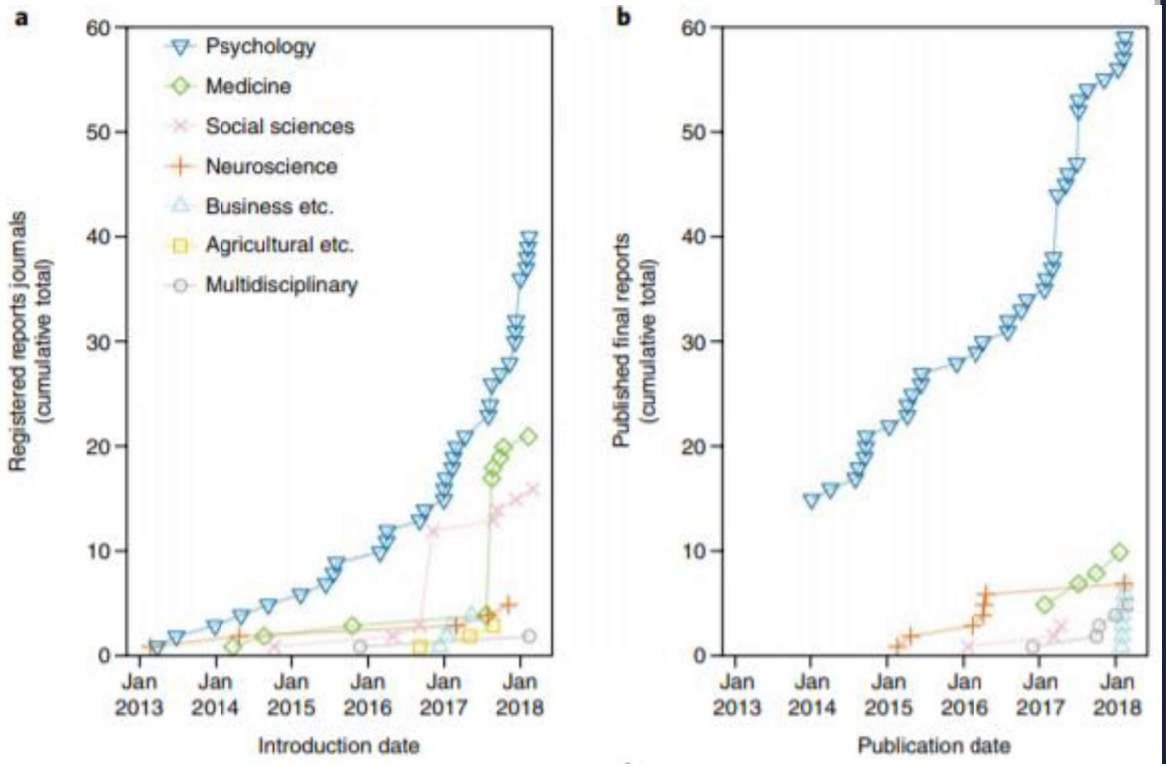
Currently, **209** journals use the Registered Reports

nature human behaviour

Comment | Published: 01 October 2018

Mapping the universe of registered reports

Tom E. Hardwicke & John P. A. Ioannidis



Registered reports and null findings

PLOS BIOLOGY

PERSPECTIVE

Open science challenges, benefits and tips in early career and beyond

Christopher Allen[1]*, **David M. A. Mehler**[1,2]*

1 Cardiff University Brain Research Imaging Centre (CUBRIC), Wales, United Kingdom, 2 Department of Psychiatry, University of Muenster, Germany

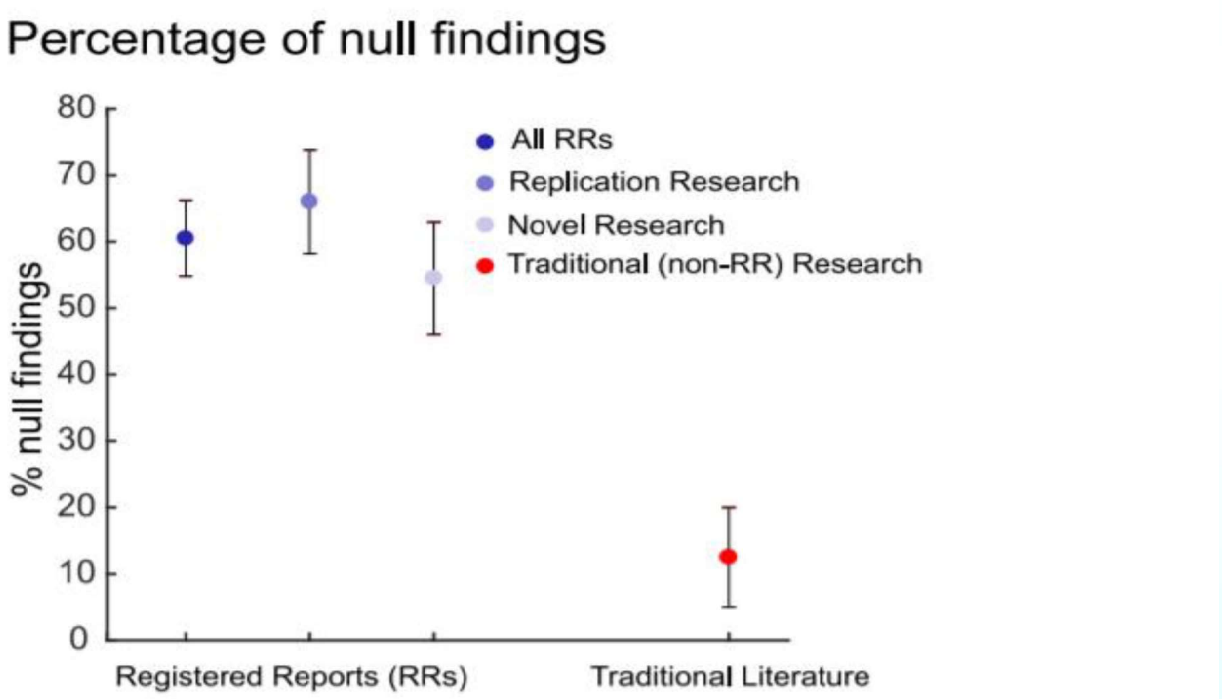


Fig 1. Percentages of null findings among RRs and traditional (non-RR) literature [46,47], with their respective 95% confidence intervals. In total, we extracted $n = 153$ hypotheses from RRs that were declared as replication attempts and $n = 143$ hypotheses that were declared as original research. The bounds of the confidence intervals shown for traditional literature were based on estimates (5% and 20%, respectively) of null findings that have been previously reported for traditional literature [46,47]. Data is available on the Open Science Framework (https://osf.io/wy2ek/) and in S1 Data. RR, registered report.

Registered reports, publication bias and meta-analyses

American Psychological Association

Journal of Experimental Psychology: General

© 2019 American Psychological Association
0096-3445/19/$12.00

2019, Vol. 148, No. 4, 688–712
http://dx.doi.org/10.1037/xge0000570

A Comprehensive Meta-Analysis of Money Priming

Paul Lodder and How Hwee Ong
Tilburg University

Raoul P. P. P. Grasman
University of Amsterdam

Jelte M. Wicherts
Tilburg University

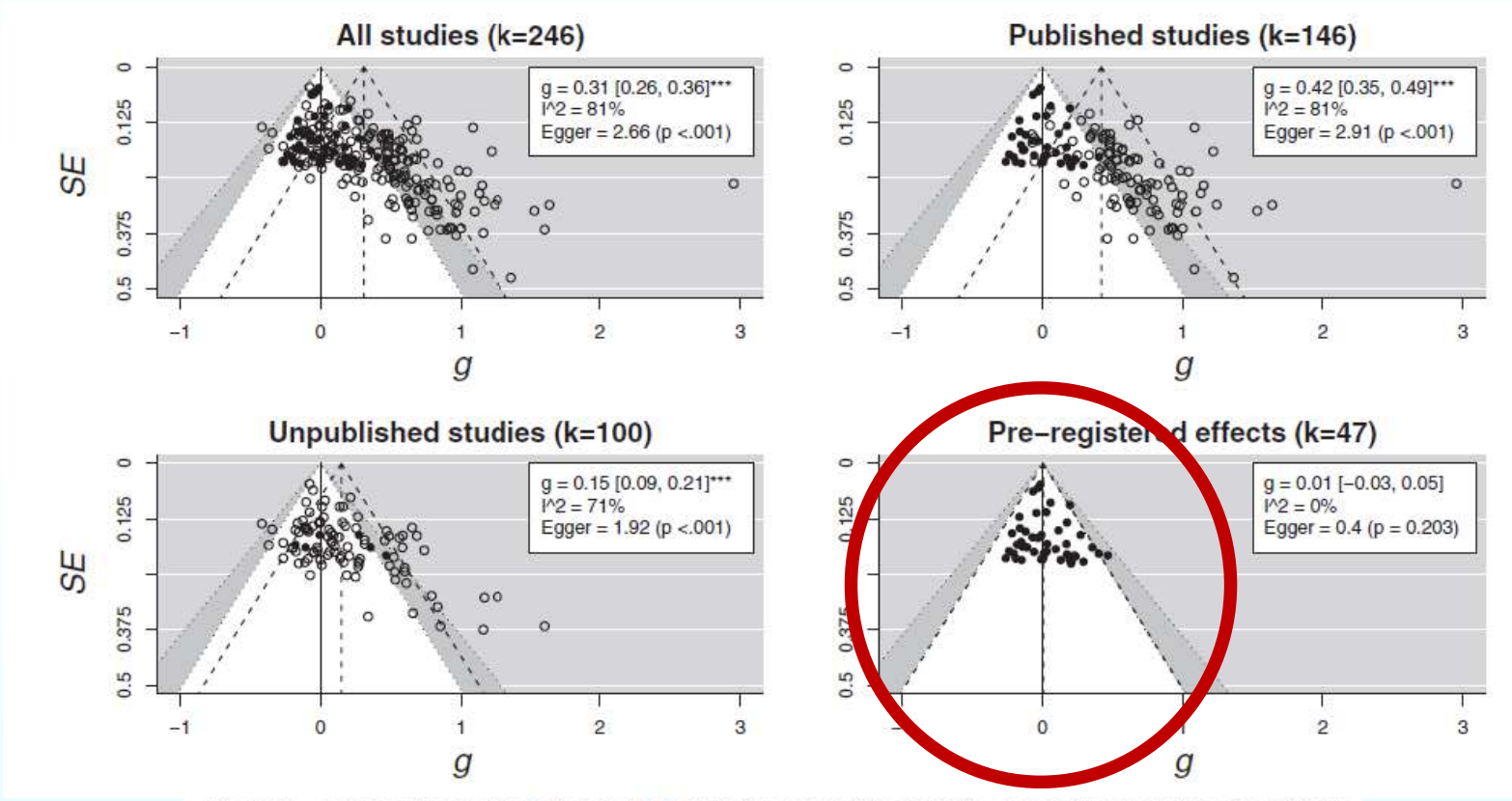


Figure 1. Funnel plots of all studies, published studies, unpublished studies, preregistered studies, main effects, and interaction effects. k = number of included effects. g = Hedges' g random effects model estimate (center of dotted funnel), including 95% confidence interval. I^2 = heterogeneity measure; Egger = Egger's test regression coefficient and p value; The white- and gray funnel represent a 95% and 99% confidence level, respectively. Black dots represent preregistered studies. *** $p < .001$.

Pointers Recap

- Power and Publication bias
- Confirmatory vs. Exploratory studies
- Meta-analytic thinking
- Meta-conditional theorizing
- Increase sample size for stable estimates (signal ↑)
- Simple/clean design to reduce error variability (noise ↓)
- Parameter estimate ≠ statistical inference
- Get it right ≠ I am right
- Don't QRPs
- Be wary of easy fixes
- Think about lowering α to .005 (for novel findings)
- Look around you (things are happening now)

Conclusions

- Increase sample size if you want to get it right
- Decrease **noise** and increase **signal** in the study
- It is not just a statistical issue (Bayes is no miracle cure)
- Clear methodological thinking matters a lot
- Dichotomous thinking does not help
- Ask what, when, how much, how something happens
- Get it right ≠ I am right
- To get it right means to reduce False positives (Type I error), False negatives (Type II error) and to have reasonably precise estimates

We don't want to fall in the nine circles of scientific hell..

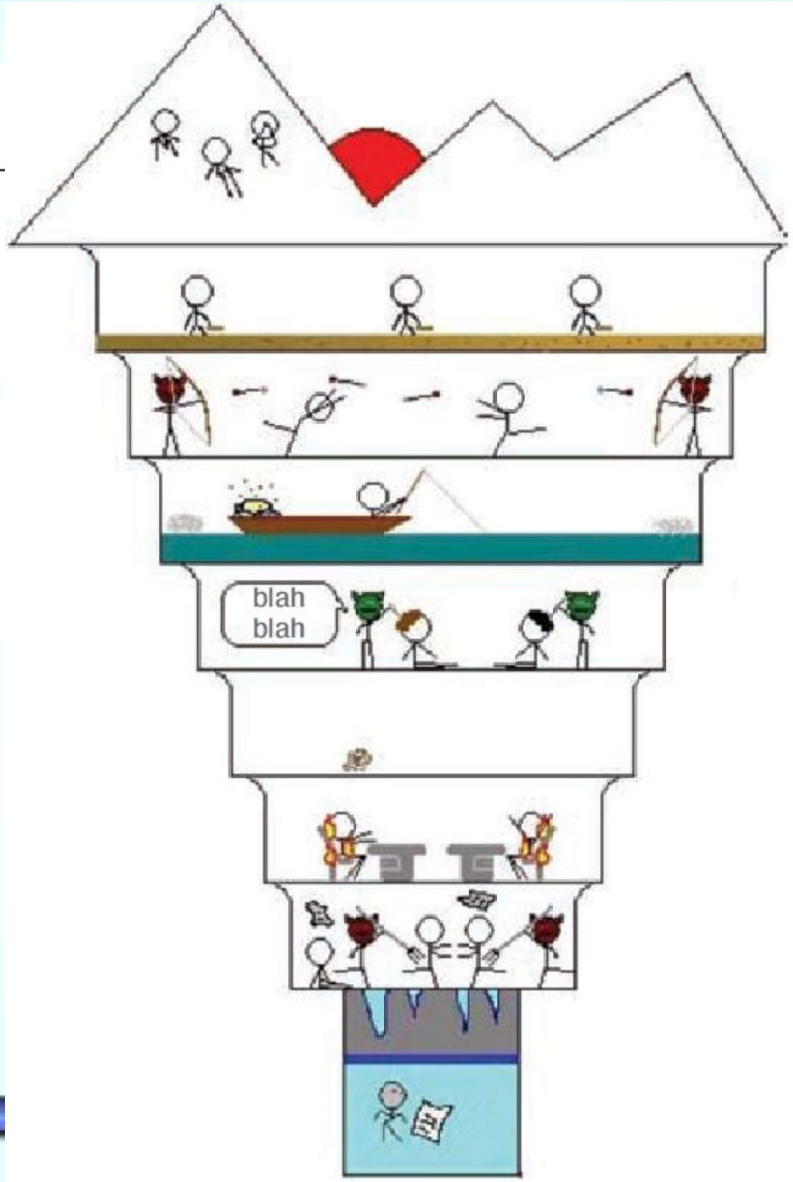
The Nine Circles of Scientific Hell

Neuroskeptic[1]

aps
ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

Perspectives on Psychological Science
7(6) 643–644

DOI: 10.1177/1745691612459519
http://pps.sagepub.com
SAGE



I	Limbo
II	Overselling
III	Post-Hoc Storytelling
IV	P-Value Fishing
V	Creative Outliers
VI	Plagiarism
VII	Non-Publication
VIII	Partial Publication
IX	Inventing Data

We want to go from the Middle Age…

…to the Renaissance

ANNUAL REVIEWS

Annual Review of Psychology

Psychology's Renaissance

Leif D. Nelson,[1] Joseph Simmons,[2]
and Uri Simonsohn[2]

[1]Haas School of Business, University of California, Berkeley, California 94720;
email: Leif_Nelson@haas.berkeley.edu

[2]The Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania 19104;
email: jsimmo@upenn.edu, urisohn@gmail.com

INTRODUCTION

If a team of research psychologists were to emerge today from a 7-year hibernation, they would not recognize their field. Authors voluntarily posting their data. Top journals routinely publishing replication attempts, both failures and successes. Hundreds of researchers preregistering their studies. Crowded methods symposia at many conferences. Enormous increases in sample sizes. Some top journals requiring the full disclosure of measures, conditions, exclusions, and the rules for determining sample sizes. Several multilab replication efforts accepted for publication before any data were collected. Overall, an unprecedented focus on replicability. What on earth just happened?

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA

so...