Open Science

A *must* for modern 21st century researchers

Jorge N. Tendeiro
HIRAKU-Global
Starter Course 2021

24 August 2021

jorgetendeiro.com/talk/2021_openscience/

Content

Open Science

What went wrong?

Maybe it's not that bad?...

What's new (depending on your field...)

Wrapping up

Open Science

- Global approach to science.
- It is a philosophy of behavior more than anything else.
- Make research findings available, free of charge.
- Emphasis on openness, reproducibility, replicability, transparency, integrity.
- Several OS principles are now mandatory at major funding boards:
  - EU's Horizon 2020 (here, here).
  - U.S.'s National Institutes of Health (NIH; here, here).
  - U.S.'s National Science Foundation (NSF; here).
  - JSPS and MEXT over open access (here, here).

Background: By Artem Beliaikin at Pexels, license.

- Contribute to robust and speedy scientific discovery.
- Sharing materials allows getting constructive feedback.
- Improve quality of published research.
- Increase societal relevance, maximize public benefit, avoid resource waste.
- Meet expectations from funders.

Background: By Markus Winkler at Unsplash, license.

See Crüwell et al. (2019), also here.

- Open data (FAIR principles; Wilkinson et al., 2016).
- Open materials, code.
- Open methodology (preregistratin, registered reports).
- Open access.
- Reproducibility, replicability (Penders, Holbrook, & de Rijcke, 2019).
- Open review.
- Open educational resources.

Background: By Trish H-C at Unsplash, license.

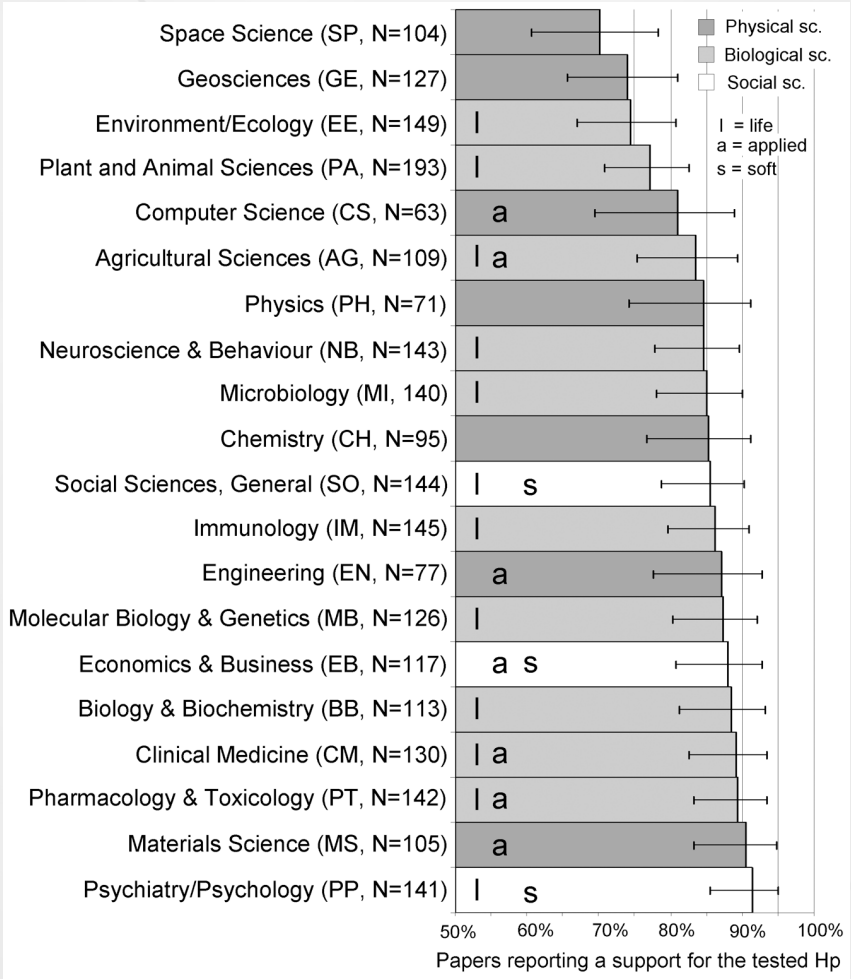
What went wrong?

What went wrong?
Publishing positive results

- Journals often prioritize publishing novel and exciting results.
- Not all such results are based on well-designed and executed experiments.
- “False positive” literature, “bias against the null.”
- This has led to a distortion in the literature.
- Many published results failed to replicate.

Background: By Marcelo Moreira at Pexels, license.

From Fanelli (2010).



Background: By Marcelo Moreira at Pexels, license.

Negative results: Those failing to support the research hypotheses.

- Hard to publish, even for well-designed and executed experiments (e.g., Fanelli, 2012).
- Perceived neither as ‘novel’ nor ‘exciting’.
- File-drawer problem (Rosenthal, 1979).

But there is a lot of good information in negative findings!

Background: By Steve Johnson at Unsplash, license.

"Take nobody's word for it"

Image from Royal Society, CC BY-SA 4.0 license via Wikimedia Commons.

What went wrong?
Sprinting marathons

- Prioritize fast and low-powered studies, over longer and high-powerd studies (e.g., Button et al., 2013, but the list is endless).
- Journals dismiss replication papers.

Background: By Nothing Ahead at Pexels, license.

We have been complacent for way to long.

> *"It is not unusual that (. . .) this ad hoc challenging of auxiliary hypotheses is repeated in the course of a series of related experiments, in which the auxiliary hypothesis involved in Experiment 1 (. . .) becomes the focus of interest in Experiment 2, which in turn utilizes further plausible but easily challenged auxiliary hypotheses, and so forth. In this fashion a zealous and clever investigator can slowly wend his way through (. . .) a long series of related experiments (. . .) without ever once refuting or corroborating so much as a single strand of the network."*
>
> *Meehl (1967)*

Background: By Pixabay at Pexels, license.

"(. . .) It was found that the average power (probability of rejecting false null hypotheses) over the 70 research studies was .18 for small effects, .48 for medium effects, and .83 for large effects. These values are deemed to be far too small."

"(. . .) it is recommended that investigators use larger sample sizes than they customarily do."

Cohen (1962)

Not so long ago (Ioannidis, 2005b):

Essay

Why Most Published Research Findings Are False

John P. A. Ioannidis

Summary

factors that influence this problem and some corollaries thereof.

is characteristic of the field and can vary a lot depending on whether the

Background: By Pixabay at Pexels, license.

What went wrong?
Are researchers to blame??

- Sometimes: Yes.
- Some bad scientists distort or downright break the rules.
- Lies, fabricated results, misconduct.
- Examples:
  - Diederik Stapel, social psychologist. Suspended in 2011. Fabricating and manipulating data.
  - Marc Hauser, psychologist at Harvard. Resigned in 2011. Scientific misconduct.
  - Jens Förster, social psychologist. Resigned in 2017. Data tampering.
  - Jan Hendrik Schön, physicist, 2002. All sorts of wrongdoing.
  - Anil Potti, cancer research, 2007. Lied about CV, fake data.
  - ...

See Retraction Watch for a sad wall of shame.

Background: By Skitterphoto at Pexels, license.

Open Science can contribute to minimize outright fraud.

But wrongdoers will always try their luck anyway, I guess.

Background: By Skitterphoto at Pexels, license.

What went wrong?
Are researchers to blame??

- But most of the times: NO.
- Often researchers are unaware about their actions.
- Also, consequences of mispractice are dire and we all know about it.

Q: So how can we explain many mistakes being done?
A: Combination of poor training, embedded bad practices in their field, current publication system, incentive to publish, wrong career incentives.

There is an expression *en vogue* for this: Questionable research practices.

Background: By Alvan Nee at Unsplash, license.

QRP: Term coined by John, Loewenstein, and Prelec (2012) (see also Simmons, Nelson, & Simonsohn, 2011).

- Not *necessarily* fraud.
- Includes the (ab)use of actually acceptable research practices.
- Problem with QRPs:
  - Introduce bias (typically, in favor of the researcher's intentions).
  - Inflated power at the cost of inflated Type I error probability ($\gg 5\%$).
  - Results not replicable.

Background: By Julia Filirovska at Pexels, license.

(John et al., 2012, Schimmack, 2015).

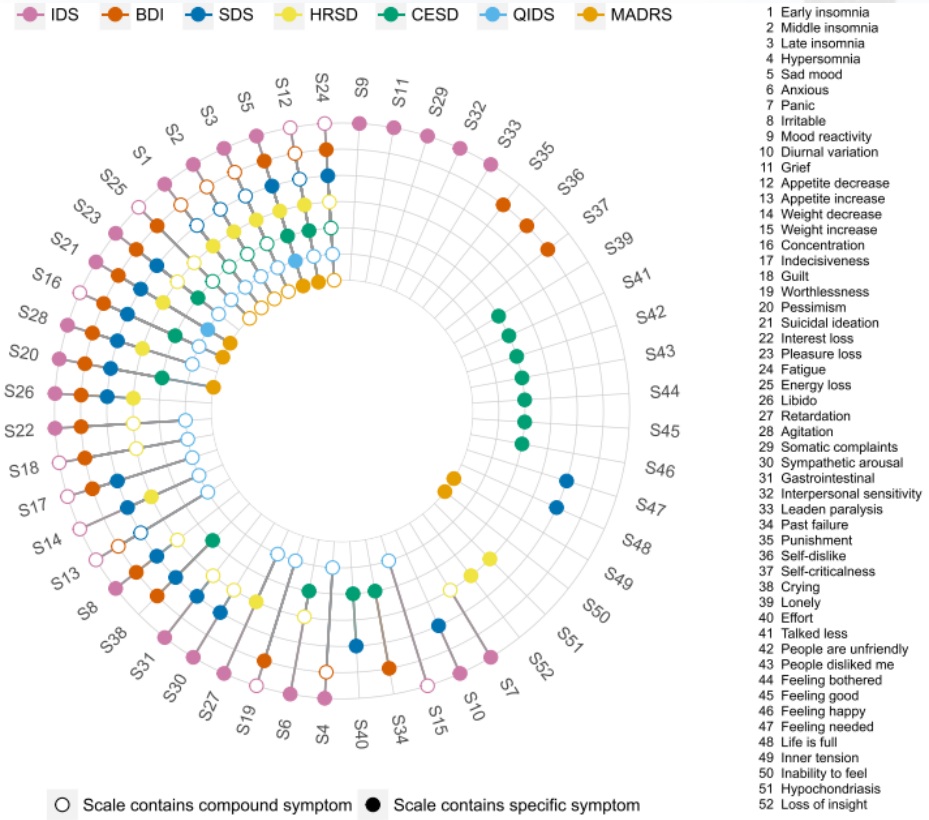
- Omit some DVs.
- Omit some conditions.
- Peeking through sequential testing — Look and decide:
  - $p > .05$: Collect more.
  - $p < .05$: Stop.
- Only report $p < .05$ results.
- p-hacking: E.g.,
  - Exclusion of outliers depending on whether $p < .05$.
  - $p = .054 \longrightarrow p = .05$.
- HARKing (Kerr, 1998): Convert exploratory results into research questions.
- ...

Background: By Julia Filirovska at Pexels, license.

- Researchers have a multitude of decisions to make (experiment design, data collection, analyses performed; Wicherts et al., 2016, Steegen, Tuerlinckx, Gelman, & Vanpaemel, 2016).
- It is very possible to manipulate results in favor of one's interests.
- This is now known as researcher's degrees of freedom (Simmons et al., 2011).
- Consequence: Inflated false positive findings (Ioannidis, 2005b).

Background: By Julian Jagtenberg at Pexels, license.

Example from Fried (2017).



- The 7 most common depression scales contain 52 symptoms.
- That's 7 different sum scores.
- Yet, all are interpreted as 'level of depression'.

Background: By Julian Jagtenberg at Pexels, license.

See Gelman and Loken (2013).

Related to researcher's degrees of freedom:

- Different data may have led to different analysis.
- Related to a multiverse of analytical options (Steegen et al., 2016).
- Not necessarily p-hacking.

Background: By Quang Nguyen Vinh at Pexels, license.

HARKing: Turning exploratory into confirmatory analysis.

From Bem (2004):

> *"(. . .) [L]et us (. . .) become intimately familiar with (. . .) the data. Examine them from every angle. Analyze the sexes separately. Make up new composite indices. If a datum suggests a new hypothesis, try to find further evidence for it elsewhere in the data. If you see dim traces of interesting patterns, try to reorganize the data to bring them into bolder relief. If there are participants you don't like, or trials, observers, or interviewers who gave you anomalous results, drop them (temporarily). Go on a fishing expedition for something– anything– interesting."*

This is not OK unless the exploration is explicity stated.
Daryl Bem is the author of the now infamous 2011 precognition paper.

Background: By WikiImages at Pixabay, license.

Bad incentives explain a lot (Nosek, Spies, & Motyl, 2012; Schönbrodt, 2015):

- "Publish or perish": Publish a lot, at highly prestigious journals.
  But...
  - Journals only publish a fraction of all manuscripts...
  - Journals don't like publishing null findings...
- Get tenured.
- Get research grant.
- Fame (prizes, press coverage, ...).
- ...

But, very importantly, it also happens in spite of the researcher's best intentions:

- Deficient statistics education (yes, statisticians need to acknowledge this!...).
- Perpetuating traditions in the field.

Background: By Anete Lusina at Pexels, license.

Maybe it's not that bad?...

From Ioannidis (2005a).

Contradicted and Initially Stronger Effects in Highly Cited Clinical Research

John P. A. Ioannidis, MD

Context Controversy and uncertainty ensue when the results of clinical research on the effectiveness of interventions are subsequently contradicted. Controversies are most prominent when high-impact research is involved.

LINICAL RESEARCH ON IMPOR-

From Ioannidis (2005a).

Contradicted and Initially Stronger Effects in Highly Cited Clinical Research

John P. A. Ioannidis, MD

Context Controversy and uncertainty ensue when the results of clinical research on the effectiveness of interventions are subsequently contradicted. Controversies are most prominent when high-impact research is involved.

CLINICAL RESEARCH ON IMPOR-

Results Of 49 highly cited original clinical research studies, 45 claimed that the intervention was effective. Of these, 7 (16%) were contradicted by subsequent studies, 7 others (16%) had found effects that were stronger than those of subsequent studies, 20 (44%) were replicated, and 11 (24%) remained largely unchallenged. Five of 6 highly-cited nonrandomized studies had been contradicted or had found stronger effects vs 9 of 39 randomized controlled trials ($P = .008$). Among randomized trials, studies with contradicted or stronger effects were smaller ($P = .009$) than replicated or unchallenged studies although there was no statistically significant difference in their early or overall citation impact. Matched control studies did not have a significantly different share of refuted results than highly cited studies, but they included more studies with "negative" results.

Background: By Johannes Plenio at Pexels, license.

From Begley and Ellis (2012).

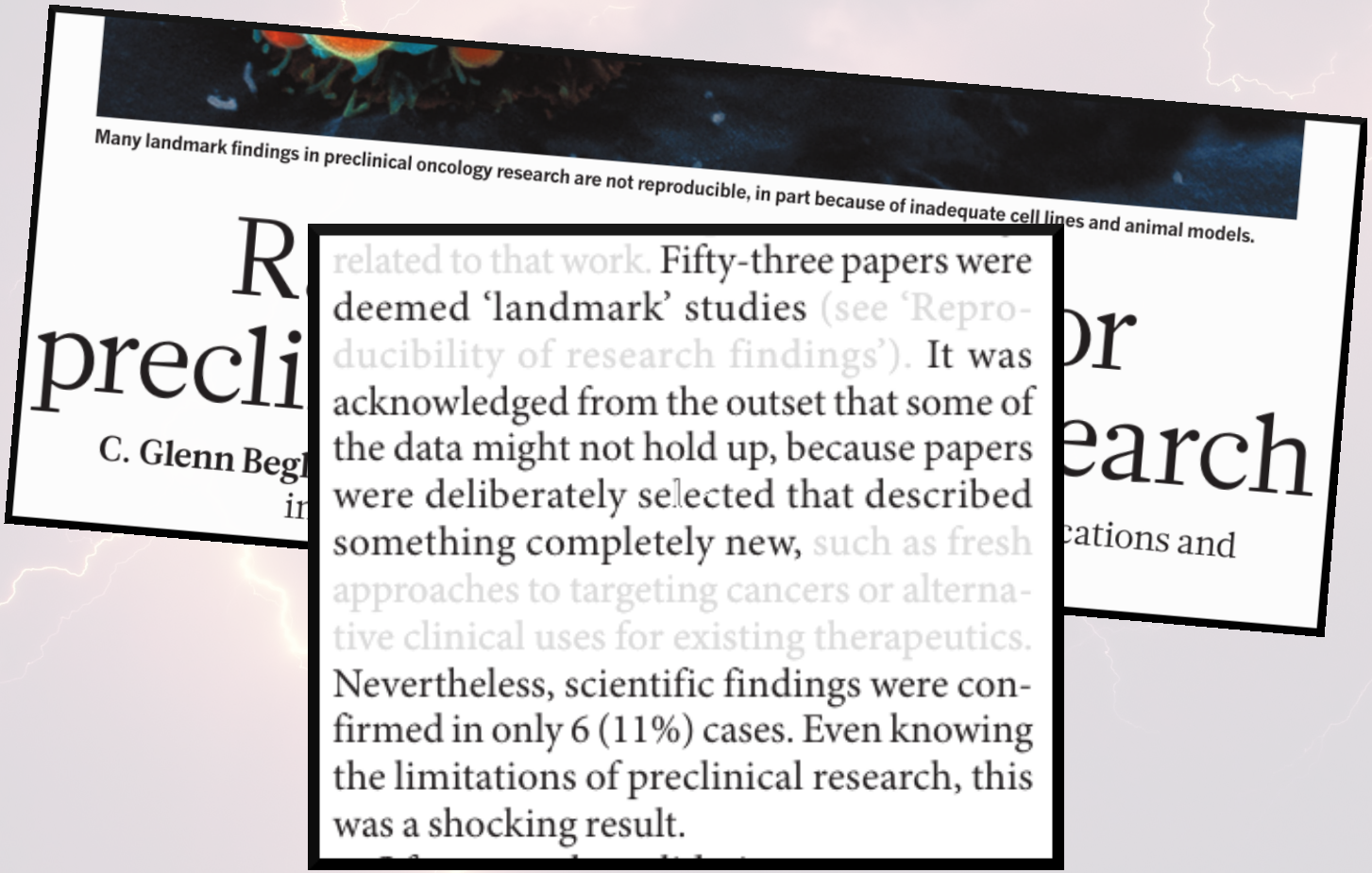
Many landmark findings in preclinical oncology research are not reproducible, in part because of inadequate cell lines and animal models.

Raise standards for preclinical cancer research

C. Glenn Begley and **Lee M. Ellis** propose how methods, publications and incentives must change if patients are to benefit.

Background: By Johannes Plenio at Pexels, license.

From Begley and Ellis (2012).



See also Errington et al. (2014), Prinz et al. (2011).

Background: By Johannes Plenio at Pexels, license.

From Camerer et al. (2016).

ECONOMICS

Evaluating replicability of laboratory experiments in economics

Colin F. Camerer,[1*†] Anna Dreber,[2†] Eskil Forsell,[2†] Teck-Hua Ho,[3,4†] Jürgen Huber,[5†] Magnus Johannesson,[2†] Michael Kirchler,[5,6†] Johan Almenberg,[7] Adam Altmejd,[2] Taizan Chan,[8] Emma Heikensten,[2] Felix Holzmeister,[5] Taisuke Imai,[1] Siri Isaksson,[2] Gideon Nave,[1] Thomas Pfeiffer,[9,10] Michael Razen,[5] Hang Wu[4]

From Camerer et al. (2016).

ECONOMICS

Evaluating replicability of laboratory experiments in economics

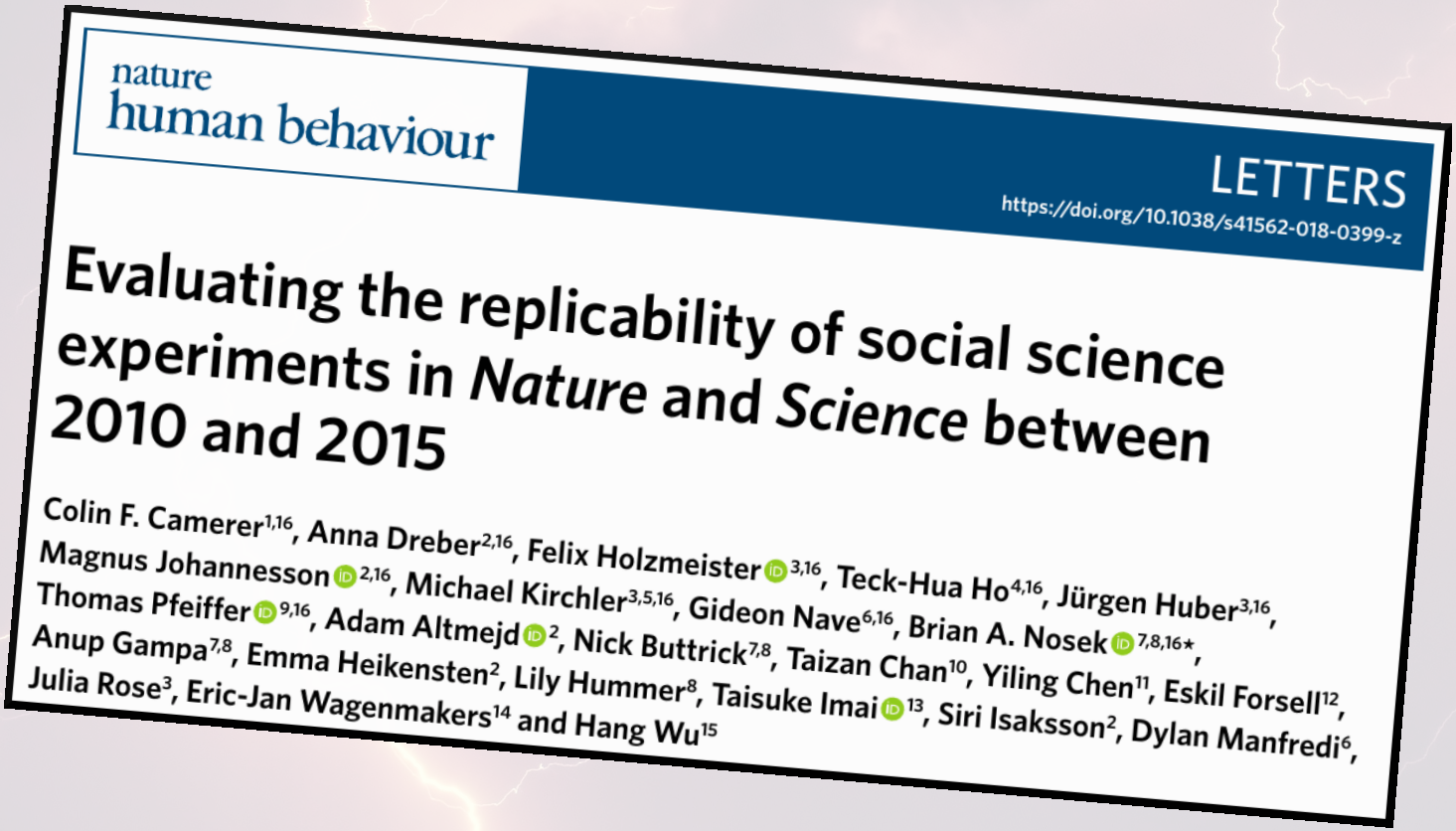
Colin F. Camerer,[1*†] A... Huber,[5†]
Magnus Johannesson
Taizan Chan,[8] Emma
Gideon Nave,[1] Thoma

The replicability of some scientific findings has recently been called into question. To contribute data about replicability in economics, we replicated 18 studies published in the *American Economic Review* and the *Quarterly Journal of Economics* between 2011 and 2014. All of these replications followed predefined analysis plans that were made publicly available beforehand, and they all have a statistical power of at least 90% to detect the original effect size at the 5% significance level. We found a significant effect in the same direction as in the original study for 11 replications (61%); on average, the replicated effect size is 66% of the original. The replicability rate varies between 67% and 78% for four additional replicability indicators, including a prediction market measure of peer beliefs.

See also Chang and Li (2021), Duvendack et al. (2017).

Background: By Johannes Plenio at Pexels, license.

From Camerer et al. (2018).



nature human behaviour

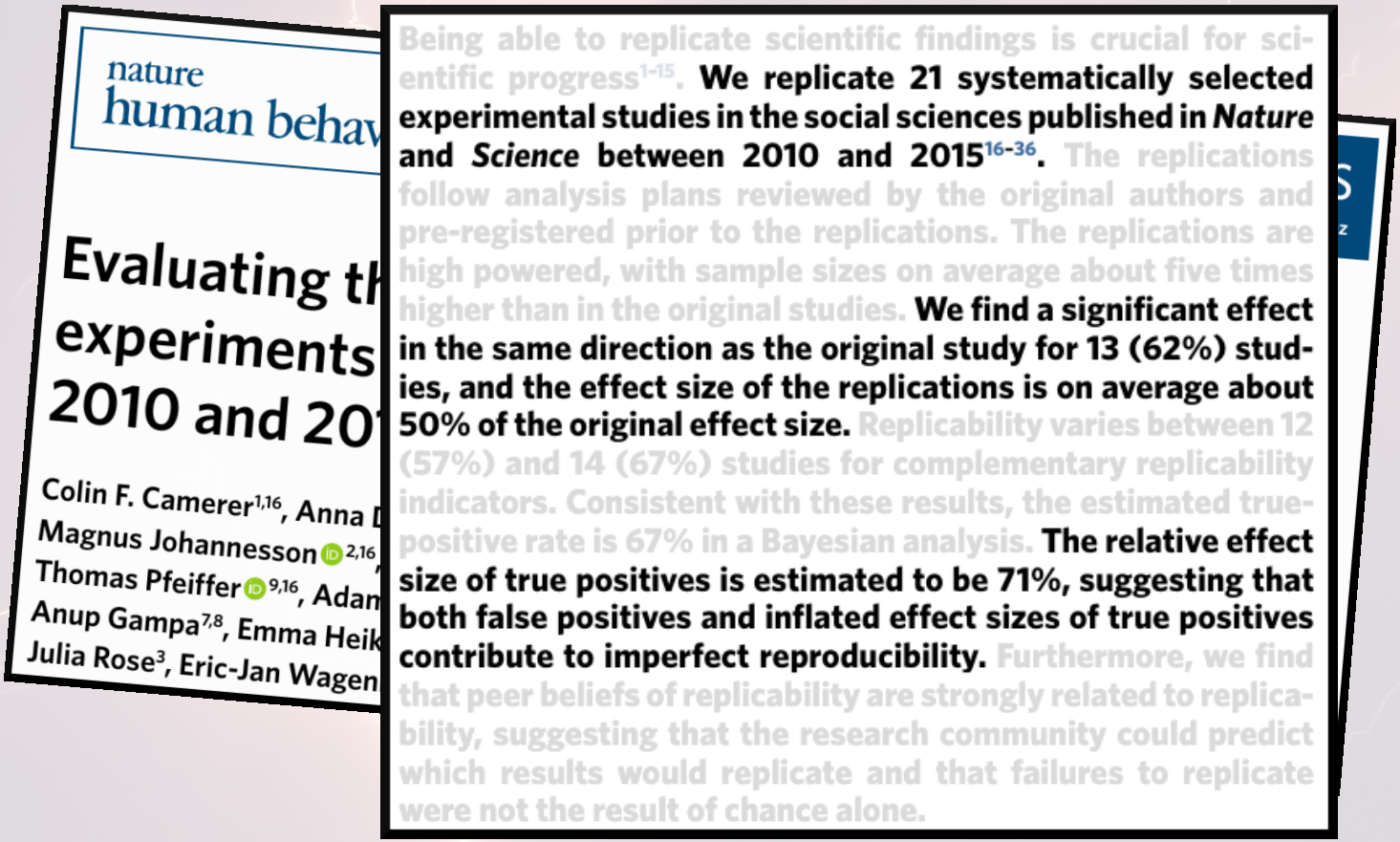
LETTERS

https://doi.org/10.1038/s41562-018-0399-z

Evaluating the replicability of social science experiments in *Nature* and *Science* between 2010 and 2015

Colin F. Camerer[1,16], Anna Dreber[2,16], Felix Holzmeister[3,16], Teck-Hua Ho[4,16], Jürgen Huber[3,16], Magnus Johannesson[2,16], Michael Kirchler[3,5,16], Gideon Nave[6,16], Brian A. Nosek[7,8,16*], Thomas Pfeiffer[9,16], Adam Altmejd[2], Nick Buttrick[7,8], Taizan Chan[10], Yiling Chen[11], Eskil Forsell[12], Anup Gampa[7,8], Emma Heikensten[2], Lily Hummer[8], Taisuke Imai[13], Siri Isaksson[2], Dylan Manfredi[6], Julia Rose[3], Eric-Jan Wagenmakers[14] and Hang Wu[15]

From Camerer et al. (2018).



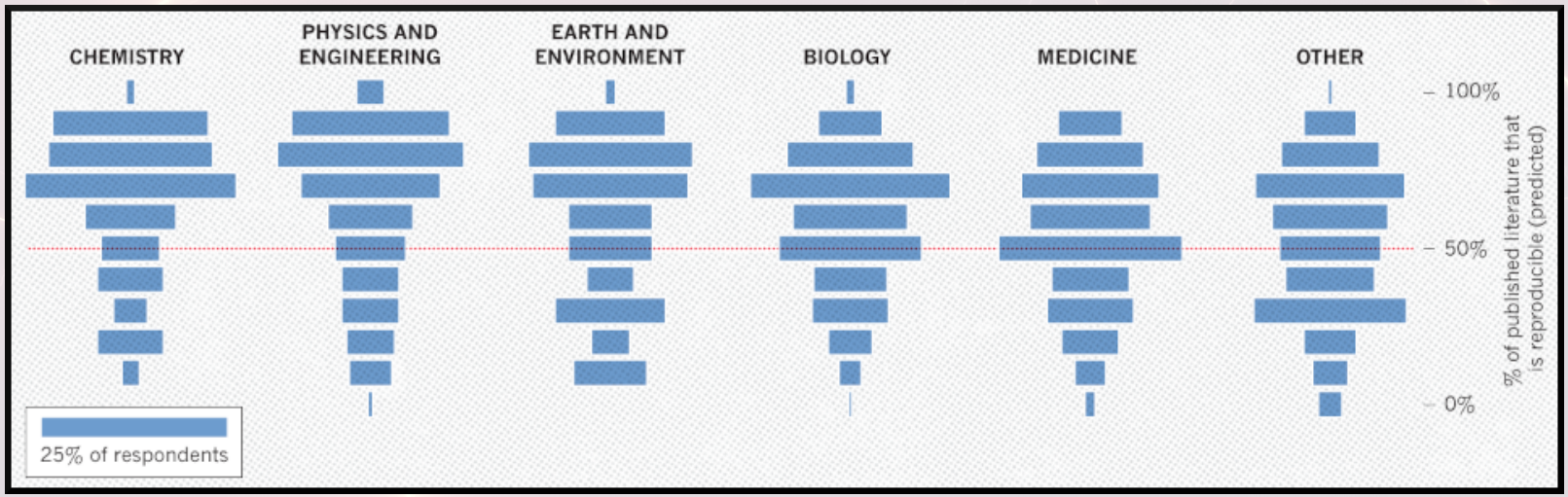
See also Klein et al. (2018), OSC (2015).

Background: By Johannes Plenio at Pexels, license.

Maybe it's not that bad?...

Failed replications – Various fields

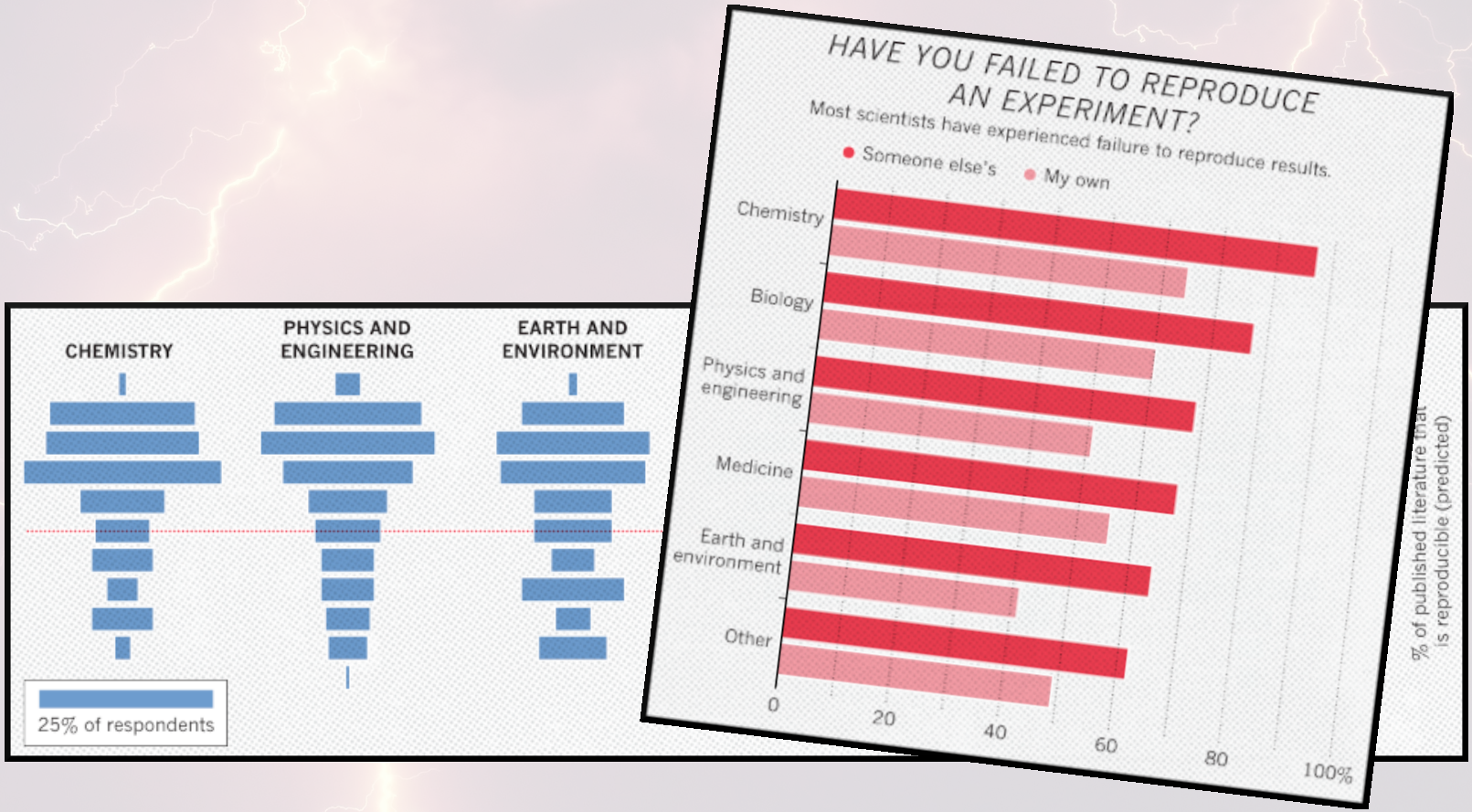
From Baker (2016).



Maybe it's not that bad?...

Failed replications – Various fields

From Baker (2016).



Background: By Johannes Plenio at Pexels, license.

From Baker (2016).

Background: By Johannes Plenio at Pexels, license.

Q: Is it really *that* bad?
A: Yes.

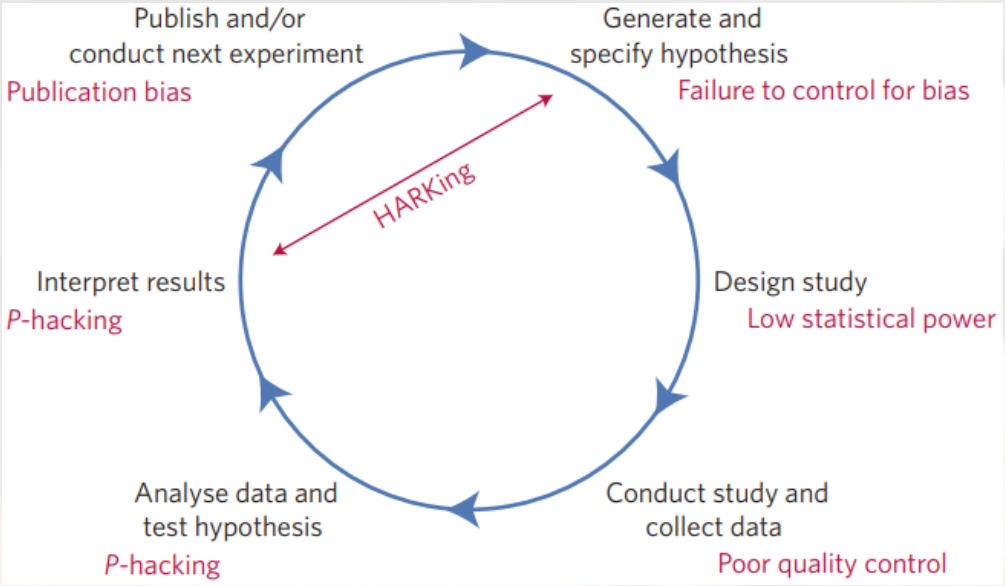
- Martinson, Anderson, and de Vries (2005): "Scientists behaving badly".
- Fanelli (2009): Meta-analysis shows evidence of science misconduct.
- John et al. (2012): Evidence for QRPs.
- Mobley, Linder, Braeuer, Ellis, and Zwelling (2013): Reported evidence of pressure to find significant results.
- Agnoli, Wicherts, Veldkamp, Albiero, and Cubelli (2017): More evidence of QRPs.
- Fraser, Parker, Nakagawa, Barnett, and Fidler (2018): More evidence from various fields of science.
- .....

Interestingly, science misconduct has been a longtime concern (see Babbage, 1830).

And for the sake of balance:
There are also some voices against this description of the current state of affairs (e.g., Fiedler & Schwarz, 2016).

Background: By OpenClipart-Vectors at Pixabay, license.

From Munafò et al. (2017).



- Hypothetico-deductive model of the scientific method.
- In red: Potential threats to this model.

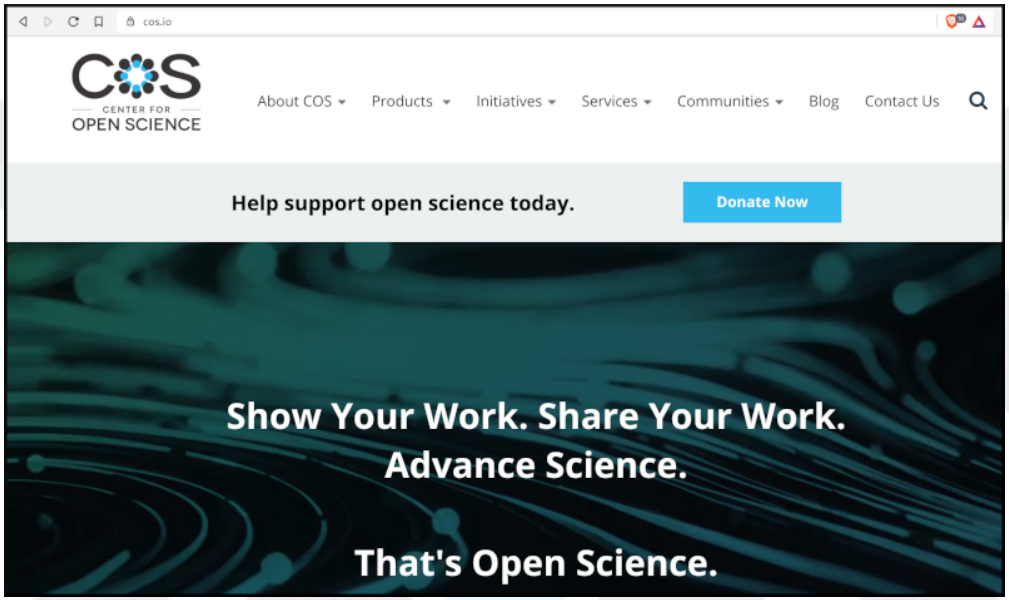
Background: By Niek Verlaan at Pixabay, license.

Maybe it's not that bad?
Distrust in science

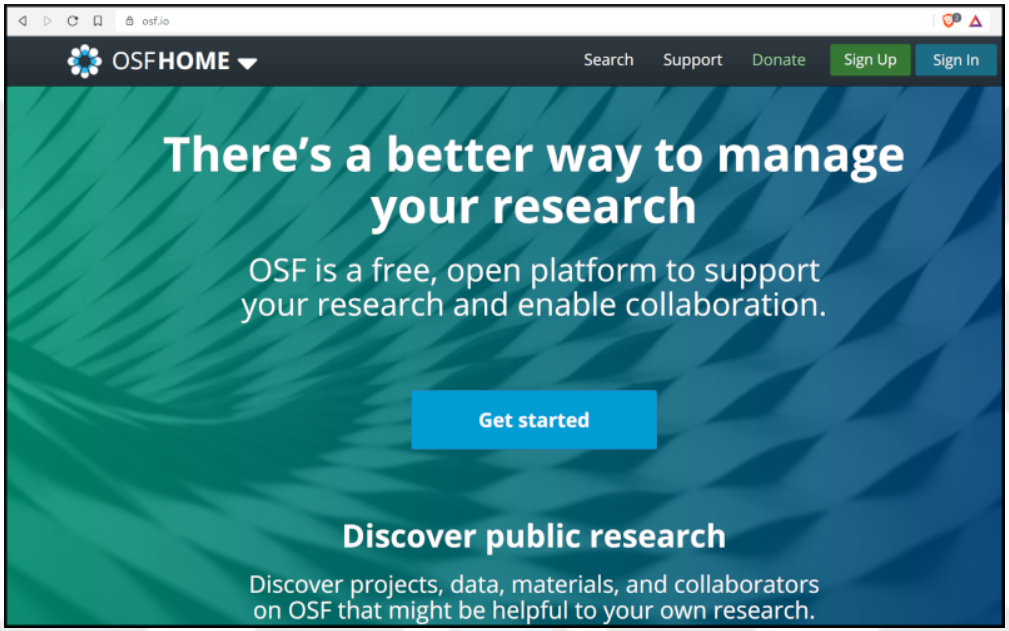
- Public becomes skeptic about the work of researchers.
- Affects allocation of public resources to research.

Background: By Niek Verlaan at Pixabay, license.

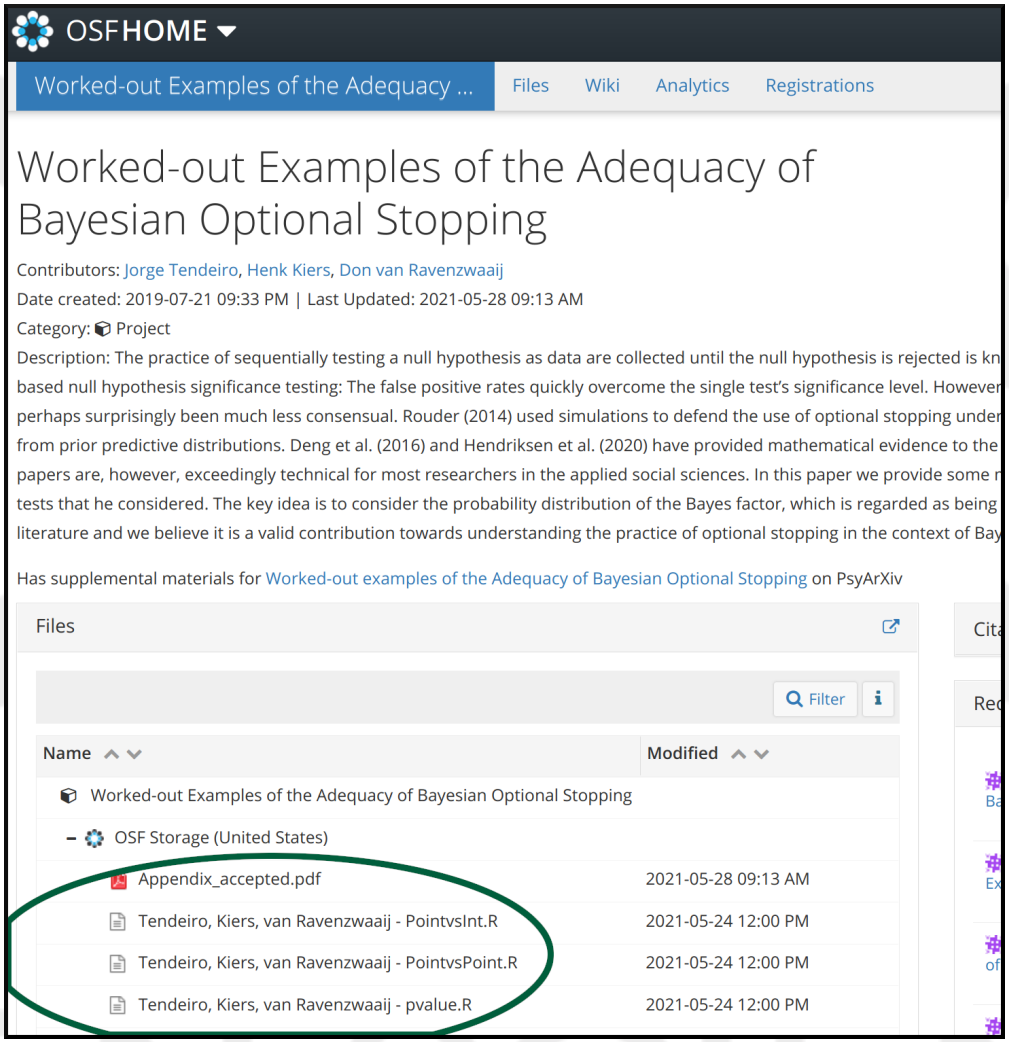
What's new (depending on your field...)



- See Center for Open Science.
- Offers many services:
  - Open Science Framework (OSF) for collaborative projects, share data, preprints...
  - Preregistrations.
  - Registered reports.
  - Open Science badges.
  - ...



- See Open Science Framework
- Allows sharing of data, study materials, research proposals.
- Easy access to preprints and effectively bypass publisher's unnaceptably expensive paywalls (please see this movie!!).



Well, new at least in some fields...

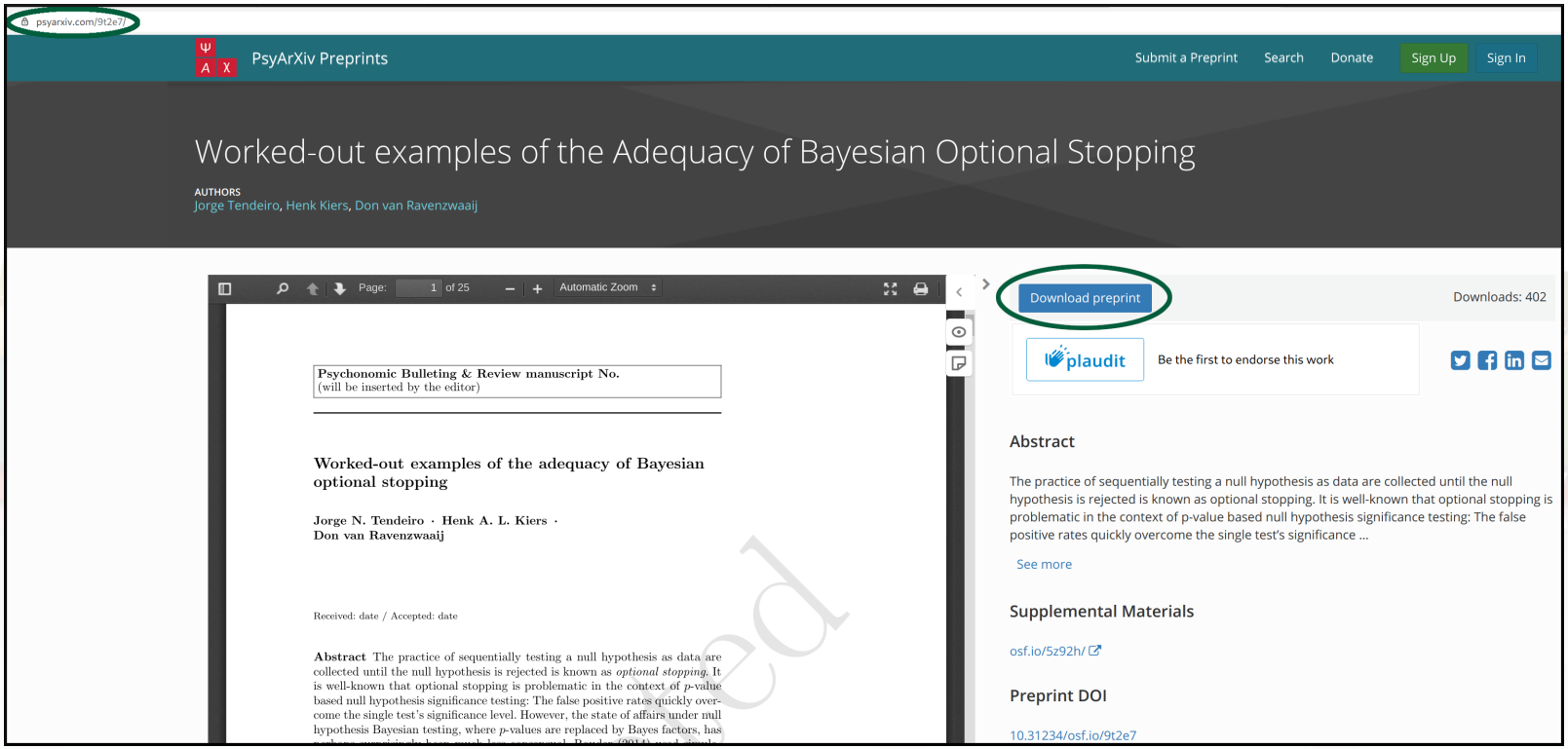
- Upload manuscripts, pre- and/or post-reviewed.
- Free access for everyone to read.
- Common in some fields for years, but still new to many others.

Examples (besides OSF already mentioned):
arXiv (since 1991!), bioRxiv (2013), ChemRxiv (2017), PsyArXiv (2016), PeerJ (2013),...

Do share preprints!

Background: By B S K at FreeImages, license.

What's new
Preprints



Background: By B S K at FreeImages, license.

"Bahh, preprints are of low quality!. . ."

Well, one of the most famous math problems of all times (the Poincaré Conjecture) has a published solution exclusively on arXiv.

Worthy of a long-standing $1,000,000 prize *and* a Fields Medal (both turned down!).

Background: By mohamed Hassan at Pixabay, license.

See Nosek, Ebersole, DeHaven, and Mellor (2018).

Document your research plan online:

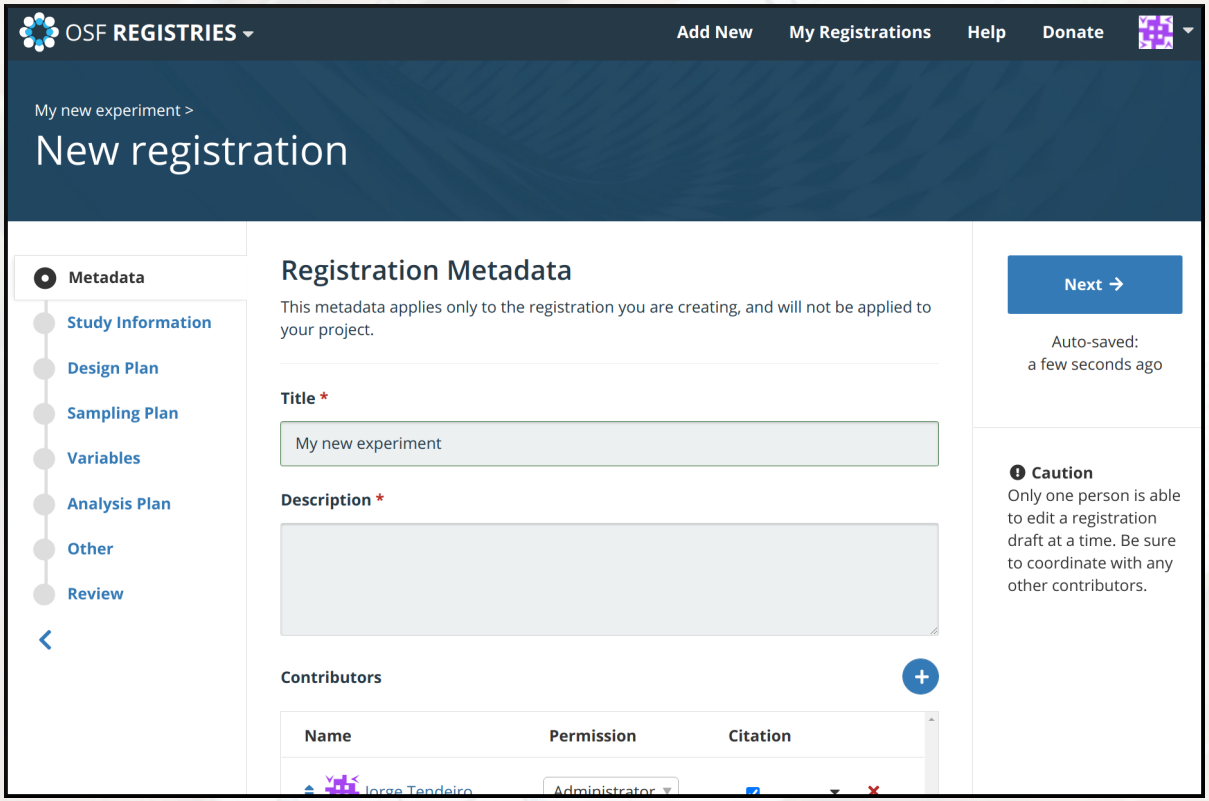
- read-only
- time-stamped
- with pre-analysis plan
- (include as much detail as possible).

Advantages:

- Distinguish exploratory from confirmatory research.
- Reduce researcher df's.
- No p-hacking, HARKing.
- Not a waste of time, just a time-reversed heurisitc.

Background: By Bich Tran at Pexels, license.

Examples: OSF, AsPredicted, ClinicalTrials
(and various options for clinical trials, where this is done for years).



Background: By Bich Tran at Pexels, license.

See Nosek and Lakens (2014).

Main ideas:

- Peer review the RQs and methodology before collecting data: Stage 1 Peer Review.
- Upon *in-principle acceptance*, complete the study by following the protocol.
- Publication is assured upon ascertaining adherence to the registered protocol (or providing compelling reasons to deviate from it): Stage 2 Peer Review.

Background: By Jazella at Pixabay, license.

Major advantage on top of those for preregistrations:

- Avoid publication bias.
- Quality of the study over novely or positive results.

Q: How popular are Registered Reports these days?
A: At the moment, about 300 journals (!) already offered this possibility (see here for a full list).

Background: By Jazella at Pixabay, license.

Wrapping up

For me, it's all about taking little steps.
Trying to do all of it at once is just crazy.
Adapt things to your field and needs.

Background: By Bruno Scramgnon at Pexels, license.

A selection of extra resources you can consider looking at, complementing what was shown before (Robson et al., 2021):

- Check if your journal is/offers open access: Sherpa/Romeo.
- Database of Open Access journals.
- FAIR data principles.
- Data repositories: Nature, Zenodo.
- Request a paywalled article (legally!).
- Peer reviewers' Openness Initiative.

Background: By Bruno Scramgnon at Pexels, license.

I still don't know much. This is what I found:

- https://openscience.jp/. But it seems outdated.
- Research Center for Open Science and Data Platform (RCOS) for research data management.
- JST also has some directives for a few years now.
- A Twitter Open Access account, but it seems inactive.
- JUSTICE (is the name a homage to the Knight Rider?)
  Includes an Open Access roadmap.

Background: By Tomáš Malík at Pexels, license.

Embrace Open Science!

References I

Agnoli, F., Wicherts, J. M., Veldkamp, C. L. S., Albiero, P., & Cubelli, R. (2017). Questionable research practices among italian research psychologists. *PLOS ONE*, *12*(3), e0172792.

Babbage, C. (1830). *Reflections on the Decline of Science in England: And on Some of Its Causes.* http://www.gutenberg.org/files/1216/1216-h/1216-h.htm.

Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature*, *533*(7604), 452–454.

Begley, C. G., & Ellis, L. M. (2012). Raise standards for preclinical cancer research. *Nature*, *483*(7391), 531–533.

Bem, D. J. (2004). Writing the empirical journal article. In *The compleat academic: A career guide, 2nd ed* (pp. 185–219). Washington, DC, US: American Psychological Association.

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*(5), 365–376.

Camerer, C. F., Dreber, A., Forsell, E., Ho, T.-H., Huber, J., Johannesson, M., … Wu, H. (2016). Evaluating replicability of laboratory experiments in economics. *Science*, *351*(6280), 1433–1436.

Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T.-H., Huber, J., Johannesson, M., … Wu, H. (2018). Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. *Nature Human Behaviour*, *2*(9), 637.

Chang, A. C., & Li, P. (2021). Is Economics Research Replicable? Sixty Published Papers FromThirteen Journals Say “Often Not”. *Critical Finance Review*, *10*.

References II

Cohen, J. (1962). The statistical power of abnormal-social psychological research: A review. *The Journal of Abnormal and Social Psychology*, *65*(3), 145–153.

Crüwell, S., van Doorn, J., Etz, A., Makel, M. C., Moshontz, H., Niebaum, J. C., … Schulte-Mecklenbeck, M. (2019). Seven Easy Steps to Open Science: An Annotated Reading List. *Zeitschrift für Psychologie*, *227*(4), 237–248.

Duvendack, M., Palmer-Jones, R., & Reed, W. R. (2017). What Is Meant by "Replication" and Why Does It Encounter Resistance in Economics? *American Economic Review*, *107*(5), 46–51.

Errington, T. M., Iorns, E., Gunn, W., Tan, F. E., Lomax, J., & Nosek, B. A. (2014). An open investigation of the reproducibility of cancer biology research. *eLife*, *3*, e04333.

Fanelli, D. (2009). How Many Scientists Fabricate and Falsify Research? A Systematic Review and Meta-Analysis of Survey Data. *PLOS ONE*, *4*(5), e5738.

Fanelli, D. (2010). "Positive" Results Increase Down the Hierarchy of the Sciences. *PLoS ONE*, *5*(4), e10068.

Fanelli, D. (2012). Negative results are disappearing from most disciplines and countries. *Scientometrics*, *90*(3), 891–904.

Fiedler, K., & Schwarz, N. (2016). Questionable Research Practices Revisited. *Social Psychological and Personality Science*, *7*(1), 45–52.

Fraser, H., Parker, T., Nakagawa, S., Barnett, A., & Fidler, F. (2018). Questionable research practices in ecology and evolution. *PLOS ONE*, *13*(7), e0200303.

Fried, E. I. (2017). The 52 symptoms of major depression: Lack of content overlap among seven common depression scales. *Journal of Affective Disorders*, *208*, 191–197.

References III

Gelman, A., & Loken, E. (2013). The garden of forking paths: Why multiple comparisons can be a problem, even when there is no “fishing expedition” or “p-hacking” and the research hypothesis was posited ahead of time. *Unpublished manuscript*.

Ioannidis, J. P. A. (2005a). Contradicted and Initially Stronger Effects in Highly Cited Clinical Research. *JAMA*, *294*(2), 218.

Ioannidis, J. P. A. (2005b). Why Most Published Research Findings Are False. *PLOS Medicine*, *2*(8), e124.

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the Prevalence of Questionable Research Practices With Incentives for Truth Telling. *Psychological Science*, *23*(5), 524–532.

Kerr, N. L. (1998). HARKing: Hypothesizing After the Results are Known. *Personality and Social Psychology Review*, *2*(3), 196–217.

Klein, R. A., Vianello, M., Hasselman, F., Adams, B. G., Adams, R. B., Alper, S., ... Nosek, B. A. (2018). Many Labs 2: Investigating Variation in Replicability Across Samples and Settings. *Advances in Methods and Practices in Psychological Science*, *1*(4), 443–490.

Martinson, B. C., Anderson, M. S., & de Vries, R. (2005). Scientists behaving badly. *Nature*, *435*(7043), 737.

Meehl, P. E. (1967). Theory-Testing in Psychology and Physics: A Methodological Paradox. *Philosophy of Science*, *34*(2), 103–115.

Mobley, A., Linder, S. K., Braeuer, R., Ellis, L. M., & Zwelling, L. (2013). A Survey on Data Reproducibility in Cancer Research Provides Insights into Our Limited Ability to Translate Findings from the Laboratory to the Clinic. *PLOS ONE*, *8*(5), e63221.

References IV

Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C., Percie du Sert, N., … Ioannidis, J. P. A. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, *1*(1), 0021.

Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences*, *115*(11), 2600–2606.

Nosek, B. A., & Lakens, D. (2014). Registered reports: A method to increase the credibility of published results. *Social Psychology*, *45*(3), 137–141.

Nosek, B. A., Spies, J. R., & Motyl, M. (2012). Scientific Utopia: II. Restructuring Incentives and Practices to Promote Truth Over Publishability. *Perspectives on Psychological Science*, *7*(6), 615–631.

OSC. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716.

Penders, Holbrook, & de Rijcke. (2019). Rinse and Repeat: Understanding the Value of Replication across Different Ways of Knowing. *Publications*, *7*(3), 52.

Prinz, F., Schlange, T., & Asadullah, K. (2011). Believe it or not: How much can we rely on published data on potential drug targets? *Nature Reviews Drug Discovery*, *10*(9), 712–712.

Robson, S. G., Baum, M. A., Beaudry, J. L., Beitner, J., Brohmer, H., Chin, J., … Tangen, J. M. (2021). *Nudging Open Science* (Preprint). PsyArXiv.

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, *86*(3), 638–641.

Schimmack, U. (2015). *Questionable Research Practices: Definition, Detect, and Recommendations for Better Practices.*

Schönbrodt, F. (2015). *Red flags: How to detect publication bias and p-hacking.*

References V

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. *Psychological Science*, *22*(11), 1359–1366.

Steegen, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing Transparency Through a Multiverse Analysis. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *11*(5), 702–712.

Wicherts, J. M., Veldkamp, C. L. S., Augusteijn, H. E. M., Bakker, M., van Aert, R. C. M., & van Assen, M. A. L. M. (2016). Degrees of Freedom in Planning, Running, Analyzing, and Reporting Psychological Studies: A Checklist to Avoid p-Hacking. *Frontiers in Psychology*, *7*.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*(1), 160018.