Open Science

Usman Afzali, PhD - Postdoctoral Fellow and Lecturer

University of Canterbury

2024-07-25

Outline

- Scientific publishing
- Open science
- Preregistration
- Conclusion

Scientific Publishing

The Evolution of Scientific Publishing

- Philosophical Transactions of the Royal Society (1655) by Henry Oldenburg. (https://royalsocietypublishing.org/rsta/about)

The Evolution of Scientific Publishing

- Philosophical Transactions of the Royal Society (1655) by Henry Oldenburg. (https://royalsocietypublishing.org/rsta/about)
- Also, Journal des Sçavans (1665) (https://en.wikipedia.org/wiki/Journal_des_sçavans).

The Evolution of Scientific Publishing

- Philosophical Transactions of the Royal Society (1655) by Henry Oldenburg. (https://royalsocietypublishing.org/rsta/about)
- Also, Journal des Sçavans (1665) (https://en.wikipedia.org/wiki/Journal_des_sçavans).
- Nature (1869) by Sir Joseph Norman Lockyer (https://www.nature.com/nature-portfolio/about) (*IF = 54.4*).

The Evolution of Scientific Publishing

- Philosophical Transactions of the Royal Society (1655) by Henry Oldenburg. (https://royalsocietypublishing.org/rsta/about)
- Also, Journal des Sçavans (1665) (https://en.wikipedia.org/wiki/Journal_des_sçavans).
- Nature (1869) by Sir Joseph Norman Lockyer (https://www.nature.com/nature-portfolio/about) (*IF = 54.4*).
- 19th Century: Growth of specialized journals.

The Evolution of Scientific Publishing

- Philosophical Transactions of the Royal Society (1655) by Henry Oldenburg. (https://royalsocietypublishing.org/rsta/about)
- Also, Journal des Sçavans (1665) (https://en.wikipedia.org/wiki/Journal_des_sçavans).
- Nature (1869) by Sir Joseph Norman Lockyer (https://www.nature.com/nature-portfolio/about) (*IF = 54.4*).
- 19th Century: Growth of specialized journals.
- 20th Century: Rise of electronic publishing and faster dissemination and broader access through digital means

Paywalls and Access to Scientific Articles

- Initially, published by scientific societies or university presses

Paywalls and Access to Scientific Articles

- Initially, published by scientific societies or university presses
- Funded through memberships, subscriptions, or institutional support

Paywalls and Access to Scientific Articles

- Initially, published by scientific societies or university presses
- Funded through memberships, subscriptions, or institutional support
- Commercial publishing: subscription fees and profit motivation

Paywalls and Access to Scientific Articles

- Initially, published by scientific societies or university presses
- Funded through memberships, subscriptions, or institutional support
- Commercial publishing: subscription fees and profit motivation
- Open Access journals

Paywalls and Access to Scientific Articles

- Initially, published by scientific societies or university presses
- Funded through memberships, subscriptions, or institutional support
- Commercial publishing: subscription fees and profit motivation
- Open Access journals
- Hybrid journals

Paywalls and Access to Scientific Articles

- Initially, published by scientific societies or university presses
- Funded through memberships, subscriptions, or institutional support
- Commercial publishing: subscription fees and profit motivation
- Open Access journals
- Hybrid journals
- Some examples: Wiley, Routldeg/Taylor and Francis, Elsevier, Nature, APA.

Questionable Research Practices (QRPs)

- False positive psychology (Simmons, Nelson, and Simonsohn 2011)

Questionable Research Practices (QRPs)

- False positive psychology (Simmons, Nelson, and Simonsohn 2011)
- Researchers degrees of freedom

Questionable Research Practices (QRPs)

- False positive psychology (Simmons, Nelson, and Simonsohn 2011)
- Researchers degrees of freedom
- Data manipulation, selective reporting, lack of transparency (e.g., in terms of excluding outliers)

Questionable Research Practices (QRPs)

- False positive psychology (Simmons, Nelson, and Simonsohn 2011)
- Researchers degrees of freedom
- Data manipulation, selective reporting, lack of transparency (e.g., in terms of excluding outliers)
- Choosing the DV, choosing sample size, covariates, reporting a subset only

Listening to "Why I'm Sixty-Four" makes you younger

Table 3. Study 2: Original Report (in Bolded Text) and the Requirement-Compliant Report (With Addition of Gray Text)

Using the same method as in Study 1, we asked ~~20~~ 34 **University of Pennsylvania undergraduates to listen** only **to either "When I'm Sixty-Four" by The Beatles or "Kalimba"** or "Hot Potato" by the Wiggles. We conducted our analyses after every session of approximately 10 participants; we did not decide in advance when to terminate data collection. **Then, in an ostensibly unrelated task, they indicated** only **their birth date (mm/dd/yyyy) and** how old they felt, how much they would enjoy eating at a diner, the square root of 100, their agreement with "computers are complicated machines," **their father's age**, their mother's age, whether they would take advantage of an early-bird special, their political orientation, which of four Canadian quarterbacks they believed won an award, how often they refer to the past as "the good old days," and their gender. **We used father's age to control for variation in baseline age across participants**.

An ANCOVA revealed the predicted effect: According to their birth dates, people were nearly a year-and-a-half younger after listening to "When I'm Sixty-Four" (adjusted *M* = 20.1 years) rather than to "Kalimba" (adjusted *M* = 21.5 years), *F*(1, 17) = 4.92, *p* = .040. Without controlling for father's age, the age difference was smaller and did not reach significance (*M*s = 20.3 and 21.2, respectively), $F(1, 18) = 1.01$, $p = .33$.

p-Hacking

- Manipulation of data analysis or reporting practices to find statistically significant results

p-Hacking

- Manipulation of data analysis or reporting practices to find statistically significant results
- Inflates the likelihood of finding false positives, undermines the reliability of research

Possible reasons for *p*-Hacking

- **Researcher's Intent:** Desire to publish significant findings.

Possible reasons for *p*-Hacking

- **Researcher's Intent:** Desire to publish significant findings.
- **Ambiguity in Data Analysis:** Lack of clear analytical plans.

Possible reasons for *p*-Hacking

- **Researcher's Intent:** Desire to publish significant findings.
- **Ambiguity in Data Analysis:** Lack of clear analytical plans.
- **Lack of Pre-planning:** Not specifying methods before data collection.

Changing the false positive rate

Table 1. Likelihood of Obtaining a False-Positive Result

Researcher degrees of freedom	Significance level		
	$p < .1$	$p < .05$	$p < .01$
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%

HARKing

- Hypothesizing After the Results are Known

HARKing

- Hypothesizing After the Results are Known
- Presenting a *post hoc* hypothesis as if it were an *a priori* hypothesis

HARKing

- Hypothesizing After the Results are Known
- Presenting a *post hoc* hypothesis as if it were an *a priori* hypothesis
- Dataset: significant correlation between X and Y, $p = .02$

HARKing

- Hypothesizing After the Results are Known
- Presenting a *post hoc* hypothesis as if it were an *a priori* hypothesis
- Dataset: significant correlation between X and Y, $p = .02$
- What are the chances of finding this correlation in this dataset?

HARKing

- Hypothesizing After the Results are Known
- Presenting a *post hoc* hypothesis as if it were an *a priori* hypothesis
- Dataset: significant correlation between X and Y, $p = .02$
- What are the chances of finding this correlation in this dataset?
- HARKing ruins the meaning of *p*-values

Why would one engage in QRP's?

- Pressure to Publish ("Publish or Perish")

Why would one engage in QRP's?

- Pressure to Publish ("Publish or Perish")
- Career advancement

Why would one engage in QRP's?

- Pressure to Publish ("Publish or Perish")
- Career advancement
- Research funding

Why would one engage in QRP's?

- Pressure to Publish ("Publish or Perish")
- Career advancement
- Research funding
- Some examples: *p*-Hacking, HARKing, selective reporting, data fabrication/falsification, inadequate reporting of methods and procedures, publication bias, over-interpretation of results

Avoiding QRP's

Table 2. Simple Solution to the Problem of False-Positive Publications

Requirements for authors

1. Authors must decide the rule for terminating data collection before data collection begins and report this rule in the article.
2. Authors must collect at least 20 observations per cell or else provide a compelling cost-of-data-collection justification.
3. Authors must list all variables collected in a study.
4. Authors must report all experimental conditions, including failed manipulations.
5. If observations are eliminated, authors must also report what the statistical results are if those observations are included.
6. If an analysis includes a covariate, authors must report the statistical results of the analysis without the covariate.

Guidelines for reviewers

1. Reviewers should ensure that authors follow the requirements.
2. Reviewers should be more tolerant of imperfections in results.
3. Reviewers should require authors to demonstrate that their results do not hinge on arbitrary analytic decisions.
4. If justifications of data collection or analysis are not compelling, reviewers should require the authors to conduct an exact replication.

Open Science

- Involves practices that make research processes and outputs more transparent and accessible

- Involves practices that make research processes and outputs more transparent and accessible
- Enhances reproducibility, accountability, and collaboration in research

- Involves practices that make research processes and outputs more transparent and accessible
- Enhances reproducibility, accountability, and collaboration in research
- Premise: Science works best in sunlight

- Involves practices that make research processes and outputs more transparent and accessible
- Enhances reproducibility, accountability, and collaboration in research
- Premise: Science works best in sunlight
- Science paid for by public, given to journals, sold back to public

- Involves practices that make research processes and outputs more transparent and accessible
- Enhances reproducibility, accountability, and collaboration in research
- Premise: Science works best in sunlight
- Science paid for by public, given to journals, sold back to public
- Publication model relies on universities paying for access to journals

- Involves practices that make research processes and outputs more transparent and accessible
- Enhances reproducibility, accountability, and collaboration in research
- Premise: Science works best in sunlight
- Science paid for by public, given to journals, sold back to public
- Publication model relies on universities paying for access to journals
- Science should be publicly available

- Involves practices that make research processes and outputs more transparent and accessible
- Enhances reproducibility, accountability, and collaboration in research
- Premise: Science works best in sunlight
- Science paid for by public, given to journals, sold back to public
- Publication model relies on universities paying for access to journals
- Science should be publicly available
- Preprints (PsyArXiv: https://osf.io/preprints/psyarxiv)

- Involves practices that make research processes and outputs more transparent and accessible
- Enhances reproducibility, accountability, and collaboration in research
- Premise: Science works best in sunlight
- Science paid for by public, given to journals, sold back to public
- Publication model relies on universities paying for access to journals
- Science should be publicly available
- Preprints (PsyArXiv: https://osf.io/preprints/psyarxiv)
- Open access/hybrid journals

Training students for the Open Science future

Felix Schönbrodt

Nature Human Behaviour **3**, 1031 (2019) | Download Citation

289 Accesses | **75** Altmetric | Metrics »

Many PhD students are enthusiastic about robust scientific practices, but afraid that 'doing good science' will jeopardize their chances on the job market, argues Felix Schönbrodt, Managing Director of the LMU Open Science Center. Aligning incentives and preparing students for a job market that values contributions to Open Science will be key.

Access options

Rent or Buy article

Get time limited or full article access on ReadCube.

from $8.99

Rent or Buy

Subscribe to Journal

Get full journal access for 1 year

£104.40

only £8.70 per issue

Subscribe

Open Materials

- Journal articles have word limits, often lack key details
- Unpublished studies left in file drawer – generally, published on OSF, or linked it from PsyArXiv to OSF

Open Data

- Allows researchers to check for mistakes, fraud
- Be careful with confidentiality and privacy
- Enhances potential for meta-analysis, re-analysis

Preregistration

- Design, analyses pre-planned
- Preregistration = writing up your hypotheses, conditions, data analytic plan, etc.
- Prevents *p*-hacking, HARKing
- Does not prevent fraud

Registered Reports

- Peer-review *before* data collection: e.g., https://osf.io/nru4x/?view_only=bc189174a1cf4ca8b1dc83cf7967cd9e

Open Access (OA) refers to making research outputs freely available online, without subscription or payment barriers.

The Open Access movement advocates for free and unrestricted access to research publications. Many researchers and institutions support this model, arguing that publicly funded research should be freely available to everyone. There are various Open Access models, including Gold Open Access (where authors or institutions pay publication fees) and Green Open Access (where authors deposit preprints or postprints in repositories).

cOAlition S: This is an initiative launched in 2018 that requires research funded by participating organizations to be published in compliant open access journals or platforms.

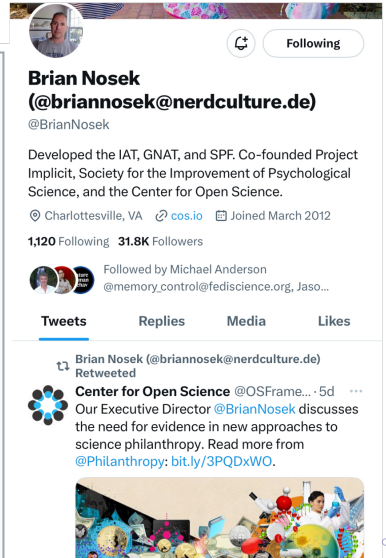
Open Science Manifesto

- Openness adds credibility
- Openness means mistakes are visible (doesn't mean mistakes don't happen!)

Centre for Open Science

CENTER FOR OPEN SCIENCE

URL	cos.io, osf.io
Commercial	No
Launched	2013; 11 years ago
Current status	Active



Open Science Framework

Open Science Framework [edit]

Reproducibility project [edit]

The Open Science Framework (OSF) is an open source software project that facilitates open collaboration in science research. The framework was initially used to work on a project in the reproducibility of psychology research,[11][12] but has subsequently become multidisciplinary.[13] The current reproducibility aspect of the project is a crowdsourced empirical investigation of the reproducibility of a variety of studies from psychological literature, sampling from three major journals: *Journal of Personality and Social Psychology*, *Psychological Science*, and *Journal of Experimental Psychology: Learning, Memory, and Cognition*.[14] Scientists from all over the world volunteer to replicate a study of their choosing from these journals, and follow a structured protocol for designing and conducting a high-powered replication of the key effect. The results were published in 2015.[15]

Preregistration

How to Preregister?

- Various platforms: https://aspredicted.org, https://osf.io
- Information to provide: Hypothesis, Dependent variable, Conditions, Analyses, Outliers and Exclusions, Sample size, Other.
- Example: https://osf.io/7ghcz

Details

- **Hypothesis.** What's the main question being asked, or hypothesis being tested in this study?
- **Dependent variable.** Describe the key dependent variable(s) specifying how they will be measured.
- **Conditions.** How many and which conditions will participants be assigned to?
- **Analyses.** Specify exactly which analyses you will conduct to examine the main question/hypothesis.
- **Outliers and Exclusions.** Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.
- **Sample Size.** How many observations will be collected or what will determine sample size?
- **Other.** Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)

Preregistration FAQ's

- **After running preregistered analyses, discover something surprising!**
- *Totally fine to run exploratory follow-up analyses*
- *Label them as exploratory in the paper*
- **Learn better analysis technique after study is preregistered**
- *Disclose in paper*
- *Usually, should also run planned analyses*
- *Reporting in supplemental materials okay*
- *If you don't run, need to explain why*
- **Data already collected**
- *OK to preregister late, but...*
- *Ideal to preregister before collection*
- *Important to preregister before looking at data*
- *Vital to preregister before running focal test*

Preregistration FAQ's

- **Data collection doesn't go as planned**
- *Power analysis indicates n = 200 N are needed to detect our effect with 80% power. We will collect n = 200."*
- *n = 87 participants failed inclusion criteria*
- *What do you do?*
- *Make reasonable decision, and disclose*
- **Vague analyses**
- *"ANOVA will test the difference between conditions"*
- *Hypothesized direction of differences?*
- *Post hoc comparisons? Planned comparisons? Which type?*
- *Disclose decision-process, be more specific in future*
- **Vague hypothesis**
- *"We predict that people will report lower levels of free will in A and B conditions than in C condition, and furthermore that free will will be lower in the A condition than in B condition."*
- *What does this predict? A > B > C: (A & B) > C, A > B*

A small diversion

- Predatory journals
- Cloned journals

Conclusion

Some useful references:

- *p*-Hacking: Crash Course Statistics #30: https://youtu.be/Gx0fAjNHb1M
- The Replication Crisis: Crash Course Statistics: https://youtu.be/vBzEGSm23y8
- Falsification:
- Academic misconduct:

Any questions?

References

Simmons, Joseph P., Leif D. Nelson, and Uri Simonsohn. 2011. "False-Positive Psychology." *Psychological Science* 22 (11): 1359–66. https://doi.org/10.1177/0956797611417632.