

## POINT OF VIEW

# Open exploration

**Abstract** Arguments in support of open science tend to focus on confirmatory research practices. Here we argue that exploratory research should also be encouraged within the framework of open science. We lay out the benefits of ‘open exploration’ and propose two complementary ways to implement this with little infrastructural change.

WILLIAM HEDLEY THOMPSON<sup>†\*</sup>, JESSEY WRIGHT<sup>†\*</sup> AND PATRICK G BISSETT<sup>†\*</sup>

## Introduction

The aim of open science is to increase the accessibility and transparency of scientific research by making changes to the way research is conducted (Watson, 2015; Nosek et al., 2015; Levin and Leonelli, 2017). To date many of the changes proposed and introduced to make research more open have been concerned with confirmatory research: examples include pre-registering studies (Wagenmakers et al., 2012), redefining statistical thresholds (Benjamin et al., 2018), increasing precision by focusing on statistical power (Button et al., 2013), and conducting replications (Open Science Collaboration, 2015).

These changes are all welcome, but we fear that this focus on confirmation gives the impression that confirmatory research is the sole part of the empirical process worth conducting openly, and that other parts – notably exploratory research – do not belong within the open-science movement. In this article we argue that that exploratory research does have a place in open science, but we do not believe open exploration will automatically follow from current initiatives that advocate open confirmatory practices (Munafò et al., 2017). Instead, exploratory research requires its own protocols and practices to bring it into the open. We start by outlining the problems that arise when exploration is closed and demonstrate that exploration is compatible with the ideals of open science. Thereafter, we outline concrete suggestions to integrate it into empirical practice.

## Moving from closed to open exploration

Data exploration can serve a variety of purposes, including: i) getting a holistic overview and understanding of the data; ii) generating a theory or hypothesis; iii) exploring degrees of freedom within the data. Sociological research detailing emotional labor (Hochschild, 2012), the discovery of penicillin (Ligon, 2004) and the training of AlphaGo (Chen et al., 2018) are a few examples of how exploratory research practices have been critical catalysts for discoveries throughout history. Routinely, advocates of exploratory research champion its value for science (Behrens, 1997; Franklin, 2005; Hollenbeck and Wright, 2017; Jebb et al., 2017; Tukey, 1980) and argue that it is insufficiently funded (Aragon, 2011; Haufe, 2013; Rozin, 2009; Wagner and Alexander, 2013). With the rise of ‘data science’, there has been an increased focus on data mining and extracting exploratory information and patterns (Donoho, 2017). In summary, while the exploration of data is necessary for scientific progress, it is not generally incentivized.

Exploration can be understood through degrees of freedom: the more degrees of freedom a researcher leverages, the more exploratory their work is. The leveraging of degrees of freedom can occur in study design (e.g., piloting experiments), during post-hoc analyses, and as an analysis strategy itself. Moreover, exploratory data analysis has been recognized as a practice distinct from confirmatory analyses since at least 1980 (Tukey, 1980). Following Tukey, we conceive of exploratory research as an interactive process in which a researcher adopts an attitude

\*For correspondence: william.thompson@stanford.edu (WHT); jessey.wright@gmail.com (JW); pbissett@stanford.edu (PGB)

<sup>†</sup>These authors contributed equally to this work

**Competing interests:** The authors declare that no competing interests exist.

**Funding:** See page 4

**Reviewing editor:** Peter Rodgers, eLife, United Kingdom

© Copyright Thompson et al. This article is distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use and redistribution provided that the original author and source are credited.

of flexibility. This stands in a clear contrast with the characteristics of good confirmatory research. To put the contrast succinctly: productive exploratory research involves taking advantage of degrees of freedom, while good confirmatory practices involve limiting and reducing degrees of freedom.

While exploration is central to progress in many domains of science, it is rarely made public. The current blueprint for a scientific paper encourages the author to feign omniscience: the exact correct hypothesis was known a priori, only a small number of confirmatory statistical tests were run to address that hypothesis, and the tests came out exactly as predicted (*Grinnell, 2013*). This blueprint is further perpetuated by funding agencies and journals that reward these practices. In reality, before any confirmatory tests are run, there is often significant exploratory science. Therefore, perhaps in service of this omniscience myth, exploration is generally closed.

## Productive exploratory research involves taking advantage of degrees of freedom, while good confirmatory practices involve limiting and reducing degrees of freedom.

This poses several problems. First, by keeping exploration closed, there is little opportunity to improve one's own exploratory practice by studying the methods and strategies used by peers. This is a missed pedagogical opportunity because a big challenge for many young scientists is moving a research program from a blooming, buzzing confusion of infinite scientific possibilities to a specific question worth attempting to confirm. Second, because the standard for current papers is a tight, clean, linear, novel, and confirmatory narrative, researchers are encouraged to mischaracterize their exploratory science as confirmatory science (*Wagenmakers et al., 2012*). Third, many exploratory studies go unreported, contributing

to the 'file drawer problem' (*Rosenthal, 1979*). Fourth, a further consequence of 'closed exploration' is that a research group can waste time and resources on studies that others know will not work. Fifth, while closed exploration remains the norm, confirmation will be perceived as the only scientific process worth doing, thus discouraging researchers from doing good or thorough exploration. For these reasons, we advocate for open exploration.

In broad terms open exploration involves making all aspects of the exploratory research pipeline available for others to see what has been done, with what, and how. Such practices are compatible with current guidelines and initiatives that promote open research. Specifically, the transparency and openness promotion (TOP) guidelines were created to outline how research practices, broadly speaking, can be conducted openly (*Nosek et al., 2015*; <https://cos.io/top/>). Exploratory research can adhere to the sections of these guidelines that concern transparency of the data, code, research methods, design and analysis. The sections in the TOP guidelines that are not relevant for open exploration are pre-registration (as this demarcates confirmatory from exploratory research) and replication (as this is a type of confirmatory research). Finally, in advocating for and thinking about open exploration, it is critical to recognize that open exploration complements confirmatory research, rather than somehow being in opposition to it.

## Proposed implementations of open exploration

We propose two concrete ways to instantiate open exploration: i) include the exploratory science as a distinct section in published papers; ii) place exploratory analyses (regardless of the outcome) on citable public repositories. Before elaborating on these proposals, we present four important additional details. First, we recognize that these solutions will not rectify the entire problem. We offer them as concrete steps forward that are low cost. Second, our proposals aim to promote exploration that does not currently see the light of day. Notably, there is some exploratory research that is already in the open, such as post hoc analyses and fully exploratory studies – see, for example, the 'exploratory reports' format adopted by the journal *Cortex*. Third, we admit that these proposals are not necessarily novel but they are stated here to

encourage their adoption. Fourth, our experience is primarily in psychology and neuroscience, so our proposals may be best suited to these and related fields. However, we hope that they may be of value to other disciplines that are facing similar challenges as there is nothing discipline-specific about them.

One way to implement the first suggestion would be to include separate sections in scientific papers for exploration. To fit this neatly into current structures, this may involve an exploratory methods section and an exploratory results section. Regardless of the specific implementation, such section(s) would give researchers an opportunity to provide details about pilot studies, initial but incorrect hypotheses, and descriptive and exploratory data analyses that were conducted prior to the confirmatory part of the study. Such section(s) do not need to focus on, if at all, inferential statistical results or power: instead, they should describe the exploratory work that was done and how it led to the hypothesis generation, methods refinement, or any additional insight for the confirmatory research. The subsequent (confirmatory) methods and results sections will then be informed by the exploratory results reported in the paper and would be held to the strict standards of confirmatory science. This re-imagining of the scientific paper will demonstrate that exploratory research is as an important part of the scientific process, instead of discouraging it or pretending it does not occur.

**This re-imagining of the scientific paper will demonstrate that exploratory research is as an important part of the scientific process, instead of discouraging it or pretending it does not occur.**

Adopting the first proposal will only make the tip of the exploratory iceberg open. To address the file drawer problem and avoid redundant work by researchers, other exploratory work that is currently considered to be 'non-publishable' must also be placed in the open. However, we do not want to flood the scientific literature with

exploratory work that does not necessarily warrant a publication, so we propose that researchers place their exploratory analysis – ideally including the code, data (if it is shareable), and descriptions of the study – on a repository with a citable DOI. (Suitable repositories include figshare, Github and osf.io, and these can be integrated with DOIs generated from Zenodo or protocols.io). There are already some instances where exploratory analyses are shared in this way (see, for example, *Konkiel, 2016*). For wider adoption of this suggestion, and to ease readability of exploratory analyses in repositories, developing a standardized template to describe the exploratory repository could be beneficial.

## Challenges and conclusions

We have outlined two proposals that implement open exploration with little infrastructural change. Together they would help to curtail the problems of closed exploration outlined previously: for example, by giving a place to exploration in journal articles, more parts of the scientific process will be conducted in the open, which would help to dispel the omniscience myth, and would also generate material that can be used for future investigations into good exploratory practices. Moreover, open exploration may make it possible to conduct meta-exploratory analyses that identifies both successful and under-exploited places in the parameter space to guide future work. This could be done by, for example, reviewing exploratory analysis repositories and presenting a summary of the parameter configurations that have been explored (noting where researcher degrees of freedom appear to make a difference) and the null and positive findings contained in the analyses.

We would also like to address some concerns that researchers might have about our proposals. First, some researchers may be hesitant to publicly display exploratory work that is incomplete, such as rough or messy code. However, sharing any code, even messy or rough code, is better than having no code accompanying an article: indeed, it has been shown that sharing any code at all increases scientific engagement with articles, as measured by citations, in image processing research (*Vandewalle, 2012*). See *Gorgolewski and Poldrack (2016)* for more advice on sharing code (and data). Second, it is important that people do not overlook that null exploratory findings could be type-II errors (that is, false negatives). Third,

there is also a possibility of more type-I errors (that is, false positives) in open exploration. However, as more exploratory analyses are shared, it may become possible to discriminate between findings worth confirming (be they negative or positive) and type-I or type-II errors due to factors such as a poor experimental design. At present, news of negative results encountered in unpublished work due to poor experimental design can only be spread through word-of-mouth, dissuading others from pursuing this line of research. Open exploration offers the possibility for these results to be evaluated. In sum, it is crucial that exploratory work should be seen as a guide for future work, not as definitively confirmed hypothesis.

Finally, there is the concern that positive or negative exploratory results that have not been confirmed may be adopted as facts or falsehoods by the public. However, misleading information already exists and gets proliferated as facts (see, for example, *Nordenstedt and Rosling, 2016*). With greater appreciation of exploration, the distinction between researchers producing tentative exploratory results and hypothesis-driven research will become more apparent to both scientists and a wider audience.

To conclude, open science has focused on confirmatory science and neglected exploratory science. Exploration exists, has scientific value, and is mostly closed. We propose open exploration, which involves bringing exploratory science out of the shadows into scientific papers and public repositories, and in doing so unlocks the potential of this essential part of the scientific process.

### Acknowledgements

We thank Russell Poldrack and James Mac Shine for helpful comments and discussions.

**William Hedley Thompson** is in the Department of Psychology, Stanford University, Stanford, United States, and the Department of Clinical Neuroscience, Karolinska Institute, Stockholm, Sweden  
william.thompson@stanford.edu  
 <https://orcid.org/0000-0002-0533-6035>

**Jessey Wright** is in the Department of Psychology and the Department of Philosophy, Stanford University, Stanford United States  
jessey.wright@gmail.com

**Patrick G Bissett** is in the Department of Psychology, Stanford University, Stanford United States  
pbissett@stanford.edu

**Author contributions:** William Hedley Thompson, Jessey Wright, Patrick G Bissett, Conceptualization, Writing - original draft, Writing - review and editing

**Competing interests:** The authors declare that no competing interests exist.

**Received** 24 September 2019

**Accepted** 11 December 2019

**Published** 09 January 2020

### Funding

Funder	Grant reference number	Author
Knut och Alice Wallenbergs Stiftelse	2016.0473	William Hedley Thompson
National Institute of Mental Health	R01MH117772	Patrick G Bissett
National Science Foundation	1655839	Jessey Wright
Social Sciences and Humanities Research Council of Canada	756-2017-0766-SSHRC	Jessey Wright
National Institute on Drug Abuse	UH3DA041713	Patrick G Bissett

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

### Decision letter and Author response

Decision letter <https://doi.org/10.7554/eLife.52157.sa1>

Author response <https://doi.org/10.7554/eLife.52157.sa2>

### Additional files

#### Data availability

No data is used in this article.

#### References

- Aragon R.** 2011. Thinking outside the box: fostering innovation and non-hypothesis-driven research at NIH. *Science Translational Medicine* **3**:70cm5. DOI: <https://doi.org/10.1126/scitranslmed.3001742>, PMID: 21325614
- Behrens JT.** 1997. Principles and procedures of exploratory data analysis. *Psychological Methods* **2**: 131–160. DOI: <https://doi.org/10.1037/1082-989X.2.2.131>
- Benjamin DJ,** Berger JO, Johannesson M, Nosek BA, Wagenmakers EJ, Berk R, Bollen KA, Brembs B, Brown L, Camerer C, Cesarini D, Chambers CD, Clyde M, Cook TD, De Boeck P, Dienes Z, Dreber A, Easwaran K, Efferson C, Fehr E, et al. 2018. Redefine statistical significance. *Nature Human Behaviour* **2**:6–10. DOI: <https://doi.org/10.1038/s41562-017-0189-z>, PMID: 30980045

- Button KS**, Ioannidis JP, Mokrysz C, Nosek BA, Flint J, Robinson ES, Munafò MR. 2013. Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience* **14**:365–376. DOI: <https://doi.org/10.1038/nrn3475>, PMID: 23571845
- Chen Y**, Huang A, Wang Z, Antonoglou I, Schrittwieser J, Silver D, Freitas N. 2018. Bayesian optimization in AlphaGo. *arXiv*. <https://arxiv.org/abs/1812.06855>.
- Donoho D**. 2017. 50 years of data science. *Journal of Computational and Graphical Statistics* **26**:745–766. DOI: <https://doi.org/10.1080/10618600.2017.1384734>
- Franklin LR**. 2005. Exploratory experiments. *Philosophy of Science* **72**:888–899. DOI: <https://doi.org/10.1086/508117>
- Gorgolewski KJ**, Poldrack RA. 2016. A practical guide for improving transparency and reproducibility in neuroimaging research. *PLoS Biology* **14**:e1002506. DOI: <https://doi.org/10.1371/journal.pbio.1002506>, PMID: 27389358
- Grinnell F**. 2013. Research integrity and everyday practice of science. *Science and Engineering Ethics* **19**:685–701. DOI: <https://doi.org/10.1007/s11948-012-9376-5>, PMID: 22740035
- Haufe C**. 2013. Why do funding agencies favor hypothesis testing? *Studies in History and Philosophy of Science Part A* **44**:363–374. DOI: <https://doi.org/10.1016/j.shpsa.2013.05.002>
- Hochschild AR**. 2012. *The Managed Heart: Commercialization of Human Feeling*. Oakland, CA: University of California Press.
- Hollenbeck JR**, Wright PM. 2017. Harking, sharking, and tharking: making the case for post hoc analysis of scientific data. *Journal of Management* **43**:5–18. DOI: <https://doi.org/10.1177/0149206316679487>
- Jebb AT**, Parrigon S, Woo SE. 2017. Exploratory data analysis as a foundation of inductive research. *Human Resource Management Review* **27**:265–276. DOI: <https://doi.org/10.1016/j.hrmr.2016.08.003>
- Konkiel S**. 2016. What kinds of research are cited quickly in policy documents? Results from an exploratory study. *Figshare* **16**:25. DOI: <https://doi.org/10.6084/m9.figshare.3175168.v1>
- Levin N**, Leonelli S. 2017. How does one "open" science? Questions of value in biological research. *Science, Technology, & Human Values* **42**:280–305. DOI: <https://doi.org/10.1177/0162243916672071>, PMID: 28232768
- Ligon BL**. 2004. Penicillin: its discovery and early development. *Seminars in Pediatric Infectious Diseases* **15**:52–57. DOI: <https://doi.org/10.1053/j.spid.2004.02.001>, PMID: 15175995
- Munafò MR**, Nosek BA, Bishop DVM, Button KS, Chambers CD, Percie du Sert N, Simonsohn U, Wagenmakers E-J, Ware JJ, Ioannidis JPA. 2017. A manifesto for reproducible science. *Nature Human Behaviour* **1**:0021. DOI: <https://doi.org/10.1038/s41562-016-0021>
- Nordenstedt H**, Rosling H. 2016. Chasing 60% of maternal deaths in the post-fact era. *Lancet* **388**:1864–1865. DOI: [https://doi.org/10.1016/S0140-6736\(16\)31793-7](https://doi.org/10.1016/S0140-6736(16)31793-7), PMID: 27751384
- Nosek BA**, Alter G, Banks GC, Borsboom D, Bowman SD, Breckler SJ, Buck S, Chambers CD, Chin G, Christensen G, Contestabile M, Dafoe A, Eich E, Freese J, Glennerster R, Goroff D, Green DP, Hesse B, Humphreys M, Ishiyama J, et al. 2015. Promoting an open research culture. *Science* **348**:1422–1425. DOI: <https://doi.org/10.1126/science.aab2374>
- Open Science Collaboration**. 2015. Estimating the reproducibility of psychological science. *Science* **349**:aac4716. DOI: <https://doi.org/10.1126/science.aac4716>
- Rosenthal R**. 1979. The file drawer problem and tolerance for null results. *Psychological Bulletin* **86**:638–641. DOI: <https://doi.org/10.1037/0033-2909.86.3.638>
- Rozin P**. 2009. What kind of empirical research should we publish, fund, and reward? A different perspective. *Perspectives on Psychological Science* **4**:435–439. DOI: <https://doi.org/10.1111/j.1745-6924.2009.01151.x>, PMID: 26158991
- Tukey JW**. 1980. We need both exploratory and confirmatory. *American Statistician* **34**:23–25. DOI: <https://doi.org/10.2307/2682991>
- Vandewalle P**. 2012. Code sharing is associated with research impact in image processing. *Computing in Science & Engineering* **14**:42–47. DOI: <https://doi.org/10.1109/MCSE.2012.63>
- Wagenmakers EJ**, Wetzels R, Borsboom D, van der Maas HL, Kievit RA. 2012. An agenda for purely confirmatory research. *Perspectives on Psychological Science* **7**:632–638. DOI: <https://doi.org/10.1177/1745691612463078>, PMID: 26168122
- Wagner CS**, Alexander J. 2013. Evaluating transformative research programmes: a case study of the NSF small grants for exploratory research programme. *Research Evaluation* **22**:187–197. DOI: <https://doi.org/10.1093/reseval/rvt006>
- Watson M**. 2015. When will 'open science' become simply 'science'? *Genome Biology* **16**:101. DOI: <https://doi.org/10.1186/s13059-015-0669-2>