

Fallibility in Science: Responding to Errors in the Work of Oneself and Others

D. V. M. Bishop

Department of Experimental Psychology, University of Oxford

Received 1/3/18; Revision accepted 4/20/18

Imagine the following scenario:

Ph.D. student David has run a series of studies trying to find a positive effect of brain stimulation on language comprehension in stroke patients. After three studies with null findings, he has changed the design in various ways and is overjoyed when the fourth study gives a statistically significant effect. His article is published in a prestigious, high-impact journal, with David as first author and his eminent supervisor as last author.

The university press office promotes the study, and it is featured on National Public Radio. Two weeks later, when preparing slides for a talk at the Society for Neuroscience, David finds that the groups were miscoded, and in fact the sham-treatment group obtained higher posttraining scores than the stimulation group.

When I use fictitious examples like this in seminars and ask the audience, “What should David do?” the usual response is that, of course, David should come clean, admit the error, and ask for the article to be retracted. But there is typically nervousness in the room. It is pointed out that there are massive pressures on him not to do so: The general perception is that admission of error will mean that the reputation of both David and his supervisor will be in tatters, with David’s prospects for a future career badly damaged.

Yet there are real-life examples of scientists admitting to honest errors that show that this doom-laden scenario is unrealistic. In a recent study, Azoulay, Bonatti, and Krieger (2017) considered how reputation is affected by retraction, by comparing subsequent citations of earlier published articles for authors who had and who had not had an article retracted. Retraction due to researchers’ misconduct led to a drop in

subsequent citations of their earlier work, but there was a smaller effect when honest error was involved—and no evidence of reputational damage for junior researchers. In an interview study of 14 authors whose articles were retracted after they notified the journal of errors, Hosseini, Hilhorst, de Beaufort, and Fanelli (2018) found that, contrary to the interviewees’ expectations, self-retraction did not damage their reputation and in some cases improved it. This fits with more informal evidence suggesting that there can be reputational advantage from going public in correcting an error: You demonstrate that you are someone who values scientific accuracy over your success in publishing (Retraction Watch, 2017). Nevertheless, there may be pressures from institutions or senior colleagues to hide errors, and journal editors are not always supportive. Hosseini et al noted: “Many authors expected rapid, empathic and detailed responses from journal editors, but reported receiving short, unsympathetic and sometimes unpleasant ones instead” (p. 200).

The thought of having to retract an article can instill fear into the heart of scientists, who see it as equivalent to being named and shamed. There are currently few incentives for honesty, and keeping quiet about an error will often seem the easiest option. Recognizing that the threat of bad consequences could act as a deterrent to honest admission of error, Retraction Watch instituted the Doing the Right Thing award to “honor those who clean up the scientific literature” (Oransky & Marcus, 2017). I give some examples of researchers who have publicized their own errors in Box 1.

There are two further points to take from the David scenario. As awful and embarrassing as it is to admit to error, the alternative, hiding a known error, has to

Advances in Methods and
Practices in Psychological Science
2018, Vol. 1(3) 432–438
© The Author(s) 2018
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/2515245918776632
www.psychologicalscience.org/AMPPS



Corresponding Author:

D. V. M. Bishop, Department of Experimental Psychology, University of Oxford, OX2 6GG, United Kingdom
E-mail: dorothy.bishop@psy.ox.ac.uk

Box 1. Examples of Researchers Who Highlighted Errors in Their Own Work

- With six coauthors, Richard Mann, a postdoctoral researcher using statistical methods to study behavioral ecology, had published an article on behavior in prawns in *PLOS Computational Biology*. He shared the data set with a colleague who was looking for data to test out some ideas on numerical integration. On his blog, Mann (2013) described the moment when the colleague phoned him to tell him of a fatal error in his analysis. As stated in the retraction notice (Mann et al., 2012),

Where each of 102 experiments should have been down-sampled to half the original size for computational efficiency, instead the number of experiments in the data set was repeatedly halved 102 times. . . . results and conclusions were based on only one experimental study, rather than the 102 reported in the paper.

The article was retracted, and the analysis was redone, giving similar findings. Mann (2013) stated that although he had a terrible few months, he did not suffer any long-term stigma.

- A story in *Nature News* (Gewin, 2015) documented how Pamela Ronald, a professor in plant pathology, became concerned when two of her postdocs could not replicate findings she had published, in 1995, in two high-profile articles on the basis of the immune response in rice. She notified the journal editors and then devoted the next 18 months to trying to locate the source of the discrepancy. It turned out that the strains of microbes she had been using were mislabeled, and in 2013 the articles were retracted. In 2015, Ronald published an article correctly identifying the source of the immune response. She has changed her lab procedures so that three independent researchers now validate new experimental approaches.
- Senior neuroscientist Russ Poldrack wrote computer code to classify a set of brain images into classes according to the task being performed. He had submitted a manuscript based on this analysis for publication, when a student collaborator told him that after obtaining far lower classification accuracy on the same data set, he found an error in the code. Poldrack's (2013) response was to write a blog post about this experience, encouraging everyone to share code, use better methods for checking code, and talk about their errors.

be worse. The person who does this is entering into a Faustian pact to reject science in favor of personal ambition. As data fraudster Diederik Stapel openly admitted, once you embark on this process, it is difficult to stop, but it creates considerable internal conflict (Stapel, 2014, pp. 128–131).

The second point is that although errors can never be eliminated, they can be reduced by adoption of open-science practices. Even in situations in which the raw data cannot be made completely open in a repository, usually because of confidentiality issues, it is often possible to deposit a version that has been modified to remove identifiable information, so that other researchers can reproduce what was done (UK Data Service, n.d.). For sensitive data, a data-sharing agreement may be needed in addition to anonymization (Medical Research Council, 2017). Practical suggestions for data sharing in psychology were recently proposed by Gilmore, Kennedy, and Adolph (2018).

Regardless of the level of security that is required, there should be no barriers to researchers making their analysis code open, so that the analysis steps can be checked. The example from Russ Poldrack in Box 1 illustrates how easy it is even for an experienced scientist to make an error in coding that has serious consequences for results. People often worry that if they make their code and data open, errors will be found, but that is really the whole point: We need to make code and data open because this is how the errors can be found. Also, if you know that your code and data will be open, you are likely to check and double-check

them far more rigorously than if you know they will never be seen by anyone else. So open practices reduce the likelihood of error. Furthermore, errors in analysis scripts are extremely common among scientists who have taught themselves to program (Merali, 2010), so errors are likely to be present. But to encourage people to share scripts, we must remove any stigma associated with detection of errors. This is not condoning sloppy science: It is just accepting the reality that we are all fallible.

Of course, making analysis programs open does not guarantee that they are free from errors. A result may be reproducible—in the sense that we obtain consistent results when the same data are run through the same program—but it may still be wrong. An example of widely used neuroimaging software that was discovered to include a bug only after many years of use was reported by Eklund, Nichols, and Knutsson (2016a). The authors noted in a subsequent correction that they were not implying that all analyses using the software were erroneous, but rather meant that it was not possible to establish which were. They explained, “Due to lamentable archiving and data-sharing practices, it is unlikely that problematic analyses can be redone” (Eklund, Nichols, & Knutsson, 2016b, para. 3). Quite simply, making code and data open does not prevent errors, but it does make it possible to detect them. And as amply documented elsewhere, it brings other benefits to researchers, in terms of improving their science as well as enhancing recognition of their work (Markowitz, 2015; McKiernan et al., 2016).

Errors in Someone Else's Work: How to Respond

The prior discussion of errors in one's own work should give clues about how to respond when you find errors in another person's work. You would not want to be pilloried for an honest error, so do not pillory others for simple mistakes. In a comment on a blog post on this topic, Weil (2014) put it very well:

. . . my first prominent publication was a note tearing down someone else's work. That work had appeared in a major journal and caused quite a stir — but the apparent results were the product of a careless (not dishonest, just careless) mistake in the analysis. The note pointing this out was not derogatory in tone, nor was it intended to shame, but was doubtless embarrassing to the authors.

Now that I am much older, a little wiser, and a little kinder (and a lot more employed, and thus less vulnerable to jerks) I would send the authors my analysis of their math first and give them the opportunity to correct. And I hope that my colleagues would give me the same consideration if (when?) I make a stupid mistake.

Life, however, is not always so simple. The researcher whose error is remarked on may respond with anger, denial, or silence. This is, of course, a normal human reaction, but it is not a sensible response if the error is unambiguous, as it can damage the author's reputation for integrity. In theory, it should be possible to resolve such issues via the journal that published the original article, but in practice, this process seldom proceeds smoothly. Allison, Brown, George, and Kaiser (2016) described how their own attempts to correct substantial errors in other researchers' work met with inaction or delaying tactics by authors and editors, and even demands for payment to publish a letter pointing out the errors. At the time of writing this Commentary, it was possible to put the record right by adding a comment in PubMed Commons (Bastian, 2014). The comment was linked to the abstract of the original article on PubMed and became part of the scientific record. The first two examples in Box 2 illustrate how both authors and other researchers have used PubMed Commons to record a correction. However, despite its utility, PubMed Commons was not widely used by commenters and was discontinued in February 2018, though the comments remain archived (National Center for Biotechnology Information, 2018).

Errors in Interpretation of Data

Research results may seem suspect because of concerns about methodology, rather than straightforward errors in calculation or scripting. For instance, a study may lack a control group, be underpowered, use an unreliable measure, or have a major confound. There may be strong suspicion that the author has engaged in *p*-hacking. These are not simple errors that can be corrected, but they affect the conclusions that can be drawn. All of these are situations in which PubMed Commons provided a venue for raising the concerns, as illustrated in Box 2, Examples 3 through 5. With the disappearance of PubMed Commons, there are limited options remaining to researchers who want to engage in postpublication peer review, given that few journals have options for commenting. For researchers who do not have access to a blog, an alternative platform, PubPeer, is likely to become the method of choice for postpublication peer review. An important difference from PubMed Commons is that commenters can be anonymous. Probably because of this, PubPeer has been far more popular than PubMed Commons, but it is also noted for a harsh style of criticism that can include accusations of malpractice (Dolgin, 2018). This is unfortunate because it leads to the impression that postpublication peer review typically involves a personal attack. Harsh criticism can polarize debate and make many people reluctant to engage. PubMed Commons was also used to draw attention to malpractice, but typically such comments described the problem without engaging in personal attack (see Box 2, Example 6).

My recommendation is that when errors are found, the starting position should be that methodological weaknesses are due to ignorance rather than bad faith. Consider, for instance, *p*-hacking. The dangers of this practice were pointed out many years ago (de Groot, 2014), but it has been normative for decades in many branches of science, including psychology. Before he moved on to fraud, Stapel (2014) engaged in *p*-hacking, noting:

What I did wasn't whiter than white, but it wasn't completely black either. It was grey, and it was what everyone did. (p. 102)

Even now that it has been prominently demonstrated that *p*-hacking is a major cause of false positive findings (Simmons, Nelson, & Simonsohn, 2011), many researchers still do not recognize how seriously it can distort results (Nuzzo, 2014). Furthermore, it is likely that *p*-hacking is deemed acceptable, because it involves paltering, that is, using a truthful statement (e.g., that the *p* value associated with a contrast is $< .05$)

Box 2. Examples of Postpublication Commentary on PubMed Commons

1. Author adding minor corrections: <https://hypothes.is/search?q=tag%3APubMedCommonsArchive+28436345>
Jim van Os noted some numerical errors in a table in an article he had published.
2. Reviewer correcting an error: <https://hypothes.is/search?q=tag%3APubMedCommonsArchive+28461468>
Pavel Nesmiyanov noted that β -endorphin, oxytocin, and dopamine were wrongly described as neuropeptides in a journal article. Although the authors did not respond on PubMed Commons, an erratum was published in the journal.
3. Reviewer critiquing methods: <https://hypothes.is/search?q=tag%3APubMedCommonsArchive+29153326>
Franck Ramus criticized the small sample size of a study on neurobiological correlates of dyslexia. The authors responded, defending the small sample size and arguing that their analyses were driven by an a priori hypothesis derived from a previous study.
4. Reviewer critiquing methods: <https://hypothes.is/search?q=tag%3APubMedCommonsArchive+28706072>
Serge Ahmed suggested that a study of planning in ravens needed an additional control for learning of the affective value of objects.
5. Reviewer noting overhyped interpretation of results: <https://hypothes.is/search?q=tag%3APubMedCommonsArchive+28735725>
Clive Bates noted that a study finding an association between vaping and smoking tobacco in adolescents had been widely interpreted in the media as showing a causal link. He added a link to a more detailed critique of the study.
6. Reviewer raising more serious concerns: <https://hypothes.is/search?q=tag%3APubMedCommonsArchive+17688420>
David Nunan noted previously raised concerns about duplicate data in an article on the role of diet in congestive heart failure.

to mislead by failing to provide relevant contextual information (e.g., that this comparison was one of numerous comparisons and would not be statistically significant if correction were made for multiple contrasts; Rogers, Zeckhauser, Gino, Norton, & Schweitzer, 2017).

Scientists are particularly prone to paltering when it comes to citing the results of other researchers. The process of conducting a literature review is likely to be affected by confirmation bias, that is, seeking and remembering evidence that supports one's position, and ignoring or forgetting evidence that does not (Nickerson, 1998). Rogers et al. (2017) showed that people judge such omission as less dishonest than inclusion of untrue information, and it is often unwitting, but the consequences can be substantial (Greenberg, 2009). One way of counteracting bias in literature reviews is to require that they follow the format of a systematic review, in which criteria for deciding which reports to include are specified in advance (Gough, Oliver, & Thomas, 2017; Wicherts, 2017).

Failure to Replicate: An Unreliable Indicator of Fallibility

I have focused so far on situations in which there are either honest errors in the data or analysis or methodological weaknesses that compromise conclusions that can be drawn. A much more complicated scenario arises when there is difficulty in replicating a published result. This has become a hot topic in science in recent years (Munafò et al., 2017), and failure to replicate findings in psychology was brought to the fore by an influential study published in *Science* (Open Science

Collaboration, 2015). These developments coincided with growing awareness of *p*-hacking as an endemic problem for psychology (Simmons et al., 2011), which made it easy to conclude that results that were not replicated were indicative of bad science. The key point to note is that although erroneous data, erroneous inferences, and failure to control bias can lead to results that are not replicated, one cannot assume that failure to replicate is necessarily the result of any of these types of error. In psychology, we are dealing with probabilistic phenomena, so random noise is always a factor affecting results: Our statistical methods are designed to guard against Type I and Type II errors, but there is an inevitable trade-off, so some statistically significant differences will be false positives, and some failures to find an effect will be false negatives (see Box 3 for a list of possible reasons for failure to replicate). Replication is important precisely because our confidence in the robustness of a given finding cannot depend on a single study.

So, the question arises as to how researchers should respond when there is a failure to replicate prior work. Given the range of reasons for nonreplication, it should not be assumed that a failure to replicate a result is evidence of poor science in the original study. Nevertheless, it is important to uncover reasons for discrepant findings. Ideally, the two sets of researchers should work together to consider how to reconcile the discrepancy. If the original researchers believe that contextual factors or researcher expertise are critical to obtaining their result, then it is up to them to specify more carefully the conditions under which the effect obtains, rather than simply put forward hypothetical explanations for a null result. When there is a failure to

Box 3. Possible Reasons for Failure to Replicate a Scientific Result

- The initial result was a false positive due to chance variation (Type I error)
- The replication study failed to detect a true effect because of chance variation (Type II error)
- The results are sensitive to contextual factors
- The method requires specific expertise that the researcher conducting the replication lacks
- The initial results rested on data-entry, computational, or statistical errors
- The initial results were obtained using questionable research practices, such as *p*-hacking

replicate a finding, it is bad if the first response is to disparage the original researchers as incompetent, malign, or fraudulent, but it is just as bad if researchers whose findings were not replicated dismiss the critics as lacking in expertise or having malevolent motives. Again, the kudos will go to the researchers who show integrity in putting scientific truth before their own career ambitions. As a positive example, consider Finkel's (2016) reflections on a failure to replicate one of his studies: "Although I am surprised by the failure of the manipulation check and disappointed that the results of the [Registered Replication Report] did not confirm the causal effects my colleagues and I originally reported, I deeply respect the process" (p. 766).

Deliberate Omission, Misrepresentation, and Misconduct

I turn now to those unfortunate situations in which it is hard to avoid concluding that a researcher is acting in bad faith. A particularly insidious kind of behavior involves deliberate selective citation of the literature, or cherry-picking. As is the case with other methodological errors, it can be difficult to distinguish deliberate misconduct from unwitting omission. No person should be pilloried for occasional bias in a review's coverage: Even if one strenuously attempts to avoid bias, searches to identify publications on a topic may miss relevant articles because positive findings garner far more citations than null findings (Greenberg, 2009). Citation bias morphs into misconduct when there is a persistent pattern of an author ignoring contrary evidence, even when it is readily available and drawn to his or her attention. Worse still are cases in which cited studies are inaccurately portrayed. These are standard ploys by authors promoting pseudoscientific views (Grimes & Bishop, 2017) and need to be robustly challenged. However, to do so effectively, it may be necessary to trawl through a huge amount of material to reveal the distortion and lack of substance in the claims, and meanwhile, amplified by confirmation bias and social media, the original article may have propagated a wildfire of misinformation that is hard to extinguish (Lewandowsky, Ecker, & Cook, 2017).

The next level after distortion of research findings is outright invention of fake data. It is generally assumed

that this is rare, though it is difficult to get accurate estimates of the frequency of this deception because of its very nature. A researcher who suspects misconduct by another scientist is placed in an uncomfortable position, and there is little formal guidance as to how to proceed. Simonsohn (2013, p. 1886), who used statistical methods to uncover the fraudulent work of two psychologists, summarized his recommended steps as follows:

- Replicate the analyses across multiple studies before suspecting foul play
- Compare suspect studies with similar ones by other authors
- Extend the analyses to raw data
- Contact the authors privately and transparently, and give them ample time to consider your concerns
- Offer to discuss matters with a trusted statistically savvy advisor
- Give the authors more time
- If suspicions remain, convey them only to entities tasked with investigating such matters, and do so as discreetly as possible

Investigating suspected misconduct is extremely important work, but it is not for the fainthearted. An accusation of fraud is serious business and requires rock-solid evidence, which can take hours of careful work to discover. Although one would hope that academic institutions would take seriously an accusation of misconduct against a staff member, they can be slow to act; it is, of course, important that they consider the possibility that they are dealing with an unjustified attack by people with vested interests or fixed ideas. Such attacks do occur, but malign intent should not be the default assumption, unless there are several "red flags" (Lewandowsky & Bishop, 2016). Although there are some notable cases of good practice by institutions (e.g., Høj, 2013), there are also many historical instances of their closing ranks to protect an eminent researcher (Judson, 2004). This is shortsighted, as the ultimate reputational damage from being revealed to be supporting a dishonest researcher is far worse than any bad publicity from early disclosure of a problem. But the scientist who is trying to put things right can find

it to be a lonely and dispiriting process, as Heathers (2017) documented on his blog. Furthermore, when we are dealing with genuine fraudsters, we can expect them to use every method possible to avoid discovery, because they have built a career on deceit. They are likely to be obstructive and may well attack back, accusing the people who are raising questions of ulterior motives. As do whistle-blowers in other areas of life, the people who detect fraud tend to get little thanks from the community whose interests they serve.

General Principles for Responding to Fallibility

Thankfully, accusations of deliberate misconduct in science are rare, but the spotlight has started to shine increasingly on fallibility in psychology, and some hitherto well-established findings are now looking less solid (e.g., O'Donnell et al., 2018). My general rule is that we should never use mockery or personal abuse against other scientists who make honest errors: Such behavior just reinforces people's unwillingness to be open about errors. Nor should we assume that failure to replicate a result is a sign of poor science in the original study; rather, it is an indication that more work needs to be done to establish whether, and under what conditions, the result is robust. But good researchers will not hesitate to note flaws in their own scientific work and the work of others. Criticism is the bedrock of the scientific method. It should not be personal: If one has to point to problems with someone's data, methods, or conclusions, this should be done without implying that the person is stupid or dishonest. This is important, because the alternative is that many people will avoid engaging in robust debate because of fears of interpersonal conflict—a recipe for scientific stasis. If wrong ideas or results are not challenged, we let down future generations who will try to build on a research base that is not a solid foundation. Worse still, when the research findings have practical applications in clinical or policy areas, we may allow wrongheaded interventions or policies to damage the well-being of individuals or society. As open science becomes increasingly the norm, we will find that everyone is fallible. The reputations of scientists will depend not on whether there are flaws in their research, but on how they respond when those flaws are noted.

Action Editor

Daniel J. Simons served as action editor for this article.

Author Contributions

D. V. M. Bishop is the sole author of this article and is responsible for its content.

Acknowledgments

This Commentary is based on a talk given on July 7, 2017, at a meeting on Reproducible Science for Early Career Researchers, at the University of Cardiff. I thank David Mehler for inviting me to present the talk at the meeting and for proposing this topic. I am also grateful to Kendal Smith for constructive comments on a preprint version of this article.

Declaration of Conflicting Interests

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

Prior Versions

An earlier version of this article was posted as a preprint at <https://peerj.com/preprints/3486/>.

References

- Allison, D. B., Brown, A. W., George, B. J., & Kaiser, K. A. (2016). Reproducibility: A tragedy of errors. *Nature*, *530*, 27–29. doi:10.1038/530027a
- Azoulay, P., Bonatti, A., & Krieger, J. L. (2017). The career effects of scandal: Evidence from scientific retractions. *Research Policy*, *46*, 1552–1569. doi:10.1016/j.respol.2017.07.003
- Bastian, H. (2014). Editorial: A stronger post-publication culture is needed for better science. *PLOS Medicine*, *11*(12), Article 1001772. doi:10.1371/journal.pmed.1001772
- de Groot, A. D. (2014). The meaning of “significance” for different types of research [translated and annotated by Eric-Jan Wagenmakers, Denny Borsboom, Josine Verhagen, Rogier Kievit, Marjan Bakker, Angelique Cramer, Dora Matzke, Don Mellenbergh, & Han L. J. van der Maas]. *Acta Psychologica*, *148*, 188–194. doi:10.1016/j.actpsy.2014.02.001
- Dolgin, E. (2018, February 2). PubMed Commons closes its doors to comments. *Nature.com*. doi:10.1038/d41586-01591-4
- Eklund, A., Nichols, T. E., & Knutsson, H. (2016a). Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. *Proceedings of the National Academy of Sciences, USA*, *113*, 7900–7905. doi:10.1073/pnas.1602413113
- Eklund, A., Nichols, T. E., & Knutsson, H. (2016b). Correction for Eklund et al., Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. *Proceedings of the National Academy of Sciences, USA*, *113*, E4929. doi:10.1073/pnas.1612033113
- Finkel, E. J. (2016). Reflections on the commitment-forgiveness Registered Replication Report. *Perspectives on Psychological Science*, *11*, 765–767. doi:10.1177/17456916166664695
- Gewin, V. (2015, July 24). Rice researchers redress retraction. *Nature.com*. doi:10.1038/nature.2015.18055
- Gilmore, R. O., Kennedy, J. L., & Adolph, K. E. (2018). Practical solutions for sharing data and materials from psychological research. *Advances in Methods and Practices in Psychological Science*, *1*, 121–130. doi:10.1177/2515245917746500

- Gough, D., Oliver, S., & Thomas, J. (2017). *An introduction to systematic reviews*. London, England: Sage.
- Greenberg, S. A. (2009). How citation distortions create unfounded authority: Analysis of a citation network. *British Medical Journal*, *339*, Article b3680. doi:10.1136/bmj.b2680
- Grimes, D. R., & Bishop, D. V. M. (2017). Distinguishing polemic from commentary in science: Some guidelines illustrated with the case of Sage and Burgio, 2017. *Child Development*. Advance online publication. doi:10.1111/cdev.13013
- Heathers, J. (2017, October 9). The buck stops nowhere: When research goes wrong, who's responsible? [Web log post]. Retrieved from <https://medium.com/@jamesheathers/the-buck-stops-nowhere-8284a57c88c9>
- Høj, P. (2013, September 3). *UQ investigates events leading to retraction: Statement from The University of Queensland President and Vice-Chancellor Professor Peter Høj* [Press release]. Retrieved from <https://www.uq.edu.au/news/article/2013/09/uq-investigates-events-leading-retraction>
- Hosseini, M., Hilhorst, M., de Beaufort, I., & Fanelli, D. (2018). Doing the right thing: A qualitative investigation of retractions due to unintentional error. *Science and Engineering Ethics*, *24*, 189–206. doi:10.1007/s11948-017-9894-2
- Judson, H. F. (2004). *The great betrayal: Fraud in science*. Orlando, FL: Harcourt.
- Lewandowsky, S., & Bishop, D. V. M. (2016). Don't let transparency damage science. *Nature*, *529*, 459–461.
- Lewandowsky, S., Ecker, U. K. H., & Cook, J. (2017). Beyond misinformation: Understanding and coping with the 'post truth' era. *Journal of Applied Research in Memory and Cognition*, *6*, 353–369.
- Mann, R. (2013, March 8) Rethinking retractions [Web log post]. Retrieved from <http://prawnsandprobability.blogspot.com/2013/03/rethinking-retractions.html>
- Mann, R. P., Perna, A., Strömbom, D., Garnett, R., Herbert-Read, J. E., Sumpter, D. J. T., & Ward, A. J. W. (2012). Retraction: Multi-scale inference of interaction rules in animal groups using Bayesian model selection. *PLOS Computational Biology*, *8*(8). doi:10.1371/annotation/7bc3a37e-db82-4813-8242-7d34877125c5
- Markowitz, F. (2015). Five selfish reasons to work reproducibly. *Genome Biology*, *16*, Article 274. doi:10.1186/s13059-015-0850-7
- McKiernan, E. C., Bourne, P. E., Brown, C. T., Buck, S., Kenall, A., Lin, J., . . . Yarkoni, T. (2016). How open science helps researchers succeed. *eLife*, *5*, Article e16800. doi:10.7554/eLife.16800
- Medical Research Council. (2017). *Using information about people in health research* (MRC ethics series). Retrieved from <https://www.mrc.ac.uk/documents/pdf/using-information-about-people-in-health-research-2017/>
- Merali, Z. (2010). Computational science: . . . Error . . . Why scientific programming does not compute. *Nature*, *467*, 775–777. doi:10.1038/467775a
- Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Percie du Sert, N., . . . Ioannidis, J. P. A. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, *1*(1), Article 0021. doi:10.1038/s41562-016-0021
- National Center for Biotechnology Information. (2018, February 1). PubMed Commons to be discontinued. *NCBI Insights*. Retrieved from <https://ncbiinsights.ncbi.nlm.nih.gov/2018/02/01/pubmed-commons-to-be-discontinued/>
- Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, *2*, 175–220.
- Nuzzo, R. (2014). Scientific method: Statistical errors. *Nature*, *506*, 150–152. doi:10.1038/506150a
- O'Donnell, M. O., Nelson, L. D., Ackermann, E., Aczel, B., Akhtar, A., Aldrovandi, S., . . . Zrubka, M. (2018). Registered Replication Report: Dijksterhuis and van Knippenberg (1998). *Perspectives on Psychological Science*, *13*, 268–294. doi:10.1177/1745691618755704
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*, aac4716. doi:10.1126/science.aac4716
- Oransky, I., & Marcus, A. (2017, May 5). Introducing the Doing the Right Thing award, honoring those who clean up the scientific literature. *Stat*. Retrieved from <https://www.statnews.com/2017/05/05/dirt-award-cleaning-scientific-literature/>
- Poldrack, R. (2013, February 20). Anatomy of a coding error [Web log post]. Retrieved from <http://www.russpoldrack.org/2013/02/anatomy-of-coding-error.html>
- Retraction Watch. (2017, March 27). *Authors who retract for honest error say they aren't penalized as a result*. Retrieved from <https://retractionwatch.com/2017/03/27/authors-retract-honest-error-say-arent-penalized-result/>
- Rogers, T., Zeckhauser, R., Gino, F., Norton, M. I., & Schweitzer, M. E. (2017). Artful paltering: The risks and rewards of using truthful statements to mislead others. *Journal of Personality and Social Psychology*, *112*, 456–473. doi:10.1037/pspi0000081
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology. *Psychological Science*, *22*, 1359–1366. doi:10.1177/0956797611417632
- Simonsohn, U. (2013). Just post it: The lesson from two cases of fabricated data detected by statistics alone. *Psychological Science*, *24*, 1875–1888. doi:10.1177/0956797613480366
- Stapel, D. (2014). *Faking science: A true story of academic fraud* (N. J. L. Brown, Trans.). Retrieved from <https://errorstatistics.files.wordpress.com/2014/12/faking-science-20141214.pdf>
- UK Data Service. (n.d.). *Anonymisation*. Retrieved from <https://www.ukdataservice.ac.uk/manage-data/legal-ethical/anonymisation>
- Weil, A. (2014, May 10). Re: Co-rex-ions [Web log comment]. Retrieved from <https://whatsinjohnsfreezer.com/2014/05/10/co-rex-ions/>
- Wicherts, J. (2017). The weak spots in contemporary science (and how to fix them). *Animals*, *17*(12), Article 90. doi:10.3390/ani7120090