Commentary:
Fallibility in science: Responding to errors in the work of oneself and others

D. V. M. Bishop

Department of Experimental Psychology
University of Oxford
OX1 3UD, UK
dorothy.bishop@psy.ox.ac.uk

Abstract

Fallibility in science cuts both ways: it poses dilemmas for the scientist who discovers errors in their own work, and for those who discover errors in the work of others. The ethical response to finding errors in one's own work is clear: they should be claimed and corrected as rapidly as possible. Yet people are often reluctant to 'do the right thing' because of a perception this could lead to reputational damage. I argue that the best defence against such outcomes is adoption of open science practices, which help avoid errors and also leads to recognition that mistakes are part of normal science. Indeed, a reputation for scientific integrity can be enhanced by admitting to errors. The second part of the paper focuses on situations where errors are discovered in the work of others; in the case of honest errors, action must be taken to put things right, but this should be done in a collegial way that offers the researcher the opportunity to deal with the problem themselves. Difficulties arise if those who commit errors are unresponsive or reluctant to make changes, when there is disagreement about whether a dataset or analysis is problematic, or where deliberate manipulation of findings or outright fraud is suspected. I offer some guidelines about how to approach such cases. My key message is that for science to progress, we have to accept the inevitability of error. In the long run, scientists will not be judged on whether or not they make mistakes, but on how they respond when those mistakes are detected.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3486v1 | CC BY 4.0 Open Access | rec: 24 Dec 2017, publ: 24 Dec 2017

Errors in your work: how to respond

Imagine the following scenario:

PhD student, David, has run a series of studies trying to find an impact 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. The paper 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 the groups were miscoded, and in fact the sham treatment group obtained higher post-training scores.

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 paper to be retracted. But there is typically nervousness in the room. It is pointed out that 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. A recent study considered how reputation is affected by retraction, by comparing subsequent citations of earlier published papers for authors who had a paper retracted vs. a control group who had not (Azoulay, Bonatti, & Krieger, 2017). Retraction of a paper due to researcher misconduct led to a drop in subsequent citations of their earlier work, but there was a smaller effect when honest error was involved – with no evidence of reputational damage for junior researchers. Indeed, more informal evidence suggests that there can be reputational advantage from going public in correcting an error: you demonstrate you are someone who values scientific accuracy over your success in publishing (Retraction Watch, 2017). I give some examples from online sources in Box 1. The thought of having to retract a paper can instil fear into the heart of scientists, who see it as equivalent to being named and shamed. Recognising that this 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).

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3486v1 | CC BY 4.0 Open Access | rec: 24 Dec 2017, publ: 24 Dec 2017

Box 1

Examples of researchers who highlighted errors in their own work

Richard Mann, a postdoctoral researcher using statistical methods to study behavioural ecology, had published a paper on behaviour in prawns in PLOS Computational Biology with six co-authors. He shared the prawn dataset 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 rang him to tell him of a fatal error in his analysis. As stated in the retraction notice: "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 paper was retracted, the analysis redone giving similar findings, and Mann states that, although he had a terrible few months, he did not suffer any long-term stigma.

Pamela Ronald, a professor in plant pathology, became concerned when two of her postdocs could not replicate findings she had published in two high-profile papers on the basis of the immune response in rice. She notified the journal editors and then devoted the next 18 months to try and locate the source of the discrepancy. It turned out that the strains of microbes she had been using were mislabelled, and in 2013 the papers were retracted. The story was covered by Nature News (Gewin, 2015), who noted that this year Ronald published a paper 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 based on the task being performed. He had submitted a paper based on this analysis for publication, when a student collaborator told him that after obtaining far lower classification accuracy on the same dataset, he found an error in the code. Poldrack's (2013) response was to write a blogpost about this experience, encouraging everyone to share code, use better methods for checking code, and talk about their errors.

There are two further points to take from the David scenario. Awful and embarrassing as it is to admit to error, the alternative, hiding a known error, has to be worse. The person who does this is entering into a Faustian pact to reject science in favour 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 where the raw data cannot be made completely open, usually because of confidentiality issues, it is often possible to deposit a version that has been modified to remove identifiable information, so others can reproduce what was done (UK Data Service, undated). For sensitive data, a data-sharing agreement may be needed in addition to anonymization (Medical Research Council, 2017). Regardless of which level of security is required, there should be no barriers to researchers making

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3486v1 | CC BY 4.0 Open Access | rec: 24 Dec 2017, publ: 24 Dec 2017

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 code and data open, then errors will be found. Yet that is really the whole point: we need to make code and data open so that the errors will be found. But to encourage people to do that, we must remove any stigma associated with detection of those errors.

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's work. You would not want to be pilloried for an honest error, so don't pillory others for simple mistakes. In a comment on a blogpost on this topic, Anne 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. Fortunately, there is a mechanism for putting the record right, by adding a comment in PubMed Commons (Bastian, 2014). This option is open to anyone who has themselves published in a journal indexed by PubMed. The comment is linked to the abstract of the original paper on PubMed and becomes part of the scientific record. Box 2, examples 1 and 2 illustrate how both authors and other researchers can use PubMed Commons to record a correction.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3486v1 | CC BY 4.0 Open Access | rec: 24 Dec 2017, publ: 24 Dec 2017

Box 2

Examples of post-publication commentary on PubMed Commons

1. Author adding minor corrections
https://www.ncbi.nlm.nih.gov/pubmed/28436345
Jim van Os notes some numerical errors in a table.

2. Reviewer correcting an error
https://www.ncbi.nlm.nih.gov/pubmed/28461468
Pavel Nesmiyanov noted that β-endorphin, oxytocin, and dopamine were wrongly described as neuropeptides. Although authors did not respond on PubMed Commons, an erratum was published in the journal.

3. Reviewer critiquing methods
https://www.ncbi.nlm.nih.gov/pubmed/29153326
Franck Ramus criticises small sample size of paper on neurobiological correlates of dyslexia. Authors respond defending the small sample size and arguing their analyses were driven by a priori hypothesis derived from previous study.

4. Reviewer critiquing methods
https://www.ncbi.nlm.nih.gov/pubmed/28706072
Serge Ahmed suggests that a study of planning in ravens needs an additional control for learning of affective value of objects.

5. Reviewer noting over-hyped interpretation of results
https://www.ncbi.nlm.nih.gov/pubmed/28735725
Clive Bates notes that a study on association between vaping and smoking in adolescents has been widely interpreted in the media as showing causal link. Bates adds a link to a more detailed critique of the study.

6. Reviewer raising more serious concerns
https://www.ncbi.nlm.nih.gov/pubmed/17688420
David Nunan notes prior concerns about duplicate data in a paper on diet in congestive heart failure.

Another scenario is when research results 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 where PubMed Commons can provide a venue for raising the concerns, as illustrated in Box 2, examples 3-5. PubMed Commons has not been widely used for post-publication review in psychology, but in best cases it can be used to initiate useful discussion about a paper or to make suggestions about methodology, and in other cases can simply raise concerns or put the record straight. But again, this should be done as far as possible in a constructive fashion, avoiding any personal attack, alerting the

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3486v1 | CC BY 4.0 Open Access | rec: 24 Dec 2017, publ: 24 Dec 2017

author to the comment and inviting them to reply. The default assumption should be that methodological weaknesses are due to ignorance rather than bad faith. In particular, although the dangers of p-hacking were pointed out many years ago (de Groot, 2014), the practice 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 still do not recognise how seriously it can distort results (Nuzzo, 2014).

I turn now to those unfortunate situations when it is hard to avoid concluding that a researcher is acting in bad faith. A particularly insidious kind of behaviour involves selective citation of the literature, or 'cherry-picking'. Unless an author has specified clear criteria for which studies are included in a review, it can be hard to detect distortion of evidence, unless one is an expert in the area. Worse still are cases where the cited literature is selectively or inaccurately portrayed, giving the impression of a large body of work supporting a given position. This is a standard ploy by those promoting pseudoscientific views (Grimes & Bishop, 2017) and needs to be robustly challenged.

The next step after distortion of research findings is outright invention of fake data. Table 1 shows the advice of Uri Simonsohn (2013), who uncovered the fraudulent work of two psychologists.

Table 1

Simonsohn's (2013) recommendations when fraud is suspected

Replicate analyses across multiple studies before suspecting foul play by a given author.

Compare suspected studies with similar ones by other authors.

Extend analyses to raw data.

Contact 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.

p. 1886

Uncovering fraud is extremely important work, but it is not for the faint-hearted. For a start, 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 fraud 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 those with vested interests or fixed ideas. These do occur, but malign

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3486v1 | CC BY 4.0 Open Access | rec: 24 Dec 2017, publ: 24 Dec 2017

intent should not be the default assumption, unless there are several 'red flags' of the kind noted by Lewandowsky and Bishop (2016). Although there are some notable cases of good practice (e.g. Høj, 2013), there are also many historical instances where institutions closed ranks to protect an eminent researcher (Judson, 2004). This is short-sighted, as the ultimate reputational damage of being revealed to be supporting a fraudster 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 James Heathers (2017) documented on his blog. Furthermore, one can expect the fraudster 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 those who are raising questions of ulterior motives. Just like whistle-blowers in other areas of life, those who detect fraud tend to get little thanks from the community whose interests they serve.

Challenges associated with lack of reproducibility

Reproducibility has become a hot topic in psychology in recent years (Munafò et al., 2017), with failure to reproduce findings in psychology being brought to the fore by an influential study published in Science (Open Science Collaboration, 2015). Failures to reproduce a specific result can arise for different reasons, as shown in Table 2. It should not be assumed that a failure to reproduce a result is evidence of poor science in the original study. Rather, both sets of researchers should work together to consider possible explanations. If the original researcher believes that contextual factors or researcher expertise are critical to obtaining the result, then it is up to them to specify more carefully the conditions under which the effect obtains, rather than simply putting forward hypothetical explanations for a null result. When there is a failure to reproduce a finding, it is bad if the first response is to disparage the original researchers as incompetent or malign, but it is just as bad if those whose finding were not reproducible assume bad motives or lack of expertise in the replicators. Again, the kudos will go to the researchers who show integrity in putting scientific truth before their own career ambitions.

Table 2
Possible reasons for failure to reproduce a scientific result

Initial result was a false positive due to chance variation
Results are sensitive to contextual factors
Lack of expertise of replicator
Initial results obtained using questionable research practices such as p-hacking
Researcher committed fraud

More generally, we should never use mockery or personal abuse against other scientists who make honest errors: such behaviour just reinforces people's unwillingness to be open about errors. But a good scientist will not hesitate to note flaws in the scientific work of other researchers. Criticism is the bedrock of the scientific method. It should not be personal: it is perfectly feasible to point to problems with someone's data, methods or conclusions without challenging their integrity. A failure to engage in robust debate because

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3486v1 | CC BY 4.0 Open Access | rec: 24 Dec 2017, publ: 24 Dec 2017

of fears of interpersonal conflict leads to scientific stasis. If wrong ideas or results are not challenged, we let down future generations who try to build on a research base that is not a solid foundation. Worse still, where the research findings have practical applications in clinical or policy areas, we may allow wrongheaded interventions or policies to damage the wellbeing of individuals or society. As open science becomes increasingly the norm, we will find that everyone is fallible. The reputation of a scientist will depend not on whether there are flaws in their research, but on how they respond when those flaws are noted.

References

Azoulay, P., Bonatti, A., & Krieger, J. L. (2017). The career effects of scandal: Evidence from scientific retractions. *Research Policy*. 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), e1001772. 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, and Han L. J. van der Maas]. *Acta Psychologica, 148*, 188-194. doi:http://dx.doi.org/10.1016/j.actpsy.2014.02.001

Gewin, V. (2015). Nature News: Rice researchers redress retraction (24 July 2015). *Nature*. doi:10.1038/nature.2015.18055

Grimes, D. R., & Bishop, D. (2017). Distinguishing polemic from commentary in science: Some guidelines illustrated with the case of Sage and Burgio, 2017. *Child Development, in press*. doi:10.1111/cdev.13013

Heathers, J. (2017). The buck stops nowhere: When research goes wrong, who's responsible? Retrieved from https://medium.com/@jamesheathers/the-buck-stops-nowhere-8284a57c88c9

Høj, P. (2013). UQ investigates events leading to retraction: Statement from The University of Queensland President and Vice-Chancellor Professor Peter Høj Retrieved from https://www.uq.edu.au/news/article/2013/09/uq-investigates-events-leading-retraction

Judson, H. F. (2004). *The great betrayal: Fraud in science.* Orlando, Florida, USA: Harcourt Inc.

Lewandowsky, S., & Bishop, D. V. M. (2016). Don't let transparency damage science. *Nature, 529*(7587), 459-461

Medical Research Council. (2017, 6th September). 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/

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 Behavior, 1*(1: 0021). doi:10.1038/s41562-016-0021

Nuzzo, R. (2014). Scientific method: statistical errors. *Nature, 506*, 150-152. doi:10.1038/506150a

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Nature, 349*(6251). doi:10.1126/science.aac4716

Oransky, I., & Marcus, A. (2017, 5th May). Introducing the Doing the Right Thing award, honoring those who clean up the scientific literature. Retrieved from https://www.statnews.com/2017/05/05/dirt-award-cleaning-scientific-literature/

Poldrack, R. (2013, 20th February 2013). Anatomy of a coding error Retrieved from http://www.russpoldrack.org/2013/02/anatomy-of-coding-error.html

Retraction Watch. (2017, 27th March). Authors who retract for honest error say they aren't penalized as a result. Retrieved from http://retractionwatch.com/2017/03/27/authors-retract-honest-error-say-arent-penalized-result/ - more-48973

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology. *Psychological science, 22*(11), 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*(10), 1875-1888. doi:10.1177/0956797613480366

Stapel, D. (2014). *Faking Science: A True Story of Academic Fraud. Translated by Nicholas J. L. Brown*. Retrieved from https://errorstatistics.files.wordpress.com/2014/12/fakingscience-20141214.pdf

UK Data Service. (undated). Anonymisation. Retrieved from https://www.ukdataservice.ac.uk/manage-data/legal-ethical/anonymisation

Weil, A. (2014). Comment on blog by John Hutchinson: Co-rex-ions. Retrieved from https://whatsinjohnsfreezer.com/2014/05/10/co-rex-ions/ - comment-22328

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3486v1 | CC BY 4.0 Open Access | rec: 24 Dec 2017, publ: 24 Dec 2017