

# When things go wrong: correcting the scientific record

Bernd Pulverer

**A reliable scientific literature is crucial for an efficient research process. Peer review remains a highly successful quality assurance mechanism, but it does not always prevent data and image aberrations and the publication of flawed data. Journals need to be in a position to detect such problems and take proportionate action. Publishers should apply consistent policies to correcting the published literature and adopt versioning. The scientific community ought to encourage corrections.**

## The rise of retraction

Peer review at scientific journals has been much maligned. There is no doubt that editorial processes can be improved and we have discussed a number of constructive enhancements that this journal has adopted, including the EMBO transparent publishing process (published anonymous referee reports and author/editor communication) (Pulverer, 2010). Referees and editors may be subject to some degree of conscious or unconscious bias, but the referee reports we receive tend to be confined to reasonable, factual queries or suggested corrections. Bias or overzealous recommendations can be tempered effectively by referee cross-commenting, listening to author responses and ensuring that editors are informed and neutral in their treatment of reports and author responses (Pulverer, 2010). The majority of our referees invest a large amount of their time—for little tangible personal benefit—and they almost invariably significantly improve the studies we publish. As long as everyone keeps a close eye on ensuring that requirements for revisions are clear, constructive and realistic, the system works well.

Yet, peer review alone is not always sufficient to ensure papers report findings in a

transparent and reproducible manner. This is why progressive editorial policies are essential. For example, we have encouraged posting “source data” alongside figures for several years, and in *The EMBO Journal*, more than 60% of papers now contain minimally processed data or replicates (Pulverer, 2014a). We also recently developed an author checklist—in consultation with other leading journals—that aims to improve the reporting of experimental procedures and statistical information (Pulverer, 2014a).

Nevertheless, not a month goes by without another prominent retraction case; examples abound across fields, countries, and journals. Much has been made of the dramatic—in the view of some commentators—increases in retractions in recent years, especially at the more prominent journals. However, given the high level of attention prominent papers receive and the increased scrutiny from dedicated blogs such as *Retraction Watch* or the anonymous discussion forum *PubPeer*, one should neither be surprised nor be alarmed: The more we look, the more we find. In fact, the retractions, which often relate to papers that have long been published, ought to be seen as evidence that the much-touted self-corrective nature of the scientific literature is at work more effectively these days. The ethicist Nicolas Steneck commented: “I don’t think there’s any doubt that we’re detecting more fraud, and that systems are more responsive to misconduct. It’s become more acceptable for journals to step in [...] it is [...] probable that the growth in retractions has come from an increased awareness of research misconduct” (Van Noorden, 2011).

Despite the increase, retraction rates remain below 0.02% (Van Noorden, 2011). That number appears to be in stark contrast to the significant number of pre-publication

issues uncovered by the few journals in the biosciences that offer systematic image integrity screens to authors (Cyranoski, 2014; Pulverer, 2014b; Van Noorden, 2015; Yamada & Hall, 2015). Journals including *J Cell Biol* and *The EMBO Journal* invariably find that around one-fifth of manuscripts otherwise due for acceptance post peer review require further revision. While almost all of the issues are properly corrected and there is little doubt that they derive from genuine mistakes or oversights, this implies that retraction and correction rates would rise significantly in case of more systematic screening post publication. We are convinced that prepublication screens are worthwhile, as mistakes are more effectively corrected in emergent manuscripts, where access to data, reagents, and knowledge remain within reach. Once published, journals have to ensure that serious mistakes in papers are corrected in a manner that renders the changes tractable by the reader.

To ensure that EMBO Press journals through their pre-publication screening do not in any way abet the tiny minority of researchers who engage in intentional deception, we have instituted a three-tiered classification for image aberrations (Table 1; more details at: <http://embopress.org/imageaberrationlevels>).

In our view, it would neither be reasonable nor be currently feasible to report every single issue at level 1 or 2 formally to the authors’ institutions. Of course, we are always open to discuss cases with institutional investigations, and, in fact, we look to such investigations to inform our actions.

## Correction as a virtue

Issues embedded deeper in datasets and deliberate data and image manipulation remain invisible to routine image integrity analysis, and issues often arise on papers that were published before such screens

**Table 1. A three-tiered classification for image aberrations.**

Level	Attributes	Action	%
I	Cosmetic aberrations or mistakes with supporting source data and satisfactory author explanation	Allow revision, no report to institution	12
II	Data “beautification” and undeclared manipulation that changes conclusions; available source data or new data	May allow revision, may report to institution	8
III	Image manipulation with digital obfuscation (splicing, cloning, insertion, and selective deletion); no explanation; no source data	Reject and report	< 0.5
Total <sup>a</sup>			20.4

<sup>a</sup>Percent of manuscripts screened post review that require follow-up (2014).

were instituted. Once a dataset has been “set in stone” in a published paper—once it has entered the canon of the scientific literature—corrective measures have to be formalized and the changes have to become tractable to avoid confusing or misleading readers.

Over one million peer-reviewed papers are published each year in the biosciences, and as we are beginning to drill deeper into the molecular mechanistic understanding of biology, papers become increasingly complex and specialized. Unless we make it a virtue to correct the literature diligently, we risk getting lost in this information flood. Understandably, airing the “dirty laundry” through a highly visible corrigendum, or indeed a retraction, can be embarrassing. This is why it is important that the whole community embraces the concept of correction as a positive: scientists who are willing to correct their published work in a transparent manner should be encouraged, and any employer should see this as a positive attribute. Consider that in IT incremental enhancements and bug fixes are released as a matter of course. Anyone who reports potential problems in the literature constructively—that is with due diligence and in a manner sensitive to the scientists affected—should not have to fear negative consequences. Anonymous reporting such as on *PubPeer* may, in reality, be unavoidable, although if unchecked, the lack of accountability may encourage abuse and vigilantism. Journals should see it as an obligation to correct serious mistakes and the community should appreciate journals that act upon retractions and corrections, not stigmatize them.

Mistakes happen: A scientific paper is a highly intricate beast; complex findings and protocols are often published under considerable time pressure through multiple rounds of revision and in coordination with many colleagues in disparate locations. Also, there are few research findings in the

biosciences that present immutable, absolute facts—a thing that seemed clear one day may soon be open to question. As a community, we need to reduce the stigma of retraction. As Richard Van Noorden noted: “Scientists would like to separate two aspects of retraction that seem to have become tangled together: cleaning up the literature, and signaling misconduct” (Van Noorden, 2011). A transparent explanation as to what happened seems the best way to reduce the stigma for genuine mistakes, even if they are sometimes hard to swallow. In this respect, publishing the referee reports alongside all papers helps to add transparency and accountability about the editorial process leading up to a correction or retraction (Pulverer, 2010).

### Correct or Retract? In search of a more diverse toolkit

Traditionally, the published record is seen as immutable—at least once a paper is published in a journal issue and posted on the prevalent bibliographic databases such as PubMed. Corrections should explicitly state what was incorrect and ideally why the information was changed in a non-judgmental manner. Indeed, the publishers of *Retraction Watch* have campaigned for years for more transparency in correction and retraction notes. Statements from both the authors and the editors of a journal can be beneficial. If there is no consensus between authors, all the affected authors should have a voice. It may be helpful to state whether and how far a corrected or retracted finding has been confirmed by the subsequent literature. Indeed, in our view, it can be useful to post new data that clarifies cause and consequence of a correction/retraction, even if the data were generated in response to the uncovering of a problem in a paper. At this journal, we are reluctant to embed such data

in correction/retraction notices, as full peer review is often not feasible and some distance to such data can be advisable. We suggest that a constructive mechanism is to post non-peer-reviewed data on sharing platforms such as bioRxiv with a link—after all, interested readers can judge for themselves (see Ross, 2015). Alternatively, even if a paper was subject to retraction, publication of the correct data/interpretation in a peer-reviewed journal may be appropriate (Gewin, 2014).

Given the wide spectrum of underlying causes for corrective measures, and of how profoundly a correction may affect the message of a given research paper, we seem to be left with far too blunt a set of corrective tools, namely either a corrigendum/erratum or a retraction. We pursue a correction (labeled “corrigendum” if the author was in error and “erratum” if we were) in cases where there is a clear and reasonable explanation from the authors—ideally supported by the findings of an institutional investigation—where the authors can provide unmodified data (source data) and the central conclusions of a research paper are not fundamentally affected (analogous to “level I–II,” Table 1). We decide to retract a research paper when the trust in the affected data is undermined—including the inability of authors to provide the original source data and a clear explanation as to why manipulations happened (akin to “level III” in Table 1)—and where that data are central to the conclusions of the paper. The level of cooperation of the authors, the apparent motives, the significance of the problem, and the age of the paper may affect this decision (is there evidence for a systematic problem? can we expect data retention and the availability of lab protocols?). Importantly, we issue retractions irrespective of evidence that a conclusion stands based on the subsequent literature, although we certainly allow reference to supporting findings in the subsequent literature in retraction notes. As the editors of *Nature* noted in a recent editorial: “In the end, it comes down to an issue that is at the very heart of the practice and communication of science: the question of trust. After all, if researchers and editors cannot safely assume [...] that scientific results are essentially true as reported, then the advancement of science is in serious trouble” (*Nature*, 2006).

What to do when the problem is clearly contained to an “excisable” subset of a dataset? If we have a strong sense that the

integrity of the rest of the paper is not in question, and if the problem is serious but “self contained”—that is, it does not undermine key conclusions of the paper—our view is that a surgical excision of the affected information can be appropriate: that is, a selective retraction at the figure level. Such a retraction ought to be labeled as clearly as a full retraction, with a highly visible watermark stating “retracted” across the relevant data and a clear correction of any associated claims. We note that this mechanism has been discussed by the *Committee on Publication Ethics* (COPE) and that some commentators question whether it is appropriate (Gewin, 2014). In our view, it allows for a proportionate response in serious cases where key findings nevertheless remain reliable. After all, a retraction removes every bit of data in a paper from the literature. This blunt corrective surgery can muddy the literature considerably, and, importantly, it undermines the efforts and academic credit of all the authors on a paper, even in cases where problems clearly derive from the actions of a single individual; the consequences for this researcher are for the research institution to decide. Authorship at the level of the figure panel (sometimes called “microattribution”) is a useful mechanism to add not only credit, but also accountability for individual experiments (Pulverer, 2010).

## Versioning

It is high time that research journals embraced what is already standard practice in correcting news reporting: the use of online only, in text corrections, but with clear annotation as to what was changed and why (this can take the form of a footnote or an inline comment box). Similar mechanisms are routine in the software industry; versioning is already available on preprint servers (see <https://arxiv.org/help/versions>) and indeed some journals (Lawrence, 2012). Versioning can extend to the addition of crucial information by the authors (currently labelled “addendum”). Versioning has been anathema to many publishers in the past because it leads to a discrepancy between the printed paper and the online paper, and because publishers aim to ensure that all online records of the paper are updated, including PubMed. Notably, this is rarely achieved and many retracted articles remain posted on repositories and PubMedCentral (Davis, 2012). Thus, we are currently left with separately published, delayed corrigenda and

retraction notices that are at best loosely connected to the original article. These issues are unavoidable in print, but in an online world, we can and should do better.

The case for a more fluid, agile way to publish can also be extended to include annotations and comments mapped onto the text of published papers. At the moment, we are stuck with either comments tacked on to the end of research papers (often uncured for veracity) or formal refutations published separately. Community commenting can be important to be sure, but in a world of rapidly advancing, often highly competitive research, the cautious editor pauses to question how we can apply the quality assurance mechanisms that we believe make at least the high-quality segment of the scientific literature such a powerful mode of communication.

## Who calls the shots?

Journals are the last checkpoint and serve as gatekeepers of the scientific literature, but they are not—and cannot be—responsible for carrying out formal investigations, let alone be expected to adjudicate and judge ethical violations.

The general recommendation by institutions such as the *US Office of Research Integrity* (ORI) and COPE is for journals to await institutional investigations and to adhere to the advice given, but there is no standardization of the quality and impartiality of such investigations, the level of detail of the information released (if any), and the turnaround times (full investigations can take years). There is also no international (or even national) consensus, nor any guidelines on how such investigations are to be carried out and by whom, what they should report to journals or the public, and what the consequences for the affected research and researchers should be. I am not an advocate of overregulation, but in my view there is a strong case for setting up an independent body to audit, advise on, or indeed carry out such investigations. The ORI in the United States has such a role in principle, but its remit is confined to federally funded research and it is not nearly of the scale that would allow it to take on investigations systematically.

Journals should certainly strive for consistency with thorough institutional investigations when making a decision on an appropriate correction to the literature. However, in the marked absence of any standardization or quality guarantees, the

conclusions of institutional investigations cannot be binding for journals. Notably, the priorities of institutional investigations may include considerations beyond correcting the scientific record. Pragmatically, we find again and again that the journal is where the buck stops and that we have to make decisions based on the often partial evidence available to us. Sometimes, investigations seem to wait to see how journals react. Given the fundamental importance of quality assurance for the scientific literature, there is certainly room for improvement toward a consistent, independent, and professional process. Coordination between research institutions and journals is a start. In recent cases involving multiple journals, we have exchanged information between journals during the evaluation process; we believe this resulted in a more consistent response.

We have embraced many of these suggested refinements to the correction process of the scientific literature in the most recent corrigenda and retractions in *The EMBO Journal* (Ross, 2015). We would invite the interested reader to comment and discuss.

## References

- Cyranoski D (2014) Research integrity: cell-induced stress. *Nature* 511: 140–143
- Davis PM (2012) The persistence of error: a study of retracted articles on the Internet and in personal libraries. *J Med Libr Assoc* 100(3): 184–189
- Editorial *Nature* (2006) Correction or retraction? *Nature* 444: 123–124
- Gewin V (2014) Retractions: a clean slate. *Nature* 507: 389–391
- Lawrence R (2012) Citations and indexing: a novel approach from F1000 Research. <http://blog.f1000research.com/2012/07/16/citations-and-indexing-a-novel-approach/>
- Pulverer B (2010) A transparent black box. *EMBO J* 29: 3891–3892
- Pulverer B (2014a) Transparent, reproducible data. *EMBO J* 33: 2597
- Pulverer B (2014b) STAP dance. *EMBO J* 33: 1285–1286
- Ross K (2015) Retraction Watch. <http://retractionwatch.com/2015/08/19/embo-investigation-yields-two-more-retractions-and-three-corrections-for-voinnet/#comments>
- Van Noorden R (2011) Science publishing: the trouble with retractions. *Nature* 478: 26–28
- Van Noorden R (2015) The image detective who roots out manuscript flaws. *Nature* doi: 10.1038/nature.2015.17749
- Yamada K, Hall A (2015) Reproducibility and cell biology. *J Cell Biol* 209: 191–193