

III

Interventions: Notes from the Field

The detection of new forms of manipulation exceeds the reach, resources, and perhaps even the mandate of traditional governmental agencies like the US Office of Research Integrity. These constraints, however, have been balanced by the emergence of a new generation of collaborative grassroots watchdogs, carnivalesque pranksters, independent organizations, and bloggers. The watchdogs often operate websites featuring public discussion or anonymous tips about publication misconduct that is often propelled by metrics. While watchdogs are often humorous, pranksters use humor and satire as their primary mode of critique to reveal spammish journals and easily gamed metrics. This section offers “notes from the field” by and about some of these new actors—notes that give us a glimpse of how they have emerged, how they operate, and how they have assumed increasingly important roles and credible voices despite a lack of governmental or university affiliations. They also tell us about the criticism, challenges, and legal threats they face.

Ivan Oransky co-founded the Retraction Watch to tell the stories behind article retractions, their typical opacity, and their increasing frequency. Virtually unknown a few decades ago, retractions have become a daily occurrence, their quantity allowing for qualitative and quantitative studies of their patterns. The Retraction Watch website, coupled with its daily email updates, looks at the retractions themselves and what their nature, distribution, and causes can tell us about general, even global, trends in misconduct and metrics manipulations. Brandon Stell’s and Boris Barbour’s chapter shows that PubPeer—initially fashioned as a web-based journal club—shares many of the motivations that animate Retraction Watch, but focuses on providing a publicly accessible platform for the detailed discussion of scientific articles chosen by the participants. Some questions and doubts are quickly answered by the authors of the articles (who are welcomed to jump into the public discussion), but in

other cases, the public questioning unearths more serious problems that may eventually lead to retractions. According to PubPeer, many of these problems are directly relatable to the craze for publishing in high-impact factor journals.

PubPeer has been very successful largely because it ignores metrics and engages in good old reading and critiquing. And while it does not take a position on the claims and allegations put forward by its largely anonymous commentators, these claims and the evidence they are based on is open for everybody to see, even when they may not be too eager to see it. For instance, Catherine Guaspere and Emmanuel Didier document a high-profile case—the “Voinnet affair”—that unfolded largely due to discussions on PubPeer and Retraction Watch. They argue that that case (in which a renowned French scientist lost an academic post and was banned from funding after the discovery of his frequent image manipulations) exemplifies a new “ecology” of misconduct detection in which the new watchdogs, despite having virtually no institutional authority, are able to create such a widespread publicity that forces the authorities to intervene or else appear to condone the misconduct.

Unlike the probing but formally polite discussions that animate PubPeer, Paul Brookes’s blog “Science Fraud” chose to deploy snarky language, and to do so from behind the mask of anonymity. This was a strategic choice, hoping that the somewhat “carnavalesque” tone of his commentary would draw additional attention, and thus put pressure on the authorities to investigate its claims. Anonymous and pseudonymous readers emailed him their discoveries of potential misconduct, which he blogged about humorously and sarcastically, while crediting his (masked) tipsters. After six months, however, Brookes’s identity was exposed. As a result, he was threatened with lawsuits by those accused of misconduct (a fate that has since been shared by PubPeer)¹ while his university—fearing liabilities—demanded that he choose between the blog and his professorship. The forced termination of “Science Fraud” and the recent legal challenges to PubPeer show how independent watchdogs, thanks to their “crowdsourcing” approach, can be quicker and more effective than traditional “misconduct police” agencies like the Office of Research Integrity. At the same time, their lack of institutional resources and authority, or of a governmental mandate, forces them to develop creative uses of anonymity (for themselves or their sources), which can backfire not only because anonymity can be breached, but also because its use may lead to accusations of opacity and unaccountability. Analogies between the predicament of PubPeer and WikiLeaks are not too far-fetched.

Through a humorous experiment, Cyril Labbé confronts the problem of quantifying academic excellence through ranking scholars according to their citation counts. He created an imaginary persona—“Ike Antkare” (I Can’t Care)—as the author of dozens of computer-generated papers, which he then proceeded to publish to demonstrate how easily Google Scholar’s ranking could be gamed. Thanks to citations from these and other computer-generated articles by other fictional scholars, Ike Antkare rose to the twenty-first position of most cited scientists on Google’s Scholarometer, beating Albert Einstein in the process.

Burkhard Morgenstern describes possibly the most hilarious prank involving the creation of fake scientists (with fake bios) to join editorial boards of spammish journals. He first became “Peter Uhnemann” (a character created by a German satirical magazine to troll a conservative politician), whose bio listed his research interest as “oximological microbiology, nonlinear submorphological endosaccharomorphosis, and applied endoplutomatics.” After Dr. Uhnemann was welcomed on a variety of editorial boards, Morgenstern decided to raise the bar by mobilizing “Hoss Cartwright” from the television show *Bonanza*. Dr. Cartwright was immediately welcomed on the editorial board of the *Journal of Primatology*, no doubt because of an impressive vitae featuring his affiliation with the Ponderosa Institute of Bovine Research, which he joined after extensive training in “Dunnowhat,” a postdoctoral fellowship at “Cowboy University” followed by a second postdoc at “Some shitty place in the middle of nowhere.” His last listed position was “senior cattle manager.”

In a different, more serious register, independent organizations—like CrossRef and Sideview—also work to critique gaming metrics in scholarly publishing. Not keen to make metrics history, Jennifer Lin describes the successes, future potential, and challenges of “altmetrics”—a large family of indicators that map references to publications and statements well beyond journal downloads, citations, and so on. It also does so across different media and platforms (including social media) rather than just journals or books. Perhaps because it is not one thing but many, and because it is rarely used by universities to evaluate faculty performance, altmetrics has not shared the criticism leveled against traditional metrics, nor has it provided widespread incentives for its gaming.² Furthermore, because of its complexity, the gaming of altmetrics “requires manipulating measurements across a diverse set of independent platforms [which] involves extensive coordination of multiple methods specific to each metric.” So far so good, but what will happen when altmetrics develops

a more clearly defined market and a specific set of users and uses, like today's impact factor?

Elizabeth Wager too targets the downsides of metrics, but sees journals as part of the problem, not the solution. She proposes—simply and radically—to terminate the use of journal articles as the canonical genre of research dissemination—a convention that, while seemingly set in stone, is in fact nonobvious. The standard length of an article is unlikely to match the nature and scope of the topic, thus creating both “salami-sliced” and redundant publications. Also, the expectations of the genre tend to disincentivize the publication of large datasets and trials. Finally, the growing pressure to publish in high-impact journals may lead to hyping the importance of one's claim, or even to misleading reporting and data manipulation. Focusing specifically on data-intensive research, Wager argues that all these problems could be avoided by switching to research reports (rather than articles) structured in a way to make their data machine readable. Operating outside of any regime of metrics, these would be “outputs” rather than “articles”—outputs to be used and built upon rather than evaluated (in the sense that metrics give to evaluation). Of course, articles discussing and commenting on published data would still be produced, but they would not be the “main course” of science publishing anymore, thus forcing universities and government agencies to articulate “new methods for measuring research productivity.”

Notes

1. Ivan Oransky, “Lawsuit Involving PubPeer Unmasks Commenter as Pseudonymous Whistleblower Clare Francis,” *Retraction Watch*, April 13, 2015. <http://retractionwatch.com/2015/04/13/lawsuit-involving-pubpeer-unmasks-commenter-as-pseudonymous-whistleblower-clare-francis/> (accessed October 10, 2016).
2. But there are exceptions, like Yves Gingras, *Bibliometrics and Research Evaluation: Uses and Misuses* (Cambridge, MA: MIT Press, 2016), pp. 67–68.