



PERSPECTIVES



SCIENTIFIC INTEGRITY

Self-correction in science at work

Improve incentives to support research integrity

By Bruce Alberts,¹ Ralph J. Cicerone,² Stephen E. Fienberg,³ Alexander Kamb,⁴ Marcia McNutt,^{5*} Robert M. Nerem,⁶ Randy Schekman,⁷ Richard Shiffrin,⁸ Victoria Stodden,⁹ Subra Suresh,¹⁰ Maria T. Zuber,¹¹ Barbara Kline Pope,¹² Kathleen Hall Jamieson^{13,14}

WEEK after week, news outlets carry word of new scientific discoveries, but the media sometimes give suspect science equal play with substantive discoveries. Careful qualifications about what is known are lost in categorical headlines. Rare instances of misconduct or instances of irreproducibility are translated into concerns that science is broken. The Octo-

ber 2013 *Economist* headline proclaimed “Trouble at the lab: Scientists like to think of science as self-correcting. To an alarming degree, it is not” (1). Yet, that article is also rich with instances of science both policing itself, which is how the problems came to *The Economist’s* attention in the first place, and addressing discovered lapses and irreproducibility concerns. In light of such issues and efforts, the U.S. National Academy of Sciences (NAS) and the Annenberg Retreat at Sunnylands convened our group to examine ways to remove some of the current disincentives to high standards of integrity in science.

Like all human endeavors, science is imperfect. However, as Robert Merton noted more than half a century ago “the

activities of scientists are subject to rigorous policing, to a degree perhaps unparalleled in any other field of activity” (2). As a result, as Popper argued, “science is one of the very few human activities—perhaps the only one—in which errors are systematically criticized and fairly often, in time, corrected” (3). Instances in which scientists detect and address flaws in work constitute evidence of success, not failure, because they demonstrate the underlying protective mechanisms of science at work.

Still, as in any human venture, science writ large does not always live up to its ideals. Although attempts to replicate the 1998 Wakefield study alleging an association between autism and the MMR (measles,

POLICY

ILLUSTRATION: DAVIDE BONAZZI

mumps, and rubella) vaccine quickly discredited the work, it took far too long—12 years—for *The Lancet* to retract that fatally flawed article. By contrast, problems flagged in the January 2014 Obokata pluripotent stem cell papers led to a prompt investigation by the research institute and *Nature* retracting the papers by July 2014.

Leaders in the research community are responsible for ensuring that management systems keep pace with revolutions in research capacity and methods. Consistent with their self-correcting norm, scientists are actively addressing the disconcerting rise in irreproducible findings and retractions. As the *Economist* article itself noted, *PLOS One* and *Science Exchange* had begun an initiative “through which life scientists can pay to have their work validated by an independent lab”; *Nature* had initiated an 18-point checklist for authors “to ensure that all technical and statistical information that is crucial to an experiment’s reproducibility or that might introduce bias is published”; and *Perspectives on Psychological Science* developed “a section devoted to replications” (1). Conferences such as the Association for Computing Machinery’s Special Interest Group on Management of Data have begun to require reproducibility in accepted papers. After Obokata, *Nature* announced measures to increase the likelihood that misrepresented visuals will be detected in the review process (4). Other efforts include tighter financial disclosure rules, journal guidelines mandating increased transparency [see Nosek *et al.*, page 1422 in this issue (5)], and increased data disclosure (6). Innovations such as Cross Mark and Retraction Watch have made it easier for scholars to purge retracted work from the scholarly dialogue.

INCENTIVES FOR QUALITY AND CORRECTION. Nosek and his colleagues have argued that some key incentive structures embraced by academia are counterproductive. Researchers are encouraged to publish novel, positive results and to warehouse any

negative findings (7). Cash bonuses paid in a number of countries have increased the number of submissions to prestigious journals (8), and in some instances that work has been fraudulent. Growing numbers of biomedical students funded by the U.S. National Institutes of Health contribute to a pipeline of researchers in serial post-doctoral positions, as well as to declining morale and ever-lengthening time until researchers obtain their first independent research grant.

We believe that incentives should be changed so that scholars are rewarded for publishing well rather than often. In tenure cases at universities, as in grant submissions, the candidate should be evaluated on the importance of a select set of work, instead of using the number of publications or impact rating of a journal as a surrogate for quality. This practice is used in nominations for election to the NAS and selecting honorees in professional societies.

The peer-review process should do a better job of mentoring young reviewers,

“Ensuring that the integrity of science is protected is the responsibility of many stakeholders.”

increasing the clarity and quality of editorial response, and uncovering instances in which a reviewer is biased for or against a particular work. Although this may add to the reviewers’ burden, one solution, used by *eLife*, is for reviewers to share their comments with each other and collaborate on a response before sending comments to the author with the editorial decision. *Science’s* cross-review, in which reviews are exchanged between reviewers, allows editors to capture broader input on the reviews before returning them to the authors. Were some reform to be implemented to address bias in reviewing, along with an incentive structure that rewards publishing quality rather than quantity, we suspect that the integrity of science could be better protected at no net increased cost in reviewers’ time.

Reliance on the term “retraction” may create a disincentive to act in the best interests of science. The word “retraction,” with its negative connotation, covers withdrawal of scholarship both for inadvertent error and for misconduct. Yet, voluntary withdrawal of findings by a researcher eager to correct an unintended mistake is laudatory, in contrast to involuntary withdrawal by a duplicitous researcher who has published fraudulent

claims. Alternative nomenclature such as “voluntary withdrawal” and “withdrawal for cause” might remove stigma from the former while upping it for the latter. “Voluntary withdrawal” would be for papers that are wrong in major respects, cannot be fixed with just a correction, but are not a result of fraud or misconduct. Authors would be encouraged to take this route to avoid leaving confusing papers in the literature and to preserve their reputations. Ideally, some statute of limitations should be placed on voluntary withdrawals. “Withdrawal for cause” would be invoked any time fraud or misconduct taints the published literature. There should be no statute of limitations on withdrawal for cause.

In a similar vein, “conflict of interest” implies that disclosed relationships are corruptive. Adopting more neutral language such as “disclosure of relevant relationships” may encourage more complete compliance without implying that all disclosed associations are sinister.

INVESTIGATION AND EDUCATION. Authoritative and timely investigations into allegations of misconduct are critical to ensuring that flawed findings, which because of fraud or misconduct cannot be redeemed, are formally decertified. Because journals lack the wherewithal to investigate allegations of misconduct in published research, they rely on scholars’ home institutions to address such concerns. Funding agencies often play an oversight role. In cases in which the institution is unwilling, conflicted, or incapable of investigating, consequential flawed findings might linger in the literature. A more robust structural solution is needed.

Even when institutions, funders, and journals work in good faith to address misconduct and ensure the accuracy of the research record, they can be overwhelmed by the growing number of allegations, the expense and complexity of investigations, the difficulties in addressing misconduct in international and cross-disciplinary collaborations, and the emergence of technologies that make it easier both to commit misconduct and to detect it. The investigating unit should be respected, neutral, nimble, have wide access to necessary expertise, and be capable of responding regardless of the funding source.

Ensuring that the integrity of science is protected is the responsibility of many stakeholders. In 1992, the NAS called for an independent Scientific Integrity Advisory Board to exercise leadership in addressing ethical issues in research conduct (9). Although not implemented, if such an entity were instituted, it could serve as a respected

¹University of California, San Francisco, CA, USA. ²National Academy of Sciences, Washington, DC, USA. ³Carnegie Mellon University, Pittsburgh, PA, USA. ⁴Discovery Research, Amgen Inc., Thousand Oaks, CA, USA. ⁵Science, American Association for the Advancement of Science, Washington, DC, USA. ⁶Georgia Institute of Technology, Atlanta, GA, USA. ⁷Howard Hughes Medical Institute and University of California, Berkeley, CA, USA. ⁸Indiana University, Bloomington, IN, USA. ⁹University of Illinois at Urbana-Champaign, Champaign, IL, USA. ¹⁰Carnegie Mellon University, Pittsburgh, PA, USA. ¹¹Massachusetts Institute of Technology, Cambridge, MA, USA. ¹²National Academies Press, National Academy of Sciences, National Academy of Engineering, Institute of Medicine, National Research Council, Washington, DC, USA. ¹³Annenberg Retreat at Sunnyslands, Rancho Mirage, CA, USA. ¹⁴Annenberg Public Policy Center, University of Pennsylvania, Philadelphia, PA, USA. ^{*}Corresponding author. E-mail: mmcnuitt@aaas.org

and neutral resource that supports and complements efforts of the research enterprise and its key stakeholders.

Universities should insist that their faculties and students are schooled in the ethics of research, their publications feature neither honorific nor ghost authors, their public information offices avoid hype in publicizing findings, and suspect research is promptly and thoroughly investigated. All researchers need to realize that the best scientific practice is produced when, like Darwin, they persistently search for flaws in their arguments. Because inherent variability in biological systems makes it possible for researchers to explore different sets of conditions until the expected (and rewarded) result is obtained, the need

“Instances in which scientists detect and address flaws in work constitute evidence of success, not failure.”

for vigilant self-critique may be especially great in research with direct application to human disease. We encourage each branch of science to invest in case studies identifying what went wrong in a selected subset of nonreproducible publications—enlisting social scientists and experts in the respective fields to interview those who were involved (and perhaps examining lab notebooks or redoing statistical analyses), with the hope of deriving general principles for improving science in each field.

Industry should publish its failed efforts to reproduce scientific findings and join scientists in the academy in making the case for the importance of scientific work. Scientific associations should continue to communicate science as a way of knowing, and educate their members in ways to more effectively communicate key scientific findings to broader publics. Journals should continue to ask for higher standards of transparency and reproducibility.

We recognize that incentives can backfire. Still, because those such as enhanced social image and forms of public recognition (10, 11) can increase productive social behavior (12), we believe that replacing the stigma of retraction with language that lauds reporting of unintended errors in a publication will increase that behavior. Because sustaining a good reputation can incentivize cooperative behavior (13), we anticipate that our proposed changes in the review process will not

only increase the quality of the final product but also expose efforts to sabotage independent review. To ensure that such incentives not only advance our objectives but above all do no harm, we urge that each be scrutinized and evaluated before being broadly implemented.

Will past be prologue? If science is to enhance its capacities to improve our understanding of ourselves and our world, protect the hard-earned trust and esteem in which society holds it, and preserve its role as a driver of our economy, scientists must safeguard its rigor and reliability in the face of challenges posed by a research ecosystem that is evolving in dramatic and sometimes unsettling ways. To do this, the scientific research community needs to be involved in an ongoing dialogue. We hope that this essay and the report *The Integrity of Science* (14), forthcoming in 2015, will serve as catalysts for such a dialogue.

Asked at the close of the U.S. Constitutional Convention of 1787 whether the deliberations had produced a republic or a monarchy, Benjamin Franklin said “A Republic, if you can keep it.” Just as preserving a system of government requires ongoing dedication and vigilance, so too does protecting the integrity of science. ■

REFERENCES AND NOTES

1. Trouble at the lab, *The Economist*, 19 October 2013; www.economist.com/news/briefing/21588057-scientists-think-science-self-correcting-alarming-degree-it-not-trouble.
2. R. Merton, *The Sociology of Science: Theoretical and Empirical Investigations* (University of Chicago Press, Chicago, 1973), p. 276.
3. K. Popper, *Conjectures and Refutations: The Growth of Scientific Knowledge* (Routledge, London, 1963), p. 293.
4. Editorial Board, *Nature* **511**, 5 (2014); www.nature.com/news/stap-retracted-1.15488.
5. B. A. Nosek et al., *Science* **348**, 1422 (2015).
6. Institute of Medicine, *Discussion Framework for Clinical Trial Data Sharing: Guiding Principles, Elements, and Activities* (National Academies Press, Washington, DC, 2014).
7. B. Nosek, J. Spies, M. Motyl, *Perspect. Psychol. Sci.* **7**, 615 (2012).
8. C. Franzoni, G. Scellato, P. Stephan, *Science* **333**, 702 (2011).
9. National Academy of Sciences, National Academy of Engineering, and Institute of Medicine, *Responsible Science, Volume I: Ensuring the Integrity of the Research Process* (National Academies Press, Washington, DC, 1992).
10. N. Lacetera, M. Macis, *J. Econ. Behav. Organ.* **76**, 225 (2010).
11. D. Karlan, M. McConnell, *J. Econ. Behav. Organ.* **106**, 402 (2014).
12. R. Thaler, C. Sunstein, *Nudge: Improving Decisions About Health, Wealth and Happiness* (Yale Univ. Press, New Haven, CT, 2009).
13. T. Pfeiffer, L. Tran, C. Krumme, D. Rand, *J. R. Soc. Interface* **2012**, rsif20120332 (2012).
14. Committee on Science, Engineering, and Public Policy of the National Academy of Sciences, National Academy of Engineering, and Institute of Medicine, *The Integrity of Science* (National Academies Press, forthcoming). <http://www8.nationalacademies.org/cp/projectview.aspx?key=49387>.

10.1126/science.aab3847

SCIENTIFIC STANDARDS

Promoting an open research culture

Author guidelines for journals could help to promote transparency, openness, and reproducibility

By **B. A. Nosek,* G. Alter, G. C. Banks, D. Borsboom, S. D. Bowman, S. J. Breckler, S. Buck, C. D. Chambers, G. Chin, G. Christensen, M. Contestabile, A. Dafoe, E. Eich, J. Freese, R. Glennerster, D. Goroff, D. P. Green, B. Hesse, M. Humphreys, J. Ishiyama, D. Karlan, A. Kraut, A. Lupia, P. Mabry, T. A. Madon, N. Malhotra, E. Mayo-Wilson, M. McNutt, E. Miguel, E. Levy Paluck, U. Simonsohn, C. Soderberg, B. A. Spellman, J. Turitto, G. VandenBos, S. Vazire, E. J. Wagenmakers, R. Wilson, T. Yarkoni**

Transparency, openness, and reproducibility are readily recognized as vital features of science (1, 2). When asked, most scientists embrace these features as disciplinary norms and values (3). Therefore, one might expect that these valued features would be routine in daily practice. Yet, a growing body of evidence suggests that this is not the case (4–6).

A likely culprit for this disconnect is an academic reward system that does not sufficiently incentivize open practices (7). In the present reward system, emphasis on innovation may undermine practices that support verification. Too often, publication requirements

(whether actual or perceived) fail to encourage transparent, open, and reproducible science (2, 4, 8, 9). For example, in a transparent science, both null results and statistically significant results are made available and help others more accurately assess the evidence base for a phenomenon. In the present culture, however, null results are published less frequently than statistically significant results (10) and are, therefore, more likely inaccessible and lost in the “file drawer” (11).

The situation is a classic collective action problem. Many individual researchers lack