

# COMMENT

**PHYSICS** The standard model is broken, whether or not the Higgs boson exists **p.24**

**SOCIOLOGY** Collective quality control is central to digital-era science **p.25**

**CHEMISTRY** The new president of the American Chemical Society on education **p.28**

**CONSERVATION** Field work is limited to office hours in Mauritius **p.29**

ILLUSTRATIONS BY PETE ELLIS/DRANGOOD.COM



## How to stop plagiarism

Duplication is easily detected by software, yet it remains a problem. Ten experts explain how to stamp it out.

### HAROLD GARNER Flag plagiarized studies

*Founder of HelioText, and creator of eTBLAST plagiarism detection software*

When my colleagues and I introduced an automated process to spot similar citations in Medline, we uncovered more than 150 cases of potential plagiarism in March 2009 (ref. 1). Subsequent ethics investigations resulted in 56 retractions within a few months. However, as of December 2011, 12 (20%) of those 'retracted' papers are still not tagged as

such in PubMed. Another two were labelled with errata pointing to a website that warns readers that the papers are 'duplicate' — but more than 95% of the text is identical, and the papers share no co-authors. Without clear retractions, a casual reader may not realize the extent of the problem — 'duplicate' could be interpreted as a mild infraction.

The PubMed Central archived article citation resource indicates that 3 of the 56 retracted papers are cited in books — one after being retracted. Eight papers were cited in other archived articles before retraction, and seven were cited after retraction.

It may take years before papers are found and retracted. Simply put, we need to establish a better system, with a faster process for identifying and labelling papers that need retracting. Of course, this project will

require time and effort to develop, and therefore may need dedicated funding — it is a worthy cause, one that will ensure the quality of the research corpus.

Editors and researchers will also need to agree on a clear definition of plagiarism. Detection software does not define it — instead, it can only say whether a scanned text exceeds a threshold of similarity to another text. In our studies thus far, we have used a similarity threshold of approximately 50%; we then compared the full text of any articles that exceeded this threshold, line by line and figure by figure. Ultimately, plagiarism comes down to human judgement. Like other questionable practices, you will know plagiarism when you see it.

### BERND PULVERER Spot subtle forms

*Head of scientific publications, European Molecular Biology Organization*

Every manuscript that the European Molecular Biology Organization (EMBO) receives undergoes a plagiarism screen supported by text-comparison software before formal acceptance. Significant text duplications are rare and often confined to the materials and methods section. Usually, it is clear that there is no intent to copy; duplications are usually resolved before publication by ensuring appropriate editing and citation.

Sometimes, we encounter less-obvious forms of plagiarism. Most of the few manuscripts that we have had to reject on the basis of duplication were extensively 'self-plagiarized' — authors re-used their own text without quotation or citation. Authors don't always realize that repeating their own text can be considered plagiarism — but why would journals republish concepts, let alone verbatim text or unattributed data?

Indeed, plagiarism extends beyond the unattributed copying of the published literature to grants, patents, preprints and even talks. Figures, images and data are subject to the same rules as text. But what about blogging and online commenting — should ▶

► those be subject to the same plagiarism criteria? In our view, any publicly available, permanent record can and should be cited.

Detection software will not spot all forms of plagiarism. The unattributed rehashing of original ideas in an author's own words is much harder to detect. Consequently, we rely on a high-level peer review process and careful editing to spot such plagiarism. With rising 'publish or perish' pressures, it is also essential to teach high ethical standards. A thorough refereeing process remains the best guarantee for a robust scientific record.

## ANA MARUŠIĆ & MLADEN PETROVEČKI

### Check all manuscripts

*University of Split School of Medicine;  
University of Rijeka School of  
Medicine, Croatia*

Although we always considered publishing integrity in the *Croatian Medical Journal*, we struggled with our first plagiarism allegations, which involved a member of the local medical community. The journal's editors were pressurized to close the case, and even accused of misconduct themselves, charges that were rejected by the relevant authorities.

This experience taught us that it was better to prevent misconduct than to deal with it after publication. We established a clear policy on research misconduct and retractions, including a standard operating procedure for scanning submitted manuscripts for plagiarism. Since 2009 we have checked all submitted manuscripts using eTBLAST

and CrossCheck, a plagiarism-detection service from the publishing technology company CrossRef. So far, about 10% of the manuscripts have been flagged due to content similarity with other items, with a few serious cases of plagiarism. We deal with each case using the Committee on Publication Ethics Flowcharts. Very often the cases involve authors who do not speak English, who say that they were unaware they could not copy text from other authors or republish their own text.

Currently, many journals with a large number of submissions only check non-research articles for plagiarism. We believe that every journal should check all submissions, including original research. If anything, that should be the priority, as research articles present new knowledge and thus should be of the highest integrity.

## JOHN LOADSMAN

### Use professional translators

*Editor, Anaesthesia and  
Intensive Care*

Authors preparing a scientific manuscript in a non-native language sometimes use 'patch writing', surrounding their own data with words taken, usually without attribution, from the work of others. This form of plagiarism is among the most common, and dealing with it imposes a heavy workload on editors. Embarrassment — or worse — can be avoided if authors write in their native language and use a professional translator. To be safe, these authors should then run the translated text through online plagiarism

detection tools to be certain it doesn't match anything else — which is good advice for everyone, not just those who are writing in a foreign language.

## YUEHONG ZHANG & IAN MCINTOSH

### Blacklist repeat offenders

*Managing editor, and English editor,  
Journal of Zhejiang University  
Science A/B/C*

In October, the US Office of Research Integrity announced that Scott Weber, a nursing researcher at the University of Pittsburgh, Pennsylvania, had admitted to plagiarizing more than 90% of a manuscript submitted for publication, and roughly two-thirds of another manuscript — including tables and figures. One such offence is bad enough, but 16 years ago, a journal found his paper contained portions of a previously published paper, as well. (Weber has denied any knowledge of this previous incident.)

Clearly, the current system of policing plagiarism isn't sufficient. Weber has agreed to a three-year penalty in which he will neither apply for nor receive government funds. We propose an additional measure: an international database that blacklists frequent offenders. In many European countries, US states and China, a driver's licence comes with a point system. If you are caught breaking the law, by speeding, for example, you are issued points. Too many points, you lose your licence, and getting it back is expensive and time-consuming.

Of course, the devil is in the details. Who would set up the database and monitor it? How many instances of plagiarism would be needed for someone to be blacklisted? All major publishers — commercial and non-profit — should sign up to the project so they can work out the answers to such questions.

## SANDRA TITUS

### Invest in prevention

*Health science administrator, US  
Office of Research Integrity*

If I had to choose between buying software to detect plagiarism and directing resources to prevent it, I would choose the latter. That is not to say that detection is unimportant, but honesty and integrity are better served if





misappropriating long word strings. But it can be done:

“A hemisection was performed with a microblade beginning at the midline of the subject’s left spinal cord and caudal to the C2 dorsal roots, and ending at its lateral most extent of the cord. The same interventions minus the lesion were used with the sham hemisectioned controls.”

The message behind this exercise is that good scientific writing requires a solid command of the language and of the knowledge domain in question, and, importantly, a considerable amount of time and effort.

## MELISSA ANDERSON Catch system gamers

*University of Minnesota, Minneapolis*

To a generation raised on electronic games, getting past a plagiarism checker is simple: change the text just enough to pass detection. Students in my course on the responsible conduct of research at the University of Minnesota in Minneapolis have told me that all they need to do is run the text through a plagiarism checker, then keep modifying the text until the checker no longer links it with the original passage. The process, they say, takes the guesswork out of text alteration. Concerned instructors can try to replace key words in students’ writing with likely substitutions to increase the chances that the detection software will identify an original source.

Detection software will also miss some instances of plagiarism. It cannot catch what it cannot access, so plagiarizers can take advantage of journals that do not post materials online. Likewise, translation plagiarism, which involves publication of translated articles without acknowledgment of the original authors, can be difficult to catch, depending on the languages involved. Instructors and editors may get preliminary leads by using an online translation service to convert the material into the suspected language, then run it through a plagiarism checker. The best remedies, though, are scholarly vigilance and steadfast insistence on good citation practices. ■

1. Long, T. C., Errami, M., George, A. C., Sun, Z. & Garner, H. R. *Science* **323**, 1293–1294 (2009).
2. Parry, M. NYU Prof Vows Never to Probe Cheating Again — and Faces a Backlash. *The Chronicle of Higher Education* (21 June 2011).
3. Christensen Hughes, J. M. & McCabe, D. L. *Can. J. Higher Educ.* **36**, 1–21 (2006).
4. McCabe, D. L. *Liberal Education* 26–31 (Summer/Fall, 2005).
5. Allain, W. J. et al. *Nature* **475**, 196–200 (2011).

plagiarism and cheating are prevented.

Software alone is not enough. Last autumn, when Panagiotis Ipeiritis, a computer scientist at New York University, scanned assignments from his Introduction to Information Technology class with the plagiarism detection software Turnitin, about 20% of his students admitted cheating. He then had to spend an enormous amount of time handling those cases, and his policing efforts resulted in deterioration in the class environment, lower student evaluations, and a subsequent hit to his yearly salary increase<sup>2</sup>.

Prevention efforts need to be directed to students, faculty members and institutions. Institutional leaders must convey a consistent message on the importance of integrity. This can be done through rallies, seminars and presentations. Signing honour codes in public is sometimes part of the process.

During students’ first year at college, or better yet in high school, a compulsory course could discuss the process of writing, define plagiarism and teach correct use of citations — including Internet resources.

Most faculty members have never confronted a student suspected of plagiarism, but there is likely to be at least one case in any class<sup>3,4</sup>. Workshops that allow faculty members to rehearse talking to a student suspected of cheating can empower them to intervene. Consistent enforcement efforts are needed to convey that cheating has consequences.

In short, multiple and ongoing strategies are needed, otherwise passivity reinforces the unacceptable behaviour, and there will be pervasive cheating and lack of integrity in future generations of scientists and other professionals.

The views in this article are personal, and do not necessarily represent those of the Department of Health and Human Services or of the US federal government.

## MIGUEL ROIG Teach scientists to paraphrase

*St Johns University, New York, author  
of guide to avoiding plagiarism*

Plagiarism is incredibly common — 40% of students admit to doing it in written assignments<sup>4</sup>. Some offenders rationalize the practice by claiming ignorance about what distinguishes acceptable paraphrasing from plagiarism, or by complaining that “there are only so many ways to say the same thing”. Providing a footnote to verbatim text won’t suffice.

We need to convince authors — particularly students — that their writing demands the same patience, attention to detail, honesty and transparency as the research they are trying to describe. Most writers know how to paraphrase correctly, but tend to plagiarize when faced with technical text. So in my workshops on avoiding plagiarism, I ask participants to paraphrase difficult-to-read text with unique terminology.

For example: “Using a microblade, a hemisection was made on the animal’s left spinal cord, caudal to the C2 dorsal roots and starting at the midline and extending to the lateral most extent of the spinal cord. Sham hemisectioned animals received all procedures but the lesion.”<sup>5</sup>

Many writers, particularly students, some early-career researchers and those who are not fluent in written English, will struggle to paraphrase this paragraph without

➔ **NATURE.COM**  
Shipping timetables  
clear Darwin of  
plagiarism:  
[go.nature.com/xoqqp2p](http://go.nature.com/xoqqp2p)