

My Words or Your Words? Detecting and Investigating Plagiarism

Nancy R. Gough

When I started my editorial career as an associate editor for the online-only journal *Science's STKE* in 2000, it was difficult to detect plagiarism. The current online plagiarism-detection tools that are widely available and used by an increasing number of publishers did not exist. Or, if they did exist, I did not have access to them, and the journal did not use them. Even with such tools, editors need to be able to properly investigate and identify cases of suspected plagiarism or self-plagiarism (also known as self-similarity). Here, I describe experiences I had and provide suggestions for how to detect and confirm cases of plagiarism of text.

Even without such tools, I identified a clear-cut case of self-plagiarism early in my career as an editor. The article in question was an invited review article. What alerted me was the quality and style of the writing from a nonnative-English speaker with whom I had been corresponding: The writing did not match the writing in the correspondence we had exchanged regarding contributing the review article.

To determine if the contributed manuscript was similar to another published article, I started searching for reviews on the same topic. It never occurred to me the plagiarized review I would eventually find might be by the same author. I had never considered self-plagiarism as a possibility. Before I could be sure the published article and the submitted manuscript matched, which from the abstracts of each seemed likely, I had to obtain a copy of the full text of the published article. Then I compared several aspects of the two documents: (i) the overall organization in terms of the sections; (ii) the beginnings and endings of the paragraphs; and (iii) the complete text of one entire section, including the references cited in that section. I expected the references to overlap in any review article published within a close time frame (approximately 6 months) on this topic. However, I found that not only was the text nearly identical with only trivial changes, but the references were almost identical, with only a few additions in the submitted manuscript (less than 10% of the total references were different), and the



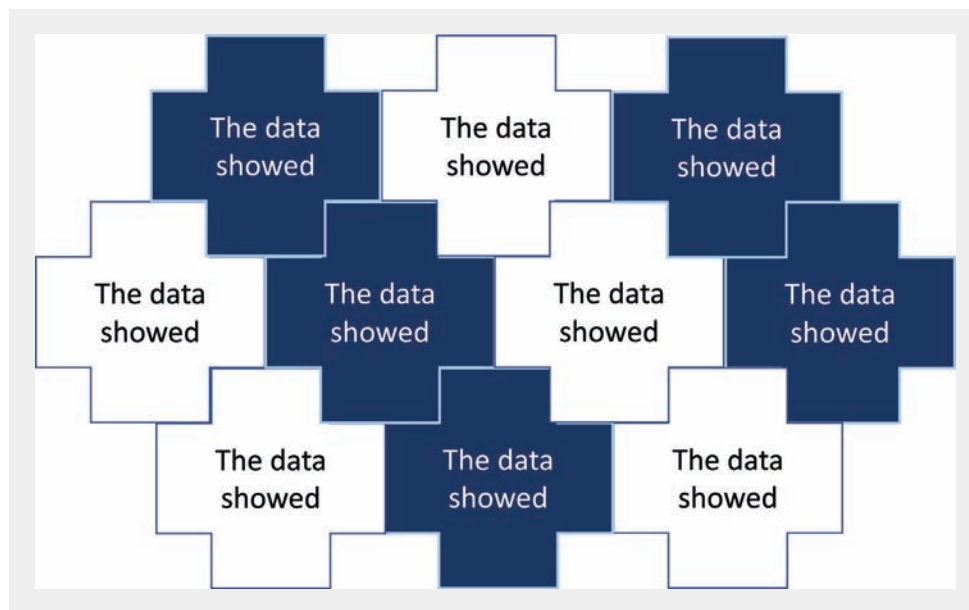
order of the cited references in the two files was exactly the same. At this point, it became clear to me that I could not proceed with the submitted article, and I contacted the author.

I was quite surprised at the response I received: The author did not realize republishing a nearly identical review with a small number of new references was not allowed.

I informed the author by email that I had determined the submitted manuscript was nearly identical to the previously published article in another journal, and I provided the exact details of the other article. I explained this is not permitted, and we were rejecting the submitted manuscript. I told the author a submission of a new manuscript that was substantially different from others they had authored would be considered if they wished to submit a new review for consideration. I was quite surprised at the response I received: The author did not realize republishing a nearly identical review with a small number of new references was

NANCY R. GOUGH is with BioSerendipity, LLC.

CONTINUED



allowed. They thought that because it was their own work it could be submitted and published in multiple journals.

This case occurred in the early days of online-only journals, which may have contributed to the confusion about this being a case of illegal self-plagiarism. The journal where I worked was online only, without any print component; the journal also had an unconventional title (including the phrase “knowledge environment”) that represented the entire online site. *Science’s STKE* was the abbreviation for *Science’s Signal Transduction Knowledge Environment*, which was published under that title from 1999 to 2007. It is possible the author thought the online-only format did not truly represent republication. I would hope this potential source of confusion is no longer an issue. However, I think many authors reuse their own text in various ways—this is not always inappropriate. The context and the extent and type of the self-plagiarized material are all factors that must be considered. Some text in grants may be used repeatedly. Descriptions of procedures are often very similar in many instances from grants and lab protocols available online to materials and methods sections of primary research articles. Authors may have a very similar or even identical way of describing their research in a biosketch or on their lab or departmental websites. A good rule of thumb is that, if the author signs a license to publish that has an exclusive publication clause, publishes under a Creative Commons license, or signs a copyright transfer agreement, then the text is not directly reusable without quoting or citing the original publication, or both. Grants are not subject to this kind of legal limitation and, generally speaking, neither are research descriptions used online for websites or inclusion in meeting programs.

In 2008, the title of *Science’s STKE* changed to *Science Signaling*, and the journal began to publish primary research. By this time, many journals were online with some moving to having online-only options for access. I was serving as Editor of the journal and handling my own assigned manuscripts as well as all ethical issues. The other form of plagiarism I encountered much more frequently than self-plagiarism was text copied directly from abstracts of cited literature. This was not typically self-plagiarism and was especially common in, but not limited to, review submissions. In this case, I had to use a different method to detect the plagiarism. Again, I did not use plagiarism-detection programs. I am not sure such programs would find these examples or, if they did, the amount of text involved would be sufficiently large to raise a red flag for the editor. Instead, the clues that plagiarism had occurred came from the writing itself.

I would notice a few sentences written in an unusual style compared with the rest of the manuscript. Even more revealing was the introduction of a new name for a molecule (protein, gene, or RNA), when in other parts of the manuscript the molecule was consistently written with a single name. This was a major red flag and was easy to investigate because the sentence or section included one or more references. I would find the references in a database, such as PubMed, and discover the sentence that triggered the warning in one of the abstracts. I also detected plagiarism of abstracts when I was trying to help authors be more precise in their presentation: I determined the authors had taken complete sentences directly from an abstract of one of the cited articles using the same process (finding the abstracts of the cited articles for the section that lacked

CONTINUED

sufficient detail). This was quite worrisome, because finding such plagiarized content suggested the authors had not actually read the articles they had cited. So, not only did the submitted manuscript have the problem of plagiarism, but it seemed to lack scholarly integrity: the authors had not read the cited articles in sufficient depth to be able to rephrase the findings in their own words or to realize the article was not actually making or supporting their claims.

In addition to detecting plagiarism, a new challenge now exists for editors—using plagiarism-detection software appropriately. Online tools are now widely available and used by many publishers to detect plagiarism. Properly using the output from such tools is a new challenge editors face. Relying solely on a simple score of similarity or identity is insufficient to gauge plagiarism in most cases. Editors also need to consider the context to decide if plagiarism of any kind has occurred. Lifting entire sentences or long scientific phrases from abstracts of the cited literature is inappropriate in a review article, especially when this is done without quotation marks or a citation to clearly indicate the text was taken from the cited article.

In a research article, the authors may have sections that are similar among their published papers. These may be close enough to trigger plagiarism flags in automated detectors. The flag may detect self-plagiarism or similarities with other authors' published work. For example, defining a protein or gene or describing the symptoms of a disease or condition is often presented similarly across publications. Papers describing case studies or clinical trials may have similar formats with consistent language—this is desirable and should not be considered plagiarism. Indeed, some journals have highly structured, almost formulaic abstracts that could trigger a high similarity score in a plagiarism-detection process.

Materials and methods sections are often similar. Although some journals prefer to have the authors use the language “performed as previously described” with a citation to a previous article, other journals are moving toward increasingly detailed materials and methods sections so the reader does not have to chase down a copy of the article containing the methods used in the paper. Industry-standard procedures or methods that exactly follow the manufacturer's protocols or instructions need not be reproduced. Conciseness is a virtue, but not at the expense of making the reader hunt for information necessary to reproduce or extend the findings of the study.

A clue that materials and methods may have been reproduced from another publication is the inclusion of sections that do not correspond to any data shown in the submitted manuscript. However, this also occurs as manuscripts are revised and reorganized after rejection or review and resubmitted for consideration. Methods that are completely identical to previous publications can also be an indication the authors have not adequately detailed any changes from previously presented methods or procedures or may be a tip the methods are incomplete. Although missing methods have nothing to do with plagiarism, sometimes they can be discovered when plagiarism of the materials and methods is detected. Authors who copy materials and methods from another publication (their own or someone else's) may fail to include descriptions of materials and methods specific to experiments performed in the current manuscript that were not part of the other publication. An inability to provide sufficiently detailed methods, relying instead on “as previously described” for most or all of the materials and methods, can be an indication the authors lack detailed information about how the experiments were conducted. Querying the authors about methods that are identical or highly similar to those that have been previously published, methods that lack any description, or missing methods for data presented is key to ensuring any specific modifications, reagents, or conditions used in the described research are presented for the reader.

As with most aspects of an editor's job, detecting and investigating plagiarism is a complex task. There are no absolute rules or a similarity threshold that will allow this process to be completely automated. Although technology makes detection easier in some cases, my 17 years of experience suggests paying close attention to the writing is critically important to properly identifying plagiarism and self-plagiarism. Furthermore, an editor needs to decide an appropriate course of action—requesting revisions by the authors, rejecting the manuscript, or reporting unethical behavior. Thus, discovering and handling cases of plagiarism will continue to require editors to read submissions carefully, have the skills necessary to investigate, and be able to exercise judgment. Ensuring the scientific literature conforms to the standards for scientific discourse, including knowing when a situation represents plagiarism, is just one of the many ways editors add value to the scientific enterprise.

Resource Nook

The **Conscious Style Guide** (<https://consciousstyleguide.com/>) is a resource for inclusive writing and editing, including questions around pronoun use and preferred terms.