Tristan A. Horrom

The perils of copy and paste: Plagiarism in scientific publishing

INTRODUCTION

Although research misconduct in all its forms can damage the integrity and prestige of the scientific community, it has often not been taken as seriously as it should be. However, leaders in science and publishing are beginning to regard misconduct as the threat that it really is. In an editorial in *Nature*, Macilwain explains that "scientists' instinctive defensiveness has produced general denial that misconduct constitutes a serious problem," but also that these attitudes are changing and misconduct is being addressed more often and with less leniency [1]. Research misconduct should be a paramount concern in the scientific publishing industry. The goal of any scientific or technical journal is to contribute to the advancement of the literature base of its target subject, and publishing fraudulent, unethical, or incorrect research damages not only the reputations of the journal and the authors but also the research field itself. Research misconduct can range from unethical treatment of research subjects to fabrication and falsification of data to plagiarism [2]. It is therefore the responsibility of every scientific journal to monitor all material to be published for research misconduct of all kinds, a responsibility that *JRRD* takes seriously. As pointed out by the Committee on Publication Ethics (COPE), journal editors are accountable for everything published in their journals, such that they should attempt to maintain the integrity of the academic record and uphold ethical standards [3].

Despite the efforts of the publishing community, research misconduct is on the rise. In the past 30 years, retractions in scientific journals have increased 10-fold [4]. Fanelli reports that nearly 34 percent of scientists and doctors admitted to employing questionable research practices [5]. This increase in misconduct is due largely to an unprecedented growth in the number of researchers and scientific publishers worldwide. As research efforts rise in parts of the world that did not previously have as large of a stake in the scientific community, such as China and India, the volume of published literature continues to increase. The number of journal articles published annually jumped from 1.09 million in 2002 to 1.94 million in 2010 [6]. *JRRD* has seen a drastic increase in our article submissions, leading to both more articles per issue and a higher rejection rate. The growth of the Internet has also led to an increase in the number of publication venues. With this increase in scholarship and publication outlets comes a pressure to publish, which can sometimes lead to misconduct.

The widespread use of computers and the Internet has also led to an increase in misconduct, specifically plagiarism. Most published material is now available at the click of a mouse over the Internet, and the Copy and Paste functions allow broad swathes of text to be duplicated with very little effort by the copier. When it is so easy to find and duplicate previously published materials, we must be especially vigilant to root out plagiarism in all its forms.

The publishing world has been taking notice of widespread plagiarism lately, and prominent publications and scholars are not immune to the problem. There have been recent plagiarism allegations against figures such as a well-known writer and journalist Jonah Lehrer [7], writer and television host Fareed Zakaria [8], and even a member of the Romanian government [9]. This renewed attention to possible plagiarism has included concerns about self-plagiarism, highlighted by the discovery that a well-known chemist had published multiple papers containing nearly identical text [10]. While plagiarism in scientific and technical writing is indeed a growing problem, we also have new tools and an increasing focus within the publishing industry with which to fight back.

DEFINITIONS AND EXPLANATIONS

Plagiarism

As stated in the *JRRD* Editorial Policies, plagiarism occurs when authors use material(s) that are not their original work without documentation (accurately citing the source, using quotation marks if necessary, and obtaining appropriate permissions) and extends to text, figures, and other unique materials. At its root, plagiarism is an attempt to pass off work that has been previously published as new and original. Not only is plagiarism unethical, but it often violates copyright law and is therefore also illegal. Plagiarism is considered misconduct and will be treated as such.

Any text or ideas that come from another source must be properly cited and may require permission from the original source, depending on the nature of the use. If text, images, or uncommon equations are used verbatim from a different source, the borrowed material must be quoted in addition to being cited. Ideas that are paraphrased but that represent borrowed concepts that are not common knowledge must also be cited [11]. Furthermore, changing a few words and creating a patchwork text using a variety of sources does not clear one from committing plagiarism. Every source used in the writing of a paper must be acknowledged even if the content is paraphrased or summarized rather than directly quoted [12].

No set rule exists for how much of a text needs to be copied for it to be considered plagiarism. While a similarity rate of 20 to 30 percent is sometimes used as a guide, the specifics of each case must be examined. The editor must consider the length of the article when evaluating what percentage match is unacceptable, as well as things such as the nature of the source material and the distribution of the copied text. For example, matching text from abstracts, from papers presented at conferences but not published in full, or from clinical trial registries or grant applications is not usually considered plagiarism, whereas matching text from another scientific journal is. For these reasons, potential plagiarism needs to be evaluated on a case-by-case basis.

Self-Plagiarism

Self-plagiarism occurs when an author reuses parts of his or her own previously published work in a new article. Self-plagiarism can take several forms: duplicate publication, "salami slicing" of research, and textual reuse. Duplicate publication is when the same or nearly the same article is published in multiple publications and is explicitly forbidden by the policies of most publishers, including *JRRD*. Duplicate publication dilutes the literature base and is often an unscrupulous way to pad one's résumé. Salami slicing refers to the practice of researchers "publish[ing] separate parts of the same study with near identical introduction and methods sections in different journals" [13]. This practice may again be an attempt to pad one's résumé by stretching one study into multiple publications, and it once again distorts the research record.

Textual reuse is by far the most common type of self-plagiarism encountered, and opinions are still divided in the publishing industry as to what constitutes misconduct. Many argue that plagiarism cannot occur when an author is using his or her own words. However, this argument does not take into account copyright law. Furthermore, many argue that once a researcher or laboratory has landed on the best way to say something, they should be able to use boilerplate language, especially in the Introduction and Methods sections of papers. Indeed, most self-plagiarism encountered at *JRRD* is in the Methods section. But the fact remains that if text has already been published, it must be properly cited because that work belongs with the research record of the previously published paper. In a 2009 case report, COPE stated that replication of whole sentences and paragraphs is not acceptable and that self-plagiarism may present a problem with transparency as well as copyright violation [14]. While self-plagiarism of Methods may not constitute a deliberate attempt to deceive the reader, it nevertheless violates most editorial policies, including those put forth by COPE, the International Society of Managing and Technical Editors [15], and the Office of Research Integrity [12]. Copying Introduction or Methods sections that you have previously written may not be deceptive, but it creates redundancy in the literature and is often the result of intellectual laziness [13].

As with use of the works of others, researchers must properly quote and cite their own work that has been published previously. Permissions may be required from the publisher of the original text even though the writers are the same. While the people involved in the research and the writing may be the same for multiple published pieces, the copyrights and the responsibility of maintaining the literature base falls on the publishers. Additionally, when researchers borrow from their own previous work, they often do not take into account that the entire author lists are not the same between papers. Claiming work to be original when some of those who took part in the previous

work are not given credit or when some of the authors listed on a paper were not involved in the previous work is misleading and unacceptable. The misconduct involved in self-plagiarism may be more subtle than plagiarism of another from an ethical standpoint, but it is nonetheless there. The self-plagiarized work may not be pirated from someone who is not being given credit, "but it implies that the work the reader currently sees is new and original and not copied from previous work" [16].

Self-plagiarism is by far the most common form of plagiarism encountered by *JRRD*. Of the 38 conditionally accepted articles deemed to contain possible plagiarism since we began using plagiarism detection software in the summer of 2010, all but 3 have been cases of self-plagiarism. There is a general feeling within many scientific and medical circles that this type of text repetition is acceptable. I have repeatedly heard that reusing Introduction and Methods sections "is what everyone does" or that "we have been doing this for years." While this may be true, the fact that self-plagiarism is a widespread practice does not mean that we can allow it. Reusing text may have long been standard practice, but that does not mean it was acceptable in the past. The new focus on self-plagiarism within the publishing industry does not represent a change in policy; rather, we now simply have better tools to catch behavior that was never acceptable but was hard to identify in the past.

Copyright, Public Domain, and Fair Use

Copyright is a legal protection provided by the laws of the United States or other countries for "original works of authorship" [17]. Once a work has been copyrighted, it is illegal for someone not the owner of the copyright to reproduce, distribute, or prepare derivative works of the original work without the permission of the copyright holder. Multiple factors are considered when determining whether a copyright has been infringed: the manner in which the material has been used, the nature of the copyrighted work, the amount of the original work that has been taken, and the potential for harm to the markets of the holder of the copyright [16]. While copyright assignment policies differ internationally, in the United States many journals require authors to assign copyright of their work to the publishers as a condition of publication. If copyright has been assigned to the journal, the publisher owns the rights of use and reproduction rather than the author.

Permission to reuse a copyrighted work may not be necessary in specific circumstances under the doctrine of "fair use." Fair use allows reproduction in cases such as criticism, comment, news reporting, teaching, scholarship, and research [18]. It is through fair use that materials may be quoted and used as sources in research papers without first gaining permission from the copyright holders. Reproduction through the fair use doctrine still requires proper citation, however. No set number of words or lines is defined to distinguish between what is fair use and what is copyright infringement. Rather, infringement is determined on a case-by-case basis. Authors should also be aware that reusing large portions of text can constitute copyright infringement even if the text in question is quoted and cited [19]. Note, also, that fair use is a legal defense rather than a right. Fair use rules are only applied once an allegation of copyright infringement has been made.

Works that are not protected under copyright are considered in the public domain. Public domain works may be used and reproduced without permission from the author because they are not legally owned by any individual. A work may belong to the public domain if no copyrights were ever obtained or if the copyright on the work has expired. Authors may also release their materials under a Creative Commons license. When a work is released under the Creative Common license, permission is given to share and use the work under conditions set by the author [20]. Additionally, most government documents and publications are public domain, including all *JRRD* articles.

While public domain works may be reused without permission from the author, it is still unethical to claim public domain works as your own without proper citation. For this reason, *JRRD* requires the same level of citation and attribution for copyrighted and noncopyrighted sources. We do not have the manpower to check the copyright status of every reference used in a submitted paper, and therefore treat plagiarism and self-plagiarism from public domain sources, including other *JRRD* articles, in the same manner that we treat plagiarism from copyrighted sources.

JRRD'S RESPONSE TO PLAGIARISM

Screening Process

JRRD is committed to catching plagiarism before it is published. As explained in a recent *Nature* commentary, it is "better to prevent misconduct than to deal with it after publication" [21]. To this end, we screen every article for plagiarism using several steps. Further details regarding submission guidelines can be found on the *JRRD* Web

site. First, all submitting authors must sign a statement of originality before manuscripts are sent for peer review. On this form, all authors on the paper affirm that the manuscript is an original work that has not been submitted or published elsewhere. After this form has been signed, manuscripts go through the peer review process, during which our reviewers will alert us if they suspect that any part of a manuscript has been plagiarized.

After a manuscript has passed through peer review and has been conditionally accepted, it is screened for plagiarism using the iThenticate plagiarism software. Before this process is complete, manuscripts are only conditionally accepted, meaning that they may still be rejected if problems arise. All manuscripts must pass through iThenticate before they are published, as indicated by the iThenticate logo that appears at the end of every *JRRD* article. iThenticate is an online plagiarism checker that compares the text of uploaded manuscripts against an extensive database that includes CrossRef, EBSCOhost, Elsevier, IEEE, and many other scholastic publishers, aggregators, and government bodies, as well as more than 24 billion Web sites and more than 122 million content items [22]. iThenticate presents the manuscript with all matching text highlighted and annotated with links to the source of the matching. Once the iThenticate checker is run, each manuscript is checked individually by an editor to determine the nature of the matches and whether the iThenticate findings require further action. All determinations on whether a manuscript contains plagiarism are solely at the discretion of the Editor; we do not use a set percentage for determining when plagiarism has occurred.

Plagiarism of Others

If a manuscript is determined to contain plagiarism, we follow COPE guidelines on how to proceed. A flowchart of this process can be found on the COPE Web site (http://publicationethics.org/files/u2/02A_Plagiarism_Submitted.pdf). First, the corresponding author is contacted and given the opportunity to explain his or her actions. If the author gives a satisfactory answer, the manuscript may be rejected or a revision may be requested. If the author does not respond, does not give a satisfactory explanation, or admits to misconduct, that author's superiors, including department heads, university deans, or funding agencies, may be contacted. Further details on this process can be found on the flowchart mentioned previously or on the COPE Guidelines page (http://publicationethics.org/resources/guidelines). This process can be expensive and can cause long-lasting harm to the individuals committing the misconduct, but this is far simpler and preferable to the action we are required to take when plagiarism is discovered after publication.

Self-Plagiarism

In the past, we have handled most cases of self-plagiarism off the record. Since the concept of self-plagiarism is new to many in the research community, we have done our best to work with and educate our authors on what is and is not acceptable behavior. Beginning with the first issue in 2013, we will take a more standardized and formal approach to dealing with self-plagiarism.

When self-plagiarism is discovered through iThenticate, we will first contact the corresponding author with a report of the matches found and a request for the author to revise the affected text. Authors will be required to substantially revise the affected section and return a revision within 2 weeks. If a revision is not received in the allotted time, the manuscript will be rejected. Manuscripts must be substantially rewritten to remove self-plagiarism; replacing a few words and moving sentences around are not sufficient to alleviate the problem and will still return matching results in iThenticate. Once a revision is received, the manuscript will be rerun through iThenticate to determine whether the revision was sufficient to resolve the problem. If the revision is not sufficient to satisfy the Editor that all plagiarism concerns have been met, the manuscript will be rejected. Depending on the author's willingness to cooperate and the nature of the self-plagiarism, the author's institution may be contacted if we cannot reach an acceptable resolution with the author.

RECOMMENDATIONS

The best way to avoid plagiarism is to make sure that all material in a submitted manuscript is new and original. Copy and paste should be avoided. There are several things that authors can do to make sure they do not hear from the Editor with an allegation of misconduct. First, authors should make sure that all sources are properly cited and quoted if necessary. Sections that contain text similar to text in another work should be rewritten in the author's own words. Even Methods sections that contain similar procedures from previous studies must be reworded to avoid copyright complications. See the **Figure** for options for including Introductions or Methods material that has previously been published.

If your paper contains a previously published procedure or section—

- Paraphrase the text and add a citation to the original.
- Put matching text in quotes and add a citation to the original.
- Refer to the original publication without repeating the text.
- Include a properly cited appendix with the relevant text for online publication.

Figure.
Recommended options to avoid plagiarism in Introduction and Methods sections.

If a researcher or laboratory uses the same procedure for many different studies, or if the Methods for a study have already been presented in another article, we recommend that authors simply reference the previously published procedures. If the exact procedures for a study are already part of the research records, then it is not necessary to publish the same text again. Rather, it is sufficient to include a statement such as "This study has been described in detail previously in [name of the previous study]" and include a citation. This will avoid duplicate text by eliminating the redundancy of multiple similar Methods. If authors are concerned that this will make it difficult for readers to find the procedures and understand the article, the previously published Methods may be included online as an appendix with proper citation.

Furthermore, we strongly recommend that all authors check their manuscripts using plagiarism detection software such as iThenticate before submitting for publication. This will ensure that authors know about potential problems and are able to resolve them before the publication staff gets involved. iThenticate offers authors and researchers a service with which they can upload an article and up to five revisions for a small fee. Checking your manuscript with plagiarism detection software before submitting will allow you to see exactly what the editors see when we check for plagiarism. Beyond iThenticate, many authors with university affiliations have access to services such as Turnitin, which belongs to the same parent company as iThenticate, iParadigms. Other detection services are available online as well. By using a plagiarism detection program before submitting your article for publication, you can catch any potential problems right at the beginning and give your manuscript a greater chance at being published.

Stopping misconduct such as plagiarism and self-plagiarism is the responsibility of both the publisher and the researcher. *JRRD* is putting into place these new policies in order to provide our stakeholders with the best possible content and to promote the best science and best practices in the field of rehabilitation research. By working with our authors, we hope to continue to improve and to catch plagiarism at the beginning so that we can avoid ever having to print a retraction.

Tristan A. Horrom
Technical Writer-Editor, *JRRD*

Email: tristan.horrom@va.gov

REFERENCES

1. Macilwain C. The time is right to confront misconduct. Nature. 2012;488(7409):7. [PMID:22859166] http://dx.doi.org/10.1038/488007a
2. Council of Science Editors. CSE's white paper on promoting integrity in scientific journal publications, 2012 update. Wheat Ridge (CO): CSE; 2012.
3. Committee on Publication Ethics. Code of conduct and best practice guidelines for journal editors [Internet]. London (UK): COPE; 2011. Available from: http://publicationethics.org/files/Code_of_conduct_for_journal_editors_Mar11.pdf
4. True costs of research misconduct [Internet]. Oakland (CA): iThenticate; 2012. Available from: http://www.ithenticate.com/research-misconduct-report/
5. Fanelli D. How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. PLoS ONE. 2009;4(5):e5738. [PMID:19478950] http://dx.doi.org/10.1371/journal.pone.0005738
6. iThenticate. Research misconduct infographic [Internet]. Oakland (CA): iParadigms; 2012. Available from: http://www.ithenticate.com/research-misconduct-infographic/#.UDvDmqDzjgk
7. Myers S. Jonah Lehrer's publisher is reviewing all of his books [Internet]. St. Petersburg (FL): The Poynter Institute; 2012. Available from: http://www.poynter.org/latest-news/mediawire/184174/jonah-lehrers-publisher-is-reviewing-all-of-his-books/
8. Haughney C. CNN and Time suspend journalist after admission of plagiarism [Internet]. New York (NY): The New York Times; 2012. Available from: http://mediadecoder.blogs.nytimes.com/2012/08/10/time-magazine-to-examine-plagiarism-accusation-against-zakaria/
9. Abbott A. Plagiarism charges for Romanian minister. Nature. 2012;485:289. http://dx.doi.org/10.1038/485289a
10. Cressey D. Eminent chemist denies self-plagiarism in 'space dinosaurs' paper [Internet]. London (UK): Nature News Blog; 2012. Available from: http://blogs.nature.com/news/

2012/04/eminent-chemist-denies-self-plagiarism-in-space-dinosaurs-paper.html/

11. Plagiarism.org. Plagiarism FAQs [Internet]. Oakland (CA): iParadigms; 2012. Available from: http://www.plagiarism.org/plag_article_plagiarism_faq.html
12. Office of Research Integrity. Avoiding plagiarism, self-plagiarism, and other questionable writing practices: A guide to ethical writing [Internet]. Washington (DC): U.S. Department of Health & Human Services; 2011. Available from: http://ori.dhhs.gov/education/products/plagiarism/5.shtml
13. Self-plagiarism: unintentional, harmless, or fraud? Lancet. 2009;374(9691):664. [PMID:19716942] http://dx.doi.org/10.1016/S0140-6736(09)61536-1
14. Committee on Publication Ethics. Self plagiarism. Case number: 09–21 [Internet]. London (UK): COPE; 2009. Available from: http://publicationethics.org/case/self-plagiarism
15. Publishing ethics 101: A guide for the editorial office. Mantua (NJ): International Society of Managing and Technical Editors; 2011.
16. Samuelson P. Self-plagiarism or fair use? Commun ACM. 1994;37(8):21–25. http://dx.doi.org/10.1145/179606.179731
17. Copyright basics [Internet]. Washington (DC): U.S. Copyright Office; 2012. Available from: http://www.copyright.gov/circs/circ01.pdf
18. Fair use [Internet]. Washington (DC): U.S. Copyright Office; 2012. Available from: http://www.copyright.gov/fls/fl102.html
19. iThenticate. White paper: The ethics of self-plagiarism. Oakland (CA): iParadigms; 2011.
20. What is Creative Commons [Internet]. Mountain View (CA): Creative Commons; 2012. Available from: http://creativecommons.org/about
21. Maruši A, Petroveki M. How to stop plagiarism: Check all manuscripts. Nature. 2012;481:22.
22. iThenticate. Largest comparison database for plagiarism detection [Internet]. Oakland (CA): iParadigms; 2012. Available from: http://www.ithenticate.com/plagiarism-detection-database/

This article and all supplementary material should be cited as follows:
Horrom TA. The perils of copy and paste: Plagiarism in scientific publishing. J Rehabil Res Dev. 2012;49(8):vii–xii. http://dx.doi.org/10.1682/JRRD.2012.09.0165