PERSPECTIVES

COPYCAT: SIMILARITIES AND ACADEMIC PRODUCTION IN THE DIGITAL AGE

INTRODUCTION

My experience as a *RAE* editor allowed me to learn about the difficulties of dealing with the issue of similarity of scientific publishing in the digital age. I had to deliberate on sensitive issues constantly, and in several opportunities I expressed myself, in oral presentations and in RAE's editorials, how I was dealing with this complex subject. In this brief article of the Perspectives section, I discuss some concepts using editorials of that period.

I will avoid, whenever possible, the expression, "plagiarism," as this discussion aims at informing and educating authors. Plagiarism presupposes deceit, and in most cases I faced, the main problems were errors of interpretation and lack of understanding of the academic publishing process. In cases where an explicit bad intention is identified, the problem has a different nature, and I will not discuss them here because they happen less frequently and must be addressed according to legal precepts, which is not my specialty. The purpose of this article is to help well-meaning authors in situations they may face when preparing papers for publication in high-level journals.

Firstly, to better understand the complexity of the problem of similarity and the diversity of situations in which it occurs, I recommend consulting the infographic Did I plagiarize? (The Visual Communication Guy, 2014). This infographic introduces a typology that comprises 13 different situations in which some degree of similarity in an paper is possible. The gravity of the problem varies according to the situation faced. Authors can learn about the limits of reusing texts in scientific works by answering the questions in the infographic.

Secondly, we are in the process of cultural adaptation to an increasingly digitized world. Until a few years ago, copying text from one place to another was much more laborious than writing an original text. The copy & paste practice is relatively recent in our lives, and due to technical issues was only introduced by the spread of digital technologies. As it is an easy and convenient practice, it has spread rapidly, and we are not yet fully aware of the risks it involves.

EDUARDO H. DINIZ

eduardo.diniz@fgv.br

ORCID: 0000-0002-7950-9146

Fundação Getulio Vargas, Escola de Administração de Empresas de São Paulo, São Paulo, SP, Brazil

Thirdly, and herein lies the most complicated problem, we (scholars) are under pressure by the publish or perish logic. The perverse thing is that the evaluation of scientific production is better fit to compute "quantity", with much less consensus on how to identify the "quality" of this production. This is the origin of so-called "productivism," a harmful behavior for the evolution of scientific knowledge, which has been discussed constantly in the academy, and generated good papers, some published in *RAE* (Machado & Bianchetti, 2011).

After this introduction, we will advance the discussion on similarities by addressing two topics. The first is about differences and similarities in full papers, i.e., how to consider two papers that are actually one or a paper that has been divided in two. The second is about similarities found in parts of papers, i.e., material that is copied from one paper and used in another.

REPUBLICATION AND SALAMI

Initially, I will present a discussion from the editorial, "Ethics, and discernment against productivism" (Diniz, 2013), written in response to a question addressed to the editorial team on whether the journal would accept a second version of one previously published, either in another language or with modified text.

Although the obvious answer should be "no," the question allowed me to reflect more deeply on the situation of a hypothetical author who, having originally published a paper in Chinese, would like to share ideas with other audiences, for example, in Brazil. On the one hand, it cannot be denied that that paper would be original in the Portuguese, what could justify its republication. On the other hand, by accepting this paper, the journal could tarnish its image for publishing a paper already published, even though it was in a language unfamiliar to most of its readers. Although this may seem a somewhat Byzantine question, it gains relevance if we consider the hot debate in the academic environment about which language national journals should adopt.

However, the editorial moves in another direction and poses a different question: Who does a magazine intend to serve by republishing a paper? If the answer concerns needs of the hypothetical author to publish in a more relevant journal because the Chinese version was published in a lower ranked journal, then we are facing the productivist logic. Alternatively, if the paper is explicitly identified as a reproduction, preventing it from being registered as a new production of the author, and the editorial team evaluates that it is of interest to the journal's community, it is possible to accept the idea of republication.

What is important in this case is to understand that one journal's primary commitment must always be to its readers, and not to the logic of productivism. Although scientific journals are committed to their authors, they should be predominantly committed to their readers and the principles of the ethical dissemination of knowledge. Thus, the best response to the question addressed to the editorial is, "It depends." Publishing simply to meet the hypothetical author's needs should receive a resounding "No," while republishing a paper, communicating to readers the existence of the original version, and emphasizing the relevance of the dissemination of that knowledge in an ethical manner is acceptable.

A similar situation occurs when a modified version of another paper is published. Would it be possible to extract more than one paper from a single different empirical investigation analysed on the same theoretical approach? Or, otherwise, to publish different papers from the same empirical basis? Once again, we have to isolate the logic of productivism and evaluate the possible contributions of the paper to the primary target audience of the journal, its readers. If the similar version brings no new contribution, there is no interest in the publication. Alternatively, if using a new conceptual approach over the same data or using the same conceptual approach on different empirical data sets brings a new contribution to a given academic area, the second version could be accepted.

Avoiding what is popularly called "salami research," a practice in which the author divides content for the sole purpose of producing a greater number of papers, it is perfectly acceptable to imagine one single research work as a source for different papers. Again, the important thing is not the number of papers but the diversity of contributions. Although any publisher might prefer a more complete contribution in a single paper instead of two partial contributions in different papers, what should be avoided is the logic of productivism.

Identifying whether two papers are actually a single paper is not always an easy task. Many authors would like journals to state in their publication guidelines the conditions for acceptance of modified versions. However, as noted above, not everything is crystal clear. Although journals are obliged to define what is acceptable within their editorial line, each case is unique. Moreover, as the editorial concludes, "more than any rule, ethics and discernment are the best medicine against productivism."

LIMITS OF SIMILARITY IDENTIFICATION TOOLS

Another editorial (Diniz, 2015b) addressed the implementation in *RAE* of a new tool for identifying similarities. As an increasingly common editorial instrument for scientific journals, I nominated similarities detection systems as the "second stage of the digital revolution in scientific journals" (Diniz, 2015a).

Given the higher level of digitization in our society, journals must care about ensuring integrity on what they publish and felt pressure to identify similarities in papers they receive. Although ethics is most obvious issue to influence adoption of similarities detection tools, dealing with intellectual property rights in an academic context cannot be ignored.

When using a similarity detection tool, journals must make decisions on conflicting situations they have not experienced in the past. With the system's implementation in *RAE*, I realized that classifying and addressing different cases of similarity had become a much more complex activity and therefore caused potentially rising attrition between the editorial team and authors.

These systems aim to prevent journals from publishing excerpts from third-party texts without referring to the original, although these cases are not very problematic, since many authors are trained to avoid such situations. Moreover, if this happens, authors usually acknowledge their error, and generally accept the request for a change, revising the citation to bring it in line with ethically acceptable ways of referring to the work of others.

Concerning the problem of incorrect, incomplete, or misquoted citations, in another editorial (Diniz, 2014), I drew attention to an article in the Times Higher Education journal that describes a situation in which the respected Polish intellectual, Zygmunt Bauman, the creator of the concept of "liquid modernity," reacted to an accusation of plagiarism by a doctoral student. According to the opinion editor of this journal, in his response to the indictment, Bauman claimed that "high-quality scholarship does not depend on obedience to technical rules on referencing", an opinion not supported by most journals and probably by the scientific community at large. Although Bauman also made clear that he "never once failed to acknowledge the authorship of the ideas or concepts that inspired the ones" he coined, this episode would not have happened without the existence of similarities detection tools.

There are other cases in which attrition between journals and authors may increase because of the use of these tools. It is very common, for example, to identify similarity in different texts by the same author. Authors usually reuse parts of their previous work, a practice known as "self-plagiarism," and do not generally consider this a wrong practice, as the text they reuse is their own. However, from the perspective of journals, this situation is not so simple.

Depending on the original source, the innocent and well-intentioned author's practice of reusing parts of his own text can cause problems for the journal. In less serious cases, when the original does not imply assigning copyright to a publishing entity, as in a conference paper, for example, many journals even take in account the similarity and adopt the understanding that in some cases previously published versions are only natural stages in the development of a paper.

It is different when the original source is associated to a copyright contract, which is common in most journals. In such cases, the contract between author and publisher usually limits its use, in whole or in part, by another publication. Thus, when this type of similarity is found, the journal is obliged to ask the author to rewrite the text to eliminate similar excerpts, under penalty of being sued for misappropriation. Authors who have difficulty understanding the separation between "authorship" and "ownership" of an intellectual work feel offended, as they believe that they have done nothing wrong. However, by signing a contract with the original publisher, the issue ceases to be purely ethical and becomes related to commercial law.

The extension of problems resulting from the identification of similarity goes beyond what has already been exposed. What if similarity is found in different papers published by the same publisher? Take as an example this article you are reading now. I set out to write about a topic I had previously discussed in *RAE* editorials, and I reported this in the first paragraph of this article. After being analyzed by *RAE's* similarity detection tool, I discovered that this article has a 9% similarity with the editorials I wrote myself and published in the same journal. Even if *RAE* does not wish to sue itself, there will be people claiming that I am "rehashing" old texts, because they would prefer to read new unpublished ones.

This is only a small sample of the enormous variety of situations that occur when a journal begins to check for similarity in its editorial processes. Aside from the origin of the text, many other issues are also considered in similarity assessment. One of these issues is the amount of similar content. How much is "acceptable": a sentence, a paragraph, a page, or half of the paper? Moreover, in which part of the text is the similarity

found: in the introduction, theoretical review, methodology or conclusions? All of these questions open doors to many situations that do not always have easy solutions.

Other issues emerge with the increasing use of automatic similarity detection tools. By exclusively verifying textual similarity, i.e., identifying similarities between a submitted text and others already published, these tools automatically check word by word, and are unable, for example, to compare published texts in different languages or even identical ideas written differently, something that humans are perfectly capable of doing.

A case that is worth reporting, and which illustrates the limits of these tools, is it was when a reviewer suspected similarities in an paper that had not been detected by the *RAE's* system. The reviewer requested further checks, and it was discovered that the identification included a large number of double quotations (such as Author 1, 1900; Author 2, 2000), which is quite unusual. The author could not escape the reviewer's watchful eye, despite not having been detected by the tool of translating excerpts verbatim from another paper, and copying even the references. The similarity was high but in another language, and the paper was not published.

As similarity detection tools continue to evolve, artificial intelligence capabilities will soon be incorporated as well. I suspect that the process will become more complex for journals, and they will have to adapt to an even more sophisticated typology of similarities.

RAE explains the whole process of tracking similarities in its guidelines (RAE Guidelines, n.d.), clearly stating that the report generated by the tool is used in addition to decision making on the acceptance of an paper. Following this process, the paper is evaluated by editorial staff before any communication is forwarded to the authors. Regardless of the results reported on similarities, the tone used in communication with the authors should never be one of accusation. It is always better to ask for clarification on the points where similarity exceeds reasonable limits, and give the author the floor.

CONTINUOUS LEARNING ON A COMPLEX AND EVOLVING SUBJECT

As has been affirmed from the beginning of this brief article, dealing with the issue of similarities is not a simple task, because it involves cultural issues in a constantly evolving technological environment. The mechanisms for scientific knowledge dissemination, including academic journals, are changing rapidly for technical and economic reasons.

Furthermore, the traditional perception of the evolution of science is that it rests on the pillars of knowledge that is freely shared, new concepts developed from a preexisting base, and community identification of what is new in a given area. Although these principles may look simple and obvious, their applicability is complicated due to economic, social, and even political factors.

As Professor Maggiolini (2014) teaches us, even ethics have been affected by the growing digitalization of society. The Bauman case mentioned previously is especially emblematic, because it is associated with a thinker who greatly contributed to the knowledge of our interconnected society.

Academic authors, pressed by the wave of productivism, use digital resources the best as they can to expand their impact on respective scientific communities. Organizations such as the Committee on Publication Ethics (COPE) help us to better understand the limits of what we can do and what we should not do to reach this goal, and everyone engaged in academic activity should be in touch with its principles of transparency and best practices for scientific publications (COPE, 2015).

To conclude, the key point is to avoid "copycat" behavior (whereby one copies ideas from others and presents them as one's own) on any of its possible levels. Let us use all available digital resources, endeavoring to always act ethically and avoid productivism.

REFERENCES

Committee on Publication Ethics. (2015). *Principles of transparency and best practice in scholarly publishing*. Retrieved from http://publicationethics.org/files/Principles_of_Transparency_and_Best_Practice_in_Scholarly_Publishingv2.pdf

Diniz, E. H. (2013). Ética e bom senso contra o produtivismo. *RAE-Revista de Administração de Empresas*, *53*(4), 331. doi:10.1590/S0034-75902013000400001

Diniz, E. H. (2014). Editorial. *RAE-Revista de Administração de Empresas*, *54*(4), 351.doi:10.1590/S0034-759020140401

Diniz, E. H. (2015a). Tradução automática, terceiro estágio da revolução digital dos periódicos. *RAE-Revista de Administração de Empresas*, *55*(2), 119. doi:10.1590/S0034-759020150201

Diniz, E. H. (2015b, Maio/Junho). Similaridade e plagiarismo: Novos desafios para a gestão de periódicos científicos. Editorial. *RAE-Revista de Administração de Empresas*, *55*(3), 239.doi:10.1590/S0034-759020150301

Jump, P. (2014). Zygmunt Bauman rebuffs plagiarism accusation. *Times Higher Education*. Retrieved from https://www.timeshighereducation.com/

Machado, A. M. N., & Bianchetti, L. (2011). (Des)fetichização do produtivismo acadêmico: Desafios para o trabalhador-pesquisador. *RAE-Revista de Administração de Empresas*, *51*(3), 244-254. doi:10.1590/S0034-75902011000300005

Maggiolini, P. (2014). Um aprofundamento para o conceito de ética digital. *RAE-Revista de Administração de Empresas*, *54*(5), 585-591. doi:10.1590/S0034-759020140511

RAE Guidelines. (n.d.). Rastreamento de similaridades. Retrieved from http://rae.fgv.br/en/manual-rae

The Visual Communication Guy. (2014). *Did I plagiarize?: The types and severity of plagiarism violations*. Retrieved from https://visual.ly/community/infographic/education/did-i-plagiarize-types-and-severity-plagiarism-violations