Journal of Psychosomatic Research xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Journal of Psychosomatic Research

Commentary

A generalized view of self-citation: Direct, co-author, collaborative, and coercive induced self-citation

John P.A. Ioannidis*

Stanford Prevention Research Center, Department of Medicine, Stanford University School of Medicine, Stanford, CA, USA
Department of Health Research and Policy, Stanford University School of Medicine, Stanford, CA, USA
Department of Statistics, Stanford University School of Humanities and Sciences, Stanford, CA, USA
Meta-Research Innovation Center at Stanford (METRICS), Stanford University, Stanford, CA, USA

ARTICLE INFO

Available online xxxx

ABSTRACT

The phenomenon of self-citation can present in many different forms, including direct, co-author, collaborative, and coercive induced self-citation. It can also pertain to the citation of single scientists, groups of scientists, journals, and institutions. This article presents some case studies of extreme self-citation practices. It also discusses the implications of different types of self-citation. Self-citation is not necessarily inappropriate by default. In fact, usually it is fully appropriate but often it is even necessary. Conversely, inappropriate self-citation practices may be highly misleading and may distort the scientific literature. Coercive induced self-citation is the most difficult to discover. Coercive Induced self-citation may happen directly from reviewers of articles, but also indirectly from reviewers of grants, scientific advisors who steer a research agenda, and leaders of funding agencies who may espouse spending disproportionately large funds in research domains that perpetuate their own self-legacy. Inappropriate self-citation can be only a surrogate marker of what might be much greater distortions of the scientific corpus towards conformity to specific opinions and biases. Inappropriate self-citations eventually affect also impact metrics. Different impact metrics vary in the extent to which they can be gamed through self-citation practices. Citation indices that are more gaming-proof are available and should be more widely used. We need more empirical studies to dissect the impact of different types of inappropriate self-citation and to examine the effectiveness of interventions to limit them.

Introduction

Self-citation is an interesting feature of scientific discourse. While the typical paradigm is direct self-citation from an author to his/her own work, there are many other forms where self-referential motives exist, often in a not easily recognizable manner. Table 1 presents four case studies that offer a sense of the wide breadth, variety, and potential impact of self-citation phenomena [1–8]. Here, I will try to review the different types of self-citation (Table 2) along with potential challenges that they create.

Direct self-citation

The classic type of self-citation is direct self-citation. The author cites his/her previous work in subsequent scholarly works. The prevalence of direct self-citations varies across authors, journals, scientific fields, countries, rank, and age of the scientists and it is relatively more prominent in the few years after the publication of a paper. There is already a rich literature on direct self-citation, e.g. see [9–19]. Thomson Reuters Web of Science readily generates citation counts for all citations and excluding direct self-citations.

Co-author and collaborative self-citation

In co-author self-citation, one or more co-author(s) of scientist Y write another paper without Y and that paper cites their common paper [20]. These are direct self-citations for the co-author(s), but the scientist of interest has not directly self-cited himself/herself. This process can take substantial dimensions, especially when there are many co-authors. There is some correlation between the number of co-authors and proportion of self-citations [21]. Scopus readily generates citation counts for all citations and excluding both direct self-citations and co-author self-citations.

A further possibility arises when scientists participate in large densely-connected collaborative teams that publish many papers within the framework of their collaboration, with only a subset of the investigators represented as authors in each of these papers. Each of these papers may cite many papers by the same collaborative team. For a given scientist, some of these citations will look as direct self-citations (scientist Y self-citing another paper where Y is an author), others will

DOI of original article: http://dx.doi.org/10.1016/j.jpsychores.2014.09.015.
* Corresponding author at: Stanford Prevention Research Center, 1265 Welch Rd, Medical School Office Building, Room X306, Stanford CA 94305, USA.

http://dx.doi.org/10.1016/j.jpsychores.2014.11.008
0022-3999/

Table 1
Some case studies of self-citation.

Case study 1. The 14th scientist in impact in Medicine in 2008–2012
According to Microsoft Academic Search, 2,066,208 scientists have published papers in Medicine in the last 5 years (2008–2012, given last database update in 2012). #14 in impact rank (h-index) is D.H. Roukos (http://academic.research.microsoft.com/RankList?entitytype=2&topdomainid=6&subdomainid=0&last=5). Independently verified, according to Scopus, that author has h-index = 68 and 7045 citations as of October 18, 2014. Typically for h = 68 one would expect a total of citations roughly 3–5 times the h^2 [1], i.e. 2–4 times as many citations. However, no paper is very highly-cited (the most-cited one has 165 citations in Scopus) and after the top-70 cited papers, citations for the other papers fall sharply. This citation pattern optimizes the h-index. The majority of the citations come either from a core of self-citing papers (137 self-citing papers) or from very few colleagues such as Ziogas D (trainee of Roukos, 69 mentor-citing papers) or Hottenrott C (mentor of Roukos, 51 mentee-citing papers) and a few others. The majority of these papers are letters or editorials/comments with very long lists of citations to Roukos. For example, the last English-language Scopus-indexed paper by Christ of Hottenrott in the 20th century was in 1994, then he published 5 more German-language Scopus-indexed papers in 1999–2002 and nothing that was Scopus-indexed in 2003–2007. Then apparently after 2008, he published within 6 years over 50 letters/comments that extensively cite Roukos. For example, the latest 3 published letters/comments of Hottenrott [2–4] cite 37, 32, and 32 items, respectively, of which 31, 26, and 25 (consistently over 80% of the cited references) are by Roukos. Overall, the recent letters of Hottenrott cite Roukos over a thousand times. The corresponding author e-mail for Hottenrott is the e-mail of a website launched by Roukos (info@gastricbreastcancer.com). Interestingly, Roukos papers typically do not cite Hottenrott papers.
Case study 2. One of the top-10 most-cited papers in 2010–2011 among papers published in 2009
The paper by the editor-in-chief AJS Coates on "Ethical authorship and publishing" was published in 2009 in the International Journal of Cardiology (IJC) and it has received 1615 citations by October 2014 in Scopus. It received 796 citations in 2010 and 587 citations in 2011, making it one of the top-10 most-cited papers across the entire scientific literature among the 2,246,377 papers published in 2009. Then it tapered to 128 citations in 2012, 8 in 2013, and 6 in 2014. The paper is a little over 1 page and it only contains the short statement of authorship that IJC is asking of all its authors to place in their papers along with citing that reference. 1555 of the 1615 citations (97%) are from papers published in IJC. Given that all IJC papers had to cite this article in the two years that count towards the impact factor calculation, IJC gained almost 1 point in impact factor by this paper alone. In 2010 a new version of "Ethics in the authorship and publishing of scientific articles" was publishing in the same journal and was cited mostly in 2011 and 2012 (668 citations-to-date). These are by far the two most-cited papers in the entire publication history of IJC. A newer version on "Statement on authorship and publishing ethics in the International Journal of Cardiology" was published in 2011 and was cited mostly in 2012 and 2013 (173 citations to-date). Opthof has estimated [5] that the 2010 impact factor of IJC increased by 57% from self-citations. The impact factor calculation is based on the citations received by papers published in the previous two years, thus as the journal self-cited paper was re-published every year, this maintained boosting of the impact factor.
Case study 3. The university with the highest number of highly-cited faculty in the world
According to the highly reliable database of Highly-Cited Researchers issued by Thomson Reuters in 2014 (www.highlycited.com), the university that has the highest number of highly-cited researchers among its faculty in the whole world is King Abdulaziz University in Saudi Arabia: 160 highly-cited researchers have declared this university as their primary or secondary affiliation, followed by 146 for Harvard, 97 for NIH, and 60 for Stanford [6]. King Abdulaziz would become second in rank only if all University of California campuses were merged (176 highly-cited researchers). The number of highly-cited faculty is one of the key criteria for ranking universities by the Shanghai system for world ranking of universities. Not surprisingly, given its outstanding performance in this metric alone, King Abdulaziz is ranked 10th in Mathematics in the world (better than MIT) and 38th in Chemistry (http://www.shanghairanking.com/World-University-Rankings/King-Abdulaziz-University.html). Bhattacharjee [7] has alerted the scientific community that Saudi Arabian universities offer highly-cited researchers financially lucrative contracts in which the researchers commit themselves to citing the Saudi Arabian university as one of their institutional affiliations in their publications or on highlycited.com.
Case study 4. Ike Ankara, a scientist more cited than Albert Einstein
In 2010, Cyril Labbe used the software scigen to generate 110 fictitious papers supposedly authored by Ike Antkare [8]. Each of these papers included self-citations to the other papers of Ike Antkare. As a result Ike Antkare obtained such a citation presence in Google Scholar that the software Scholarometer gave him an h-index of 94, making him one of the most cited scientists of all times, way ahead of poor Albert Einstein who only had h = 84 at that time [8].

look as co-author self-citations (some team scientist(s) authoring a paper without scientist Y and self-citing a paper where both this team scientist(s) and Y are authors), and some others will be neither (some team scientist(s) authoring a paper within the same collaboration framework without scientist Y and citing a paper where Y but not this team scientist(s) is an author). This latter category may be called collaborative self-citation.

Table 2
Classification of types of self-citation: a generalized view.

According to who cites whom and where
Direct self-citation
Co-author self-citation
Collaborative self-citation
Coercive induced self-citation
By peer-reviewers of single papers
By editors of journals
By peer-reviewers of grants
By scientific advisors of research agendas
By leaders of funding agencies
By institutions (listing of institutional affiliations)
According to beneficiary of self-citation
Single scientist
Groups of scientists
Journals
Institutions
According to nature of citing unit
Citing papers are genuine
Fabricated citing papers
According to appropriateness
Appropriate
Inappropriate

Collaborator networks can be anywhere from very small to very large. In case study 1 (Table 1), the citing collaborator network is very small and the published products of the collaboration network are primarily citation-loaded letters/comments. Large collaborator networks are becoming frequent in many domains. For example, a search for "European Prospective Investigation into Cancer and Nutrition" (EPIC) in title/abstract/keyword in Scopus (October 26, 2014) yields 978 papers by this prolific collaborative group. A total of 159 authors have authored at least 25 of the 978 papers. Only 9 authors have co-authored more than a quarter of the 978 papers (range 304–447 EPIC papers co-authored by these 9 scientists). EPIC papers unavoidably cross-cite previous EPIC papers. In physics, Thomson Reuters had to revise its rules for identifying Highly-Cited Researchers (www.highlycited.org) and to exclude papers with over 500 authors, because otherwise all highly-cited authors would be selected from the collaborator network centered at CERN where typically hundreds of authors appear in each paper.

Coercive induced self-citation

In coercive induced self-citation, the citing scientists are neither co-authors nor collaborators of the cited scientist [22,23]. They are induced to cite a paper with some degree of coercive pressure: they face potential negative consequences unless they cite the paper and/or they expect rewards if they cite it. Coercive induced self-citation cannot be detected

Please cite this article as: Ioannidis JPA, , J Psychosom Res (2014), http://dx.doi.org/10.1016/j.jpsychores.2014.11.008