Self-correction of science: a comparative study of negative citations and post-publication peer review

Frédérique Bordignon

Self-correction of science: a comparative study of negative citations and post-publication peer review

Frederique Bordignon
Ecole des Ponts, Marne-la-Vallée, France
frederique.bordignon@enpc.fr
0000-0002-4918-9137

Abstract

This study investigates whether negative citations in articles and comments posted on post-publication peer review platforms are both equally contributing to the correction of science. These 2 types of written evidence of disputes are compared by analyzing their occurrence in relation to articles that have already been retracted or corrected. We identified retracted or corrected articles in a corpus of 72,069 articles coming from the Engineering field, from 3 journals (*Science*, *Tumor Biology*, *Cancer Research*) and from 3 authors with many retractions to their credit (Sarkar, Schön, Voinnet). We used Scite to retrieve contradicting citations and PubPeer to retrieve the number of comments for each article, and then we considered them as traces left by scientists to contest published results. Our study shows that contradicting citations are very uncommon and that retracted or corrected articles are not more contradicted in scholarly articles than those that are neither retracted nor corrected but they do generate more comments on Pubpeer, presumably because of the possibility for contributors to remain anonymous. Moreover, post-publication peer review platforms, although external to the scientific publication process contribute more to the correction of science than negative citations. Consequently, post-publication peer review venues, and more specifically the comments found on it, although not contributing to the scientific literature, are a mechanism for correcting science. Lastly, we introduced the idea of strengthening the role of contradicting citations to rehabilitate the clear expression of judgment in scientific papers.

Acknowledgments

We are indebted to Joshua Nicholson from Scite for making citation polarity data available.

Keywords

In-text citations – Negative citations – Post-publication peer review – Retracted articles – Corrected articles – Self-correction of science

Article Highlights

- Negative citations are very uncommon.
- Retracted or corrected papers are not more contradicted than others in scholarly articles.
- Post-publication peer review platforms contribute more to the correction of science than in-text negative citations to papers.

Introduction

Science needs both innovation and self-correction. This has been ensured for centuries by academic controversy (*disputatio*[1]), in other words, free and contradictory discussion within scientific peers. Scientists should both be sceptical of any new claim and provide statements, hypotheses and theories that can be falsified (refuted) (Popper 1959). Science self-correction can result in the production of new knowledge with publications that rely on previous statements, oppose to them or just ignore them. It can also take a more drastic way with the official purge of the scientific literature and the withdrawal of what is recognized as false science. Retraction notices or errata are then published to correct the scientific record.

Our study focuses on the process of challenging past works and investigates 2 mechanisms: negative citations and post-publication peer review. Negative or contradicting citations can be found in scientific writing but post-publication peer review comments do not contribute to the scientific literature as such. Nevertheless, both target and cite a preexisting work by expressing a disagreement.

We have made the methodological choice to compare them by analyzing their occurrence in relation to articles that have already been retracted or corrected, in other words, publications whose contentious nature has already been acknowledged by a correction of the scientific record. This work is therefore at the boundary of 3 research topics: retracted or corrected papers, negative in-text citations, and post-publication peer review. It will address the following research question: are negative citations and post-publication peer review comments both equally contributing to the correction of science?

Background

Retracted and corrected papers

Retractions are a way to alert the scientific community (and beyond) that flawed research has been published. The problematic pieces of work are therefore supposed to be purged from the literature or at least flagged because they are partially or fully inaccurate. They should not be used to build new research. Retractions are reserved for circumstances in which significant portions of an article are incorrect or cannot be substantiated whereas errata are published when isolated inaccuracies have been identified (Furman et al. 2012).

In both cases, the reasons why a paper is retracted or corrected are various: fabrication or falsification of data and plagiarism are commonly agreed to be scientific misconduct, but honest errors can also lead to the correction of published literature. The frontier is sometimes hard to establish but the intention to deceive is the key aspect (Bar-Ilan and Halevi 2018; Fanelli 2009). (Bar-Ilan and Halevi 2018) coined the term "scientific distortion" to describe a category of articles including intentional and unintentional errors, both being considered hurdles for the advancement of science.

Although (Ioannidis 2005) claims that "most research findings are false for most research designs and most fields", retracted and corrected papers are very scarce. Data from the RetractionWatch[2] database show that the absolute number of annual retractions has grown over the past decade (from 100 annual cases before 2000 to nearly 1,000 in 2014) but even if the rate doubled from 2003 to 2009, it has remained stable since 2012 and only 4 of every 10,000 papers are now retracted (Brainard 2018). Although different from the ordinary process of self-correction of science (when new knowledge updates the old), retractions and errata are also a way to correct the scholarly record by changing the status of previously published works (Dougherty 2019).

[1] Formalized procedure of debate dating from the Middle Ages.

[2] RetractionWatch is a blog launched in 2010 by two science journalists and is dedicated to the tracking of retracted papers. In 2018, they released an online database of all the cases they have recorded: of all the cases they have recorded http://retractiondatabase.org.

The identification of retracted and corrected papers is difficult and despite COPE guidelines[3] and the Crossmark initiative[4], publishers and bibliographic databases do not adopt any standards and are not consistent in their reporting of cases.

Contradicting or negative citations

(De Solla Price 1963) discusses the phenomenon of cumulation of papers ("the way in which each paper is built on a foundation of previous papers then, in turn, is one of several points of departure for the next") and stated that "the most obvious manifestation of this scholarly bricklaying is the citation of references". A citation is, therefore, the explicit reference to another scientific piece of work within the full-text of a scholarly paper. And as (Gross et al. 2002) stated, from the late 20th century it has become a routine to conclude scientific articles with the list of references to past literature cited in the preceding text.

Even if it has been done widely for decades, merely counting citations is controversial (Cano 1989; Chubin and Moitra 1975; Kaplan 1965) and it is partly because "citation-in-context analysis" should also be taken into account (Small 1980). There is indeed a wide variety of reasons for a researcher to cite a previous piece of work, and it is obvious that citing is not always performed in a positive or supportive manner. In the early 1980s, (Garfield 1964) already discussed adding useful markers such as "critique", "data spurious" or "conclusions wrong" to describe the kind of relationship between the citing and cited documents. It is the same idea that is taken up by (Peroni and Shotton 2012) by proposing the CiTO ontology (Citation Typing Ontology): it provides the authors with the ability to capture their intent when citing previous works, as it allows them to add specific metadata to annotate a citation with its reasons.

Table 1 shows that whereas negative citations of all types are always mentioned as one or several specific categories in citation classification schemes, the studies that rely on them show consensually that they are very uncommon, most of the times representing less than 5% of all references cited in a document. This very low rate is quite surprising since, as (MacRoberts and MacRoberts 1984) stated, "criticism is the life blood of science". (Catalini et al. 2015) suggest that criticisms expressed through citations could also be part of the "falsification" process defined by (Popper 1959). But making a negative criticism towards another researcher requires a greater effort than making a perfunctory citation; it implies creating a context in which to justify an attack and explain why previous studies fell short (Rousseau et al. 2018; White 2001). It also means to take a risk by openly stating one's disagreement. Moreover (Catalini et al. 2015) have shown that the authors were more willing to criticize researchers located further away geographically and explain it may be socially costly to negatively cite the work of a local colleague.

Among those studies presented in Table 1, early ones aimed at describing the motives for citations and demonstrating that they were not equal. The corpora they analysed were quite small. But recently with the help of Natural Language Processing tools and Machine Learning, the automatic analysis of in-text citations relies on sentiment analysis techniques with schemas based upon basic polarity sentiment: positive, neutral and negative.

[3] Committee on Publication Ethics, http://publicationethics.org/files/retraction%20guidelines.pdf

[4] "Crossmark gives readers quick and easy access to the current status of an item of content. With one click, you can see if content has been updated, corrected or retracted and access valuable additional metadata provided by the publisher" (source: https://www.crossref.org/services/crossmark)

	Categories dedicated to negative or contradicting citations	Corpus description	Negative citation rate
(Garfield 1964)	Correcting one's own work; correcting the work of others; criticizing previous work; substantiating claims; disclaiming the work or ideas of others (negative claims); disputing priority claims of others (negative homage)	-	-
(Moravcsik and Murugesan 1975)	Confirmative–negational (and also evolutionary–juxtapositional)	30 articles (Phys. Rev. Jour.)	14%
(Chubin and Moitra 1975)	Partially negated, totally negated	443 papers (phys.)	5%
(Spiegel-Rosing 1977)	Cited source is negatively evaluated; Results of citing article disprove, put into question the data as interpretation of cited source; Results of citing article furnish a new interpretation/explanation to the data of the cited source.	2,309 citations	0.11%
(Oppenheim and Renn 1978)	Theory or method not applicable or not the best one	23	1.5%
(Small 1980)	Negative or refuted	-	-
(Brooks 1985)	Negative credit	26 researchers interviewed about their motivations for citing references	-
(Cano 1989)	Negational (according to (Moravcsik and Murugesan 1975) classification)	42 articles	2%
(Shadish et al. 1995)	Negative citations	283 articles	5%
(Wouters 1998)	Negational vs Confirmative	25,617 citations (soc. sci.)	2.83%
(Garzone and Mercer 2000)	Citing work totally disputes some aspect of cited work; Citing work partially disputes some aspect of cited work; Citing work is totally not supported by cited work; Citing work is partially not supported by cited work; Citing work disputes priority claims; Citing work corrects cited work; Citing work questions cited work	20 articles (phys. & biochem.)	-
(Teufel et al. 2006)	Weakness of cited approach; Contrast/Comparison in Goals or Methods (neutral); Author's work is stated to be superior to cited work; Contrast/Comparison in Results (neutral); Contrast between 2 cited methods	360 conference articles (comp. ling.)	4.1%
(Peroni and Shotton 2012)	Negative (corrects, qualifies, disagrees with, disputes, refutes, critiques, parodies, ridicules)	-	-
(Stremersch et al. 2015)	Negation (according to (Baumgartner and Pieters 2003) classification)	24,632 papers	0.6%
(Catalini et al. 2015)	Negative citations	15,731 articles (Journal of Immunology)	2.40%
(Bertin and Atanassova 2016)	Negational citations	3,601,842 sentences with in-text reference extracted (from 80,000 PLoS articles)	Ratio of negative citations through the IMRaD article sections

Table 1.Summary of literature on citation classification schemes and their focus on contradicting or negative citations

Citations of retracted and corrected papers

The studies about citations of retracted papers have mainly explored temporal characteristics, focusing on post-retraction rate of citations, and concluded that citations continue long after retraction date, mostly because authors are unaware of the new status of the document they once read and now cite (Bornemann-Cimenti et al. 2015; Garfield and Welljams-Dorof 1990; Kochan and Budd 1992; Korpela 2010; Mott et al. 2019; Oppenheim and Renn 1978; Teixeira da Silva and Bornemann-Cimenti 2017; Wouters et al. 2018). Some studies went further and classified citations received by retracted papers, but they examined only a few papers, sometimes focusing on particular high-profile misconduct cases (Bornemann-Cimenti et al. 2015; Gabehart 2005; Garfield and Welljams-Dorof 1990; Kochan and Budd 1992; Korpela 2010). There is variance in the percentage of negative citations they identify: from 0 to 15% (Gabehart 2005; Kochan and Budd 1992), and even up to 32% in (Garfield and Welljams-Dorof 1990), the unique study (Breuning case) not focusing on post-retraction period. When revisiting this study, (Korpela 2010) discovered negative citations were overrepresented but also signalled an upsurge of positive citations after several years. For these authors, the continuous citations of retracted articles after the publication of the retraction notice is a serious problem. But we believe that citations of these articles are problematic both before and after the retraction notice. And of course, it is even worse if the citation is supportive. The problem is indeed the dissemination of false statements and results at any time. What is more, some studies acknowledge that it is hard to determine a retraction date (Bar-Ilan and Halevi 2017; Neale et al. 2010), so comparative studies before and after retraction might not be accurate.

Post-publication peer review

Our study also investigates post-publication peer review comments as another type of traces left by scientists to contest published results. Traditional peer review is necessary to a sound process of publication and can be one of the mechanisms involved in the self-correcting nature of science. But peer review is unable to detect all problems, neither scientific flaws nor misconduct.

Comments or suggestions can be posted on academic social networks (e.g. ResearchGate, ScienceOpen) or with the help of dedicated tools (e.g. Hypothes.is). Some publishers also allow commenting on each publication on their website (e.g. PLoS, F1000) but most journals do not welcome correspondence or comments that criticize their publications (Barbour and Stell 2020). That is why post-publication peer review platforms have recently developed (Dubois and Guaspare 2019; Teixeira da Silva and Bornemann-Cimenti 2017). The most famous is PubPeer where comments can be made anonymously.

Although positive and praising comments can be made, they are in practice mainly negative or critical. Post-publication peer review venues are to be considered as tools to identify erroneous works that went through the traditional peer review process (Teixeira da Silva et al. 2017).

Methods and data

We introduce the idea to consider comments left on PubPeer as contradicting citations. Indeed, commenting on Pubpeer is the same principle as making a citation, with the consequence of creating a link between the comment and the target publication. However, what radically differentiates a comment on PubPeer from a citation in an article is that the former can be anonymous.

In order to evaluate negative citations in articles and comments on PubPeer as mechanisms for correcting science, we have built up a corpus of articles associated with retraction or correction notices. It is a way of having a material that has already been officially contested and corrected. From a methodological point of view, this saves us from having to individuate false science, and, therefore, it allows us to build a corpus of contested articles in an objective manner.

We used Scopus and the Web of Science to constitute a corpus of 72,069 articles (metadata), focusing on a particular discipline, on some particular journals and even individuals. We used 3 different criteria (discipline, journal, author) and constituted 7 datasets for comparison (Table 2):

- Discipline: we chose the Engineering field which seems to have not been the focus of as much attention as biomedical sciences. We retrieved the ISSN of all journals classified in the Engineering subject area according to the ASJC classification scheme[5] and made two queries to retrieve the corresponding articles from the Web of Science and Scopus databases, limiting the results to 2012-2015 publication years.
- Journal:
    - we chose 2 journals in biomedical field, *Cancer Research* and *Tumor Biology*, which are journals with high rates of retractions according to the RetractionWatch database,
    - we chose a multidisciplinary high-impact journal, *Science*, because some studies hypothesized that articles in high-profile journals are more prone to retractions (Fang and Casadevall 2011; Furman et al. 2012) and also because the most prestigious journals publish the least reliable science according to (Brembs 2018).

  We used their ISSN to retrieve articles from the Web of Science and Scopus databases, limiting the results to 2012-2015 publication years.
- Author: with 3 high-profile researchers whose substantial part of their work has been officially retracted or corrected: Sarkar, Schön and Voinnet. We made 3 queries and retrieved all published articles of these authors with no publication date limit.

The corpus compilation is long and painstaking because retractions and errata are uncommon. It is therefore, necessary to have a very important initial database to obtain a sufficiently representative corpus. And they are also difficult to identify and locate online (Hesselmann et al. 2017; Poworoznek 2003; Teixeira da Silva and Bornemann-Cimenti 2017) and in bibliographic databases. That is the reason why we used both Scopus and Web of Science to identify retraction notices and cross-checked with RetractionWatch database. To be as exhaustive as possible, we retrieved the erratum notices for the corpus we delineated, extracted the DOI or the article title from each notice title (see example), and then tagged the original articles. When the same article is associated to one or several erratum notices and is also retracted, we tagged it as retracted.

> Example of a title erratum notice: "Erratum: Recruitment of RNA polymerase II by the pioneer transcription factor PHA4 (Science (2015) 350 (aad5928) DOI: 10.1126/science.aad5928)"
>
> Example of a title retraction notice: "Retraction notice to "Preparation and characterization of vertically aligned ZnO microrods on glass substrate" [Mater. Lett. 97 (2013) 71-73]"

Then we used data from two websites to retrieve two kinds of critiques and enrich the corpus at article level:

- Scite[6] to retrieve the number of contradicting citations,
- PubPeer[7] to retrieve the number of comments.

Scite was launched in 2019. According to their founders, it is a platform that allows finding if a scientific article has been supported, contradicted, or mentioned by subsequent studies. Scite automatically extracts citations from papers and classifies them using deep learning models and a network of experts.
PubPeer was founded in 2012, it is a commenting website for centralized post-publication discussions; some of them have led to retractions or corrections.

We assumed that the number of contradicting citations retrieved by Scite and the number of comments posted on PubPeer are indicators of controversy and attempts to make a correction.
As we are interested in citations, we focus here on articles with at least 1 citation according to Scite database, that is 45,811 articles (the corpus is avalaible for download (Bordignon 2020)).

The corpus contains 1.28% of retracted articles (see Table 2 for distribution by sub-corpus).

[5] All Science Journal Classification, which is used to classify journals and conferences in Scopus
[6] https://scite.ai
[7] https://pubpeer.com

Scite returns 3 values (number of mentioning, contradicting, and supporting citations); we reduce each of them to the total number of citations found by Scite (not the number of citations identified by Scopus or the Web of Science). PubPeer returns a number of comments, we process them as contradicting citations and thus also reduce them to the number of citations identified by Scite.
Our objective here is not to test Scite's accuracy and we must assume that its detection of citation polarity, even if it is perfectible, gives a trend on which we can rely in this study. We also decided to group the retracted and corrected articles under a single label.

Corpus	Number of articles retrieved from WoS and Scopus (deduplicated)	Publication years	Number of articles with at least one citation in Scite database		
			Non-retracted/ corrected	Retracted/ corrected	Total
Engineering	60,919	2013-2015	38,232	278	38,510
Cancer Research	1,753	2013-2015	1,678	59	1,737
Tumor biology	2,860	2013-2015	2,518	98	2,616
Science	2,706	2013-2015	2,403	51	2,454
Sarkar	454	all	290	47	337
Schön	107	all	22	20	42
Voinnet	141	all	82	35	117
Total	68,936		45,225	586	45,811

Table 2. Corpus description (number of articles per corpus only considering those with at least one citation in Scite database)

Results and discussion

Negative citations

Consistently with previous studies, negative citations are very uncommon in the whole corpus: 0.29% (n=1904) on average citations per article. 97.02% are mere mentions and only 2.71% are supporting citations (Table 3). There are slight differences according to the corpus tested. In the Engineering domain, authors seem to be accustomed to citations of mere mentions, which is consistent with the rate found for Schön's papers, whose work is close to this domain.

	Contradicting		Mentioning		Supporting	
Corpus	Avg citation rate	Standard deviation	Avg citation rate	Standard deviation	Avg citation rate	Standard deviation
Engineering	0.23%	2.95%	97.50%	11.69%	2.30%	10.09%
Cancer Research	0.60%	2.52%	94.25%	6.94%	5.15%	6.43%
Tumor biology	0.88%	5.43%	93.40%	14.92%	5.73%	13.95%
Science	0.42%	1.45%	95.78%	5.07%	3.80%	4.77%
Sarkar	0.61%	1.95%	94.26%	7.29%	5.14%	7.14%
Schön	0.24%	1.54%	98.42%	6.32%	1.34%	6.18%
Voinnet	0.41%	1.18%	95.83%	5.31%	3.76%	5.08%
Total	0.29%	3.06%	97.02%	11.52%	2.71%	10.05%

Table 3. Average citation rate for each corpus and according to Scite citation type

On focusing on retracted or corrected articles, we aimed at checking whether papers that became officially known to be contentious generate more negative citations, either before or after they were flagged. But in fact, no such trend has been detected.

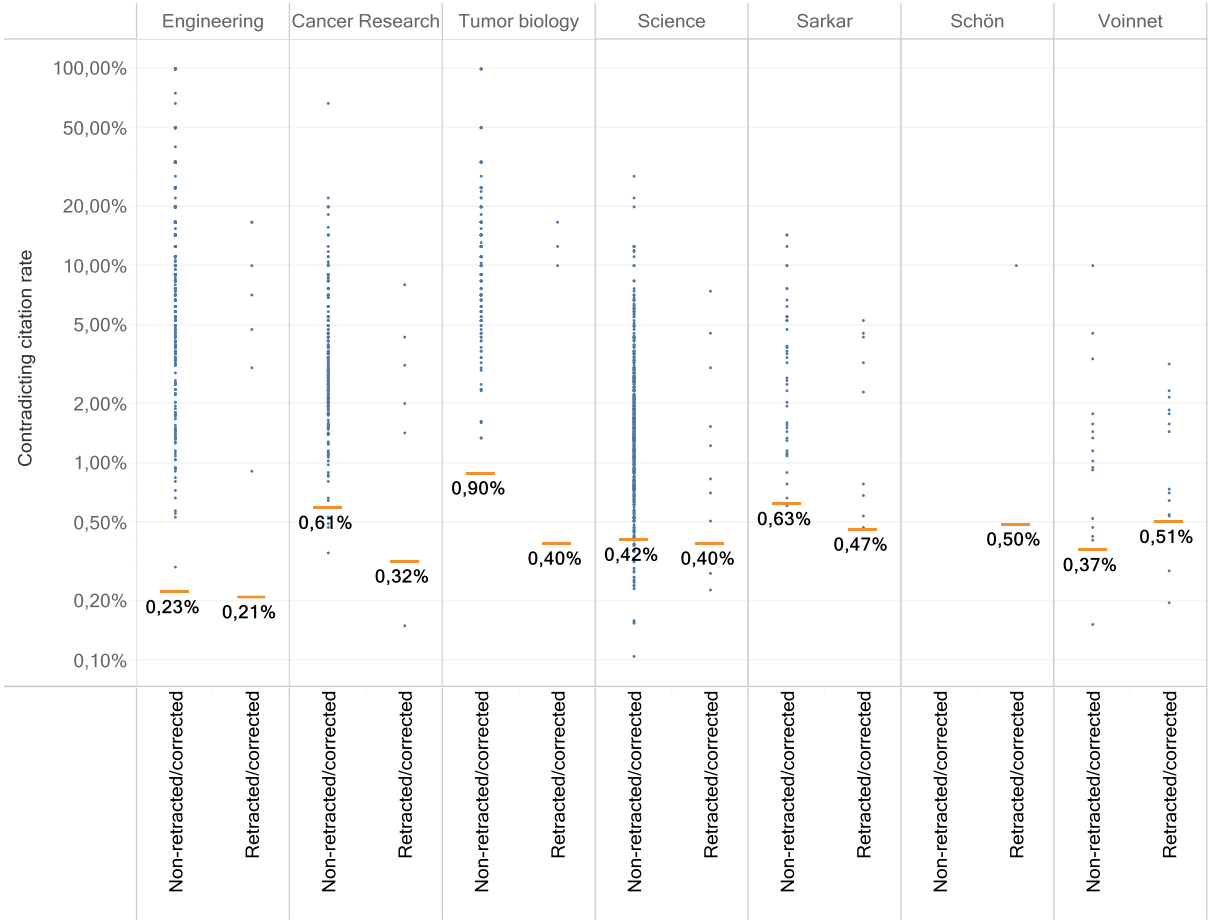

Figure 1. Average contradicting citation rates according to corpus and article status

Figure 1 shows the average contradicting citation rate for the 7 corpora and the difference between retracted/corrected articles and the others. There is no significant emerging trend. Nevertheless, here is what we can say:

- Aside from the Voinnet corpus (and the Schön corpus as well, but there is no reference point), retracted or corrected articles are less contradicted than "normal" articles. But for both Engineering and *Science* corpora, the rates are too close to each other to be interpreted as different.
- Corpora from individual cases are those for which the contradicting citation rate of retracted/corrected articles is the highest; however, the profiles of these 3 cases are completely different.

On the whole, these results show that negative citations are not more frequent towards retracted or corrected articles. Generally speaking, as they are too under-used by the authors, they do not have the impact they could have in the process of correcting science. Incidentally, the total absence of citations of papers that are supposed to be cited (because they are closely related to the topic) must be at least as significant an indicator.

Comments on PubPeer

Assuming that comments on PubPeer are a form of contestation very close to a negative citation, we tested the articles of our corpora in PubPeer to check if questioning the content of a publication is more frequent outside the usual publication process, and more particularly on a dedicated platform that allows anonymous comments. Table

4 shows the percentage of PubPeer comments received in respect to the total number of citations (in Scite database) and also displays the difference between retracted/corrected articles and the others.

	Retracted/corrected (n=586)	Non-retracted/corrected (n=45,225)	Total	Standard deviation
Engineering	4.79%	0.03%	0.06%	4.16%
Cancer Research	4.75%	0.16%	0.31%	3.14%
Tumor biology	0.00%	0.22%	0.21%	4.25%
Science	5.17%	0.34%	0.44%	7.28%
Sarkar	15.60%	1.81%	3.73%	13.40%
Schön	-	-	-	-
Voinnet	19.32%	3.13%	7.98%	23.49%
Total	5.80%	0.07%	0.15%	4.67%

Table 4. Average ratio of PubPeer comments to citations

Our data show that in general, corrected and retracted papers receive more comments on PubPeer than others. Furthermore, individual and mediatized cases (Voinnet and Sarkar corpora) generate the most comments on average in terms of the number of citations.

Comparison between negative citations and comments on Pubpeer

Lastly, Figure 2 shows no correlation between the rate of negative citations and comments on PubPeer and confirms PubPeer as a possible place where the debate and contestation of findings can be carried out.

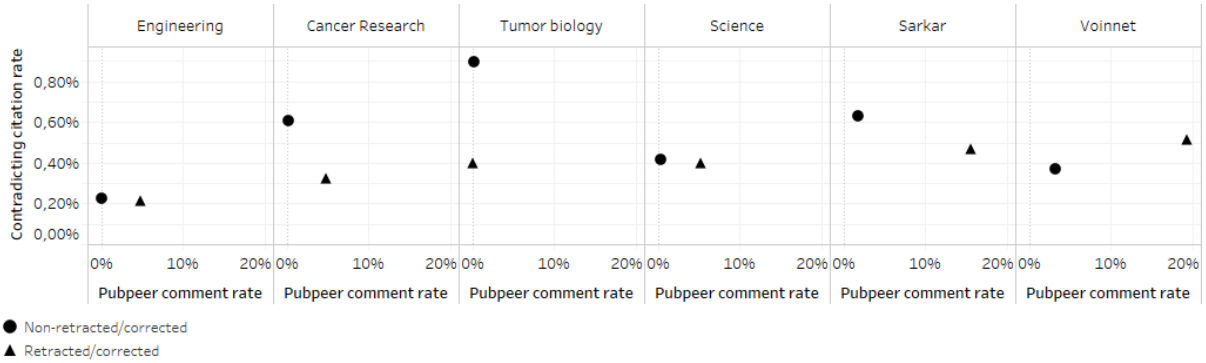

Figure 2. No correlation between contradicting citation rate and PubPeer comment rate

Our results show that Pubpeer, although external to the scientific publication process, contributes more to the correction of science than negative citations. The shallow analysis of some retraction notices also attests to this, since they often explicitly mention PubPeer as the place where the debate took place (e.g. "As raised on PubPeer, the article was found to contain images with signs of duplication and manipulation..."), thus feeding the decision to correct the record with a retraction or an official corrigendum. If negative citations are so uncommon, it is because it is difficult for an author to disagree publicly for fear of consequences. By allowing anonymity, Pubpeer overcomes this limitation, facilitating criticism and stimulating debate to improve the soundness of science. In some extreme cases, the scientific record is corrected through retraction or correction, but in other contexts, negative critiques at least provide a signal to the community that "scientific distortion" has been reported. It is essential that comments remain rigorously moderated by Pubpeer and that they stay "factual and verifiable", as stated in the platform guidelines. This is to avoid anonymous criticisms done with malicious intent in an attempt to weaken a competitor, as may be seen on non-scientific platforms, for example on consumer review websites (Wu et al. 2020).

Since our data show that negative citations are extremely scarce and supportive citations are hardly more frequent, this could lead us to question the interest in tagging the polarity of in-text citations, as suggested by (Garfield 1964) and made possible by the CiTO ontology (Peroni and Shotton 2012).
But when we consider the growing popularity of Pubpeer and the motivation of contributors to elaborate on the contradiction they provide, we could imagine on the contrary that journals should require a minimum ratio of contradicting citations to more easily see the benefit of an article in relation to previous works and to more easily identify "scientific distortions". This idea is compliant with the Leiden Manifesto (Hicks et al. 2015) decrying that performance evaluation is now led by the data rather than by judgment. It is also in the line of (Griesemer 2020)'s proposition to "reimagine a role for judgment in the face of the data-driven metrics". Contradicting citations could be seen as a first step or a mere contribution to the rehabilitation of judgment. Further work is needed to explore the implications of that suggestion so that it will not fuel new opportunities to game the metrics.

Conclusion and future work

Our study is consistent with previous works and shows that contradicting citations are very uncommon: 0.29% of citations in the whole corpus we tested. Retracted or corrected articles are not more contradicted in scholarly articles than those that are neither retracted nor corrected but they do generate more comments on Pubpeer, presumably because of the possibility for contributors to remain anonymous. Consequently, post-publication peer review venues, and more specifically the comments found on it, although not contributing to the scientific literature, are a mechanism for correcting science. Future work, including surveys involving authors, may confirm that it is the fear of expressing their disagreement formally that motivates researchers to contradict their peers anonymously on Pubpeer.

Our study deals with a twofold scarce material (retracted/corrected articles and contradicting citations), it is thus also necessary to confirm our results with other corpora and to take into consideration the reason for the retraction or erratum. Indeed, plagiarism will not trigger criticism like data manipulation or scientific mistakes would, and errata sometimes refer to errors that do not affect the soundness of science (e.g., an erratum published for an error in the bibliography). We, therefore, plan to use the classification of reasons developed by RetractionWatch in their database and to carry out automatic analysis of the texts of the retraction and erratum notices. Lastly, our study is dependent on the accuracy of Scite, a recent tool whose reliability was not the subject of our study but which will have to be measured to go further. In a future project, we aim to re-evaluate Scite results and also to explore the possibility of further refining its classification by proposing nuances in contradiction. The text-mining of the comments on Pubpeer will also enable us to better understand the expression of contradiction and to fuel future tools to detect it.
We have also cautiously introduced the idea of strengthening the role of contradicting citations to rehabilitate the clear expression of judgment in scientific papers. We will explore the implications of that suggestion in further studies so that it will not be used as a new way to game the metrics. In the meantime, the tools that allow leaving a comment straight at the article level or those that alert an article has been commented on any platform, are undoubtedly a means of increasing the expression of contradiction on the one hand and the openness to criticism on the other.

Reference list

Bar-Ilan, J., & Halevi, G. (2017). Post retraction citations in context: a case study. *Scientometrics*, *113*(1), 547–565. https://doi.org/10.1007/s11192-017-2242-0

Bar-Ilan, J., & Halevi, G. (2018). Temporal characteristics of retracted articles. *Scientometrics*, *116*(3), 1771–1783. https://doi.org/10.1007/s11192-018-2802-y

Barbour, B., & Stell, B. M. (2020). PubPeer: Scientific assessment without metrics. In M. Biagioli & A. Lippman (Eds.), *Gaming the metrics : misconduct and manipulation in academic research* (pp. 149–155). The MIT Press.

Baumgartner, H., & Pieters, R. (2003). The Structural Influence of Marketing Journals: A Citation Analysis of the Discipline and its Subareas over Time. *Journal of Marketing*, *67*(2), 123–139. https://doi.org/10.1509/jmkg.67.2.123.18610

Bertin, M., & Atanassova, I. (2016). Weak links and strong meaning: The complex phenomenon of negational citations. In *CEUR Workshop Proceedings* (Vol. 1567, pp. 14–25).

Bordignon, F. (2020). Data for: "Self-correction of science: a comparative study of negative citations and post-publication peer review." *Mendeley Data*. https://doi.org/10.17632/3gndhj3t6h.1

Bornemann-Cimenti, H., Szilagyi, I., & Sandner-Kiesling, A. (2015). Perpetuation of Retracted Publications Using the Example of the Scott S. Reuben Case: Incidences, Reasons and Possible Improvements. *Science and Engineering Ethics*. https://doi.org/10.1007/s11948-015-9680-y

Brainard, J. (2018). Rethinking retractions. *Science*, *362*(6413), 390–393. https://doi.org/10.1126/science.362.6413.390

Brembs, B. (2018). Prestigious science journals struggle to reach even average reliability. *Frontiers in Human Neuroscience*, *12*. https://doi.org/10.3389/fnhum.2018.00037

Brooks, T. A. (1985). Private acts and public objects: An investigation of citer motivations. *Journal of the American Society for Information Science*, *36*(4), 223–229. https://doi.org/10.1002/asi.4630360402

Cano, V. (1989). Citation behavior: Classification, utility, and location. *Journal of the American Society for Information Science*, *40*(4), 284–290. https://doi.org/10.1002/(SICI)1097-4571(198907)40:4<284::AID-ASI10>3.0.CO;2-Z

Catalini, C., Lacetera, N., & Oettl, A. (2015). The incidence and role of negative citations in science. *Proceedings of the National Academy of Sciences*, *112*(45), 13823–13826. https://doi.org/10.1073/pnas.1502280112

Chubin, D. E., & Moitra, S. D. (1975). Content Analysis of References: Adjunct or Alternative to Citation Counting? *Social Studies of Science*, 423–41.

De Solla Price, D. J. (1963). *Little science, big science.* Columbia University Press.

Dougherty, M. V. (2019). Correcting the scholarly record for research integrity. *Monash Bioethics Review*, *37*(1–2), 76–78. https://doi.org/10.1007/s40592-019-00093-x

Dubois, M., & Guaspare, C. (2019). "Is someone out to get me?" : la biologie moléculaire à l'épreuve du Post-Publication Peer Review. *Zilsel*, *6*(2), 164. https://doi.org/10.3917/zil.006.0164

Fanelli, D. (2009). How Many Scientists Fabricate and Falsify Research? A Systematic Review and Meta-Analysis of Survey Data. *PLoS ONE*, *4*(5), e5738. https://doi.org/10.1371/journal.pone.0005738

Fang, F. C., & Casadevall, A. (2011). Retracted Science and the Retraction Index. *Infection and Immunity*, *79*(10), 3855–3859. https://doi.org/10.1128/IAI.05661-11

Furman, J. L., Jensen, K., & Murray, F. (2012). Governing knowledge in the scientific community: Exploring the role of retractions in biomedicine. *Research Policy*, *41*(2), 276–290. https://doi.org/10.1016/j.respol.2011.11.001

Gabehart, M. E. (2005). *An analysis of citations to retracted articles in the scientific literature*. University of North Carolina, Chapel Hill, Master of Science thesis. Retrieved from https://ils.unc.edu/MSpapers/3050.pdf

Garfield, E. (1964). Can Citation Indexing Be Automated? In *Statistical Association Methods for Mechanized Documentation: Symposium Proceedings, Washington 1964. (National Bureau of Standards Mliscellaneous Publication 269)* (pp. 189–192). Washington, D.C., U.S.G.P.O. http://www.garfield.library.upenn.edu/essays/V1p084y1962-73.pdf. Accessed 26 June 2019

Garfield, E., & Welljams-Dorof, A. (1990). The Impact of Fraudulent Research on the Scientific Literature. *Journal of the American Medical Association*, *263*(10), 1424–1426. https://doi.org/10.1001/jama.1990.03440100144021

Garzone, M., & Mercer, R. E. (2000). Towards an Automated Citation Classifier. In *Canadian AI 2000. Lecture Notes in Computer Science, vol 1822* (pp. 337–346). Springer, Berlin, Heidelberg. https://doi.org/10.1007/3-540-45486-1_28

Griesemer, J. R. (2020). Taking Goodhart's Law Meta: Gaming, Meta- Gaming, and Hacking Academic Performance Metrics. In M. Biagioli & A. Lippman (Eds.), *Gaming the metrics : misconduct and manipulation in academic research* (pp. 77–87).

Gross, A. G., Harmon, J. E., & Reidy, M. S. (2002). *Communicating science : the scientific article from the 17th century to the present*. Oxford University Press.

Hesselmann, F., Graf, V., Schmidt, M., & Reinhart, M. (2017). The visibility of scientific misconduct: A review of the literature on retracted journal articles. *Current Sociology Review*, *65*(6), 814–845. https://doi.org/10.1177/0011392116663807

Hicks, D., Wouters, P., Waltman, L., De Rijcke, S., & Rafols, I. (2015, April 23). Bibliometrics: The Leiden

Manifesto for research metrics. *Nature*. Nature Publishing Group. https://doi.org/10.1038/520429a
Ioannidis, J. P. A. (2005). Why Most Published Research Findings Are False. *PLoS Medicine*, *2*(8), e124. https://doi.org/10.1371/journal.pmed.0020124
Kaplan, N. (1965). The norms of citation behavior: Prolegomena to the footnote. *American Documentation*, *16*(3), 179–184. https://doi.org/10.1002/asi.5090160305
Kochan, C. A., & Budd, J. M. (1992). The persistence of fraud in the literature: The Darsee case. *Journal of the American Society for Information Science*, *43*(7), 488–493. https://doi.org/10.1002/(SICI)1097-4571(199208)43:7<488::AID-ASI3>3.0.CO;2-7
Korpela, K. M. (2010). How long does it take for the scientific literature to purge itself of fraudulent material?: the Breuning case revisited. *Current Medical Research and Opinion*, *26*(4), 843–847. https://doi.org/10.1185/03007991003603804
MacRoberts, M. H., & MacRoberts, B. R. (1984). The Negational Reference: or the Art of Dissembling. *Social Studies of Science*, *14*(1), 91–94. https://doi.org/10.1177/030631284014001006
Moravcsik, M. J., & Murugesan, P. (1975). Some Results on the Function and Quality of Citations. *Social Studies of Science*, *5*(1), 86–92. https://doi.org/10.1177/030631277500500106
Mott, A., Fairhurst, C., & Torgerson, D. (2019). Assessing the impact of retraction on the citation of randomized controlled trial reports: an interrupted time-series analysis. *Journal of Health Services Research & Policy*, *24*(1), 44–51. https://doi.org/10.1177/1355819618797965
Neale, A. V., Dailey, R. K., & Abrams, J. (2010). Analysis of Citations to Biomedical Articles Affected by Scientific Misconduct. *Science and Engineering Ethics*, *16*(2), 251–261. https://doi.org/10.1007/s11948-009-9151-4
Oppenheim, C., & Renn, S. P. (1978). Highly cited old papers and the reasons why they continue to be cited. *Journal of the American Society for Information Science*, *29*(5), 225–231. https://doi.org/10.1002/asi.4630290504
Peroni, S., & Shotton, D. (2012). FaBiO and CiTO: Ontologies for describing bibliographic resources and citations. *Journal of Web Semantics*, *17*, 33–43. https://doi.org/10.1016/j.websem.2012.08.001
Popper, K. R. (1959). *The Logic of Scientific Discovery*. Routledge.
Poworoznek, E. L. (2003). Linking of errata: Current practices in online physical sciences journals. *Journal of the American Society for Information Science and Technology*, *54*(12), 1153–1159. https://doi.org/10.1002/asi.10320
Rousseau, R., Egghe, L., & Guns, R. (2018). *Publication and Citation Analysis*. *Becoming Metric-Wise*. https://doi.org/10.1016/b978-0-08-102474-4.00005-4
Shadish, W. R., Tolliver, D., Gray, M., & Sen Gupta, S. K. (1995). Author Judgements about Works They Cite: Three Studies from Psychology Journals. *Social Studies of Science*, *25*(3), 477–498. https://doi.org/10.1177/030631295025003003
Small, H. (1980). Co-citation context analysis and the structure of paradigms. *Journal of Documentation*, *36*(3), 183–196. https://doi.org/10.1108/eb026695
Spiegel-Rosing, I. (1977). Science Studies: Bibliometric and Content Analysis. *Social Studies of Science*, *7*(1), 97–113. https://doi.org/10.1177/030631277700700111
Stremersch, S., Camacho, N., Vanneste, S., & Verniers, I. (2015). Unraveling scientific impact: Citation types in marketing journals. *International Journal of Research in Marketing*, *32*(1), 64–77. https://doi.org/10.1016/J.IJRESMAR.2014.09.004
Teixeira da Silva, J. A., Al-Khatib, A., & Dobránszki, J. (2017). Fortifying the Corrective Nature of Post-publication Peer Review: Identifying Weaknesses, Use of Journal Clubs, and Rewarding Conscientious Behavior. *Science and Engineering Ethics*, *23*(4), 1213–1226. https://doi.org/10.1007/s11948-016-9854-2
Teixeira da Silva, J. A., & Bornemann-Cimenti, H. (2017). Why do some retracted papers continue to be cited? *Scientometrics*, *110*(1), 365–370. https://doi.org/10.1007/s11192-016-2178-9
Teufel, S., Siddharthan, A., & Tidhar, D. (2006). Automatic classification of citation function. In *EMNLP '06 Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing* (pp. 103–110). Sydney.
White, H. D. (2001). Authors as citers over time. *Journal of the American Society for Information Science and Technology*, *52*(2), 87–108. https://doi.org/10.1002/1097-4571(2000)9999:9999<::AID-ASI1542>3.0.CO;2-T
Wouters, P. (1998). The signs of science. *Scientometrics*, *41*(1–2), 225–241. https://doi.org/10.1007/BF02457980
Wouters, P., Costas, R., Franssen, T., Yegros-Yegros, A., Reyes, A., Suze Van Der Luijt-Jansen, E., et al. (2018). The Schön case: Analyzing in-text citations to papers before and after retraction. In *STI 2018 Conference proceedings*. Leiden. https://hdl.handle.net/1887/64521
Wu, Y., Ngai, E. W. T., Wu, P., & Wu, C. (2020). Fake online reviews: Literature review, synthesis, and directions for future research. *Decision Support Systems*, *132*, 113280. https://doi.org/10.1016/j.dss.2020.113280