

Chapter 5

Citations

As described in the Preface, the growth of scientific knowledge is predominately incremental—we build on past knowledge more often than we displace it. Thus, the first pillar of science—a communal collection of knowledge—requires mechanisms for preserving and disseminating knowledge within the scientific community. By far the most important mechanism in use today is the scientific publication. Although there are many forms of scientific publication, the two most common are the conference presentation (with or without some non-peer-reviewed written text) and the peer-reviewed journal paper (both in print and online).

Because virtually all scientific advances build on past knowledge, it is critical that the new work be placed in the proper context with respect to the past work upon which it builds. The primary mechanism for this is the citation (or reference). Within a scientific paper, references are placed to other works, creating points of contact with the communal collection of scientific literature in order to fit the new work into the web of knowledge. But given the skeptical attitude that is also a part of science, citations are also used to help readers verify the quality of the new work and assess the strength of its conclusions.

5.1 The Five Goals of Citations

A citation is, by definition, a reference to a source of information or data. Things that can be cited include journal articles, conference proceedings, books, student theses, newspapers, non-print sources (such as film or other recorded media), websites or other online resources, computer materials (such as a published CD-ROM of data or a piece of software), and personal communications. The citation should be located in the text in such a way that it is clear what material requires the citation. Often, this is at the end of a sentence, but sometimes it must be put in the middle of the sentence to enhance clarity. Obviously, citations must supply sufficient detail so that the referenced material can be found and uniquely identified. As such, every journal establishes a specific format for citations that must be followed. (Alas, there is no universal format that all journals follow.)

Though simple in concept, citations in a scientific paper serve many goals. The five most important goals are

- Provide sufficient context of the work to allow for critical analysis of the work by others and thus to enable the readers to gauge for themselves whether the author's conclusions are justified;
- Give the reader sources of background and related material so that the current work can be understood by the target audience (thus creating a web of science);
- Establish credibility with the reader (e.g., the authors knows the field, have done their homework, etc.) and/or inform the reader that the paper belongs within a specific school of thought;
- Provide examples of alternate ideas, data, or conclusions to compare and contrast with this work; and
- Acknowledge and give credit to sources relied upon for this work (i.e., acknowledge the use of another's ideas or data), thus upholding intellectual honesty.

Of these five goals, the most commonly mentioned is to give credit to others (the so-called normative theory of citations¹) and thus demarcate what credit is due the new work. In truth, scientists can have big egos. We are frequently motivated by the desire for peer recognition. Thus, we try to carefully stake a claim to new ideas or data in our paper, knowing full well that others will be checking to make sure we do not claim too much. Even so, there is no pretense that a list of references will provide a complete list of influences; such a list would be excessive in even the simplest of cases.

While important, the “give credit” purpose of citations is, in my opinion, less compelling than the other goals. I view citations, like all aspects of scientific writing, from a simple perspective: what best serves the needs of the reader? Thus, the primary goal of citations should be to help the reader gain the most from the paper. Imagine your paper being read by graduate students or postdocs: smart, but new to the field. If they read all of the citations, will they have enough background to understand your work? Will any of the references be unneeded or redundant (and thus a waste of the reader's time)? Chances are very good that a simple test will be sufficient to decide on most references: will adding this reference here make the paper more valuable to the reader, or less?

5.2 The Literature Search

A new research project almost always begins with a literature search, as discussed in Chapter 1. Thus, you should have a good idea what the key papers in the field are before you begin the research. This literature search should be updated during your research, especially as new ideas come or directions change. A review of your literature search results just before you begin drafting your manuscript will allow you to cite as you write. Also, other researchers are often working on similar topics and may have published papers after your original literature search was completed.

A common mistake is to save the literature search until the end of the paper writing process. Doing a literature search only at the end often generates spurious citations (a problem that will be discussed in Section 5.4) and rarely provides the most valuable citations.

5.3 Verify, Verify, Verify

One of the most pervasive problems with citations is that they are frequently incomplete or inaccurate. *It is the job of the authors to verify the accuracy of the references.* Editors, copyeditors, and reviewers are not responsible for reference accuracy and are not expected to check references for accuracy. And though copyeditors try to flag incomplete or improperly formatted references, it is the authors who must ultimately fix the errors found. Why not do the work up front to ensure that the references are complete, accurate, and properly formatted? It will only save time and effort in the end, and indicate to the editors and reviewers that you care enough to pay attention to these important details.

Alas, far too few authors take this advice seriously. Several studies have found that between 34 and 67% of references in a variety of medical and biomedical journals contained errors.² These errors can be broken down into major and minor errors. A major error means that the article could not be found given the information in the citation. One study found that major errors occur in 7% of the citations from one class of medical journals.³ Minor errors include punctuation or spelling mistakes, mistakes in the article titles, mistakes in the name and initials of the author(s), and citation style mistakes. These errors serve as irritants to the reader—they can still find the article, but they have to put more effort into it.

It is probably obvious that the main cause of errors in citations is simple sloppiness on the part of the author. There is another problem, however, that may also be at work: copying citations from other papers. In other words, some authors commit a cardinal sin of citations and add a reference without ever having read that paper. Copying citations from other papers without actually looking up and reading that paper can result in a propagation of errors that are never corrected⁴ (like a children's game of "telephone"). A slightly less egregious form is the abstract citation: citing a paper after reading only the abstract. Both types of unread citation should be avoided: cite only papers you have read.

5.4 Other Problems with Citations

There are other reasons why a specific reference does not fulfill the goals set out here and thus does not benefit the reader.

Spurious citations: citations that are not needed but are included anyway. These citations are sometimes added at the last minute, after the paper is written, to give the impression that a literature search and proper citation work have been done. They often include redundant citations, where the extra citations do not add any value beyond the first one. A simple example was given by Brian Thompson:⁵ "related work on the technique has been carried out by numerous researchers."^{1–101} The problem is obvious: an interested reader must wade through far too much

literature to get the needed background. Sometimes spurious cites are meant to give an impression of erudition by citing an obscure, historical reference (if the referenced work is in a foreign language, all the better).⁶ In all such cases, simply asking the question “If the reader looks up this reference, will it be time well spent?” will be enough to decide if that reference is spurious.

Biased citations: references added (or omitted) for reasons other than meeting the five goals of citations. Biases include overciting of friends’ or colleagues’ work, omitting cites to the work of rivals, and gratuitous citations in an attempt to curry favor with a boss or potential referee.

Self-cites: citations to one’s own work. There is nothing wrong with self-citations, *per se*. After all, the work represented in a single paper is often just the latest result of a larger ongoing project. As such, citations to one’s earlier work are often perfectly appropriate and sometimes required. Self-cites are a problem when they are either spurious or biased.⁷ Knowing as we do the tendencies of many scientists toward self-promotion, one fears that self-cites may be designed to boost the recognition of the author rather than increase the value of the paper to the reader.

Excluding contrary evidence: a form of biased citations where citations to prior work whose conclusions or data contradict the current work are omitted. Because one of the five goals of citations is to explicitly contrast the new work with prior work containing conflicting data or conclusions, avoiding such conflict (for whatever reason) does not serve the interest of science.

In the end, authors must find a balance between too many and too few citations. The literature base even on vary narrow topics is often vast, and it can be difficult to pick a small subset to cite. In general, authors can mitigate citation problems by asking two questions:

“Have I provided the references that will make this paper as useful as possible?”
“If the reader looks up a given reference, will it be time well spent?”

5.5 More on Self-Citations

Citations sometimes have significance for reasons other than the five listed above. Citations can be counted, and in a data-driven world these counts have assumed outsized importance as a proxy for the influence of a given paper. Citation counts serve as (flawed) measures of journal importance (the impact factor) and researcher clout (the h-index, among other metrics). Today, such citation counts and their metrification are used in hiring and promotion decisions, especially in academia, often as a substitute for thoughtful and informed judgment.

Be careful what you measure, because a truism of the business world is ‘what gets measured gets managed’. And measures that come with rewards often get gamed. When a person’s career or reputation depends on citation counts, the temptation to inflate those counts is never far away. Some authors are more likely to cite their colleagues’ work than their competitors’; some journals expect their submitting authors to preferentially cite work published in that journal. But the

easiest way to promote your own work (and thus yourself) is with the *self-citation*: a citation to one's own prior work.

Self-cites are not inherently problematic. Most scientific publications describe a part of a longer-term research effort, and self-citations can put the new publication in the context of that larger effort. Self-cites become a problem only when they are either spurious or biased. Because deciding that a specific citation is either spurious or biased requires a judgement based on the cited work, the paper in which the citation occurs, and the field within which the work resides, it is not always an easy evaluation to make. Some cases are obvious, as when a majority of a research paper's citations are to the author's own work in a popular field of research. Other cases are less obvious, as when the authors are nearly the only ones working on a very specialized topic. Still, I think most authors know when they are pushing into spurious or biased territory with their self-citations. So the best defense against abuse is self-regulation.

Or is it? An interesting study commissioned by the *Chronicle of Higher Education* looked at the role of gender in self-citation rates.⁸ An examination of 1.7 million *JSTOR* papers spanning disciplines and over 60 years found that nearly 10% of citations were self-citations. Further analysis showed that men were 56% more likely to cite their own work than women, with the gender disparity growing over time. Apparently, self-regulation of self-citations is more effective in women than men.

What is the cause of this gender disparity? Women in academia seem less inclined to self-promotion than men, probably to their detriment. Does society pressure women to be more "feminine" and modest about their accomplishments? Are men encouraged to be more aggressive in pursuit of career success? Do women work on smaller teams with fewer publications and fewer opportunities for self-citations? I am certainly not qualified to address such heady questions, but regardless of cause the issue of gender disparity in self-citations has consequences.

In the age of Big Data, success breeds success, and popularity snowballs. The most linked webpages, the most watched videos, and the most downloaded journal papers are "recommended" or promoted to website visitors and social media consumers, generating a handful of winners-take-all and a long tail of neglected also-rans. The bandwagon effect seems true in the world of academic citations as well. Could it be that even modest differences in self-citation rates might snowball into noticeable differences in total citations? In other words, does self-promotion through self-citation work?

One 2007 study showed that it does, with each self-citation multiplying into three other citations to that author over a five-year period.⁹ Further, the penalties for excessive self-citation seem to be small or none. Although this study looked at papers published from 1981–2000, I imagine that the higher levels of online searching and reading today have only increased this multiplying effect. Differences in self-citation rates are likely only one of many factors contributing to gender disparities in academic careers, but it may be one of the easier ones to address.

Proper citations require careful consideration of the appropriate goals of citations, aided by a simple ethos: *make the paper reader-centric, not author-centric*. Self-promotion is an author-centered way of looking at the activity of publishing, and it is neither good nor bad when considering the needs of the reader. Though self-cites should not be added to a paper solely for self-promotion, neither should self-cites be avoided for fear that they might appear self-promoting and thus unseemly. By focusing on the reader and the five proper goals of citations, most problems concerning citations can be easily avoided.

5.6 Conclusions

To do a good job of providing citations in a scientific publication, one must keep in mind the multiple goals of proper citing. But like other aspects of good scientific writing, a simple theme has emerged: *make the paper reader-centric, not author-centric*. Although it is common to choose citations that make the paper more valuable to the author (by demarcating what is novel, for example), good citations make the paper more valuable to the reader. Unfortunately, doing a good job of citing requires more work from the authors. But careful citing is worth the effort if your goal is a quality scientific publication.

References

- ¹ M. H. MacRoberts and B. R. MacRoberts, “Problems of Citation Analysis: A Critical Review”, *J. Am. Soc. Inform. Sci.* **40**(5), 342–349, (1989).
- ² M. Wright and J. S. Armstrong, “The Ombudsman: Verification of Citations: Faulty Towers of Knowledge?”, *Interfaces* **38**(2) 125–139 (2008).
- ³ A. E. Mohammad and D. M. Laskin, “Citation Accuracy in the Oral and Maxillofacial Surgery Literature”, *J. Oral Maxillofac. Surg.* **66**(1), 3–6 (2008).
- ⁴ M. V. Simkin and V. P. Roychowdhury, “Read before you cite!”, *Complex Systems* **14**(3), 269–274 (2003).
- ⁵ B. J. Thompson, “What is a Reference?”, *Opt. Eng.* **34**(7), 1861 (1995).
- ⁶ A. Lakhtakia, “Editorial: False Erudition”, *J. Nanophoton.* **3**, 039902 (2009).
- ⁷ J. Hartley, “To cite or not to cite: author self-citations and the impact factor”, *Scientometrics* **92**(2), 313–317 (2012).
- ⁸ R. Wilson, “Lowered Cites”, *The Chronicle of Higher Education* **60**(27), <http://chronicle.com/article/New-Gender-Gap-in-Scholarship/145311> Accessed 7/2/2015 (March 17, 2015).
- ⁹ J. H. Fowler and D. W. Aksnes, “Does self-citation pay?”, *Scientometrics* **72**(3), 427–437 (2007).