On the Heterogeneity and Classification of Author Self-Citations

Stephen M. Lawani
International Institute of Tropical Agriculture, P.M.B. 5320, Ibadan, Nigeria

The heterogeneity of author self-citations is highlighted and a systematic scheme for their classification is presented. Self-citations are either synchronous or diachronous and each of these classes or genera has four subclasses or species. The distribution of self-citations among the four species is governed by a number of factors including collaborative tendencies in the discipline or research specialty and the relative statuses of the collaborating authors. The classification of self-citations may be applied to study aspects of research collaboration and the matter of egotism in scholarly work.

Introduction

It has been remarked that the term self-citations is used by different people to refer to various forms of relationships between citing and cited articles [1]. Earle and Vickery [2] have defined subject field self-citation. They reported that of the references contained in social science sources, 58% belong to the social sciences; they described this by stating that the *self-citation* of the social sciences as a whole was 58%. The corresponding values for science and for technology were 70 and 81%, respectively [2]. Journal self-citation occurs when an article cites another article published in the same journal. The self-citation rates for two American agronomic journals, *Agronomy Journal* and *Crop Science*, were 52 and 60%, respectively, and for two French agronomic journals, *L'Agronomie Tropicale* and *Annales Agronomiques*, were 100 and 83%, respectively [3].

The term self-citation has also been used to describe the relationship between the citing paper and the cited paper when both originate from the same institution [4,5]. Most commonly, however, self-citation is used to describe the relationship between the authors of the citing and of the cited papers. This article deals exclusively with this kind of self-citation which one may call author self-citations.

Received August 17, 1981; revised December 29, 1981; accepted January 26, 1982

Citation analysis has become an important research technique with applications in various disciplines including sociology, history, and research policy [6]. There have been objections to certain applications, such as the use of citations to measure the quality of research productivity, because of the phenomenon of self-citations [6–8]. In other situations, self-citations are regarded merely as a "disturbing variable" [6], and presumably not worthy of careful study in their own right. And yet, self-citations, no matter how they are defined, constitute a large fraction of total citations. Tagliacozzo [1], who conducted the only systematic study of author self-citations *per se*, reported that they amount to 16.6% in plant physiology and 17.5% in neurobiology.

The objective of this article is to establish a framework for clarity in discussions of the subject of author self-citations. It highlights the heterogeneous nature of self-citations and presents a scheme for their systematic classification.

Two Genera of Self-Citations

Discussions of author self-citations encountered in the literature give the misleading impression that self-citations are homogeneous. This is not so. There are two classes of self-citations, and within each class, four categories are identifiable. Borrowing the terminology of taxonomy, we may say that there are two genera of self-citations and that each genus contains four species.

The two genera are synchronous self-citations and diachronous self-citations. An author's synchronous self-citations are those contained in the citations the author *gives*, whereas diachronous self-citations are those included in the citations an author *receives*. That is, when self-citations are considered along with the references listed by an author in his papers, they are synchronous but when the same are considered along with the citations made to the works of the author, they are diachronous. For example, there is a total of 11 references in this article; of these, three are this author's own works. The synchronous self-citation rate for this one article is therefore 27.3%. This article has not yet been cited by any-

 CCC 0002-8231/82/050281-04$01.80

body and therefore the diachronous self-citation rate for this article is as yet *undefined*.

In general, an author's synchronous self-citation rate is determined by considering all the papers he has published or coauthored, finding the number of his own papers listed in the references, and expressing this as a percentage of the total number of references in all the papers. To determine an author's diachronous self-citation rate, one needs to consult an appropriate citation index or a combination of citation indexes to find out how many times the works of the author have been cited. One notes how many of these citations were made by papers in which the author's name appeared and then calculates what percentage these self-citations constitute.

While an author may cite his own works, he normally also cites other persons' works as well. Therefore, a synchronous self-citation rate of 100% would be extremely rare. Every scholarly author has a defined synchronous self-citation rate which is either zero or a positive figure (the improbable exceptions would be those, if there are such, who habitually do not include *any* reference in their scholarly writings). On the other hand, the diachronous self-citation rates of many authors are *undefined* because their works have been cited neither by others nor by themselves. A 100% diachronous self-citation rate is conceivable but the incidence is low. In one study, the citations received by a random sample of 315 and a quality sample of 555 cancer research papers published in 1974 during the first five years following publication (1974–1978) were examined. Eleven of the 315 (3.5%) and only 4 of the 555 (0.7%) had 100% diachronous self-citations. All 15 papers had few citations; none had more than 3 [9].

As a further illustration of the distinction between synchronous and diachronous self-citation, the first four articles in the first issue of the *Proceedings of the National Academy of Sciences, USA* for the year 1977 (Vol. 74) were selected arbitrarily and the citations these papers *gave* were analyzed; the analysis yielded the synchronous self-citation rates for the four articles. The citation which these papers *received* during the first four years following their publication, i.e., 1977–1980, were ascertained by searching the *Science Citation Index* (SCI). From the SCI data, the diachronous self-citation rates were calculated. The rates for both genera of self-citations are displayed in Table 1.

From the limited illustrative data of Table 1, it is evident that a paper (or an author) may be characterized by the following relationships:

(1) a high synchronous self-citation rate *and* a low diachronous rate (e.g., articles 1 and 4),
(2) a low synchronous self-citation rate *and* a high diachronous self-citation rate (e.g., article 2),
(3) a high synchronous self-citation rate *and* a high diachronous self-citation rate (e.g., article 3).

Another possible combination, not revealed by the illustrative data, would be a low synchronous rate associated with a low diachronous rate. Each of these possibilities may have sociological significance worth investigating.

TABLE 1. Synchronous and diachronous self-citation rates for the first four articles of *Proceedings of the National Academy of Sciences*, Vol. 74, 1977.

Article	Synchronous Rate	Diachronous Rate (1977–1980)
1. Ogunmola, G. B.; Zipp, A.; Chen, F.; Kauzmann, W. *PNAS*. 74: 1-4; 1977.	Total of 22 citations of which 6 were self-citations, i.e., synchronous rate is 27.3%.	Total of 11 citations of which 1 was self-citation. Diachronous rate is thus 9.1%.
2. Hindman, J.C.; Kugel, R.; Svirmickas, A.; Katz, J.J. *PNAS*. 74: 5-9; 1977.	Total of 16 citations. No self-citation, i.e., synchronous rate is 0.0%.	Total of 17 citations of which 4 were self-citations. Diachronous rate is thus 23.5%.
3. Lee, S-H.; Jhon, M.S., Eyring, H. *PNAS*. 74: 10-12; 1977.	Total of 17 citations of which 6 were self-citations, i.e., synchronous rate is 35.3%.	Total of 1 citation which was a self-citation. Diachronous rate is 100.0%.
4. Van Wart, H.E. and Scheraga, H.A. *PNAS*. 74: 13-17; 1977.	Total of 24 citations of which 8 were self-citations, i.e., synchronous rate is 33.3%.	Total of 8 citations of which 1 was self-citation, thus Diachronous rate is 12.5%.
All four articles	Total of 79 citations of which 20 were self-citations, i.e., synchronous rate is 25.3%.	Total of 37 citations of which 7 were self-citations, thus Diachronous rate is 18.9%.

By making the distinction between synchronous and diachronous self-citations, certain sociological issues become immediately better understood. For example, take the issue of egotism. A high synchronous self-citation rate does not necessarily imply egotism whereas a high diachronous rate definitely does. A researcher's synchronous self-citation rate may be high but if he also is cited heavily by others, his diachronous rate would be low. Such a researcher is not an egotist. Indeed, a high synchronous self-citation rate coupled with a low diachronous self-citation rate may well suggest that the researcher concerned is a productive and key figure in his research specialty. Conversely, a researcher may have a low synchronous self-citation rate and yet a high, possibly 100%, diachronous rate. This would be a case of egotism. Sociological applications of the classification of self-citations need to be explored further.

The Four Species of Self-Citations

It was noted that each genus of self-citations has four species. The four species of the synchronous genus are exactly analogous to those of the diachronous genus. It is not, therefore, necessary to describe the species separately for each genus.

Species I is a self-citation in which the first-named author of the citing paper is also the first-named author of the cited paper. This may be called the classic author self-citation and it is probably the only type which most investigators consider.

Species II self-citation is one in which any of the coauthors of the citing paper is the first-named author of the cited paper.

Species III is a self-citation in which the first-named author of the citing paper is a coauthor of the cited paper. It is possible, however, to have a self-citation in which a coauthor of the citing paper is the first-named author of the cited paper (*Species II*) and the first-named author of the same citing paper is a coauthor of the same cited paper. To ensure mutually exclusive classes, the author recommends and adopts the convention that only self-citations in which the first-named author of a citing paper is a coauthor of the cited paper, *but* where the first-named author of the cited paper is *not* a coauthor of the citing paper will be described as *Species III* self-citation.

Species IV self-citation is one in which a coauthor of the citing paper is also a coauthor of the cited paper. Note that the conditions for Species I and IV, and Species II, III, and IV may hold concurrently. Therefore, again adopt the convention that a self-citation will belong to Species IV if and only if it is not Species I, II, or III.

Illustrative Data and Discussion

Collection of data on the individual species of synchronous self-citations is straightforward, but, because of the way existing citation indexes are arranged, collection of data on Species III and IV of diachronous self-citations are extremely laborious and time-consuming. The data presented here are for synchronous self-citations and are intended only to illustrate the relative abundances of the species.

A sample of 237 research papers in agronomy was selected from the 1979 literature. These papers had a total of 3469 authored references of which 511 had one or more authors in common with the citing papers. The distribution by species of the self-citations is shown in Table 2. The 237 papers had a total of 615 authors; thus, the *Collaborative Index*, defined as average number of authors per paper [9], of the agronomic literature was 2.59 (standard deviation was 0.24).

Another sample of 109 research papers was selected from the 1974 cancer research literature. The papers listed a total of 2432 authored items including 246 self-citations. The *Collaborative Index* for this sample was 3.43 (standard deviation was 0.37). The breakdown by species of the self-citations is shown in Table 3.

The relative abundances of the species of self-citations of any given genus depend on the collaborative tendencies of authors and the fields to which they belong. Single-author papers cannot have Species II or IV self-citations. A field characterized by low collaboration and high solo effort, such as the arts and the humanities, will tend to have high abundances of Species I self-citations and very low abundances for the other species. Cancer research is one of the most collaborative fields [10]; this is reflected by the relatively low percentage of Species I self-citations (2.71% of all citations), i.e., the classic self-citations in which the first-named author is the same as the first-named author of the citing paper. Agronomy, which is less collaborative, has a higher percentage of Species I self-citations. The greater the number of coauthors of a paper, the greater are the chances that it will refer to a paper in which one of the coauthors is the first-named author. That is, Species II self-citations increase with collaborative levels. Comparison of the data of Table 2 with those of Table 3 indicates that this is so.

It should be noted that the relationship between self-citations per paper and the number of authors per paper (the collaborative level) is a subtle one. When only run-of-the-mill papers are considered, there is no relationship [1]. However, when the diachronous self-citations of a random sample and two quality samples of cancer research papers were studied, statistically significant positive correlations between mean self-citations and collaborative levels were obtained for each of the quality samples [9]. But as in Tagliacozzo's study [1], the correlation for the random sample was not statistically significant [9]. Species II and IV self-citations are apparently higher for quality papers than for ordinary papers in a given research specialty.

Within a given discipline or research specialty, another factor that may affect the relative abundances of

TABLE 2. Species distribution of synchronous self-citations in agronomic literature.

Species	Species Total	Percent of All Self-Citations (N = 511)	Percent of All Citations (N = 3469)
Species I	220	43.05	6.34
Species II	152	29.75	4.38
Species III	50	9.78	1.44
Species IV	89	17.42	2.57
Total	511	100.00	14.73

TABLE 3. Species distribution of synchronous self-citations in cancer literature.

Species	Species Total	Percent of All Self-Citations (N = 246)	Percent of All Citations (N = 2432)
Species I	66	26.83	2.71
Species II	109	44.31	4.48
Species III	26	10.57	1.07
Species IV	45	18.29	1.85
Total	246	100.00	10.11

the four species of self-citations is the relative standing of the collaborating authors. In the master/apprentice type of collaboration [11], most of the self-citations are likely to be those of the "master." Whenever the "apprentice" is the first-named author, Species I and III will, in general, tend to be low while Species II and IV will tend to be high. Where the "master" is the first-named author, the relative abundances are reversed. In general, a field in which many authors are publishing for the first time and one marked by a high "author turnover" will have low Species I and III self-citations. The distribution of self-citations among the four species would probably tend to be even where the professional standings of the collaborators are comparable, unless there are wide differences in their productivities.

Summary and Conclusions

Author self-citations are heterogeneous. Any self-citation belongs to either of two broad classes or genera-synchronous and diachronous. Each class or genus consists of four subclasses or species. The two genera and the four species are described and illustrated with data from agronomic literature and cancer research.

Some of the factors which may affect the relative abundances of the species of self-citations are suggested. These include the extent of multiple authorships in the research area concerned and the relative standings of the collaborating authors. A comparative study of self-citations in cancer papers judged to be of high quality and in a random sample of cancer papers also indicate that quality may be a factor in the relative proportions of the four species [9].

A classification of self-citations in the manner suggested in this article has sociological applications. An obvious one is in the study of research collaboration. Another is in clarifying the issue of egotism and offering a possibility for its measurement. These applications should be explored further.

Acknowledgment

The author recognized the need to analyze and classify self-citations while on sabbatical and working on his doctoral dissertation at Florida State University, Tallahassee. Discussions with Professor Gerald Jahoda, his major advisor, and with Dr. Charles W. Conaway were most helpful in clarifying the author's ideas.

References

1. Tagliacozzo, R. "Self-citations in Scientific Literature." *Journal of Documentation*. 33: 251-265; 1977.
2. Earle, P.; Vickery, B. "Social Science Literature Use in the UK as Indicated by Citations." *Journal of Documentation*. 25: 123-141; 1969.
3. Lawani, S. M. "The Professional Literature Used by American and French Agronomists and the Implications for Agronomic Education." *Journal of Agronomic Education*. 6: 41-46; 1977.
4. Westbrook, J. H. "Identifying significant research." *Science*. 132: 1229-1234; 1960.
5. Wallmark, J. T.; Eckerstein, S.; Langered, B.; Holmqvist, H.E.S. "The Increase in Efficiency with Size of Research Teams." *IEEE Transactions on Engineering Management*. EM-20: 80-86; 1973.
6. Garfield, E. *Citation Indexing—Its Theory and Application in Science, Technology and Humanities*. New York: Wiley; 1979.
7. Lawani, S. M. "Citation Analysis and the Quality of Scientific Productivity." *Bioscience*. 27: 26-31; 1977.
8. Schaefer, C. W. "Citation Analysis" (Letter to the Editor). *Bioscience*. 27: 442-443; 1977.
9. Lawani, S. M. "Quality, Collaboration and Citations in Cancer Research: A Bibliometric Study." Ph.D. Dissertation, Florida State University, 1980.
10. Beaver, D. deB.; Rosen, R. "Studies in Scientific Collaboration, Part III. Professionalization and the Natural History of Modern Scientific Co-Authorship." *Scientometrics*. 1: 231-245; 1979.
11. Hagstrom, W. O. *The Scientific Community*. Carbondale, IL: Southern Illinois University Press; 1965: Chap. III.