Quantitative Measures of Communication in Science: A Study of the Formal Level

Michael H. MacRoberts and Barbara R. MacRoberts

Social Studies of Science 1986 16: 151
DOI: 10.1177/030631286016001008

DISCUSSION PAPER

• ABSTRACT

Studies of scientific communication have relied on citation indices and bibliographies for data. We examined papers to see how much influence appears as references in bibliographies. We found that very little does.

Quantitative Measures of Communication in Science: A Study of the Formal Level

Michael H. MacRoberts and Barbara R. MacRoberts

During the past two decades, references and footnotes have become a leading analytical tool in the sociology of science.[1] The invention of the *Science Citation Index (SCI)* is the main reason for this. Apart from the attraction of a large volume of data readily available at the flip of a page or the press of a button, the allure of citation analysis is that it is not only 'unobtrusive' but 'purely objective', requiring no 'personality appraisal or a reading of the works. ...'[2]

Beginning with the assumption 'that the research cited [referenced] by scientists in their own papers represents a roughly valid indicator of influence on their work',[3] citation analysis, co-citation analysis, and other bibliometric analyses are the logical outcome. As Garfield has put it: 'The *SCI* tells how each brick in the edifice of science is linked to all the others.'[4] He continues:

> In the near future a historian or sociometrist will be able to sit before a computer console and specify some starting point — a person, a word, a citation, or a place. He will then ask the computer to display a list of pertinent

Social Studies of Science (SAGE, London, Beverly Hills and New Delhi), Vol. 16 (1986), 151–172.

> papers. The computer will respond by drawing or displaying a historical road map which will show not merely a list of papers and books, but also a graphical approximation of the history of that subject.[5]

While citation analysts have discussed a number of problems associated with unqualified use of references as data,[6] as yet no one has undertaken a study of papers to determine what relation exists between what is evident as influence in a paper and what appears in the paper's bibliography, even though two decades ago Kaplan pointed out the necessity for doing this.[7] As recently as 1977, Garfield, after claiming that '... the vast majority of papers do properly cite the earlier literature', was able to say in the next sentence, 'unfortunately, there has never been a definitive study of this assertion. ... [I]t would be a formidable task to determine, in a large enough random sample, how often papers fail to cite accurately and comprehensively.'[8] Narin, in a detailed review of the various types of citation analyses, explicity points out the same deficiency,[9] and Smith, in a more recent review, gives no indication that this deficiency has been corrected.[10] Only a few analysts have paid any attention to the text of papers, and even when they have, their attention has been highly selective, involving little more than an examination of the immediate context of the citation itself.[11] Consequently, it would seem time to take the step of examining the assumptions of citation analysis against the events it hopes to describe.

Annotated Classification of Text-Bibliography Discrepancies

The purpose of this section is to classify and exemplify the types of discrepancies we encountered when comparing content of papers with their bibliographies — that is, the discrepancy between influence as evident in the text, and influence as captured in the bibliography.[12] Also we indicate how the more problematic of the discrepancies might be dealt with.

Basic Assumptions and Background Knowledge not Referenced

One of the major difficulties encountered by anyone entering a

new specialty is initial orientation. When first encountered, papers in a new specialty are more often perplexing than illuminating. The reason is simple: the author does not explain, much less reference, tacit assumptions and background knowledge that would orient the outsider. This is deemed unnecessary for the intended audience.

All scientific papers illustrate this point, so we need give only one example. The Coles published a paper entitled 'The Ortega Hypothesis' almost simultaneously with a book, *Social Stratification in Science*.[13] The paper, with minor exceptions, is Chapter 8 of the book. The book, to a certain extent, explicates the basic assumptions and background knowledge of the paper, but the paper, in its opening paragraphs, refers to no more than the book and three articles, also by the Coles. In other words, the reader is given little more orientation to the paper's context than to read these other works. Thus the key proposition of the paper — the functional theory of stratification — and the myriad individuals who were involved in its development, are not referenced.

For the citation analyst, this practice of not referencing basic assumptions and background knowledge presents a fundamental problem — he or she will be confronted, in the first paragraph of any paper, with the task of reconstructing an undetermined and largely undeterminable number of influences.

Superficial Influence not Referenced

In addition to tacit assumptions and knowledge, there is a great deal of other, more superficially influential work that is not referenced. We give a few examples.

Mendel's 1866 paper mentions five individuals. In the second paragraph, he says: 'Numerous careful observers, such as Kolreuter, Gartner, Herbert, Lecoq, Wichura, and others, have devoted a part of their lives to this problem with tireless persistence.'[14] In his discussion Medel again mentions Gartner, Kolreuter and Wichura, and their results, and in one place he says: 'Finally, the experiments performed by Kolreuter, Gartner and others. ...'[15] While we have no doubt that these individuals influenced Mendel, we also realize that Mendel used unnamed 'others'. We can speculate on who they were, and a number of people have, but we will probably never know for sure.[16] There has also been

speculation about the non-botanical influences on Mendel's work, notably from physics, mathematics, chemistry and Greek philosophy.[17]

Einstein's 1905 paper 'On the Electrodynamics of Moving Bodies', like Mendel's, has no references. The few individuals mentioned in it are cited, usually eponymously, in the text — for example, 'Maxwell's equations', 'Euclidian geometry', 'Cartesian coordinates', 'Newton's equations'. There is no mention of Michelson, Morley, Poincaré, Moriarty, Voigt, Larmor, Fizeau, Fitzgerald, Mach, Nobel, or many others who obviously influenced Einstein in the production of this work. Frank, one of Einstein's biographers, says:

> Einstein threw himself into the work of these classics of theoretical physics, the lectures of Helmholtz, Kirchoff, Boltzmann, the electricity theory of J.C. Maxwell and H. Hertz, and their exposition in the textbook of Abraham-Foppl. Einstein buried himself with a certain fanaticism day and night in these books from which he learned how one builds up the mathematical framework and then with its help constructs the edifice of physics.[18]

Can we assume that only Euclid, Descartes, Newton, Maxwell, Doppler, Hertz, Lorentz, and M. Besso represent a roughly valid indicator of influence on the production of this work?

McGervey gives another example:

> Consider a 1968 paper by Gell-Mann, Oakes and Renner. ... It cited 26 papers; all were by theorists. A check of the cited papers shows that almost all of the papers cited in them were also by theorists. The only references to experimental work were 'second generation' citations of books or review articles.[19]

The question for the citation analyst is how to handle these myriad influences. Persumably they should be added — but how many?

Textual Citations

A typical practice in scientific paperwriting is to cite textually but not in the bibliography. While the author is giving credit, such internal citations do not appear in bibliographies and the papers must be read to discover them. They never surface in citation indices.

Examples are common. We have already indicated several such

items in our discussion of Mendel's 1866 paper and Einstein's 1905 paper. We give one additional example.

Goudsmit, discussing the Coles' 'Ortega Hypothesis' paper, says:

> A striking example is the article by Edwin D. Becker and T.C. Farrar ... just preceding the article by Cole and Cole. It describes the basic features of Fourier transform spectroscopy. One gathers that its authors consider 'Fellgett's advantage' and the 'Jacquinot advantage' to be significant factors in this research technique, but the article carries no footnotes referring to Fellgett and Jacquinot. In fact, all experimental papers mention techniques without a reference to their origin. Scintillation counters and photomultipliers are generally used in experiments in nuclear and particle physics, but their inventors and the dozens of researchers who have improved these essential tools to their present perfection are rarely cited.[20]

Clearly, the analyst must add these to the bibliography.

Parsimonious Citing: The Covering Citation

Self-citing, citing review papers, citing inclusive literature (for example, 'see Smith [1980] for further literature'), citing techniques, and citing 'schools' or entire disciplines, all present problems for the analyst because of unspecified authority brought forward. Let us illustrate.

In the introduction to our monograph on Acorn Woodpeckers, we say: as the species has received only cursory attention in the past (Ritter 1938, Bent 1939), we designed the research to gather information on a wide variety of topics. ...'[21] Here we cite two reviews covering and supplanting fifty or so papers, many of which we had read and were influenced by. We do not cite them because it is more parsimonious to use the two reviews.

The Goudsmit example used above illustrates another aspect of this problem. Here we find techniques mentioned not only without referencing their inventors but without reference to the dozens of subsequent researchers who have improved them. In this category also fall such general references as 'neo-Darwinian theory', 'Mendelian genetics', and so on, that encompass the work of not dozens but hundreds of individuals.[22]

In these cases, what does the analyst do? If he is concerned with influence, then references must be added to bibliographies.

Evolved Meaning

Our method of directly comparing citations with cited items led us to the realization that a significant proportion of citations do not — to use an analogy with printing or photography — 'register' with the cited item. We found that registration ran from exact — author cited item correctly — to no registration at all — author credited cited individual with a view opposite to the one he held. At first we passed over these misregistrations as special cases, but it soon became evident from their frequency that here was a pattern that needed explanation.

One possibility was that the author simply misread the paper. This is obviously the explanation in some cases. Another was that the author intruded his interpretation on to that of the author. Undoubtedly this also happens. But neither of these explanations seemed adequate because it was obvious that many individuals were misreading and misinterpreting in the same direction for no apparent reason. This 'directed misregistration', we believe, derives from the fact that most scientists become aware of literature largely through informal channels and by means of secondary literature rather than by an independent assessment of the primary literature.

Mendel's work on peas provides a good case in point. As Olby, Brannigan, and a number of others have pointed out, almost all subsequent literature has been based, not on Mendel's original paper, but on a series of reinterpretations beginning with the 'rediscovery' in 1900.[23] This means that we are not directly influenced by Mendel but by Mendel and a host of intermediate and (usually) uncited reinterpretations.

Einstein's 1905 paper is another example. It remained uninfluential until some time around 1918 when, because of the author's increasing fame arising from the 1919 eclipse prediction, this paper became the object of intense scrutiny and interpretation, from which it emerged altered out of all recognition.[24]

Plato and Aristotle are two more examples. Neo-Platonism bears little resemblance to Platonism, and when St Thomas got through with Aristotle, the original was lost for a millennium.[25]

The question for the analyst is how much of a scientist's knowledge about his colleagues' work comes from secondary sources, such as review articles, and how much comes from interacting with the original work? The citation analyst can never

know which has occurred unless he descends to primary sources. When 'registration' is off, unknown and unnamed 'others' are implicated. How is the analyst to handle these in totting up citations? Does he add intermediate influences? Further, at what point does 'evolved meaning' cease to be considered 'direct influence'?

Ignorance of the Literature

Considering that most scientific papers go through a long process of maturation, involving informal discussion with colleagues, negotiations between co-authors, seminars, presentations at conferences and prepublication readings by colleagues in the same field, before they are formally reviewed by a number of assessors chosen by a journal editor, it is surprising how poorly integrated with the available literature many papers are. Whatever the researcher's initial ignorance of the literature may be, it should have been corrected by colleagues and expert reviewers by the time his paper is published, and probably much of it is corrected by these contacts. But if this is the case, then the scientist's knowledge of the formal literature is even more discrepant than the data from the formal (final publication) level indicate. Interesting here is a statement by Garvey and Griffith. They compared final published papers with their antecedent technical reports and found that in half of them 'the corresponding journal articles were typically better written and better related to other work in the same subject-matter areas'.[26] We give an example.

Gibson published a book entitled *The Ecological Approach to Visual Perception*. One book reviewer, while claiming that 'this is one of the most important books in psychology that has ever been written', points out that 'Gibson believes this is an entirely new viewpoint; but J.R. Kantor was making many similar points in a more integrated field approach (it could be called an ecological approach) in his *Principles of Psychology*, 1924–1926, and has continued to do so ... right up to the present.'[27] Another commentator points out that the ecological perspective of Gibson was 'definitely preceded' by elaborate presentations from 1920 on by J.R. Kantor.[28] Gibson does not cite this work yet he is no novice to the field, having published in the area for fifty years.

How is the citation analyst to handle ignorance of the literature

if citations are to be used as indicators of quality or contribution to science? Presumably if the author had been better informed, he would have cited relevant work or, if he knew that his work was a duplication of previous work, perhaps he would not have published at all. Does the analyst add references to the author's list, or does he eliminate the author's work altogether, on the grounds that ignorance cannot be used as a basis for judging quality or contribution?

Miscitation

Examples of miscitation are common, but it takes a knowledge of both the citing and cited work to detect them.

Brannigan's account of the pre-1900 citations to Mendel contains a number of such mistakes.[29] He says that 'Focke cited Mendel's *Pisum* work fifteen times. ...'[30] and attributes this to a paper by Glass. Glass, however, is innocent of this error. Glass says: 'Focke referred to Mendel's work ... fifteen times . ...'[31] Brannigan says that 'Focke's reference [to Mendel] was copied ... by Romanes and cited in a list of plant hybridists in an article entitled, "Hydridism", ... in the ... *Encyclopedia Britannica* ...', and cites Olby as the authority here.[32] However, Olby does not say that Focke's reference to Mendel was copied by Romanes — Romanes did not copy the reference; he simply mentions Mendel's name along with other names that he copied from Focke.[33] Brannigan attributes to Weinstein the contention that Jackson's reference to Mendel was 'lifted directly from the citation in Focke'.[34] Weinstein, however, explicitly says that Jackson's reference was not lifted from Focke:

> Focke's book bears the same date as Jackson's, 1881, but must have been published earlier than the *Guide* since it is listed there. It is listed, however, only in the Addenda, on page 492, so it could not have been the source of the Mendel citation, which is in the main part of the book, on page 100.[35]

What is the citation analyst to do with such miscitations? Presumably he omits them from the author's list of references.

Lifting a Reference without Consulting It

Scientists commonly include in their bibliographies references that they have not seen or that they have seen but have not read.[36] Such 'lifting' is of course hard to detect at the formal level and only becomes obvious when the copyist repeats an 'error' contained in a secondary source or attributes something to the original that is not in it.

Let us illustrate. From an examination of the dozen or so pre-1900 citations to Mendel's 'Versuche uber Pflanzen-Hybriden', it is evident that approximately half simply were copied from secondary sources. Bailey copied from the Royal Society Catalogue, Nageli and Peter copied from Besnard, Romanes copied from Focke, and so on.[37]

We found the same pattern when we examined citations to A.J. Lotka's 1926 paper 'The Frequency Distribution of Scientific Productivity'. Sociologists take their information on Lotka largely from Price's 1963 *Little Science, Big Science* and almost invariably discuss the work of Lotka and Price together. They usually cite only the first page of the Lotka paper, even when the bibliographic style of the journal in which they are publishing requires inclusive pages. (Price gives only the initial page of Lotka's paper.) While this, of course, is not proof that sociologists are lifting, it is suggestive.

When lifting is detected, how is the citation analyst to handle it? Obviously the original work is influential, but so is the secondary source from which it has been taken.

Referenced but not Influential

References are made to individuals whose work has actually not been influential. Instances such as this are common but they are hard to detect at the formal level of analysis. 'Ceremonial citations', 'paying of intellectual debts', and some 'perfunctory' citing fall into this category.[38]

In our 1976 paper on Acorn Woodpeckers, we reference a 1974 paper by Griffin and two papers by White. The former is a short paper that did not influence us. In the text, 'Griffin MS' is cited several times. We relied heavily on this unpublished, annotated plant list as well as personal contacts with Griffin. Because the list

was not published, the editor of *Ornithological Monographs* would not allow it in the bibliography. This is a general practice; much work circulated in manuscript form does not get into the formal literature. We added the 1974 paper of Griffin and the papers by White solely because they were 'something anyhow'.

Watson and Crick, in their first 1953 *Nature* paper, cite Fraser in this way:

> Another three-chain structure has also been suggested by Fraser (in the press). In his model the phosphates are on the outside and the bases on the inside, linked together by hydrogen bonds. This structure as described is rather ill-defined, and for this reason we shall not comment on it.[39]

This passage has all the marks of being an uninfluential reference. And we know that it is, but only because of Watson's later account of how it came to be included in the paper.[40]

Using Secondary Sources

As we have already mentioned, it is not uncommon to cite a secondary source instead of a primary one. For example, when Olby discusses Romanes' 1881 mention of Mendel, he does not reference Edwardson's 1962 paper, which brought this pre-1900 citation to the attention of other scholars. In turn, Brannigan uses Olby's account of Romanes' citation for his own account and also fails to mention Edwardson.[42]

Menard provides another good example. He says:

> In the sciences the quantity of output, if not the quality, can be determined by counting published papers. A quantitative measure of fame is the annual number of citations to each paper, and this can also be counted. The results of these counts are remarkably consistent. The number of authors producing n papers is proportional to $1/n^2$.[43]

Menard references Price's *Little Science, Big Science* for this material. Lotka is not mentioned, although he is the originator of this notion.

Krumland, Will and Gorry say:

> The high productivity of certain scientists caused Cole and Cole (1972) to postulate that 10 percent of all scientists produce 50 percent of all scientific papers.[44]

These authors reference the Coles instead of Lotka and Price, the originators of these statements.

A great deal of science is done from secondary sources. Presumably, the analyst must add primary sources to the author's list.

Skipping Intermediate Sources and Going to Primary Sources

Weinstein is one of the few Mendel scholars publishing in the English language who persists in going to original sources.[45] This is an admirable habit and should be commended. But for the analyst it presents problems because it leaves out intermediate sources of influence or use. For example, in Weinstein's paper mentioned above, we find no explanation of how he located the esoteric work of Schmalhausen, Besnard, Blomberg, and others. It is unlikely he did so by a diligent search of the nineteenth century literature; instead he probably used uncited intermediate sources. These intermediate influences are omitted from his bibliography, and the analyst presumably must reinstate them.

Wrong Person gets Credit in SCI

Because of our citing practices, it is often the case that the wrong person gets included in *SCI*. Let us give an example. Cole and Cole say: 'Lord Florey, a recent president of the Royal Society, expressed this point of view (5): "... Science is rarely advanced by what is known" ...'[46] When we consult reference (5), we read: '5. This quotation appears in J.G. Crowther, *Science and Modern Society* (Shocken, New York, 1968), p. 363.' When we consult *SCI*, we find that Crowther, not Florey, receives an entry in that source.

Obviously, the analyst must rectify these errors.

Informal Influences in Paper

While we are concerned with the formal level only, informal influences are evident in many papers. 'Personal communications'

as well as formal acknowledgements represent an author's way of getting around the difficulty of referencing influences from the informal level. Interesting in this regard is the Coles' remark:

> The extent to which unpublished work is being cited by leading journals is increasing rapidly, at least in physics. Second to articles in *Physical Review* ... private communications are the most-cited source of information in contemporary physics.[47]

The following paragraph in a paper by Stomps is another example of influence from the informal level.

> During the last Genetics Congress at Bellagio I got the impression, that outside Holland the true and exact story of the rediscovery of Mendel's work by Hugo deVries is not accurately known. I was even shocked by the utterance of one of the members that it appeared that in the beginning deVries had been dishonest and had tried to hush up the name of Mendel. Several members asked me to publish what I knew on the true situation through my personal acquaintanceship with deVries. For that reason I wish to say a few words here about the matter.[48]

Because these types of influences are not transferred to data bases such as *SCI*, the analyst must add these to the author's references, taking into account that some of them, notably acknowledgements, may be totally ceremonial.

Cited for Reasons Unrelated to Intellectual Content

A number of individuals have pointed out that many citations are made for reasons unrelated to the intellectual content of the cited paper. This paper is an example. Many of the papers we cite (internally as well as in the bibliography) are used to illustrate points not related to their cognitive content. Any number of substitute papers could have been used. Presumably the citation analyst must eliminate these from the authors' bibliography.

Duplicate Citing

It is often the case that an author breaks up his research into several papers and perhaps tops it off with a book in order to get as many publications as possible out of a piece of work.[49] This

practice often results in the same thing being said several different times with only a slight twist introduced to make each paper seem different. References such as 'Smith (1979a, b, c, d, e, f)' are frequently encountered, and when these are looked up, they all turn out to be essentially the same paper. An extreme example of this is the scientist who breaks up his thesis into chapters, publishes them, and simultaneously publishes the thesis. Brannigan apparently did something like this.[50] Chapter 6 in his 1981 book is his 1979 article almost verbatim. The Coles did likewise — their 1972 paper is Chapter 8 of their 1973 book.[51]

When the citation analyst discovers duplicate citing, presumably he must eliminate one or the other of the references.

Persuasive Citing

Gilbert has argued convincingly that authors cite work that their audience will regard as presenting valid or important arguments and results.[52] When faced with a number of papers of equal quality or usefulness, authors cite the ones they believe their audience will recognize and respect. This occurs in all fields. In biology, notably in the late nineteenth century, Darwin's name appears in almost every treatise whether or not his was the best work in the area. Einstein has similar status in the twentieth century and is cited even by individuals who can make no sense of his work. Each specialty has a coterie of such individuals who are cited in excess, largely because they have become symbolically linked to an idea or concept.[53] Predecessors and contemporaries whose work is equally meritorious are not cited. This process obviously inflates the citations of some individuals at the expense of others. In other words, the rich get richer.[54]

When an author cites in this manner, the analyst has some reconstructing to do — for example, to add all of those individuals who improved scintillation counters and photomultipliers mentioned by Goudsmit.

Numerical Results

The preceding should illustrate some of the problems encountered in analysing papers for influence. Before giving some numerical

results, let us analyse a passage in order to illustrate the procedure. Discussing the Mendel neglect issue, Lawani says:

> It is widely assumed that important contributions can go unnoticed. For example, it is commonly believed that Gregor Mendel's breakthrough paper in genetics was ignored (and therefore not cited) by the scientific community from the time it was published in the *Verhandlungen* in 1866 until it was 'rediscovered' in 1900. Zirkle (1964), however, suggested that Mendel's work was, in fact, not ignored. It was cited by Hermann Hoffmann in 1869 and by Blomberg in 1872. W.O. Focke cited it 15 times in his 1881 paper, and Mendel was mentioned in the article on hybridism in the ninth edition of the *Encyclopedia Britannica* (1881–1895). The *Royal Society Catalogue of Scientific Papers* (1800–1900) listed it, and L.H. Bailey included Mendel's paper in the bibliography of his 1894 book, *Plant Breeding*. Darwin's 1876 paper cited Hoffmann's 1869 work in which Mendel's paper was cited five times, but Darwin did not himself cite Mendel.[55]

Sentence 1. Lawani says that it is widely assumed that important contributions can go unnoticed but does not cite anyone, although presumably he has a number of individuals in mind. *Sentence 2.* Lawani does not tell us who believes that Mendel's paper was ignored, but presumably he was influenced by several individuals to make this statement. *Sentence 3.* Zirkle's 1964 paper is the only work referenced by Lawani in this passage, but if one reads Zirkle's paper, one finds that Lawani must either not have read it or read it only cursorily, because Zirkle emphatically and repeatedly says that Mendel was ignored. The second sentence of the paper reads: 'We know that his [Mendel's] work was ignored for 35 years. ...'[56] The closest he comes to Lawani's misinterpretation is to say that Mendel was not 'completely' ignored.[57] *Sentence 4.* Lawani does not reference Hoffmann's book or the paper by Punnett that brought the Hoffmann citation of Mendel to the attention of other scholars. Likewise, Blomberg is not referenced. The Blomberg citation was first mentioned by Larsson in 1915, but this fact did not reach the attention of English-speaking scholars until the 1960s, when Gustafsson repeatedly pointed it out. *Sentence 5.* Focke did not cite Mendel's 1866 paper 15 times. This is a mistake that Lawani apparently miscopied from Zirkle. Zirkle says: 'Focke mentioned Mendel 15 times in his *Pflanzen-Mischlinge*. ...'[58] To our knowledge, the figure '15' was first published by Roberts in 1929 and has been faithfully copied ever since.[59] It refers to the number of times Mendel's name appears in Focke. Concerning the reference to Mendel in the *Encyclopedia*

Britannica, we are not told that the article was written by Romanes and the Edwardson brought it to the attention of scholars although he was not the first to note it. *Sentence 6.* We do not know who first uncovered the fact that Mendel's papers are listed in the Royal Society Catalogue, but Lawani cites no one in this regard. L.H. Bailey did not include Mendel's papers in the bibliography of his 1895 book, *Plant Breeding*. To give the date of Bailey's book as '1894' is a common mistake, originating we know not where but sometime in the first part of this century and copied and recopied since. Lawani apparently copied it from Zirkle. *Sentence 7.* Punnett, in 1925, brought Darwin's citation of Hoffmann to the attention of scholars. Lawani does not mention Punnett, nor does he refer to Darwin in his bibliography.

One further point. A comparison of Lawani's passage with a similar passage in Garfield[60] leaves little doubt that Lawani used Garfield, but Garfield is not referenced in this context.

This is a brief and incomplete analysis, but it suffices to give the gist of the approach. Salient points should be noted. (1) While at least a dozen references would be required to do justice to the information contained in this passage, only one is given, and from what we can ascertain, only two sources were directly consulted, one of which is not cited. (2) Neither of these was the best source available at the time. (3) Information readily available when Lawani's article was written was not consulted. (4) The number of mistakes suggests that Lawani is unfamiliar with the subject. (5) Lawani misinterprets the only paper he cites. (6) Works mentioned in the text were probably not consulted, nor do they appear in Lawani's bibliography.

Ideally, one should 'randomly' select fifty or a hundred papers in some field and analyse them for total influence. This procedure, however, proved impossible to execute. A paper is not a simple item. Although it may be in a subject with which the analyst is very familiar, some part or parts of it will almost invariably encompass material that is outside his expertise. Because of this, even confining his analysis to one discipline, the analyst would have to master subject after subject as he proceeded through the sample. Nevertheless, considering this limitation, we attempt to give a *very preliminary* numerical analysis of a few papers in order to gain some idea of how much influence is captured in bibliographies.

We analysed fifteen randomly selected papers in the history of genetics (a subject which we have been studying recently)

published since 1950 and found from zero (paper had no footnotes or references) to about 64 percent influence captured in references and footnotes (Table 1). That is, after reconstructing bibliographies, we estimated that these fifteen papers taken together require some 719 references at a minimum to cover their content, when in fact they contain only 216 references. This is a coverage of about 30 percent for the entire sample. We doubt that further analysis in this field will appreciably alter this figure. Of the thirty-one additional papers on this subject we have in our file, four are heavily referenced and seven are without any references. Additionally, of thirty-four papers specifically concerned with Gregor Mendel, only eighteen (53 percent) actually reference any of his papers.

TABLE 1
Percentage Influence Captured in Bibliographies/Footnotes

Paper	Corrected References in Paper	References Needed to Cover Information in Text	Percent Captured in References
I	18	33	55
II	12	35	34
III	1	32	3
IV	0	56	0
V	23	58	40
VI	1	21	5
VII	8	69	12
VIII	6	24	25
IX	53	83	64
X	19	48	40
XI	17	62	27
XII	10	21	48
XIII	20	47	43
XIV	0	12	0
XV	28	118	24
Total	216	719	

Although such an analysis, as it must, involves 'subjective' decisions at every step — hence, no two analysts will duplicate when using the same sample — we believe that these figures probably reflect the correct general magnitude for the subject studied. Also, from our examination of papers in other fields, we

believe that these figures will not be far out for other disciplines; they even may be a little high, for in historical research, some individuals tend to be very scholarly.

But our point is not to affix a firm numerical value representing the history of genetics or any other discipline, but to indicate that if anyone was to conduct a citation analysis in this research area using bibliographies, he would lose approximately 70 percent of the influence at the outset.

Finally, it should be kept in mind that we have obtained this figure from an examination of the formal level of scientific communication only — that is, what Edge calls the 'tip of the iceberg' of scientific communication.[61] Edge has already argued convincingly that in science little of the influence at the informal level of communication finds its way into formal bibliographies, not only because these influences are so many and so easily forgotten but because the formal cannot accommodate them. If we consider the informal level to be as important as the formal, our figure is cut in half: that is, we would estimate that only about 15 percent of the influence on a paper is contained in its references — a very small figure indeed considering the assumption underlying citation analysis.

Discussion

Because our study does not confirm one of the major assumptions underlying citation analysis, it affects all types of citation studies as currently conducted which rely exclusively on references for data (for example, stratification studies, co-citation analysis, 'historiographs').[62]

Taken alone, our findings call into question the practice of culling references from bibliographies and using these as data. The mere presence of a reference is not a marker of influence, nor is the absence of a reference evidence that it is uninfluential. References are simply obvious historical leads and evidence of influence only when they have been demonstrated to be so.

This does not mean that citation analysis may not have its purposes, but it does mean that if it is to be taken seriously, investigators must first descend to the documents from which these data are derived in order to reconstruct influences before proceeding further. This in turn means that the analyst must be

more expert than the scientists whose communication practices he is studying, for he is going to have to add to, subtract from, and otherwise reconstruct original bibliographies to fit actual events. At each step he must also make 'subjective' judgements regarding such things as which references to keep and which to eliminate. Even so, it will never be possible to reconstruct all influences with any certainty, if for no other reason than that scientists — like everyone else — forget who influenced them.

This place the sociologist of science exactly where the historian has always been — immersed in the literature of a small specialty. Gone is the type of study that came hard on the heels of *SCI*, in which a lone sociologist believed he could tackle the entirety of high-energy physics simply by feeding entries from *SCI* into a computer.

Ultimately our study raises questions about the relationship among the events an investigator desires to study, the data he uses to substitute for those events, and the conclusions that he draws from the data. In the present case, the events to be examined were 'influence' or 'communication' among scientists and the relative 'quality' of scientific work. The data that were substituted for those events were numbers of references in bibliographies or in *SCI*. The conclusions were drawn from the data, not from the events. At no point was the event-data-interpretation continuum checked to determine whether or not the data actually reflected — could be substituted for — the events.

The same disparity between events and data has characterized other quantitative studies in the sociology of science — for example, studies of productivity as measured by numbers of 'papers'.[63] Here we find entries in abstracting services substituting for events (papers published), with no determination of what relationship exists between papers produced and entries in abstracting services: that is, the investigator does not begin with the events in question — the individual's publication list or, better, his actual papers — but with substitutes that he merely assumes stand in an appropriate relationship with those events.

The problem seems to derive from an unwillingness to descend to events in order to see whether data accord with them. This is understandable from the standpoint of the sheer magnitude of work this would involve — in the present case, papers have to be read and analysed — but if influence is to be determined, it is a necessary step.

Postscript

After completing this paper, we analysed it. Not more than 20 percent of the influence is captured by the references. Also, the work that most influenced us is not mentioned because it is unknown to the intended audience and thus would not be persuasively valuable or symbolically meaningful. In fact, because it is not known, it might detract from the paper's persuasiveness.

• NOTES

1. F. Narin, *Evaluative Bibliometrics: The Use of Publication and Citation Analysis in the Evaluation of Scientific Activity* (Cherry Hill, NJ: Computer Horizons, 1976); L.C. Smith, 'Citation Analysis', *Library Trends*, Vol. 30 (1981), 83–106; A.L. Porter 'Citation Analysis: Queries and Caveats', *Social Studies of Science*, Vol. 7 (1977), 257–67; see also J.S. Long, R. McGinnis and P.D. Allison, 'The Problem of Junior-Authored Papers in Constructing Citation Counts', *Social Studies of Science*, Vol. 10 (1980), 127–43.

2. E. Garfield, 'Citation Indexing for Studying Science', *Nature*, Vol. 227 (15 August 1970), 669–71, at 669 and 671; see also Smith, op. cit. note 1.

3. J.R. Cole and S. Cole, 'The Ortega Hypothesis', *Science*, Vol. 178 (27 October 1972), 368–75, at 370. See also B.C. Griffith, M.C. Drott and H.G. Small, 'On the Use of Citations in Studying Scientific Achievements and Communication', *Society for Social Studies of Science Newsletter*, Vol. 2 (1977), 9–13; D. Edge, 'Why I am not a Co-citationist', ibid., 13–19; E. Garfield, *Citation Indexing — Its Theory and Application in Science, Technology, and Humanities* (Philadelphia: ISI Press, 1979); Garfield, *Essays of an Information Scientist*, 5 Vols. (Philadelphia: ISI Press, 1979–83).

4. Garfield, op. cit. note 2, 669.

5. Ibid., 670.

6. D. Edge, 'Quantitative Measures of Communication in Science: A Critical Review', *History of Science*, Vol. 17 (1979), 102–34. See also Smith, op. cit. note 1, for review and references.

7. N. Kaplan, 'The Norms of Citation Behavior: Prolegomena of the Footnote', *American Documentation*, Vol. 16 (1965), 179–84.

8. Garfield, *Essays* op. cit. note 3, Vol. 3, 217.

9. Narin, op. cit. note 1, 78–79.

10. Smith, op. cit. note 1.

11. D.E. Chubin and S. Moitra, 'Content Analysis of References: Adjunct or Alternative to Citation Counting?', *Social Studies of Science*, Vol. 5 (1975), 423–41; E. Shearer and M.J. Moravcsik, 'Citation Patterns in Little Science and Big Science', *Scientometrics*, Vol. 1 (1979), 463–74; H.G. Small, 'Cited Documents as Concept Symbols', *Social Studies of Science*, Vol. 8 (1978), 327–40; M.H. MacRoberts and B.R. MacRoberts, 'The Negational Reference: Or the Art of Dissembling', ibid., Vol. 14 (1984), 91–94.

12. We are obviously not defining 'influence' in the most general sense imaginable but are using it precisely in the manner used by citation analysts. For example, we do not consider it 'influence' if a physicist uses geometrical analogues to solve mathematical problems because twenty-five years earlier in high school he had an excellent geometry teacher, while we do consider Mossbauer to have influenced his colleagues and hence, the course of physics, if the Mossbauer Effect is used by other physicists.

13. Cole and Cole, op. cit. note 3; J.R. Cole and S. Cole, *Social Stratification in Science* (Chicago: The University of Chicago Press, 1973).

14. Quoted in C. Stern and E.R. Sherwood, *The Origin of Genetics* (San Francisco, Calif.: W.H. Freeman, 1966), see 1.

15. Ibid., see 44.

16. C. Zirkle, 'Gregor Mendel and His Precursors', *Isis*, Vol. 42 (1951), 97–104; H. Iltis, *Life of Mendel* (New York: Hafner, 1966); H. Stubbe, *History of Genetics* (Cambridge, Mass.: MIT Press, 1972).

17. Iltis, op. cit. note 16; F. Monaghan and A. Corcos, 'Possible Influences of some 19th Century Chemical Concepts on Mendel's Ideas about Heredity', *Journal of Heredity*, Vol. 77 (1983), 297–99; E. Mayr, 'The Recent Historiography of Genetics', *Journal of the History of Biology*, Vol. 6 (1973), 125–54; H. Kalmus, 'The Scholastic Origins of Mendel's Concepts', *History of Science*, Vol. 21 (1983), 61–83.

18. Quoted by G. Holton, *Thematic Origins of Scientific Thought* (Cambridge, Mass.: Harvard University Press, 1973), 210.

19. J.D. McGervey, 'Citation Analysis', *Science*, Vol. 183 (11 January 1974), 28–30, quote on 30.

20. S.A. Goudsmit, 'Citation Analysis', *Science*, Vol. 183 (11 January 1974), 28.

21. M.H. MacRoberts and B.R. MacRoberts, 'Social Organization and Behavior of the Acorn Woodpecker in Central Coastal California', *Ornithological Monographs* No. 21 (1976), 1–115.

22. Small, op. cit. note 11. Small, in a different context, has discussed citations like these under the label 'symbolic citations'.

23. A.J. Bennett, 'Mendel's Laws?', *School Science Review*, Vol. 46 (1964), 35–42; R. Olby, 'Mendel No Mendelian?', *History of Science*, Vol. 17 (1979), 53–72; A. Brannigan, 'The Reification of Mendel', *Social Studies of Science*, Vol. 9 (1979), 423–54; F.V. Monaghan and A. Corcos, 'The True Mendelian Laws', *Journal of Heredity*, Vol. 75 (1984), 321–23.

24. D.T. MacRoberts, 'On the Discovery of Relative Simultaneity', *The Toth-Maatian Review*, Vol. 3 (1984), 1193–97.

25. J.H. Randell, *Aristotle* (New York: Columbia University Press, 1960).

26. W.D. Garvey and B. Griffith, 'Scientific Communication: Its Role in the Conduct of Research and Creation of Knowledge', *American Psychologist*, Vol. 26 (1971), 349–62, quote on 357.

27. N.W. Smith, Book Review of J. Gibson, *The Ecological Approach to Visual Perception, Psychological Record*, Vol. 30 (1980), 587–88.

28. 'Observer', 'Priority and the Pace of Scientific Progress', *Psychological Record*, Vol. 31 (1981), 285–92.

29. Brannigan, op. cit. note 23, 426–27.

30. Ibid., 426.

31. B. Glass, 'The Long Neglect of a Scientific Discovery: Mendel's Laws of

Inheritance', in Johns Hopkins — History of Ideas Club, *Studies in Intellectual History* (Baltimore, Md: The Johns Hopkins Press, 1953), 148–60, at 154.

32. Brannigan, op. cit. note 23, 426.
33. R. Olby, *Origins of Mendelism* (London: Constance, 1966), 195.
34. Brannigan, op. cit. note 23, 426.
35. A. Weinstein, 'How Unknown was Mendel's Paper?', *Journal of the History of Biology*, Vol. 10 (1977), 341–64, at 344.
36. D. Davis, 'Citation Idiosyncrasies', *Nature*, Vol. 228 (26 December 1970), 1356.
37. J.R. Edwardson, 'Another Reference to Mendel before 1900', *Journal of Heredity*, Vol. 53 (1962), 152; F. Weiling, 'Die Hieracium-Kreuzengen J.G. Mendels sowie ihr niederschlag in Literatur and Herbarien', *Zeit. fur Pflansensuchtung*, Vol. 62 (1969), 63–99; M.H. MacRoberts, 'L.H. Bailey's Citations to Gregor Mendel', *Journal of Heredity*, Vol. 75 (1984), 500–01.
38. M.J. Moravcsik and P. Murgesan, 'Some Results on the Function and Quality of Citations', *Social Studies of Science*, Vol. 5 (1975), 86–92; D.E. Chubin and S.D. Moitra, op. cit. note 11; Cole and Cole, op. cit. note 13.
39. J.D. Watson and F.H.C. Crick, 'Molecular Structure of Nucleic Acid', *Nature*, Vol. 171 (25 April 1953), 737–38, quote on 737.
40. J.D. Watson, *The Double Helix* (New York: Athenaeum, 1968), 219–20.
41. Olby, op. cit. note 33, 195.
42. Brannigan, op. cit. note 23, 426.
43. H.W. Menard, *Science: Growth and Change* (Cambridge, Mass.: Harvard University Press, 1971), quote on 10.
44. R.B. Krumland, E.E. Will and G.A. Gorry, 'Scientific Publication of a Medical School Faculty', *Journal of Medical Education*, Vol. 54 (1979), 876–84, quote on 877.
45. Op. cit. note 35.
46. Cole and Cole, op. cit. note 3, quote on 369.
47. Ibid., quote on 375.
48. T.J. Stomps, 'On the Rediscovery of Mendel's Work by Hugo deVries', *Journal of Heredity*, Vol. 45 (1954), 293–94.
49. W.J. Broad, 'The Publishing Game: Getting More for Less', *Science*, Vol. 211 (13 March 1981), 1137–39.
50. Brannigan, op. cit. note 23; A. Brannigan, *The Social Basis of Scientific Discoveries* (Cambridge: Cambridge University Press, 1981).
51. Cole and Cole, op. cit. notes 3 and 13.
52. G.N. Gilbert, 'Referencing as Persuasion', *Social Studies of Science*, Vol. 7 (1977), 113–22.
53. Small, op. cit. note 11.
54. R.K. Merton, 'The Matthew Effect in Science', *Science*, Vol. 159 (5 January 1968), 56–63.
55. S.M. Lawani, 'Citation Analysis and the Quality of Scientific Productivity', *BioScience*, Vol. 27 (1977), 26–31, at 30.
56. C. Zirkle, 'Some Oddities in the Delayed Discovery of Mendelism', *Journal of Heredity*, Vol. 55 (1964), 65–72, 65.
57. Ibid., 66.
58. Ibid.
59. H.F. Roberts, *Plant Hybridization before Mendel* (New York: Hafner,

1965), 211.

60. Garfield, op. cit. note 2, 670

61. Edge, op. cit. note 6.

62. See Narin, op. cit. note 1, Smith, op. cit. note 1, and Garfield, op. cit. notes 2 and 3, for discussion of types of studies which use citations as data.

63. M.H. MacRoberts and B.R. MacRoberts, 'A Re-Evaluation of Lotka's Law of Scientific Productivity', *Social Studies of Science*, Vol. 12 (1982), 443–50.

Michael H. MacRoberts received his PhD at the University of California, Berkeley. *Barbara R. MacRoberts* received her MA from the same institution. Neither is currently affiliated with any institution. They are presently working on other facets of citation analysis. *Authors' address:* 740 Columbia, Shreveport, Louisiana 71104, USA.