NOTES AND LETTERS

• *ABSTRACT*

An interpretation of citation practice in scientific literature is offered which regards citation of a document as an act of symbol usage. By examining the language of the text around the footnote number the particular idea the citing author is associating with the cited document may be determined: the document is viewed as symbolic of the idea expressed in the text. This analysis was done for a sample of very highly cited documents in chemistry. A high degree of uniformity is revealed in the association of specific concepts with specific documents. These documents may be seen, in Leach's terms, as 'standard symbols' for particular ideas, methods, and experimental data in chemical science. Some implications of these findings for the social determination of scientific knowledge (conceived as a dialogue among citing authors on the 'meaning' of earlier texts), and the relationship between cited documents as concept symbols and Kuhn's exemplars, are discussed.

Cited Documents as Concept Symbols

Henry G. Small

In the last few years sociologists of science have begun to explore the fine structure of citation practice by examining the contexts in which citations occur — specifically the text surrounding the footnote number. Most of these studies have attempted to develop and apply classification schemes for references.[1] Some categories in these schemes reflect the function of the reference for the author who does the citing; others evaluate how important the reference is to the citing author. These detailed examinations of citation practice appear to be motivated by the feeling that we do not know enough about why authors cite (or the various functions citations can have) to be able to interpret studies based on statistical analysis of citation counts. Some work in this area arises from doubts that citations can be used as measures of the quality or importance of scientific work, and that a breakdown of citations into various types will help sharpen our measurements.

In general, these efforts add a dimension to the use of citation data which had previously been missing in their application as indicators of various collective aspects of science. In my view, however, they have missed an important and perhaps

Social Studies of Science (SAGE, London and Beverly Hills), Vol. 8 (1978), 327-40.

crucial point. Very little, if any, attention is given in these studies to the scientific content of the citation context.[2] Presumably this is not of interest since it did not shed light on the author's motivation for citing a particular work, or the implicit value judgment rendered by the citing author. Hence these studies have missed the role citations play as symbols of concepts or methods. This cognitive function arises from the formal requirement imposed on the scientist-author of embedding his references to earlier literature in a written text. This leads to the citing of works which embody ideas the author is discussing. The cited documents become, then, in a more general sense, 'symbols' for these ideas.

To further contrast my approach, consider the formal character of references in scholarly works. The footnote number has the function of pointing to a portion of the text in which it is embedded and at the same time corresponding to a specific document usually given at the bottom of the page or grouped at the end of the article. The footnote number should unambiguously point to a word, phrase, sentence, or other unit of text to show what ideas are to be connected with the cited document. Failure to locate the footnote number properly creates difficulties for the reader; the reference may no longer make sense. (Consider the effect on meaning of moving a footnote number from one word to another in a particular passage.) The reader expects to find some idea in the text which may be connected with the cited work as a rationalization for the author's having cited it. The reader may regard the reference as inappropriate or perhaps uninterpretable, but clearly the expectation of finding some rationale is implied. Thus the footnote number establishes a link between the cited document and language in the citing text.

In the tradition of scholarship. the references are the 'sources' which the author draws upon to give further meaning to his text. Reversing this view, as I am suggesting here, the author is imparting meaning to his 'sources' by citing them. For example, if I use Lowry's method of protein determination and cite his paper, I am not only telling the reader where he can find a description of the method, but I am stating what his paper is about, i.e., a method for protein determination. Referencing viewed in this way is a labelling process. The language pointed to by the footnote number labels or characterizes the document cited — or, in other words, constitutes the author's interpretation of the cited work. In citing a document an author is creating its meaning, and this, I will argue, is a process of symbol making.

The interpretation of citations as concept symbols (taken in a broad sense, as discussed below) is a more direct interpretation of citation practice than previous 'classification' attempts, because it is more closely related to the way citations are deployed by authors in scientific papers. It is also the basis of the use of citations for information retrieval. Garfield[3] has pointed out that a cited document is formally analogous to a subject heading in an indexing system. The topic I will explore in this paper is what kinds of 'subjects' are indicated, especially by citations to very highly cited documents.

A Note on Concepts

The term 'symbolic' is used here in the sense that Leach uses it.[4] In general, I mean that an object 'stands for' an idea; for citations, the cited document is the 'object' and the 'idea' is expressed in the text which cites it. Most citations are the author's own private symbols for certain ideas he uses: such citations are, in Leach's terminology, 'nonce symbols'. Other citations are to documents whose significant content may be shared by a community or group of scientists, and such documents are likely to be frequently cited: in Leach's terminology they are 'standard symbols'. We know very little about how a document is transformed from 'nonce symbol' to 'standard symbol', and such an analysis will not be attempted here. I will focus, rather, upon demonstrating the existence of highly cited documents which behave as 'standard symbols'.

A further clarification of the problem can be achieved using Leach's scheme. A citation may be regarded as 'symbolic' at two levels. First (and most obviously), the citation as 'author, journal, volume, page and year' is a sign for the physical document itself. In the more interesting (and less obvious) sense, the cited document (or its sign) is a symbol for a concept. The kind of relationship between the document (consisting of pages) and its sign (consisting of author, journal, volume, page, year) is in Leach's terminology 'metonymic' — that is, there are physically shared characteristics. The relationship between the cited document and the concept it symbolizes, on the other hand, is 'metaphoric'. In the extreme, this means that there need not be any similarity between the document and the concept it stands for — or, to put it more directly, the perceived content of a document is independent of the document itself. This is an overstatement of the case, because certainly in most cases the document contains the ideas which it comes to symbolize. To the extent that it does, the relationship between cited document and concept is also 'metonymic' (for example, when a direct quote is made from the cited document).

The way in which I use the terms 'concept' and 'idea' requires further elaboration. Clearly, I regard ideas or concepts as residing in the mind and expressed in language. I do not, however, restrict these terms to abstract or theoretical formulations: I include experimental findings, methodologies, types of data, metaphysical notions, theoretical statements or equations — or, in general when dealing with citations, any statement which may be taken as characterizing or describing the cited document. For example, the citation context, 'Small's paper (1) presents a completely erroneous interpretation of citation practice,' would read as follows: the 'completely erroneous interpretation of citation practice' is the idea; 'Small's paper' is the symbol for that idea.

I also follow Leach in regarding an 'idea' in its written form in a scientific paper as an imperfect 'copy of an original' which resides in the mind of an individual. In the case of 'standard symbols', the 'idea' is the product of a dialogue and selection process on the part of many individuals over a period of time. It follows that any single actor's utterance cannot be used to reconstruct the 'standard symbol': we can achieve this only by aggregating many utterances. One of the hypotheses to be explored in this paper is that a scientist carries with him a repertoire of such collective concepts and their corresponding document-symbols. These are his tools-of-the-trade, and provide the conceptual and methodological framework for his work.

Highly Cited Documents in Chemistry

The approach I have taken to exploring these issues is to examine citation contexts for a set of very highly cited articles and books in the discipline of chemistry.[5] My purpose is twofold: first, I want to show that individual citation contexts may be regarded as instances of symbol deployment; second, I want to determine for the case of highly cited items the extent to which the symbolic content is shared among a number of citing authors. To do this I will introduce the notion of 'uniformity of usage' which is defined as the percentage of citing contexts which share a particular view (the most prevalent) of the cited item. Focusing upon very highly cited items is, of course, a restriction on the generality of the findings. Whether the same pattern (symbolic character and degree of uniformity) is more or less pronounced in a sample of less cited works remains to be determined. My results should be considered as only a first empirical step in the investigation of shared document-symbols.

A set of 294 journals categorized under 'chemistry' and related headings (including crystallography but not biochemistry) in the 1972 *Science Citation Index (SCI) Guide and Journal Lists*[6] was selected, and this journal set was used to create a special subset citation index of the 1972 *SCI*. This subset contained only those references made by the 294 chemistry journals. The references were cumulated and a list of the 52 most cited items (those cited 66 times or more) was produced (see Table 1). The special nature of this sample is clearly indicated by the fact that the 52 most cited items in this chemistry file comprise only a 0.009 percent sample of items cited by the 294 chemistry journals, and only a 1.2 percent sample of items cited ten or more times.[7]

Aside from the highly select nature of this group, some further characteristics should be noted. It was found, for example, that the highly cited documents were on the whole older than the less cited documents. Documents in the 1972 chemistry file in the range of 30 citations and above are about three years older than documents in the range of 10 to 15 times cited (1961 versus 1964). As has been observed in other synchronous studies,[8] the number of cited items of a given age drops off exponentially with increasing age. Hence there is a definite tendency for the relatively few old documents on the highly cited list to be cited more frequently. Another feature of the highly cited group is the increased proportion of books. In the chemistry file, 14 percent of the items in the citation range above 16 were books or monographs; in the citation range from 10 to 11, only six percent of the items were books or monographs.

For each of the 52 documents, about 12 citing papers were sampled randomly from the 1972 *SCI*. For each citing paper, the point or points in the paper where reference was made to the item was determined. Usually it was sufficient to examine two or three sentences around the point where the footnote number appeared in order to determine the concept used. For each citing context the term or phrase most closely associated with the footnote number was recorded. Finally, for all contexts citing a particular item the percentage of references attributable to the most frequently mentioned term or phrase (concept) was calculated. (This is given in Table 3 as the 'Percent Uniformity,' and measures the degree to which these documents have become 'standard symbols' for particular concepts, or the degree of consensus on their usage.)

Table 2 gives an example of a context of citation for each of the top ten items listed in Table 1. In each the sentence is quoted in which the footnote number appeared (under 'citation context') and the concept symbolized by the cited document in that specific instance is recorded (third column). These quotations were selected to illustrate how the symbolic content was reconstructed from the citing text. Consider the first paper on the list, that by Stewart et al. The quoted context indicates that the author of the citing work has used a 'scattering factor' for hydrogen given in the Stewart paper. (Table 3 gives 'hydrogen scattering factors' as the most frequently associated concept with a percent uniformity of 93.0) Hence, for the citing author, the Stewart paper is symbolic of values for hydrogen scattering factors. Another example is the paper by Lowry et al., second on the list. This paper is well known to citation analysts as by far the most cited paper in science.[9] In symbolic terms, the Lowry paper stands for a method of protein determination, itself a modification of an earlier method. (Table 3

Table 1

The 52 Most Cited Documents in the 1972 Chemistry File

Times Cited	Bibliographical Data
309	**Stewart, R.F., Davidson, E.R. & Simpson, W.T.** Coherent x-ray scattering for the hydrogen atom in the hydrogen molecule. *J. Chem. Phys* 42(9):3175, 1 May 1965.
244	**Lowry, O.H., Rosebrough, N.J., Farr, A.L. & Randall, R.J.** Protein measurement with the Folin phenol reagent. *J. Biol. Chem.* 193:265, 1951.
241	**Cromer, D.T. & Waber, J.T.** Scattering factors computed from relativistic Dirac-Slater wave functions. *Acta Cryst.* 18:104, 1965.
204	**Pauling, L.** *The Nature of the Chemical Bond* (Ithaca: Cornell University Press, 1960).
199	**Woodward, R.B. & Hoffmann, R.** The conservation of orbital symmetry. *Angew. Chemie* 81:797, 1969; and *Angew. Chemie Internat. Ed.* 8:781, 1969.
193	**Cromer, D.T.** Anomalous dispersion corrections computed from self-consistent field-relativistic Dirac-Slater wave functions. *Acta Cryst.* 18:17, 1965.
166	**Karle, J. & Karle, I.L.** The symbolic addition procedure for phase determination for centrosymmetric and noncentrosymmetric crystals. *Acta Cryst.* 21:849, 1966.
161	**Pople, J.A. & Segal, G.A.** Approximate self-consistent molecular orbital theory. III. CNDO results for AB_2 and AB_3 systems. *J. Chem. Phys.* 44:3289, 1966.
149	**Hanson, H.P., Herman, F., Lea, J.D. & Skillman, S.** HFS atomic scattering factors. *Acta Cryst* 17:1040, 1964.
140	**Hoffmann, R.** An extended Hückel theory. I. Hydrocarbons. *J. Chem. Phys.* 39:1397, 1963.
127	**Hamilton, W.C.** Significance tests on the crystallographic *R* factor. *Acta Cryst.* 18:502, 1965.
114	**Schomaker, V. & Trueblood, K.N.** On the rigid-body motion of molecules in crystals. *Acta Cryst.* B24:63, 1968.
114	**Pople, J.A. & Beveridge, D.L.** *Approximate Molecular Orbital Theory* (New York: McGraw-Hill, 1970).
114	**Sanders, J.K.M. & Williams, D.H.** Tris(dipivalomethanato) europium. A paramagnetic shift reagent for use in nuclear magnetic resonance spectroscopy. *J. Amer. Chem. Soc.* 93:641, 1971.
108	**Jaffe, H.H.** A reexamination of the Hammett equation. *Chem. Rev.* 53:191, 1953.
107	**Pople, J.A., Beveridge, D.L. & Dobosh, P.A.** Approximate self-consistent molecular-orbital theory. V. Intermediate neglect of differential overlap. *J. Chem. Phys.* 47:2026, 1967.
106	**Hinckley, C.C.** Paramagnetic shifts in solutions of cholesterol and the dipyridine adduct of trisdipivalomethanatoeuropium (III). A shift reagent. *J. Amer. Chem. Soc.* 91:5160, 1969.
102	**Johnson, C.K.** *ORTEP: A Fortran Thermal-Ellipsoid Plot Program for Crystal Structure Illustrations,* Report ORNL-3794 (Oak Ridge: Oak Ridge National Laboratory, 1965).
101	**Mulliken, R.S.** Electronic population analysis on LCAO-MO molecular wave functions. I. *J. Chem. Phys.* 23:1833, 1955.
100	**Karplus, M.** Contact electron-spin coupling of nuclear magnetic moments. *J. Chem. Phys.* 30:11, 1959.
97	**Wilson, E.B., Decius, J.C. & Cross, P.C.** *Molecular Vibrations* (New York) McGraw-Hill, 1955).
96	**Cromer, D.T. & Mann, J.B.** X-ray scattering factors computed from numerical Hartree-Fock wave functions. *Acta Cryst.* A24:321, 1968.
95	**Pariser, R. & Parr, R.G.** A semi-empirical theory of the electronic spectra and electronic structure of complex unsaturated molecules. II. *J. Chem. Phys.* 21:767, 1953.
94	**Roothaan, C.C.J.** New developments in molecular orbital theory. *Rev. Mod. Phys.* 23:69, 1951.
93	**Woodward, R.B. & Hoffmann, R.** *The Conservation of Orbital Symmetry* (New York: Academic Press, 1970).
92	**Hirschfelder, J.O., Curtiss, C.F. & Bird, R.B.** *Molecular Theory of Gases and Liquids* (New York: John Wiley, 1954).

Table 1 (Continued)

Times Cited	Bibliographical Data
90	**Pariser, R. & Parr, R.G.** A semi-empirical theory of the electronic spectra and electronic structure of complex unsaturated molecules. I. *J. Chem. Phys.* 21:466, 1953.
89	**Spackman, D.H., Stein, W.H. & Moore, S.** Automatic recording apparatus for use in the chromatography of amino acids. *Analyt. Chem.* 30:1190, 1958.
87	**Pople, J.A., Santry, D.P. & Segal, G.A.** Approximate self-consistent molecular orbital theory. I. Invariant Procedures. *J. Chem. Phys.* 43:S129, 1965.
87	**Herzberg, G.** *Spectra of Diatomic Molecules* (Princeton: D. Van Nostrand, 1950).
85	**Flory, P.J.** *Principles of Polymer Chemistry* (Ithaca: Cornell University Press, 1953).
80	**Turner, D.W., Baker, C., Baker, A.D. & Brundle, C.R.** *Molecular Photoelectron Spectroscopy* (London: John Wiley, 1970).
80	**Hatchard, C.G. & Parker, C.A.** A new sensitive chemical actinometer. II. Potassium ferrioxalate as a standard chemical actinometer. *P. Roy. Soc.* A235:518, 1956.
79	**Pople, J.A. & Segal, G.A.** Approximate self-consistent molecular orbital theory. II. Calculations with complete neglect of differential overlap. *J. Chem. Phys.* 43:S136, 1965.
77	**Sanders, J.K.M. & Williams, D.H.** A shift reagent for use in nuclear magnetic resonance spectroscopy. A first-order spectrum of n-hexanol. *Chem. Comm.* pg. 422, 1970.
75	**Nicholson, R.S. & Shain, I.** Theory of stationary electrode polarography single scan and cyclic methods applied to reversible, irreversible, and kinetic system. *Analyt. Chem.* 36:706, 1964.
75	**Pople, J.A.** Electron interaction in unsaturated hydrocarbons. *T. Faraday Soc.* 49:1375, 1953.
74	**Hammond, G.S.** A correlation of reaction rates. *J. Am. Chem. Soc.* 77:334, 1955.
72	**Huzinaga, S.** Gaussian-type functions for polyatomic systems. I. *J. Chem. Phys.* 42:1293, 1965.
72	**Busing, W.R., Martin, K.O. & Levy, H.A.** *Program ORFLS,* Report ORNL-TM-36 (Oak Ridge: Oak Ridge National Laboratory, 1962).
71	**Busing, W.R. & Levy, H.A.** The effect of thermal motion on the estimation of bond lengths from diffraction measurements. *Acta Cryst.* 17:142, 1964.
70	**Clementi, E.** *Ab initio* computations in atoms and molecules. *IBM J. Res. Develop.* 9:2, 1965.
70	**Fessenden, R.W. & Schuler, R.H.** Electron spin resonance studies of transient alkyl radicals. *J. Chem. Phys.* 39:2147, 1963.
69	**Bloembergen, N., Purcell, E.M. & Pound, R.V.** Relaxation effects in nuclear magnetic resonance absorption. *Phys. Rev.* 73:679, 1948.
69	**Abragam, A.** *The Principles of Nuclear Magnetism* (London: Oxford University Press, 1961).
68	**Davis, B.J.** Disc electrophoresis. II. Method and application to human serum proteins. *Ann. N.Y. Acad. Sci.* 121:404, 1964.
68	**Clementi, E. & Raimondi, D.L.** Atomic screening constants from SCF functions. *J. Chem. Phys.* 38:2686, 1963.
66	**Demarco, P.V., Elzey, T.K., Lewis, R.B. & Wenkert, E.** Paramagnetic induced shifts in the proton magnetic resonance spectra of alcohols using tris(dipivalomethanato) europium (III). *J. Amer. Chem. Soc.* 92:5734, 1970.
66	**Hinze, J. & Jaffe, H.H.** Electronegativity. I. Orbital electronegativity of neutral atoms. *J. Amer. Chem. Soc.* 84:540, 1962.
66	**Streitwieser, A.** *Molecular Orbital Theory for Organic Chemists* (New York: John Wiley, 1961).
66	**Wilson, A.J.C.** Determination of absolute from relative x-ray intensity data. *Nature* 150:151, 1942.
66	**McClellan, A.L.** *Tables of Experimental Dipole Moments* (San Francisco: W.H. Freeman, 1963).

Table 2

Selection Citation Contexts For The Ten Most Highly Cited Documents

Cited Document	*Citation Context*	*Cited Document Symbolic of*
Stewart, 309	'The scattering factor for the hydrogen atom was that given by Stewart, Davidson and Simpson (1965)'	Hydrogen scattering factor
Lowry, 244	'Protein was determined by the Lowry et al. modification [6] of the Folin-Ciocalteu phenol reagent method.'	Protein determination method
Cromer, 241	'Atomic scattering factors of Cromer and Waber [10] were used	Atomic scattering factors
Pauling, 204	'[The average bond length]. . . compares quite favorably with the value of 2.34A predicted by the single-bond covalent radius (Pauling 1960).'	Covalent radius
Woodward, 199	'One of the most stimulating recent developments in organic chemistry has been the enunciation of selection rules for concerted reactions — the Woodward-Hoffmann rules.[1, 2]'	Woodward-Hoffmann rules
Cromer, 193	'Real and imaginary components of the anomalous dispersion were those of D.T. Cromer (1965).'	Components of anomalous dispersion
Karle, 166	'Solution of the structure was achieved using a combined application of the symbolic addition procedure and the tangent formula.[4]'	Symbolic addition procedure
Pople, 161	'Recent calculations[1] of the dipole moment derivatives for a number of small molecules . . . by the complete neglect of differential overlap (CNDO) approximate method,[2] have been most encouraging.'	CNDO approximate method
Hanson, 149	'Atomic scattering factors for the refinement of both structures were taken from Hanson, et al.[5]'	Atomic scattering factors
Hoffmann, 140	'There are, in fact, a plethora of such semiempirical techniques, the best known and most widely used being the extended Hückel method (Mulliken 1949, Wolfsberg and Helmoltz 1952, Hoffmann 1963).'	Extended Hückel method

indicates 100 percent uniformity of this usage.) An example of a theoretical document-symbol is the Woodward paper, fifth on the list. This paper is a symbol for the enunciation of the so-called 'Woodward-Hoffmann rules', also known as 'orbital symmetry conservation rules'. Examination of these and other quoted citation contexts in Table 2 shows how each may be regarded as an instance of symbol deployment — 'ideas' in the broad sense linked to specific objects (cited documents).

Another feature to emerge from this list of quotations is the operational or procedural character of most of the contexts. The authors are performing or doing something, and the way to do it is described in the document cited. This is true even when a theory or principle is cited. This operational orientation is easily understood: the very highly cited papers and books (which this sample represents) are oriented to what might be called 'tools-of-the-trade' — that is, instructions on how to carry out certain basic operations at the lab bench or at the desk. These papers are not what we normally think of as 'research front papers', which are of a more recent origin, generally less highly cited, and often of a more conceptual (as opposed to operational) nature.

The results of the analysis of all citation contexts are presented in Table 3. This table reveals the second point I would like to make — namely, the uniformity with which these works are used. As an example, take the case of R. Hoffmann, *J. Chem. Phys.*, Vol. 39 (1963), 1397. In Table 4, I quote the phrases which appear in close proximity to the footnote number in a selected (but typical) set of citing articles. The terminology has clearly been standardized to 'extended Hückel theory' when authors cite this work, and this phrase also appears in Table 3 with a 90 percent uniformity of usage. Such a degree of consensus on a theoretical tool is characteristic of what Kuhn has called 'normal' science — science which is guided by what he calls 'exemplars'.[10]

The overall uniformity of usage for the 52 most cited documents in the chemistry file is 87 percent. Only ten of the 52 dip below 80 percent uniformity, and it is significant that seven of these ten are books. (Books are otherwise outnumbered by about four to one in the sample of 52.) Hence, books are used and cited in a wider variety of ways than journal articles. (The mean uniformity of usage for books is 68 percent, and it is 92 percent for journal articles). This is not a surprising finding when we consider the broader information content of most books.

Another observation is the frequency with which these works are involved in what Moravcsik has called 'redundant' patterns of citation.[11] This occurs most obviously when an author cites several works simultaneously with a single footnote number or a series of numbers. Redundant citations can be taken to indicate one of a number of situations. For example, they may signal simultaneous and independent discovery. Such situations are also characterized by high levels of co-citation among the redundantly cited works.[12] This is illustrated, in the sample of 52, by the papers on lanthanide shift reagents by Sanders (f = 114), Hinckley (f = 106), Demarco (f = 66) and Sanders (f = 77). These four papers reported early work on a new technique in NMR spectroscopy, and all were very highly co-cited and redundantly cited. Although Hinckley's paper is clearly the earliest the others followed quickly and extended the initial work in several important directions. Hence they are often cited in a collective (or redundant) way as initiating a new line of NMR research.

Another reason for redundant citations is the availability of more than one good source for the same concept or procedure. This is illustrated by the several works cited in connection with the theoretical techniques in molecular orbital theory, variously labelled CNDO (complete neglect of differential overlap) and INDO (intermediate

Table 3.

Context of Citation for Highly Cited Documents

Cited Item	*Pub. Type*	*Times Cited*	*Concept*	*Percent Uniformity*
Stewart	J	309	hydrogen scattering factors	93
Lowry	J	244	protein determination	100
Cromer	J	241	atomic scattering factors	93
Pauling	B	204	atomic radii	56
Woodward	J	199	orbital symmetry rules	72
Cromer	J	193	correction for anomalous dispersion	100
Karle	J	166	symbolic addition method	79
Pople	J	161	CNDO/2 method	86
Hanson	J	149	atomic scattering factors	82
Hoffmann	J	140	extended Hückel theory	90
Hamilton	J	127	R-ratio test	91
Schomaker	J	114	rigid body analysis	100
Pople	B	114	CNDO and INDO methods	100
Sanders	J	114	lanthanide shift reagents	90
Jaffe	J	108	Hammett constants	88
Pople	J	107	CNDO and INDO methods	100
Hinckley	J	106	lanthanide shift reagents	100
Johnson	R	102	stereodiagrams	92
Mulliken	J	101	electron population analysis	92
Karplus	J	100	Karplus equation	100
Wilson	B	97	Wilson GF method	64
Cromer	J	96	atomic scattering factors	91
Pariser	J	95	Pariser-Parr-Pople method	82
Roothaan	J	94	Roothaan equation	91
Woodward	B	93	orbital symmetry conservation	69
Hirschfelder	B	92	impact parameters	60
Pariser	J	90	Pariser-Parr-Pople method	82
Spackman	J	89	amino acid analyzer	90
Pople	J	87	CNDO and NDDO methods	100
Herzberg	B	87	bond lengths	36
Flory	B	85	Flory parameters	42
Turner	B	80	photoelectron spectra	100
Hatchard	J	80	ferrioxalate actinometry	92
Pople	J	79	CNDO/2	90
Sanders	J	77	lanthanide shift reagents	100
Nicholson	J	75	peak potentials	57
Pople	J	75	Pariser-Parr-Pople method	90
Hammond	J	74	Hammond's postulate	100
Huzinaga	J	72	Huzinaga basis set	100
Busing	R	72	program ORFLS	100
Busing	J	71	riding motion corrections	100
Clementi	J	70	atomic wave functions	100
Fessenden	J	70	ESR of radicals	100
Bloembergen	J	69	nuclear spin relaxation	82
Abragam	B	69	line shape equation	36
Davis	J	68	disc gel electrophoresis	100
Clementi	J	68	minimal basis set	91
Demarco	J	66	chemical shifts	100
Hinze	J	66	ionization potentials and electron affinities	93
Streitwieser	B	66	Hückel MO method	82
Wilson	J	66	Wilson plot	100
McClellan	B	66	dipole moments	100

Key: *Publication Type*
J Journal article
B Book or monograph
R Report

neglect of differential overlap). Presumably the researcher has a wide variety of equivalent formulations from which to select, sometimes written by the same individual, and one or more of these may be cited. The prevalence of molecular orbital papers on this list may be the result of the diffusion of these theoretical techniques into wider realms of application (especially organic chemistry). Often those who apply the techniques may be familiar with one or another of the formulations of the theory, and so the works cited vary from author to author. This gives rise to a family of works cited essentially for the same ideas and purposes.

Table 4
Citation Contexts to Paper by R. Hoffmann Showing Its Standardized Usage

[Cited Item: R. Hoffmann, 'An Extended Hückel Theory .I. Hydrocarbons,' *J. Chem. Phys.*, Vol. 39, No. 6 (1963), 1397.[

Citation Context

'. . . many other structures have become accessible to quantum chemistry through the introduction of semi-empirical all-valence-electron techniques such as extended Hückel theory (EHT)[3] . . .'

'Here, we report extended Hückel (EHT) calculations (3) . . .'

'The MOs of $Cu(dtc)_2$ were calculated by means of the LCAO-MO extended Hückel method[3].'

'Fahey et al. (25) employed Pople-Santry formalism (26) with extended Hückel (27) coefficients . . .'

'Both extended Hückel (EH)[16] and INDO[17] calculations have been carried out . . .'

'There are, in fact, a plethora of such semi-empirical techniques, the best known and most widely used being the extended Hückel method (Mulliken 1949. Wolfsberg and Helmholtz 1952. Hoffmann 1963.)'

'Thus, correlations of nmr chemical shifts with charge distributions calculated by the Extended Hückel Theory (12) have been demonstrated . . .'

'. . . charge densities on the individual atoms of Ach and Carbachol (fig. 1) were derived by using two quantum-mechanical techniques, the extended Hückel technique (EHT) in the original parametrization (Hoffmann, 1963), and the complete neglect of differential overlap . . .'

'The extended Hückel (22) molecular orbital wave-functions and energies have been employed in Eq. (11) to evaluate the coupling constant'.

Discussion

A theory of citation practice, if such a theory is possible, must take account of the symbolic act of authors associating particular ideas with particular documents. The analysis has shown that when a scientist cites, he or she is creating a link between a concept, procedure, or kind of data, and a document or documents. In some cases, the association of idea and document is well established by uniform practice within the community (Leach's 'standard symbol'). Recurring patterns of terminology used by citing authors when referring to these documents show that they have become standardized in their usage and meaning. In other cases, the individual author (scientist) may be making the association for the first time and the document-idea connection remains in the realm of private symbols (Leach's 'nonce symbols').

We have seen that many of the very highly cited documents in a chemistry file have very uniform or standardized usage and meaning. Books tend to have lower degrees of uniform usage than research papers, probably due to their greater diversity of content. No comparable data is available yet on less frequently cited documents that might be considered to be at the research front (although work is now underway on a group of such papers in the recombinant-DNA area). Nor can we assume that infrequently cited or uncited papers either do or do not have standardized images, even though they will undoubtedly possess scientific images for individuals familiar with them. One could postulate a lower percent uniformity for a less cited item, since the process of symbol formation is still underway and the community has yet to arrive at a standard interpretation of the work. These points remain for future research.

Gilbert has interpreted citation practice as an author's device for persuading readers of the validity of his arguments.[13] While I agree with the essentials of Gilbert's argument, I have not attempted to give a causal explanation of why author-scientists cite certain papers and not others. Such reason might include the desire to persuade, to curry favour, to publicize, to avoid offending, to favour one approach over another, to give credit, and so on. None of these, however, appears to be adequate to explain the full range of motivations for citing. What does appear to be more nearly universal is the citation as a symbol for an idea. Of course, there will be exceptions to this interpretation — for example, when an author is merely listing works on a general topic (as in a bibliography) without commenting upon them — but even then there is an implied context which says 'these documents are examples of work in this area.' The concept symbol interpretation of citation practice does not contradict the functional, social or political interpretations, but is complementary to them. Whether the motive for citing a work is politically conditioned or is merely haphazard (for example, adding references to a paper after it is written, where they 'fit in'), the work cited must be associated with specific language in the text and cannot be appended without some explicit or implicit context.

Gilbert has also pointed out that citations to frequently cited papers may be related to Kuhn's concept of the 'exemplar'. Kuhn seems to have something like this in mind when he says:

> Close historical investigation of a given specialty at a given time discloses a set of recurrent and quasistandard illustrations of various theories in their conceptual, observational, and instrumental applications.[14]

However, in this early description of the sources of the 'paradigm' Kuhn seems to be thinking more of pedagogic materials (such as textbooks, lectures and laboratory exercises) than of reports of research published in the scientific journals. The former embody perhaps more tacit elements in a scientist's mental set, which are unlikely to emerge as citations to specific documents. Nevertheless, if exemplars are illustrations of methods or theories which comprise the essential repertoire of techniques (tools-of-the-trade) for the practitioners of a specialty, then the finding that highly cited documents are symbolic of standardized procedural tools and theories suggests that Kuhn's sources of 'exemplars' can be extended to the research literature itself.

In the sense that the document-exemplars are guiding and directing research, I agree with Gilbert that their citation may be taken as evidence of an author's compliance with the general paradigm. Without such compliance, he runs the risk of being misunderstood. It would be wrong, however, to regard citation of certain works as mandatory, or as constraining research. At most, they provide a weak framework which is easily violated and fairly fluid.

The discussion of highly cited documents as 'exemplars' leads naturally to the problem of the social determination of scientific knowledge. I have stressed the importance of viewing citations as interpretations of cited works. To quote a recent citation study: 'What is thought of published work is as important as, or more important than, the act of publication itself.'[15] The interesting question is not whether the cited work is 'correct', or whether the citing author has made a 'correct' interpretation of it, but rather whether the interpretation given is in accord or at variance with the interpretations others have given it. It is the process of acquiring a standard or conventional interpretation that is crucial for the social determination of scientific ideas. Stated another way, as a document is repeatedly cited, the citers engage in a dialogue on the document's significance. The verdict or consensus which emerges (if one does) from this dialogue is manifested as a uniform terminology in the contexts of citation. Meaning has been conferred through usage and what is regarded and accepted as currently valid theory or procedure has been socially selected and defined.

A difficulty with this view is that it allows the cited document very little role in determining its fate. For the sake of discussion, consider a document which has been frequently cited and has become the standard symbol for an idea. The document could originally have been written to convey ideas other than the one it has come to symbolize. In other words, the process of becoming public property has transformed the document into something the author may not have intended. While such extreme instances are undoubtedly rare, and in most cases the intended message is the received message, the possibility of the social transformation of meaning must be recognized.

I have not examined the process by which a work acquires its standard or conventional interpretation (meaning), but perhaps the best way to study this would be by examining the contexts of citation in chronological sequence — rather than (as I have done) at one point in time.

Perhaps the most important result of the social selection of knowledge through citation is the narrowing of meaning which occurs. By condensing or 'capsulizing' a complex original text into a few standard statements, the community of scientists can more easily confirm, refute or build upon the earlier work. This serves the needs of the specialty by enabling work to go on unencumbered by the necessity of unravelling the complete meaning and implications of the earlier text, even though this may result in the distortion or oversimplification of the original.

Another important function of citations is to make more concrete that which is by its nature abstract. Thus citation practice can be seen, as Gilbert has seen it,[16] as part of a scientists's strategy for making reality more objective. This is, of course, a psychological effect. By citing a physically existing document, the idea it symbolizes is given, indirectly, physical embodiment. As Leach puts it:

> By converting ideas, products of the mind (mentifacts), into material objects 'out-there,' we give them relative permanence, and in that permanent material form we can subject them to technical operations which are beyond the capacity of the mind acting by itself.[17]

This is reminiscent of Durkheim's notion of 'collective representations' — for example, when he states:

> . . . in order to express our own ideas to ourselves, it is necessary . . . that we fix them upon material things which symbolize them.[18]

Although Durkheim would probably object to an analogy between religious symbols and the kind of scientific symbols I am suggesting, there is no doubt that the latter fit well with his general formulation of 'collective representations.'

These considerations also suggest the possibility that citations serve as a kind of language system, which can be deployed with greater flexibility than ordinary language. Cited documents as concept symbols may be freely combined and juxtaposed, unhampered by the customary rules of logic or syntax, to suggest by analogy a wide range of conceptual possibilities.

• *NOTES*

A preliminary version of this paper was presented at the American Chemical Society meeting in San Francisco on 1 September 1976. Research was supported by National Science Foundation contract C-795. I would like to thank Edwin Greenlee of ISI for valuable discussions.

1. M.J. Moravcsik and P. Murugesan, '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, 'Content Analysis of References: Adjunct or Alternative to Citation Counting?; ibid., 423-41; I. Spiegel-Rösing, '*Science Studies:* Bibliometric and Content Analysis', ibid, Vol. 7 (1977), 97-113. One of the earliest attempts to classify references was: Ben-Ami Lipetz, 'Improvement of the Selectivity of Citation Indexes to Science Literature through Inclusion of Citation Relationship Indicators', *American Documentation,* Vol. 16, No. 2 (1965), 81-90. The difference between 'citation' and 'reference' is only one of perspective on the linkage between citing and cited documents: if one is looking from the citing document to the cited document, it is a 'reference'; if one is looking from the cited to the citing, it is a 'citation'.

2. Somewhat more attention to content has been given by Stephen Cole, who has provided a cognitive analysis of the references to the work of Robert Merton on social structure and anomie: see S. Cole, 'The Growth of Scientific Knowledge: Theories of Deviance as a Case Study', in L.A. Coser (ed.), *The Idea of Social Structure* (New York: Harcourt, Brace, Jovanovich, 1975), 175-220.

3. E. Garfield, 'The Citation Index as a Subject Index' (Editorial), *Current Contents* (1 May 1974), reprinted in: E. Garfield, *Essays of an Information Scientist,* Vol. 2 (Philadelphia: ISI Press, 1977), 62-64.

4. E. Leach, *Culture and Communication: the Logic by which Symbols are Connected* (Cambridge: Cambridge University Press, 1976).

5. For a more detailed description of how this file was generated, and its characteristics, see H. Small, 'Characteristics of Frequently Cited Papers in Chemistry', *Final Report on National Science Foundation Contract C-795,* Phase I (Philadelphia: Institute for Scientific Information, 1974).

6. *Science Citation Index, 1972 Guide and Journal Lists* (Philadelphia: Institute for Scientific Information, 1972).

7. The distribution of number of cited items cited N times is, of course, hyperbolic (a straight line is obtained if both the number of cited items and N are plotted on log axes) as has been shown in previous studies: see D. Price, 'A General Theory of Bibliometric and other Cumulative Advantage Processes', *Journal of the American Society for Information Science,* Vol. 27, No. 5 (1976), 292-305.

8. M.B. Line and A. Sandison, 'Obsolescence and Changes in the Use of Literature with Time', *Journal of Documentation,* Vol. 30 (1974), 283-350.

9. O.H. Lowry, N.J Rosenbrough, A.L. Farr and R.J. Randall, 'Protein Measurement with the Folin Phenol Reagent', *J. Biol. Chem.,* Vol. 193 (1951), 265, and E. Garfield, 'Citation Frequency as a Measure of Research Activity and Performance', *Current Contents* (31 January 1973), reprinted in: E. Garfield, *Essays of an Information Scientist,* Vol. 1 (Philadelphia: ISI Press, 1977), 406-08.

10. T.S. Kuhn, *The Structure of Scientific Revolutions* (Chicago: The University of Chicago Press, second edition, 1970).

11. Moravcsik and Murugesan, op. cit. note 1.

12. H.G. Small, 'Co-Citation in the Scientific Literature: a New Measure of the Relationship between Two Documents', *Journal of the American Society for Information Science,* Vol. 24 (1973), 265-69.

13. G.N. Gilbert, 'Referencing as Persuasion', *Social Studies of Science,* Vol. 7 (1977), 113-22.

14. Kuhn, op. cit. note 10, 43.

15. S.M. Lawani, 'Citation Analysis and the Quality of Scientific Productivity', *BioScience,* Vol. 27, No. 1 (January 1977), 26-31.

16. G.N. Gilbert, 'The Transformation of Research Findings into Scientific Knowledge', *Social Studies of Science,* Vol. 6 (1976), 281-306.

17. Leach, op. cit. note 4, 37.

18. E. Durkheim (trans. J.W. Swain), *The Elementary Forms of the Religious Life* (New York: Free Press, 1965), 260.

Henry Small is Director of Grant and Contract Research at the Institute for Scientific Information. His most recent publication deals with methodological problems of cluster analysis (*Information Processing and Management,* 1977). He is currently working on citation studies of physics in the 1920s and recombinant-DNA in the 1970s. *Author's address:* Institute for Scientific Information, 325 Chestnut Street, Philadelphia, Pennsylvania 19106, USA.