Scientometrics, Vol. 3. No. 2 (1981) 135–147

SOME METHODOLOGICAL PROBLEMS IN RANKING SCIENTISTS BY CITATION ANALYSIS

G. FOLLY, B. HAJTMAN, J. I. NAGY, I. RUFF*

Institute of Inorganic and Analytical Chemistry, L. Eötvös University, P.O. Box 123, 1443. Budapest (Hungary)

(Received June 21, 1979, in revised form May 2, 1980)

A sample of 80 Hungarian scientists, authors or co-authors of a total number of 6273 papers – published between 1930–1976 – has been analysed. Citation data to each *paper* were collected form the 1964–76 SCI's by manual search. Citation counts were distinguished with respect to the following categories: (I) the set of cited authors has element(s) common with the set of citing authors (self citation), (II) condition I is not satisfied, but the cited author under study and at least one of the citing authors were co-authors prior to the publication of the cited paper, (III) none of the former criteria is satisfied. The yearly average citation frequency of a paper was not corrected for obsolescence, since there is no evidence that the decay of citation frequency with time is independent of the absolute citedness of the paper. Individual performance has been measured (a) by the sum of the yearly average type *III* fractional citation frequencies over all of the author's papers, (b) by the sum of the yearly average citation frequency normalized to one single-authored paper per year over the period of the author's activity, (c) by the same as in *a*, but summed up only over the most highly cited papers "scattering upwards" from the individual's own average, (d) by the fractional authorship, and (e) by the number of items in the author's publication list. The first three parameters seem to be applicable in measuring the utility of the individual's scientific contribution with slightly different emphasis on different aspects. These parameters are uncorrelated with those measuring the output of individuals.

Introduction

The original aim of the present work was to study the correlation of simple citation counts with various derivatives of other citation parameters which emphasize one or another feature of an individual's performance.

The paper was already under preparation when – thanks to the courtesy of the editors – the manuscript of one of Eugene *Garfield*'s recent papers was made available to us.[1] In that paper a comprehensive account is given of the controversy about citation analysis. We feel that most of our results can serve as good illustrations to several points *Garfield* made in his paper, and that they generally support his views. Thus, after a description of our sample, we are going to structure this paper so as to follow *Garfield*'s argumentation.

*To whom reprint requests should be addressed.

On the other hand, the recent appearance of Garfield's paper relieves us of the task of reviewing the literature on individual assessment based on citation analysis. In this respect we also refer to that paper and the references therein.

The sample

The selection of the 80 Hungarian scientists was not arbitrary. Based on one of our earlier observations[2] in a branch of science, that leading Hungarian scientists comprising about 5% of all authors are authors or co-authors of about 3/4 of the total Hungarian publication output of that branch, we hoped to minimize the number of first author names to be looked up in the SCI during the manual search of citations by confining ourselves to the list of publications of scientists who had held the D. Sc. degree. Since such a sample proved to be still too large, the most "visible" scientists were retained. For selection we used the criterion whether the given scientists had been considered as potential candidate for corresponding membership of the Academy of Sciences in the previous cooption procedure or he was being considered in the forthcoming one.

On the ground that 88% of the scientists selected in this way proved to be heads of university departments or research institutions, one could suspect that such a selection was biassed by peer evaluation, i.e. they formed the elite of all D. Sc.'s. To check this hypothesis, *all* D.Sc.'s and academicians in *physics* were asked for their list of publications, and the 58% providing the lists were added to the sample. The most "visible" 6 physicists ranked 3rd, 6th, 14th, 25th, 26th

Table 1
Distribution of individuals and their papers among disciplines

Discipline	Number of		Mean (and standard deviation) of		
	individuals	papers	authorship	fractional authorship	age
Mathematics	9	432	48 (23)	40.8 (20.3)	45 (9)
Physics, all	33	2178	66 (26)	41.5 (18.7)	52 (10)
["visible"]	[6]	[301]	[50 (18)]	[33.4 (14.0)]	[45 (5)]
Chemistry	9	772	86 (30)	46.3 (16.4)	46 (5)
Biology	5	378	76 (52)	40.7 (24.4)	45 (4)
Medical Sci.	14	1754	125 (54)	58.1 (33.6)	50 (5)
Agriculture	5	566	113 (83)	76.0 (58.2)	52 (4)
Technology	5	193	39 (17)	25.0 (15.7)	48 (2)
Total or gross average	80	6273	79 (37)	45.5 (21.0)	50 (8)

and 30th with respect to citations to their papers in the subset of the 33 physicists studied. This uniform distribution supports the assumption that the sampling procedure supplied an average sample of the Hungarian Doctors of Sciences.

For a general characterization of the sample, the number of individuals and their papers in the different disciplines are summarized in Table 1 together with the average and standard deviation of their authorship, fractional authorship (single author contribution to the papers), and age. Although the individual output of papers shows a slight increase from mathematics to (basic and applied) life sciences, the differences are fairly diminished when the effect of multiple authorship is eliminated by calculating the fractional authorship. It is well-known that single-author papers are the most frequent in mathematics and the least in life sciences, and multiple authorship increases the "productivity" of scientists working with several coworkers. The exceptionally low output in technology can be attributed to the special interest of technical scientists not to disclose results important in some production process.[3]

The uniformity of the data in the last two columns, their standard deviations as well the fact that none of the authors has published less then 20 papers indicate that this sample is more suitable to study correlations between individual output and citation patterns, than some arbitrary sample of authors would be, since it does not involve scientists with only one or two papers to their name for which the low citedness–low output correlation would be spurious. On the other hand, the scatter of the individual output data is not too narrow either to cause a trivially low correlation.

In determining the number of items in each individual's publication list, we took the special Hungarian custom into account that scientists often publish the same paper both in Hungarian and in foreign language. This generally accepted practice aims at facilitating communication within the Hungarian community of scientists and at supporting Hungarian scientific phrasaeology to keep pace with the development of science. Authors usually number the items in their list of publications so that duplications comprise one item. In cases when this was not unanimous, we scrutinized the lists, and unified the duplications into one item. The citations, however, were collected to each of the duplicates and ultimately unified under a single item.

The lists of papers often contained items that did not satisfy the condition of being public indeed. We have omitted conference lectures unless they were published. University textbooks, theses (mainly in Hungarian) and educational papers were also omitted as well as books *edited* by the scientist studied.

Citations were collected from the 1964–76 volumes of the *Science Citation Index (SCI)* to each *paper* irrespective of the fact whether or not the individual studied was the first author. We attempted to minimize the errors originating from mis-spelling of sometimes long and – for a foreigner, strange – Hungarian names in such a way that we also scrutinized the close vicinity of the alphabetical order in the *SCI*. In doing so we found several, quite frequent, errors. For instance some authors use one or two initials alternatively, some citing authors – or the punchers in the ISI? – are rather careless in copying the bibliographic data of the cited paper which results in a large number of variation of a name like Szentágothai. An additional source of errors is the multiple transcription of names from Latin alphabet to an other (e.g. Russian) and back again, e.g. Szabó–Caбo–Sabo. Although we tried hard in eliminating such errors, we are inclined to think that no perfect job can be done in this respect.

For most of the papers we knew both the first and the last page. Thus when the page number was not identical with what we have found in the *SCI*, we accepted it if it was within the first and last page. If either the volume or the year of publication seemed to be misprinted, we accepted the identification, when only one digit was different in either one. "To be published" or "unpublished" papers were disregarded. With this procedure, based on a search by papers instead of names, we eliminated most of the homograph problem.

The type of citation was determined according to the following criteria:

I. *self-citation:* if the set of cited authors has element(s) common with the set of citing authors.

II. *cooperational citation:* the cited author under study and one citing author were co-authors prior to the publication of the cited paper and criterion (I) does not hold.

III. *independent citation:* no detectable relations between cited and citing authors.

These conditions were checked by the use of the *Source Index* which lists all citing authors, while the cited authors were known from the lists of papers we worked with.

Thus the result of the compilation of citation data was that the number of citations to each paper (duplicates included) was known as a function of time (in years) after its publications, and broken down into the three categories mentioned above.

In addition, the year of obtaining the degree "Candidate of Sciences" and "Doctor of Sciences"* was known for almost all individuals, so we could compare this sort of peer evaluation with their citation parameters at those times.

*In Hungary these degrees can be given to a candidate by a 7-member committee of the Academy of Sciences (not by universities) on the basis of a dissertation. The C. Sc. degree is equivalent to or a bit higher than the Ph. D., while the D. Sc. thesis "has to prove that the

Discussion

In his discussion *Garfield* arrives at the conclusion that citation counts measure the *utility* of the scientific results achieved by the individual. This is supported also by the good correlations between peer evaluation and citation counts. The citation data of most of these studies were, however, contaminated by self-citations and some other methodological problems arising from the primary-author effect and some others.

This bias has been eliminated by the classification of citations and by the all-authors search in our counting procedure, thus we can attempt, first to test the correlation between these *citation counts* compared with what we know about the individuals' D. Sc. degree.

The D. Sc. dissertation is usually a 100–200 pages summary of the candidate's scientific results achieved within the previous 5–15 years by him and his coworkers. It is a general requirement that most of the results should be published previously. The average number of papers the dissertation is based on varies between 5 to 30. The dissertation is usually focussed around one or two core problems.

In light of all this we can safely assume that at least the "core idea" or the solution of the "core problem" of the dissertation must be found in one of the candidate's papers, and – if it proved to be really significant contribution to the field – it should be cited with a relatively high frequency. Setting a level to a citation frequency of a paper above which it can be considered significant is fairly arbitrary. Still, if one takes into account on one hand that the average citation frequency to a paper is about 1.7 citation per year (self-citations included), and on the other hand that the distribution of citations is a very much skewed one, one can accept a level around one citation of type III per years as the level of significance. Table 2 shows the percentage distribution of the authors studied with respect to the criterion whether they published papers significant in the above sense at the time of their D. Sc. degree award, or before. Making allowance to the possibility that the most important results were being sent for publication at the time of presentation of the dissertation, counting was carried out for dates two years later than the D. Sc. award.

Considering the difference in citation habits in various disciplines which can account for the low percentages in the case of mathematics, agriculture and tech-

candidate has contributed ***significantly*** to the development of his special branch of science". Three of committee members have to present a written criticism of the dissertation at the public meeting of the committee. After a public discussion and a secret vote the committee makes recommendation to the Council about the D. Sc. award.

nology, we can conclude that these data are not inconsistent with the assumption that *peers judge the quality of scientific work by the same criteria as citing authors do*.

Given the classification of citation described above, the hypothesis that self-citations distort citation counts can also be tested. Table 3 shows a comparison of simple citation counts and type III citation counts (both sums for the 13 years

Table 2
Percentage of individuals publishing a minimum of "significant" papers prior to or at about the time of their D. Sc. degree award

Discipline	Minimum number of "significant" papers					
	1		2		4	
	%	N	%	N	%	N
Mathematics	44	4	33	3	22	2
Physics	96	22	78	18	48	11
Chemistry	50	4	13	1	0	0
Biology	80	4	80	4	60	3
Medical Sciences	77	10	69	9	69	9
Agriculture	20	1	20	1	0	0
Technology	60	3	0	0	0	0
Average/total	68	48	51	36	36	25

Table 3
Comparison of simple citation counts (cumulative for the 13 years searched) with or without self-citations

Discipline	Number of citations		Percentage of type III citations	Number of citations/paper		Average number of citations/individual	
	all	type III		all	type III	all	type III
Mathematics	288	225	78	0.667	0.521	32	25
Physics, all	5 222	3848	74	2.398	1.767	158	116
("visible")	(978)	(690)	(71)	(3.249)	(2.292)	(163)	(115)
Chemistry	1 280	616	48	1.656	0.797	142	56
Biology	1 496	914	61	3.958	2.418	299	183
Medical Sci.	5 376	4171	78	3.068	2.381	384	298
Agriculture	163	99	61	0.287	0.175	33	20
Technology	84	58	69	0.435	0.301	17	12
Total	14 193	9931	70	2.263	1.583	173	124

searched). Roughly about 30% of all citations proved to be direct or indirect self-citations* with no significant differences between the disciplines. The correlation coefficient between simple and type III citation counts calculated with respect to the data of each individual studied is r = 0.970 which indicates that – through not their number, but – *the effect of self-citations is negligible.***

This result gives support to earlier citation analyses based entirely on simple citation counts. Since most of the individuals of our sample seem to be medium cited ones if compared to world average,*** the very good correlation between simple citation counts with or without self-citations confirms the validity of such simplified analyses even for less cited scientists, provided that all-author search is made.[1]

At the first sight this finding was rather striking to us. One would think that there is a high number of scientists who publish quite a great number of papers citing their own previous work with no attention whatsoever from the scientific community. However, when the lists of papers were reviewed with respect to the journals publishing them, it turned out that most of the low-cited scientists published many papers in periferal journals the SCI does not cover. Considering that these are low-cited journals, the possible loss of some type III citations is accompanied with a proportional loss of self-citations, too. Thus one can understand why a low-cited author is simultaneously a low-self-cited one in spite of a rather long list of papers to his/her name. In the light of these results *Garfield*[1] seems to be right in assuming that low-stadard authors are forced into periferal journals, but consequently this does not necessarily increase the self-citation rate because the SCI does not cover these journals.

**Tagliazzo* (J. *Documentation*, 33 (1977) 251–265) counted the self-references "going out" of a sample of biomedical literature, and found that about 16% of all references is self-reference. Considering the difference between the nature of the two statistics, our results is not in contradiction with *Tagliazzo*'s.

**The high correlation might be spurious due to the different citedness of the disciplines. To check this hypothesis we calculated the correlation coefficients for aech branch separately. The results are: 0.976 (math.), 0.963 (phys.), 0.869 (chem.), 0.985 (biol.), 0.989 (med.), 0.985 (agric.) and 0.972 (technol.). Thus we can conclude that the high correlation is real and generally valid.

***Unfortunately we do not know about any other studies of a similar national sample which could be compared with these results. Although the individuals of *Garfield*'s studies[1] are much better cited than those in our sample, one can only conclude that the latter do not belong to the world elite. Yet they may not be inferior to a similarly selected sample of another nation.

With our national sample at hand, we could compare the citation counts with a computer search of the ISI on papers identified by them as Hungarain on the ground of the postal address of the first author.[4] They searched the 1973 source tape of their data-base and counted the citations to these papers in 1973–76. Table 4 shows the comparison of their results with our manual search on the 289 1973 papers in our sample. It indicates on one hand a rather large uncertainty in this type of national identification, and a fairly high error even in those cases when the identification worked. (The loss is mainly due to the fact that some 1973 papers were on the 1974 tapes of the ISI.) On the basis of these results, we must warn against individual assessment in large-scale computerized treatments unless the software used is checked against a careful manual search.

Table 4
Comparison of an ISI computer search with our manual counts for the 1973 papers in our sample

Kind of papers	N	%
Papers the ISI search did not identify as Hungarian, because they were published in journals the ISI does not cover		
(a) uncited in SCI	121	42
(b) cited in SCI	21	7
Papers with foreign first-author address, but with Hungarian co-author(s) all cited	14	5
Papers the ISI search did not identify as Hungarian, although they were published in ISI source journals*		
(a) uncited in SCI	10	3
(b) cited in SCI	81	28
Papers correctly identified as Hungarian		
(a) uncited by both searches	6	2
(b) cited by both searches (r = 0.92)	36	12
Total	289	100

*The loss is due to the fact that most of these papers are on the 1974 tapes because of delayed journals receipt.

The results of the present study also agree with *Garfield*'s observation in the fact that the two most cited authors of our sample (1374 and 1036 citations over the 13 years covered) published methodological papers. One of them is a neurophysiologist, the other one a physiologist. However, – sticking to the group of medical scientists – there are also some representatives of clinical medicine (e.g.

surgeons) who published methodological papers, yet they are much less cited. Thus citation rates seem to measure *utility in further research*, but not utility in every-day practice or production.

The data in Table 3 clearly demonstrate that citation counts cannot be used in comparing individual performance in different disciplines. Moreover, some comparisons within disciplines may be misleading as well.

Within the most numerous group of our sample, *the physicists*, two subgroups can be separated without dividing them into too narrow categories: nuclear physicists and solid state physicists. With similar consideration, mathematicians could be grouped into computer scientists and basic, "classical" mathematicians. While in Table 5 no marked difference can be observed between the two subgroups of matematicians and physicists, the division of the group of medical scientists into those dealing with basic medical research and those with clinical medicine reveals significant difference. We suspect that scientists engaged partially with the application of basic results outside the bulk of science are a bit "unfavoured" by citations in a way technologists are in comparison to basic scientists.

Table 5
A break-down of some disciplines into sub-fields

Discipline	Branch	Number of		Average number of type III citations	
		individuals	papers	per individual	per paper
Mathematics	basic	4	135	19	0.548
	computer	5	297	30	0.508
Physics	nuclear	17	1156	94	1.387
	solid state	10	567	126	2.217
Medical sciences	basic	7	848	481	3.969
	clinical	7	906	115	0.889

Such uncertainties make *the concerted use of citation analysis and peer evaluation inevitable*. The necessary caution can not in principle be exercised by science administrators with no research experiences of their own in the discipline they happen to analyse. The present authors feel that the reluctance of the scientific community in accepting citation analysis as an evaluation tool may rather be due to the fear of a fluorishing new type of bureaucracy than to the fear of its results. Were there some suitable measures eliminating this danger, citation analysis would lose most of its opponents.

The rest of the methodological problems in the use of citation analysis for the assessment of individuals or small groups are connected with two main problems: (1) how to compile citation data, and (2) how to derive the resultant quantities from them?

Since the number of papers per scientist does not exceed some hundreds even in the case of the most profilic ones, this number is not high enough to allow any *a priori*, deliberate, neglect of any paper of his list of publications. Thus, *the compilation of data must attempt completeness.* The distribution of the number of citations is so much skewed (see Fig. 1.) that the omission of a single well-cited paper can make the results unreliable. In light of this there seems to be no excuse for not making a comprehensive search; if the amount of work proves to be too great due to the size of the sample, it is better to desist from doing it rather than to risk being unfair to any individual.

A similar caution is to be exercised with respect to corrections for the obsolescence of papers. In a study of the "decay" of the number of citations on the

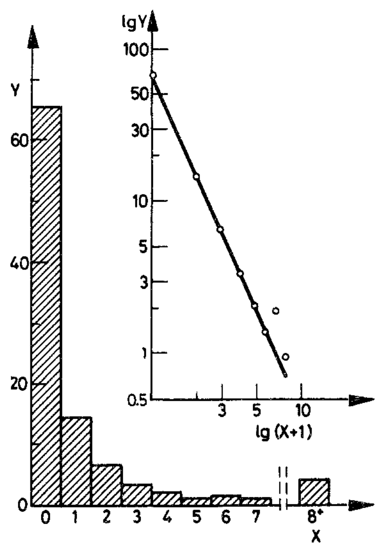

Fig. 1. Average distribution of the papers of an individual by the number of real impact citations obtained between the year of their appearance and 1976 (or – if they appeared before 1964 – between 1964–76). Y = percentage of papers, X = total number of citations. The log-log curve is plotted to show the distribution according to Lotka's law, in the form of $Y = \text{const}/[(X+1)^2]$ corresponding to the assumption that the appearance itself is worth a citation since the referees have read the paper and found some scientific information in it

same sample which will be published separately,[5] we show that the decay of citation frequency depends on the absolute citedness of the papers. The less cited the paper, the faster its obsolescence. Thus, if citations in the early years of the life-time of a paper are unknown due to the non-existence of the *SCI*, some corrections must be used, however, *not by the same* obsolescence curve as described in Ref.[6]

The more cited the paper in general, the more negligible the effect of the peak in the early years of the life-time of the paper on the average yearly citation frequency, whereas the less cited the paper, the more uncertain the rate of its obsolescence. No correction to decay could disfavour some papers published before about 1956 with an error of about 30% however a wrong estimate of the decay parameters could cause much greater errors. Thus we decided against making any correlation and characterized the citedness of a paper by the simple, un-weighted average of the number of citations it obtained divided either by the number of years of its life-time up to the year of investigation (for post-1964 papers), or by the number of years for which the SCI existed (for pre-1964 papers).

This average citation frequency has been divided equally to each co-author resulting in the fractional citation frequency attributed as the share of the studied author. The sum of this latter quantity over all papers of an individual under study comprised one ranking parameter.

We observed, however, that in some cases several of the papers of a given scientist were simultaneously cited by the same citing paper. More often the simultaneously cited papers were published in the same year. Thus the scientist who happens to break down his yearly work into several fractions and published them separately as letters or short communications had an advantage over those who published, say, a single but longer paper per year. In order to attempt the elimination of this bias, we have divided the sum of average fractional citation frequency of papers published in the same year by the sum of the fractional authorship of the given year. In this way we normalized the average fractional citation frequency to a single paper per year which could also be a ranking parameter with slightly different meaning.

This parameter, however, may "punish" those authors who – after a long era of publication-output-centered assessments – just adopted to the expectations of the scientific community, and published as many papers as they could pass through the referee barrier of the journals. To some extent almost every scientist in our sample seems to have such intentions which is indicated by the rather large proportion of uncited papers. It would, thus, be desirable to use a ranking parameter which takes only that fraction of an individual's publications into account which, in their citedness, scatters upwards from his/her average. We can apply the estimation of *Price*[3] by using only the citation frequency of the $\sqrt{N}$ number of the "best" papers (N = the total fractional authorship of the scientist in question). In calculating this quantity we ranked the papers of each individual, and added up the yearly average fractional citation frequency for the first, second, etc. papers until their fractional authorship made up (or first exceeded) $\sqrt{N}$. When papers happened to have the same yearly average fractional citation frequency, the ranking was done by the year of appearance with preference to earlier papers.

Another advantage of the use of this measure as a ranking parameter is that it distinguishes between ways how a scientist accumulated the citations by his papers: by a few but outstanding papers or by almost all of his papers with a low average citedness. In the latter case it is very likely that the citations obtained are not "organic" citations, but "perfunctory" ones.[7]

Table 6 shows the linear correlation coefficients between the possible parameters discussed above. It is clearly indicated that all quantities based on citation counts measure *something different* from those representing the output of a scientist. The highest correlation coefficient between these two groups of parameters is that of authorship and simple citation count which by all means is due to the effect of self-citations. Still, we have to emphasize, that this coefficeint is essentially lower than that found by *Cole* and *Cole*[8] for their sample of physicists in some issues of the *Physical Review*. This difefrence may stem from the fact that our sample is more uniform in publication output than theirs, and so the trivial correlation between low output – low citedness may increase the correlation coefficeint.

Table 6
Linear correlation matrix of the various parameters used in this study

x \ y	(1) Authorship	(2)	(3)	(4)	(5)	(6)
(2) Fractional authorship	0.520	–				
(3) Simple citation count	0,332	0,076	–			
(4) Type III Citation count	0.083	0.007	0.970	–		
(5) Sum over papers of yearly average citation	0.003	0.001	0.910	0.938	–	
(6) Sum over years of yearly average citation frequency normalized to a 1 paper/year/author productivity	0.000	0.000	0.730	0.766	0.634	–
(7) The same as (5) but summed only over the "best" papers	0.000	0.000	0.819	0.853	0.903	0.739

(N = 80)

It is also remarkable how strict the correlation is between type III citation counts and all of its derivated quantities [see column (4)], whereas the data in line (6) indicate that, in general, there is a superfluous amount of papers written by some scientists with no scientific utility which lowers the average citedness of their papers.

On the basis of these results we suggest that *the ranking parameters listed in Table 6 can be used in a multidimensional citation analysis scheme for individual scientists with simultaneous attention to each citation parameter in such a way that whenever they give drastically different ranks, the case is to be considered with special caution.*

References

1. E. GARFIELD, Is Citation Analysis a Legitimate Tool?, *Scientometrics*, 1 (1979) 359–375.
2. S. PAPP, I. RUFF, Scientometric Study of Branches of Science, III, An Investigation of the Literature of Hungarian Coordination Chemistry between 1934–1977, *Kém. Közl.*, 51 (1979) 345.
3. D. de SOLLA PRICE, *Little Science, Big Science*, Columbia University Press, 1973.
4. M. V. MALIN, private communication.
5. B. HAJTMAN, J. I. NAGY, I. RUFF, to be published.
6. N. L. GELLER, J. S. CANI, R. E. DAVIES, Life-time Citation Rates as a Basis for Comparisons within a Scientific Field., *Proc. Am. Stat. Assoc.*, Soc. Stat., Washington, D. C., 1975, pp. 429–433.
7. I. RUFF, Citation Analysis of a Scientific Carreer: A Case Study, *Social Studies of Science*, 9 (1979) 81–90;
   M. J. MORAVCSIK, P. MURUGESAN, Some Results on the Function and Quality of Citations, ibid., 5 (1975) 86–92.
8. J. R. COLE, S. COLE, *Social Stratification of Science*, University of Chicago Press, Chicago, 1973.