

## MATCHING AUTHORS AND READERS OF SCIENTIFIC PAPERS

M. KOCHEN and R. TAGLIACCOZZO

Mental Health Research Institute, University of Michigan, Ann Arbor, Michigan

*(Received 6 March 1974)*

**Summary**—A novel service is proposed to help authors of scientific papers to select journals to which to submit their manuscripts. The aim of the service is to optimize the common interests of authors, readers/users, and science, by providing the authors with a tool for reaching those readers who can most benefit from the news of their accomplishments. The five basic factors which intervene in the choice of a journal are analyzed: relevance, acceptance rate, circulation, prestige and publication lag. With the help of a mathematical model, we show how these five variables can be used in a plausible decision procedure leading to the selection of journals suitable for the publication of given papers.

### 1. INTRODUCTION

WHEN a scientist submits an article for publication, the choice of a suitable journal is often guided by self-interest. It helps the scientist's career if the results of his work are published in a journal representative of his discipline. It helps most if the journal is the most prestigious in the subject area, with the largest circulation, and with the shortest publication lag. If an author chooses wisely where to publish, this may also serve the interests of the potential readers of his papers. Their concern is to gain valuable information with the minimum possible delay. There is, therefore, a convergence of interests on the part of authors and readers that the proper audience for the reporting of scientific investigations be reached. The healthy growth of the scientific enterprise—and also of enterprises based on the utilization of knowledge—is dependent on the matching of these complementary interests.

Some scientists find it difficult to select wisely the journals to which to submit their papers. Scientists who are just starting their careers, who do not have experienced colleagues to advise them, and who do not yet belong to any "invisible college", may experience difficulties. So may scientists who have just switched from one specialty to another, or people who work in interdisciplinary fields or in fields without a well-developed "paradigm". Almost any scientist, even the most experienced, has probably felt, at one time or another, that he had submitted some of his papers to the wrong journal or addressed it to the wrong audience.

A good service to advise an author on how to reach the readers who can benefit most from news of his accomplishments would be helpful to him, to his potential readers, and to science. To justify such service, one should be able to show that it would fill a real, important and sizeable need, and that it would do as well or better than the advice offered by experienced professional colleagues.

Some scientific foundation is required to design such a service with assurance that it will be helpful. Prior work on the sociology of publishing [1-3], on the psychology of authorship [4-5], and on the patterns of communication in a scientific community [6-12] provides the starting point of this study. The "Bradford law" [13-14] states

that most readers as well as authors in a specialty concentrate on less than a dozen core journals, but that a fraction of their publication or of their reading time is scattered over many hundreds or thousands of widely ranging journals. This will serve as a springboard for some of what follows.

The central problem to be addressed here is that of defining the best way authors should distribute their papers over available journals to advance their own interests and those of the readers.

## 2. A PROPOSED SERVICE

Imagine an author who has just made the decision to write an article for publication. Perhaps he has already developed an idea or established a finding that he thinks is sufficiently significant, valid, and novel to be shared with his colleagues and other potential users of whatever information he has to offer. If the topic of his contribution is so specialized that there are only a few such colleagues and potential users, and he has them clearly in mind, then he has no need for the proposed service. He follows established lines of communication—report, preprint, a specialized journal—and is assured that he will reach his audience.

The topic may be such that in the intellectual community there are a few leaders through whom news of significant contributions is spread to where it does the most good. If the author has those key leaders clearly in mind, he does not need the proposed service either. The author we are considering is one whose article is not on such a topic, or who does not know who the key leaders are or how to reach them. We assume, however, that he knows that it matters whether the proper audience will be reached and that his concern will motivate him to use the proposed system.

The sooner after planning his paper the author requests aid from the proposed service, the more cost-effective such aid can be. Suppose he presents to a hypothetical “counselor” the gist of his planned paper immediately after making the decision to write for publication. The counselor examines the author’s prior publication pattern, if any. Through a dialogue, he helps the author articulate his plans for intellectual growth. He tries to represent the potential referees, editors, readers and users of the proposed paper. Together, and with the help of a data base containing information on who is concerned with what problems, they identify possible audiences to reach and means (journals) for doing so. There may be constraints, such as the impossibility of reaching a certain mix of, say, three audiences. These are made explicit. Together, author and “counselor” formulate a plan for selecting a journal that would maximize utility to both the author and users. Then, using a decision procedure which optimizes the weighted total utility without violating the constraints, a journal is recommended to the author.

What is the need for such a service? Even though scientific communities of the most advanced countries (e.g. the United States) are stabilizing their sizes due to reduced financial support, they still receive a steady stream of newcomers and field switchers. Furthermore, other advanced areas (U.S.S.R., Western Europe, Japan) are expanding, and provide a growing stream of new potential users for the author’s contributions. Recent, subtle changes in the social structure of science also make the service appear valuable. There may well be a reaction against the present climate of over-specialization, with a trend toward articles that could and should reach larger audiences. Changes in the choice of problems toward those of greater relevance to contemporary issues to

which science can be applied has been creating instabilities in the elite structures of the scientific establishment. Authors, like other citizens bewildered by the complexity of the situations they must cope with and by the variety of options available to them, will increasingly need information services that help orient and advise them. The class of authors we have described as needing the proposed service is likely to expand.

If the service we propose is needed, can it be supplied and will it meet the need and the demand competitively with alternative solutions? It is obvious that the services of an adequately qualified counselor would be both hard to obtain and too expensive for authors most in need. A less costly and more easily supplied service is to provide the author with a table, giving for each journal the size of its readership, its editorial objectives and policies, its acceptance rate, its publication lag, the fraction of its articles which are significant, and other pertinent information. Such a table may give the author more guidance than would otherwise be available to him on the journals eligible to publish his article. The number of journals is far greater than an author can be cognizant of, and there may be new journals, unknown to him, for which his article might be ideal. Of course, for the table to be useful, it must already be restricted to less than 100 or so journals likely to fit any given author. If the data base contained over 10,000 journals, screening would require some initial input from the author of the kind that would have been elicited by the counselor in a dialogue. Whether such a selective screening system can be provided at a low enough cost and high enough sensitivity is a key problem. Another crucial factor consists of selecting the kind of output data about journals which would help the author to make wise decisions.

A more complex version of the service performs higher tasks. It submits the manuscript on the author's behalf. It chooses the journal for him, with his approval. If necessary, it resubmits the manuscript to another journal. At this level, the service can also help journals by pooling referees and their comments. The referees' reactions are used in choosing the journal to which to submit a given manuscript or to advise the author on how to revise it.

At an even higher level, the service revises the manuscript for the author to increase the likelihood of its being accepted by the journal into which it is judged to fit best. At this level, the service also serves the readers by directing specific articles and journals to them. Several higher levels of the service can be envisaged. Of course, cost increases with level. So does benefit to authors, users and science.

Whether such a spectrum of services can compete with existing means whereby an author seeks the informal advice of experienced colleagues or mentors depends on how good such advice happens to be. It is, of course, not free. If experienced colleagues invest significant time or energy in providing serious counselling, the cost *to them* is considerable. Their advice may be hasty and superficial, but likely to be taken by their junior colleagues because of their status and success. Experienced authors, too, would find the service helpful, not only in providing them with up-to-date information about journals in which to place their own publications (e.g. who is on the editorial board, who are the referees, what are the latest editorial policies, the publication lag), but also in helping them to follow and guide the development of their own ideas and findings in the works of junior colleagues. From this point of view, it is in their interest to become better counselors and to have their counselling amplified so that it can reach greater numbers with greater effect.

A workable service of the kind we propose depends on at least the following:

(1) reliable knowledge of what variables to elicit from the author and what variables describing journals to supply him or his counselor;

(2) a good decision procedure for recommending an optimal choice of journal and a longer-range publication strategy to the author.

To clarify what these two conditions mean, the logical precision of a mathematical model (Section 4 of this paper) of the service is of value. Prior to discussing the model, some background data about present publication patterns, presented in Section 3, will both illuminate the above discussion and justify some of the assumptions of the model. In the final section, we use the model to provide tentative conclusions about optimal publication strategies as a basis for the proposed service. This may help to design the service, and it may also help some authors to plan their publication strategies before the service is available.

### 3. HOW AUTHORS SPREAD THEIR PUBLICATIONS OVER JOURNALS

We examined the 1966–1970 publications by scientists from four departments of the University of Michigan Medical School as listed in the annual “University of Michigan Bibliography: Scholarly and Creative Activities of the Faculty and Staff” (Office of Research Administration, The University of Michigan, Ann Arbor, Michigan). We found that a large number of articles were concentrated in a small number of journals. These “core” journals are the periodicals with which every specialist in the field is most familiar and where he looks first for the most important new research. They are also the journals whose articles are most frequently cited and, it can be assumed, have the higher “impact” on the scientific community.\*

The journals of the next category, which we called “major” journals, published a more modest number of contributions during the 5-year period under examination. This group shows an interesting mixture of very general and very specialized journals. They appear to be an important source of information for the reader working in a specific field, although they are not read as regularly or as widely as the “core” journals.

The third group of journals, which we called “peripheral”, includes journals that over the 5-year period published only one paper produced by the departments under consideration. This group of journals is very heterogeneous. The reasons for their low popularity with the authors of our sample are in some cases evident, in other cases, less so. The term “peripheral”, therefore, should not be taken literally to indicate lack of relevance or importance, but rather to illustrate their secondary position as communication media for the authors of our sample.

In Table 1 we summarize the distribution of articles in the three groups of journals defined above. We placed in Group 1 those journals which published more than five articles in the 5-year period. The maximum number of articles published by any one journal was 61 for the Department of Biological Chemistry, 43 for the Department of Surgery, 32 for the Department of Human Genetics, and 12 for the Department of Psychiatry. In Group 2 we placed journals which published between two and five articles, and in Group 3, those which published only one article in the 5-year period.

\*Seven of our eight “core” journals for the Department of Biological Chemistry and four of the six “core” journals for the Department of Human Genetics (see Table 1, Group 1) appeared in the list of the 152 most frequently cited journals in the biological sciences, ranked by “impact factor” [15]. The five higher ranking “core” journals of our biological chemistry sample also appeared as the 1st, 3rd, 5th, 10th and 7th in an updated list of high-ranking chemical journals [16].

Table 1. Journals which published articles authored by scientists in the four departments of the medical school at the University of Michigan (1966–1970)

Departments	Group 1 (published more than 5 articles)		Group 2 (published 2–5 articles)		Group 3 (published 1 article)		Totals	
	No. of journals	No. of articles*	No. of journals	No. of articles*	No. of journals	No. of articles*		
Biological chemistry	8	140	14	38	44	44	66	222
Human genetics	6	79	38	116	61	61	105	256
Psychiatry	9	74	26	76	69	69	104	219
Surgery	17	264	26	69	56	56	99	389

\*Not included in the count: Abstracts, book reviews, chapters in books, papers presented at special symposia or meetings.

Some differences in spread of publication can be noticed. It is evident, for instance, that the publications generated by the Departments of Biological Chemistry and of Surgery are concentrated in a relatively smaller number of journals than those originating from the Departments of Human Genetics and Psychiatry. The journals of Group 1 are in fact responsible for well over 60 per cent of all the papers produced in the Biological Chemistry and Surgery Departments (63 and 68 per cent respectively), whereas they account for only 31 and 34 per cent of the total in the other two departments.

The pattern of publication is shown in more detail by the graph of Fig. 1, which plots the rank of journals that published articles authored by members of the four departments (ordered according to decreasing number of published articles) against the cumulative percentages of published articles. The four higher ranking journals include 50 per cent of the total number of articles published by members of the Department of Biological Chemistry, but only 31 per cent of the total number of articles produced by the Department of Surgery, 25 per cent of those authored by members of the Department of Human Genetics, and 19 per cent of those from the Department of Psychiatry. It seems, therefore, that the distribution of papers in the journals available to scientists working in different areas of the biomedical sciences follows patterns which present some dissimilarities.

This points to the possibility that the users of the service we propose may have different needs and preferences in their choices of publication strategies. Whether authors' choices can be predicted on the basis of some general characteristic of the field in which they work—be this, for instance, a “hard”-science discipline as opposed to a “soft”-science discipline, or a basic-science area as opposed to a clinical-science area, or a specialized field rather than an interdisciplinary area—is impossible to tell at this point. Our data suggest that one of the factors responsible for the difference in spread of publication may be the seniority of the disciplines that we compared. There is some similarity, in fact, in the shape of the curves for the Department of Biological Chemistry and of Surgery—both older departments—and of those for Psychiatry and Human Genetics, which have a more recent history (see Fig. 1). It could be hypothesized that scientists from older, well-established disciplines tend to concentrate their

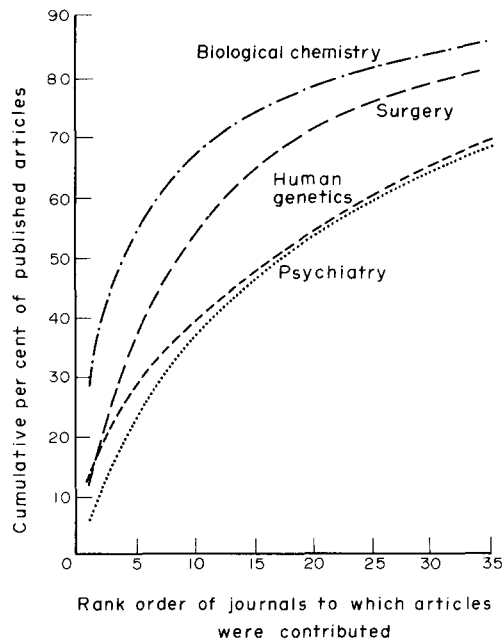


Fig. 1. Spread of journal articles generated by four departments of the University of Michigan Medical School, 1966–1970.

publications in a few, representative journals, while authors in more recently established fields, which do not yet possess a well-developed “paradigm” and a well integrated communication system, tend to spread their publications more widely, at times invading journals of other fields or interdisciplinary journals. It does not seem, at this point, that our data can support the conclusions of GOFFMAN and WARREN, suggesting a constant ratio of journals to authors for all medical literature [17].

Another comparison of some interest is that between authors who can be considered very productive and those who exhibit moderate or low productivity. We know, as stated by the inverse-square law of productivity, that the number of people producing  $n$  papers is proportional to  $1/n^2$ ; consequently, a small number of very productive authors is responsible for a large portion of the published literature [18]. It seems important, therefore, to investigate whether the highly productive scientists present publishing patterns which differentiate them from the less productive authors. To this effect, we examined the distribution of the articles published by those authors (in the four departments of our sample) whose bibliographies had been listed in at least 2 of the 5 years between 1966 and 1970. As an operational definition of higher productivity, we took the publication of five articles or more in a single year.

We noticed a tendency of the higher-productivity group to concentrate their articles in a relatively smaller number of journals than did the lower-productivity group. The average ratios of total number of published articles to number of different journals in which they were published were 4.3, 3.4, 2.3 and 4.8 for the higher-productivity authors of the Departments of Biological Chemistry, Human Genetics, Psychiatry and Surgery, and 2.8, 2.2, 2.2 and 3.3 for the lower-productivity groups. Furthermore, in three of the

four departments the higher-productivity authors showed a trend toward publishing a relatively higher percentage of articles in the “core” journals than did the lower-productivity authors. The differences did not, however, in the majority of cases reach statistical significance, and these results, although suggestive, have to be considered inconclusive at this time.

#### 4. CHOOSING A JOURNAL: A MODEL

By what criteria can we judge a *good* publication strategy, i.e. a strategy from which authors, readers and the institution of science can all benefit? Authors aim at fame, fortune, power and many other rewards of a personal nature that they share with people in other professions. But they also aim at spreading the news of their own scientific achievements for less selfish reasons; for instance, to advance science or to improve the quality of life. Readers, too, are motivated by self-interest, such as the wish to read about useful new techniques or ideas in order to improve their performance or to avoid duplication in research. Besides these personally motivated goals, however, readers want to know about other scientists’ work because they hope it will help them to make significant contributions to the progress of science. “For the development of science, only work that is effectively perceived and utilized by other scientists . . . matters” [5]. In discussing publication strategies, we wish to focus on the complementary interests of authors and readers rather than on the possible conflicts among diverging private interests.

We must assume that the most successful scientists are those who not only produce significant results but also succeed in communicating them to the appropriate audience. Nonetheless, it is possible for an author of less significant achievements to make an impact, too, by virtue of a good publication strategy. Conversely, it is possible that a scientist with significant accomplishments may fail to have appropriate impact due to a poor publication strategy. Carrying these opposite positions to extremes, we could envisage insignificant achievements being published—with consequent waste of journal space and readers’ time—and significant achievements not being published—with loss of worthy contributions to the scientific community. The implications of inefficient publication policies are serious enough to justify a careful analysis of this aspect of the scientific communication system.

Imagine a potential author  $a$  who is about to complete a manuscript  $d$  and must choose a journal to which to submit it. Let  $j, j'$  denote two journals he is considering. In order to express some plausible assumptions and deduce their implications, we need to define the key variables that we think characterize this selection process. Although many personal, subjective considerations certainly intervene in the author’s final decision, five factors of a general nature are probably at the basis of any choice of this kind. They are (a) relevance, (b) acceptance rate, (c) circulation, (d) prestige, and (e) publication lag.

(a) *Relevance*. This variable characterizes the extent to which a given document  $d$ , by being published in journal  $j$ , reaches the “right” readers.\*

Let  $R(j) = \{r: r \text{ is a person who has read at least one article in } j\}$ .

$C(r) = \{q: q \text{ is a problem with which } r \text{ is or has been concerned}\}$ .

\*For readers unfamiliar with the subsequent notation, the colon is to be read “such that.” For example,  $\{x: x > 1\}$  denotes the set of all  $x$  such that  $x$  is greater than 1.

$s_a(d, j)$ : author  $a$ 's judgment that his own concern in authoring document  $d$  falls into  $C(r)$  for some  $r \in R(j)$ . This may be just 0 or 1, or on a scale from 0 to 1.

$j \stackrel{\geq}{a,d} j'$  means: author  $a$  prefers journal  $j$  to  $j'$  or is indifferent between them as to which to submit  $d$ .

*Assumption 1*:  $j \stackrel{\geq}{a,d} j'$  if  $s_a(d, j) \geq s_a(d, j')$  for all  $a, d, j, j'$ . In plain terms, an author prefers to submit his manuscript to that journal which he judges to have at least one reader who shares his interest in the paper.

(b) *Acceptance rate*. We know that the acceptance rate of journals varies widely, some journals rejecting well over 80 per cent of the submitted manuscripts, while for other journals the rejection rate can be as low as 20 per cent. In a sample of 83 journals from various disciplines, ZUCKERMAN and MERTON [1] found that the rate of rejection is highest in the humanities, followed by the social and behavioral sciences, mathematics, and statistics. The physical, chemical, and biological sciences have the lowest rates, equivalent, on the average, to one-third the rates found in the humanities. Various factors besides the quality of the submitted manuscripts are to be considered responsible for the differences in rejection rates; for example, page space, length of articles, the page charge, the backlog of accepted manuscripts. Not only is the rate of rejection higher in the "soft" as opposed to the "hard" science journals, but also in the interdisciplinary journals as compared to specialized journals. LODAHL and GORDON detected a relationship between rejection rates of the various journals and paradigm levels of the disciplines to which the journals belong [19].

A high acceptance rate is a favorable factor in the selection of a journal for publication. No author, in fact, cherishes the thought of having to submit his paper more than once. In some cases, one rejection is enough to discourage him from further attempts to publish. The publication delay due to rejection and resubmission is not only annoying and frustrating for the author; it is also damaging for the potential readers and the scientific communication process at large.\*

Let  $a(j)$  = acceptance rate of a journal  $j$ , or fraction of all manuscripts received by  $j$  that are eventually published in  $j$ , assumed known to  $a$  or estimated by  $a$ .

*Assumption 2*:  $j \stackrel{\geq}{a,d} j'$  if  $a(j) \geq a(j')$  for all  $a, d, j, j'$ . In words, an author prefers the journal with the higher acceptance rate.

(c) *Circulation*. The diffusion and visibility of a journal article is directly related to the number of copies of the journal sold.†

*Assumption 3*:  $j \stackrel{\geq}{a,d} j'$  if  $|R(j)| \geq |R(j')|$ . By  $|R|$  is meant the number of elements in set  $R$ , i.e. the number of readers. It is possible for a journal to have a small total circulation but a large circulation in a narrow specialty, making it significantly representative of that specialty. This is true of many official publications of scientific societies

\*It is estimated, for instance, that about one-fifth of articles published in the main psychology journals have been previously rejected by one or more journals [9]. For these articles, then, the lag between completion of the manuscript and publication is considerably longer than the 9-month average lag for articles published in the journal of first choice. GARVEY *et al.* [8] found that the delay in publication due to previous rejection was 8 months for social science manuscripts as compared to 4 months for physical science manuscripts.

†Journal circulation depends on many factors, the foremost being the number of people working in the field represented by the journal. The quality of the journal and its price are also important. The only overview of circulation trends that we were able to find is in the Report of the Task Group on the Economics of Primary Publications [20]. According to this report, most circulation figures are between 1500 and 15,000 copies. Only in a few fields do journals achieve circulation of over 15,000. Medicine is one of these: circulations of the order of 50,000 or more occur, especially for journals with clinical orientation.



with a limited membership. In such cases, the question of the total number of copies printed and sold is of secondary importance. The author is sure that by publishing in that journal he will reach practically all the readers to whom his paper may be of interest.

(d) *Prestige*. The prestige of a journal depends to a large extent on its life history. The longer it has existed, the greater are its chances of being seen and cited, and of citations from it being seen. The greater also are its chances of attracting the best authors and improving its quality. Older, well-established journals with a wide circulation and a well-defined readership will be more prestigious than journals in new research areas which have “not yet been legitimized” [21]. Other factors contributing to the prestige of a journal are the status of the editor and the editorial board, and whether the journal is the official organ of a prestigious scientific association. Recent attempts have been made to provide a measure of the importance of journals by ranking them on the basis of their “impact factor”, that is, the average number of citations per item published in the journal [15].

It is a well-established fact that in each discipline there exists a hierarchy of journals, ordered according to prestige. Rarely, if at all, does an author submit his manuscript first to a journal lower in the hierarchy and then, if not accepted, to a higher one. The reverse is the usual practice, as shown by studies in the area of sociology and education [22, 23].

Let  $r_a(j)$  be a’s judgment about the rank of  $j$ ’s readers and authors.

*Assumption 4:*  $j \succcurlyeq_{a,d} j'$  if  $r_a(j) \geq r_a(j')$ . Simply put, an author will submit his manuscript to the journal that has higher-ranking authors and readers.

(e) *Publication lag*. Prompt publication is in the the interest of authors, readers, and science alike. This may be inversely related to prestige, because the great volume of submitted manuscripts causes a backlog.

Let  $T(j)$  be the mean time (in weeks) from the date a manuscript  $d$  is officially received by  $j$  to the date when it appears in circulation.

*Assumption 5:*  $j \succcurlyeq_{a,d} j'$  if  $T(j) \leq T(j')$ .

Suppose that  $T_0(j)$  is the mean time elapsed from the date at which  $d$  is submitted to  $j$  until an acceptance/rejection decision is made. (Acceptance here includes conditional acceptance.) If the manuscript is rejected, suppose that the author resubmits it to some other journal with probability  $R$ ; the probability that he abandons publication attempts is  $1-R$ . Suppose further that, if  $d$  is rejected by  $j$ , a resubmits it to  $j_1$ ; if  $j_1$  rejects it, he resubmits to  $j_2$ , etc. The expected delay in getting  $d$  published is then

$$T_a(j_1, j_2, j_3, \dots, j_n) = \sum_{k=1}^n a(j_k) R^{k-1} \sum_{e=1}^k T_0(j_e) \frac{k-1}{e-1} (1 - a(j_e)).$$

In general, we can regard resubmission as a Markov chain with the following  $j+1$  states: “submit to  $j$ ,”  $j = 1, 2, \dots, j$ , where  $j$  is the number of journals; “Abandon publication”.

We have already remarked that the publication of a paper by a prestigious journal may suffer a longer delay than if it were submitted to a less prestigious one. Relevance of an article to a particular journal may at times be in conflict with the prestige of the journal or its readership size (circulation). A known high rejection rate may keep authors from submitting to a journal which would otherwise present attractive features.

It is part of the publishing strategy of the author to evaluate these different factors and weight the tradeoff between conflicting advantages and disadvantages. We could suppose that  $a$  computes the function:

$$f_a(d, j) = w_{a,s} s_a(d, j) + w_{a,a} a(j) + w_{a,r} |R(j)| + w_{a,r} r_a(j) - w_{a,t} T(j)$$

and then selects  $j$  for which  $f_a(d, j)$  is maximum, subject to constraints, such as  $s_a(d, j) \geq s_{\min}$ . The quadruple of weights  $w_a = (w_{a,s}, w_{a,a}, w_{a,k}, w_{a,r})$  characterize an aspect of  $a$ 's publication pattern.

Similar considerations apply to how a reader  $r$  selects a journal. In place of  $s_a(d, j)$ , we have  $s'_r(j)$ . This is reader  $r$ 's judgment about the size of  $C(r) \cap C(a)$  for some author of an article in  $j$ . It is a reader's estimate of how many *problems* are of concern to *both* some author and this reader. The larger the reader's estimate of that number for authors in a given journal, the more he is likely to select that journal, if other factors are the same.

In the place of  $R(j)$ , we put  $D(j)$ . This is the set of articles per year published by  $j$ . In place of  $r_a(j)$ , we take  $r_r(j)$ , the reader's judgment about the rank or prestige of  $j$ 's authors and other readers.

Let  $g_r(j) = w_{r,s} s'_r(j) + w_{r,d} |D(j)| + w_{r,r} r_r(j) - w_{r,t} T(j) - p$  where  $p$  is the subscription price to the journal or some other measure of the cost of  $r$ 's effort to get  $j$ . Suppose that  $r$  selects  $j$  for which  $g_r(j)$  is maximum, subject to constraints such as  $s'_r(j) \geq s'_{\min}$ ,  $|D(j)| \leq D_{\max}$ ,  $p < p_{\max}$ , etc.

Assume now that the most significant accomplishments are the rarest. Does that mean that the less significant findings should not be published? Far from it. Major accomplishments are often based on a large mass of modest results, sometimes in widely scattered topics. For the significant results, assuming that they are so recognized, the key problem is to make sure that they do not fail too often to come to the attention of potential significant users. For the less significant results, the main problem is to make sure that they reach at least one significant user. To simplify the discussion, let us confine the users to those who aim at adding to the store of cumulated knowledge and exclude those who use the findings of a scientific specialty to improve the quality of life. A significant user, then, is one who produces significant results.

Let  $p$  be the probability that a significant user is informed about a significant result. Let  $q$  be the probability that a significant user is informed about a less significant result that he eventually uses. If we assume that the events of a significant user being informed about various results are independent, then the probability that a significant result reaches  $m$  or fewer significant users is

$$\sum_{k=0}^m \binom{m}{k} p^k (1-p)^{m-k},$$

where  $m$  is the number of significant users and  $1 - (1-q)^m$  is the probability of at least one significant user seeing a less significant result that he eventually uses. It is desirable that  $q$ ,  $p$  and  $m$  be as large as possible.

Let  $P(j)$  be the probability of a significant result in  $j$  in 1 year.

Let  $P'(j)$  be the probability of a significant user using journal  $j$ .

If the optimal journal choices by the authors coincided with the optimal choices by the readers, then all authors would submit to the few core journals and all readers

would confine their attention to these. The core journals would, however, receive many more submissions than they could accept without expanding their size. Hence they would cease to serve readers. An oversized journal publishing  $D$  articles/year could be split into  $n$  smaller journals, each publishing  $D/n$  articles/year. The readership would not, however, be partitioned the same way. The question then arises: which articles get into which journals?

Let  $|N(j)|$  be the number of manuscripts/year which are submitted to journal  $j$ . It is easy to see that

$$|D(j)| = a(j) |N(j)|.$$

Let  $b(j, j')$  be the fraction of all manuscripts submitted to  $j$  that are rejected and which are resubmitted (as rebounds) to journal  $j'$ , and let  $|F(j)|$  be the number of manuscripts/year which are first submissions to journal  $j$  (no rebounds). Then,

$$|N(j)| = |F(j)| + \sum_{j'} |N(j')| (1 - a(j')) b(j', j) \text{ for all } j.$$

If  $|F(j)| = \alpha |N(j)|$ , where  $0 \leq \alpha \leq 1$ , then we can write

$$\mathbf{N} = \alpha \mathbf{N} + \mathbf{M}\mathbf{N},$$

where  $\mathbf{N}$  is a vector with components  $N(1), N(2), \dots$ , and  $\mathbf{M}$  is a square matrix with element  $(1 - a(j')) b(j', j)$ . Further,

$$\mathbf{N}((1 - \alpha) \mathbf{I} - \mathbf{M}) = \mathbf{0}$$

where  $\mathbf{I}$  is the identity matrix and  $\mathbf{0}$  the vector of all 0's. There is a non-trivial solution for  $\mathbf{N}$  to this matrix equation only if the determinant  $\|(1 - \alpha)\mathbf{I} - \mathbf{M}\|$  is 0. This means that  $(1 - \alpha)$  can take on only the eigenvalues of  $\mathbf{M}$ .

An interesting problem arises when an author judges his paper to be better than the average paper published by  $j$ . Assume that his judgment is shared by most others. If enough authors like him submit their manuscripts (and they are accepted), the average quality of  $j$  goes up. If they don't submit to  $j$ , then  $j$ 's average quality goes down. By submitting his manuscript, the author does more for the journal than the journal does for him. A more established author can afford this and should do it for journals he wants to improve. A younger author might do better to get into journals where the average paper is at least as good as his.

##### 5. OPTIMAL PUBLICATION STRATEGY: A POINT OF VIEW

If each journal published articles on a well-defined, precisely specified topic – and if journals were precisely labeled by an adequate indexing and classification scheme, the goal of good author-reader matching could be reached and only the problem of quality control would be left. The counselling service is then primarily one of helping authors and journals to improve quality. Each journal can deal with the quality-control problem with the present peer-review system by improving methods of selecting referees, eliciting better judgments from them, and devising new methods of using

their judgments to reach decisions about publication or revision. Some work on this problem has already been done [24].

If we assume that a significant user's using  $j$  is independent of  $j$ 's containing a significant result (not plausible), then

$$p = \sum_j P(j)P'(j) \text{ and } q = \sum_j (1 - P(j)) P'(j).$$

If we drop this independence assumption, we must define the conditional probability  $Q(j)$  that a significant user uses  $j$ , given that it contains a significant result, and

$$p = \sum_j Q(j)P(j) \text{ and } q = \sum_j Q(j)(1 - P(j)).$$

If one of the  $m$  significant users produces an average of  $k$  significant results over a professional life of 40 years, the probability of at least one significant result appearing in some one of  $j$  journals (averaging  $D$  papers/year) is  $s = mk/40DJ$ .

If  $Q(j) = aP(j)$ , then  $p = a \sum_j P^2(j)$  and  $q = as - p$ . To maximize  $p$ ,  $P(j)$  should be as concentrated as possible; ideally, one journal should contain all the significant results. To maximize  $q$ ,  $DJ$  should be as small as possible; ideally, only the significant results should be published. But  $s$  depends on the number of less significant results as well as on the number of significant ones, perhaps with  $ds/dt$  varying as the product of these two variables. Here  $t$  stands for time. If  $ds/dt = ks(1 - s)$ , then  $s = Ce^{kt}/(1 + Ce^{kt})$ , which tends to 1 as  $t \rightarrow \infty$ . If  $k > 0$ , this is contrary to a statement by de Solla Price that the number of good (significant) articles varies as the square root of the total number of publications:  $mk/40 = K'\sqrt{DJ}$ ; this implies that  $s = K'/\sqrt{DJ}$ , and if  $DJ$  increases as  $Ae^{Bt}$ , then  $s = K'/A e^{(-Bt/2)}$ . This suggests that  $k < 0$ , and  $s \rightarrow 0$  as  $t \rightarrow \infty$ . Even if the number of publications increases only linearly,  $s$  still decreases with time. This means that the fraction of less significant papers increases. There must come a time when the ratio of less significant publications to significant ones is too great for all to be used. At that time the number of publications per year should become constant or decrease.

While the literature still grows, the significant findings should be sprinkled throughout all journals, not concentrated in a few, so as to increase the likelihood that less significant yet useful articles will be found by significant users. Concentrating *all* the significant articles in a few journals would leave significant users little motivation to examine any journals other than these few. The best strategy is probably to select journals by maximizing the functions  $f_a(d, j)$  and  $q_r(j)$  defined in Section 4, augmented by a random variable so that a chance element is deliberately introduced to distribute articles over journals outside their specialties and to mix significant with less significant items.

## 6. CONCLUSIONS

We propose the development of a novel service to help authors of scientific papers to select journals to which to submit their manuscripts. The aim of such a service is to optimize the common interests of authors, readers/users, and science, by providing the authors with a tool for reaching those readers who could most benefit from the news of their accomplishments. Developing such a service requires deeper insights into some basic questions of information science; a vigorous development activity can stimulate

and support the underlying basic research. This paper contributes to one of the key problems in information science which is also central for such a service: how can we characterize a journal and how can we optimize the performance of its key functions?

We argue that relevance, acceptance rate, circulation, prestige and publication lag are five crucial variables to be considered when selecting a journal for submission of a manuscript. We explicate these variables and show how to use them in a plausible decision procedure with the help of a mathematical model. We report some findings about the nature of these and related variables.

The primary contribution of this paper is the idea of an author–journal matching service. The mathematical model is only a beginning toward designing a good decision procedure on which the matching would be based. We cannot yet specify an optimal decision procedure. Further basic research, both theoretical and empirical, is required to be built upon the foundation we constructed. Our analyses, mathematical and empirical, have already led to some insights which can provide tentative decision procedures for a pilot version of the service we propose. Only experimentation can tell how good they are and what is their cost.

The service would operate by eliciting from an author–client *a* with a manuscript *d* a judgment as to which of several problem-classes the topic considered in *d* falls into, as well as *a*'s judgement about the relevance of his paper to a sample of authors and readers of some candidate journals. The problem-classes might be described by something akin to subject-headings or class-names in librarianship. The suitability of the journal may be ranked on the basis of a few selected people known to be contributors or constant readers. From the journal, the service would obtain data about acceptance rate, publication lag, circulation volume, etc. All this information would be combined into a measure of the attractiveness of a given journal to the author with manuscript *d*. He would then select the journal with the highest measure, or the service would recommend it.

Implementation of such a service could be started by the librarian or information specialist of a research institute, who would initiate the compiling of data about journals in which the researchers of that institution are most likely to publish. He or she then would present this information to the authors and elicit from them the judgments needed to compute the measures proposed here. Success could lead to explicit demand and to expansion of the service.

Such a pilot service should be operated under controlled experimental conditions, to the extent that is possible without disrupting its utility as a service. The various assumptions can then be tested, and the findings should result in an improved model and decision procedure.

#### REFERENCES

- [1] H. ZUCKERMEN and R. K. MERTON: Patterns of evaluation in science: institutionalisation, structure and functions of the referee system. *Minerva* 1971, 9, 66–101.
- [2] S. COLE: Professional standing and the reception of scientific discoveries. *Am. J. Sociol.* 1970, 76, 286–306.
- [3] J. R. COLE and S. COLE: The Ortega hypothesis. *Science* 1972, 178, 368–375.
- [4] R. K. MERTON: Priorities in scientific discovery. *Am. Soc. Rev.* 1957, 22, 635–659.
- [5] R. K. MERTON: The Matthew effect in science. *Science* 1968, 159, 56–63.
- [6] D. J. DE Solla PRICE: *Little Science, Big Science*. Columbia University Press, New York (1963).
- [7] D. J. DE Solla PRICE: Some remarks on elitism in information and the invisible college phenomenon in science. *J. ASIS* 1971, 22, 74–75.

- [8] W. D. GARVEY, N. LIN and C. E. NELSON: Communication in the physical and the social sciences. *Science* 1970, **170**, 1166–1173.
- [9] W. D. GARVEY and B. C. GRIFFITH: Scientific communication: its role in the conduct of research and creation of knowledge. *Am. Psychologist* 1971, **26**, 349–362.
- [10] D. CRANE: The gatekeepers of science: some factors affecting the selection of articles for scientific journals. *Am. Sociologist* 1967, **2505**, 195–201.
- [11] S. CRAWFORD: Informal communication among scientists in sleep research. *J. ASIS* 1971, **22**, 301–310.
- [12] J. VIRGO: A Statistical Measure for Evaluating the Importance of Scientific Papers. Ph.D. Thesis, University of Chicago (1974).
- [13] S. C. BRADFORD: *Documentation*. Crosby Lockwood, London (1948).
- [14] B. C. BROOKES: Bradford's law and the bibliography of science. *Nature, Lond.* 1969, **224**, 953–956.
- [15] E. GARFIELD: Citation analysis as a tool in journal evaluation. *Science* 1972, **178**, 471–479.
- [16] I. N. SENGUPTA: Recent growth of the literature of biochemistry and changes in ranking of periodicals. *J. Docum.* 1973, **23**, 192–211.
- [17] W. GOFFMAN and K. S. WARREN: Dispersion of papers among journals based on a mathematical analysis of two diverse medical literatures. *Nature, Lond.* 1969, **221**, 1205–1207.
- [18] D. J. DE SOLLA PRICE: *Little Science, Big Science*, Chapter 2. Columbia University Press, New York (1963).
- [19] J. B. LODAHL and G. GORDON: The structure of scientific fields and the functioning of university graduate departments. *Am. Sociol. Rev.* 1972, **37**, 57–72.
- [20] Report of the Task Group on the Economics of Primary Publications. National Academy of Science, Washington (1970).
- [21] R. D. WHITLEY: The Formal Communication System of Science: A Study of the Organisation of British Social Science Journals. *Sociol. Rev. Monograph* 16, *The Sociology of Sociology* (Edited by P. HALMOS). Keele University (1970).
- [22] N. LIN and C. E. NELSON: Bibliographic reference patterns in core sociological journals, 1965–1966. *Am. Sociologist* 1969, **4**, 47–50.
- [23] C. E. NELSON: The communication system surrounding archival journals in educational research. *Education Res.* 1972, 13–16.
- [24] M. KOCHEN: Quality control in the publishing process and theoretical foundations for information retrieval. *Software Engng* 1971, **2**, 19–53.