pose that a great many of these principles will coincide with those of the ethical systems of the great religions, dignity and brotherhood of man, but only as derivatives, the condition of dignity and brotherhood being most conducive to the pursuit of truth. The hope is that those activities of man which are condemned in most ethical systems but rationalized on other than ethical grounds will disappear, because their rationalizations will become untenable in the light of scientific inquiry. The same applies to quasi-ethical systems such as totalitarian ideologies and the highly specialized codes of conduct of small isolated communities. All these rest either on coercion or on exclusion of experience. Both coercion and exclusion of experience can be maintained, in the long run, only by the maintenance of sacrosanct fictions. Therefore, all coercive and provincial ethical systems depend critically on the fictions which support them. They collapse when the fictions are shattered, and their fictions are easily shattered once even the primitive elements of scientific inquiry are directed against them.

To summarize, it is possible to approach ethics scientifically in a stronger sense than simply by scientifically investigating existing ethical systems or by offering an analysis of the efficacy of means, employed to pursue given ends. The mixing of science and ethics guarantees that science will play far more than a descriptive or an instrumental role. This is so because science brings with it its own (for the time being, only "professional") ethics. This ethics, however, can be generalized to a complete ethical system which has a viability far greater than the existing ones. This greater viability is the result of the same properties that are possessed by scientific knowledge. Although this knowledge has always rested on fictions, it did not deteriorate when the fictional foundations were repeatedly shattered by reorganizations of knowledge but, on the contrary, gained from each such crisis. There is no sharp distinction between scientific outlook and scientific ethics. Both eschew authority—that is, coercion in any form —and probably for this reason are irresistibly attractive as means of liberating man from the bonds which, in his ignorance, fear, and ethnocentrism, he has imposed on himself.

Classifying and Indexing for the Special Library

Saul Herner and Robert S. Meyer

Classification is one of the most universally applied and least appreciated methods by which scientists and librarians organize and obtain information. Classification pervades practically every selection process in human experience, whether the thing being selected is an item in a supermarket or a book in a library. The primary factors that dictate the way that things are classified are physical necessity, economic necessity, and intellectual habit.

In a supermarket, the housewife shopping for a vegetable for her family's dinner goes to the vegetable department. Here, she finds an array of vegetables of various shapes, colors, and flavors. The vegetable department is an example of classification based on physical, economic, and intellectual habit factors. The odd shapes and quantities of vegetables, their perishability, and the fact that the housewife thinks of vegetables as a single concept or class of things necessitate their being displayed together or close to one another in the market. One or more of the same three factors furnishes the basis for the existence and location of all other classes of merchandise in the supermarket.

Turning to libraries, we find the same three factors dictating the manner in which books and other publications are arranged on the shelves. For the majority of libraries, the primary factor is the presumed intellectual habit of the library user and the librarian. Books are arranged on the shelves according to subject categories that are a reflection of logical or traditional relationships. Ideally, these relationships will coincide with the subject relationships that exist in the mind of the user.

In addition, there will be an index or catalog with multiple subject entries to direct the user to the part or parts of the collection containing publications on a given subject. The library uses this device to correct for the fact that a book may deal with a number of subjects but can be in only one place at a time. Ideally, the subject headings or entries in the catalog will be in the language of the users and will serve also as a key to the classification. In addition, they will be consistent and mutually exclusive.

Thus, there is a useful artifact, an alphabetical index in the language of the user, superimposed on a physical arrangement of published materials that is a reflection of the user's intellectual habits and associations. This is an ideal situation, in which the seeker of information can choose between going to the index and then to the books or going directly to the books.

Many librarians and documentalists consider it inefficient to go directly to the books without first consulting the card catalog. The logic behind this is that the library card catalog, through its multiple subject entries, can tell the searcher *all* the publications in a library that treat of a given subject, whether in a major or minor way, whereas by going directly to the shelves the searcher is likely to find only those that treat of a subject in a major way, if indeed he finds anything at all. Nevertheless, most specialist-users of the library go directly to the books when they can, and they seem to find this arrangement satisfactory. When they use the card catalog at all, they generally use it to locate items they already know about.

This rather informal approach to the literature and the limited use of the card catalog probably stem from the fact that the professional worker in a field is not likely, except on rare occasions, to want to know everything on a given subject. He is merely looking for something to supplement or help recall what he already knows. Classification favors this approach by laying before the seeker of information a group of related publications whose major content is the subject in which he is interested.

Mr. Herner and Mr. Meyer are partners in Herner, Meyer and Company, Washington, D.C., consultants who conduct research in library planning, organize and conduct informational surveys, and design information systems.

Defects of Existing Systems

In recent years, classification has fallen in the general esteem for various reasons. One reason is related to the factors of economics and physical necessity. Some of our large libraries—the John Crerar Public Library and the New York Public Library, to name two—have been forced by the increasing size of their collections and the rising cost of storage space to discard classified arrangements of their books in favor of arrangements based on the chronological order in which the books are received and processed. This results in a great saving, since it eliminates the need for leaving spaces on the shelves to allow for new books in the various subject classes. By this method, the books are packed together tightly, and additions are made at the end. The obvious shortcoming of this arrangement is that the books are arranged in random order, and the possibility of browsing is practically eliminated. However, in the case of the John Crerar and New York Public libraries, this is no real loss, since neither library permits public access to the bookstacks anyway.

Another reason for the decline in usefulness of classification systems is the fact that those presently in use are inefficient in a number of rather serious respects. Unlike our ideal classification system, the average system now in use does not reflect the logical and traditional relationships that exist in the minds of the users. Potentially useful books are often overlooked because they are shelved among books that are not related in the mind of the library user.

Another shortcoming of existing classification systems is the inadequate definition of their subject classes. This fault results in the classifier's being unable to decide which of a number of similar classes a book rightfully belongs in. It also results, sometimes, in the classifier's discovering that there is *no* category in the classification to accommodate the subject of a given book. What generally happens in such cases is that the classifier has the choice of forcing the book into the subject class that comes closest to fitting its subject or of establishing a new subject class. Confusion is compounded when the library user is forced to duplicate the perplexities of the classifier in trying to find out in which of several equally logical places the book is actually resting.

In addition to their other errors of omission, the classification systems presently in use lack the flexibility to assimilate new doctrines and new developments on every level of generality. This is probably the most important reason for the square-peg-in-the-round-hole dilemma. The presence of a rigid hierarchy and the lack of a simple means of altering or expanding its existing parts greatly diminish the current usefulness of the average classification system. This is particularly serious in science and technology, where the growth of new doctrines, new developments, new subjects, and new subject relationships is truly prodigious.

However, the criticisms we have leveled are criticisms of specific classification systems and cannot be validly applied to the entire concept of classification. To condemn library classification as a whole, and to discard it completely because some classification systems are inadequate, is to rob the scholar unjustly of one of his most valuable tools in his use of books.

Causes of the Defects

The inadequacy of present-day library classification systems is the result, primarily, of their attempt at universality. There are really only three systems that are broadly recognized and used at the present time. These are the Dewey Decimal classification, the Library of Congress classification, and the Universal Decimal classification, which is actually an extension of the Dewey Decimal classification.

All of these systems have in common the fact that they are designed for libraries where the use made of the literature is extremely broad. The classification systems reflect this breadth of use by attempting to categorize every possible subject from every possible viewpoint. In any such situation there will naturally arise difficult problems of selection for the cataloger, particularly if the subjects and viewpoints are not clearly demarcated. This will give rise to inconsistencies in cataloging, which are bound to stymie the library user.

Another cause of the inadequacy of the existing classification systems is the fact that they are generally keyed in their design to a body of knowledge rather than to a body of literature containing this knowledge. This results in a failure of these systems to reflect the growth and subject content of the literature. In the case of the Library of Congress classification, the system is based on a specific collection, but this collection is so broad and is categorized from so many viewpoints that its effectiveness and meaning for the specialized library and specialist-user are seriously diminished.

In actual practice, it is difficult if not impossible to apply general classifications to specialized bodies of literature. In most instances where this is attempted, refinements of one sort or another are necessary. In one sense, the general classification is too detailed for use in a specialized collection, because it attempts to classify all subjects from all viewpoints. The library containing a relatively narrow collection of literature must obviously treat it from the viewpoint of the specialized group that is to use it.

In another sense, the general classification is too lacking in detail for the specialized collection. In attempting to encompass all the world's knowledge, the designer of the general classification is forced to treat any given phase of this knowledge rather shallowly. This gives rise to the problem, often encountered in the specialized library, of having available a relatively small number of subject classes into which a relatively large number of publications must be categorized. The necessary solution is to subdivide the available subject classes in some fashion. Thus, without formally setting out to do so, the librarian of the specialized collection, in attempting to adapt a general classification to the needs of his library, is actually designing his own classification system.

Faced with the difficulties of applying general classification systems to specialized collections of literature, many librarians and documentalists have discarded classification altogether and have relied exclusively on various forms of indexing. Where such indexing systems have been in the language of the users, and where they have been consistent in their terminology and meanings, they have been quite useful. They have fallen down when they have attempted to include indexing terms or headings for every conceivable shade and aspect of every pertinent subject. This breadth of coverage often results in a fractionation of subjects which makes the index entries far more specific than they are in the searcher's mind. This complicates the cataloging and retrieval processes.

The retrieval aspect of the problem is perhaps illustrated by the recent experience of one of us, who, in the course of a cataloging assignment, sent to the Library of Congress for a set of printed catalog cards for a book entitled *Business Success Handbook; Your Complete Guide to Executive Growth.* When he received the cards, he looked to see under what subjects the Library of Congress had indexed the book. He found that it was indexed under only one subject: Success! Fortunately, it was classified in a subject class for business and would have been shelved and found among the books on that subject. In this way it would be possible to find a book despite the inept terminology that sometimes characterizes index entries.

In many of the more recently developed indexing systems, this alternate retrieval method is not possible, because they are keyed to physical arrangements that do not reflect subject relationships. As a result, a person doing a subject

search in the library is forced to use the catalog or index whether or not he considers it adequate. Thus, the user of the literature becomes the servant of the system by which it is indexed. This is obviously a topsy-turvy situation.

The basic shortcoming of most of the classification and indexing systems presently in use is the fact that they do not truly reflect the needs and approaches of their users. In addition, where classification and indexing systems are used in conjunction with each other, they are very often not coordinated in context and terminology, with the result that the classification views the literature in one way and the index views it in quite another way.

Tailor-Made Classifications

For the past several years, we have been concerned with the development of classification systems, and of indexes to these systems, which are truly reflective of the language, viewpoints, and requirements of their users. A basic proposition in the design of such tailor-made systems is that they be constructed for a fixed group and not for a general public. The systems presently under development are designed for groups of scientists or for other relatively specialized groups of individuals whose literature needs and approaches are similar.

At this writing, two such systems have been designed and installed, one in the library of a trade association in the construction field, and the other in the research library of a firm manufacturing a line of surface cleaners and related products. A third system, designed for the organization of literature on atomic energy and related subjects, is now under development. This system is actually experimental in nature. It is being developed, first, as a basis for testing the efficacy of tailor-made classification and indexing systems for their own sake and, second, as a basis for comparing such systems with other systems presently in use. This work is supported, in part, by a grant from the National Science Foundation.

There are seven basic requirements which must be met in the custom design of classification and indexing systems. Liberal mention has already been made of the first of these requirements: (i) The subject classes and the terms used to define these classes must be directly reflective of the viewpoints and language of the users. (ii) The system must reflect the actual literature to be organized as well as the actual purposes for which this literature is used. (iii) All classes and descriptive terms must be mutually exclusive in their content and meanings. True mutual exclusiveness is a difficult, if not impossible, thing to obtain in nature, but, in a system designed for a specific population and a specific body of literature, it can be accomplished by means of delimiting labels which define clearly the scope and content of each class and descriptive term. (iv) The number of documents within classes must be approximately equal and of such magnitude as to permit ready perusal. (v) The system must be readily and logically expandable to permit the assimilation of new documents and new subjects. (vi) The notation used to identify classes must be constant in its number of characters and otherwise simple to transmit and recognize. (vii) The classification must be constructed by means of groupings of like subjects, and any hierarchical relationships designed into the system must reflect the intellectual habits and preferences of the users rather than any philosophic laws of nature.

On a given level of generality, subjects are usually coordinate in value. It really does not matter to the person who is interested in automobile wheels whether books on tire rims are placed before or after books on tires, on the shelf. The important thing is that they be located close to one another and that the library user be able to ascertain this location with a minimum of difficulty.

There are several ways in which the requirements, viewpoints, and technical language preferences of the user population can be determined. One method is by means of detailed interviews with a cross section of the users. Another method is through observations of the users' day-to-day activities and problems over a representative period of time. A third method is through the analysis of representative samples of the writings of the users, when such writings are available. A fourth method is through the analysis of a cross section of reference questions which actually arise from the activities of the user group.

The method used in a given case is a function of the size and character of the user-audience being studied. When the group involved is located within the confines of a single organization, the first two methods—interviews or on-the-spot observations of activities—can be applied. When the user-group is ill-defined or physically scattered, the latter two methods—analysis of writings or analysis of reference inquiries—can be used. In some instances, a combination of the foregoing methods is used.

Design of Specific Systems

In the construction of the classification and indexing system for atomic energy literature, the reference inquiry analysis method is being used. Actual reference questions received by libraries and reference centers in atomic energy installations all over the country are being collected and analyzed for subject content and terminology. At this writing, approximately 5000 questions have been received and analyzed. Upon the completion of this phase of the project, a total of approximately 10,000 reference questions will have been collected and analyzed.

Sorted into categories based on similarity of content, these questions will constitute a crude classification system. At this stage, a cross section of the literature to be organized will be classified in terms of the categories established in the crude classification system. This furnishes a guide to the probable number of documents that will fall into the various subject categories.

The pilot classification process actually has three functions: (i) it insures that the classification truly reflects the literature it seeks to organize; (ii) it serves as the basis for the development of subjects on lower levels of generality; (iii) it furnishes a means of ascertaining the number of documents that will fall into the various subject classes. Counts are made at this phase to find out what classes are likely to contain so few documents that they should be combined with closely related classes. At the same time, counts are made to determine which classes are likely to be so overcrowded with documents that subdivision into smaller classes will be necessary. The foregoing operations constitute a "smoothing" of the subject classes. In the course of this "smoothing," criteria are established to serve as a basis for subdivision, where this is indicated.

The final stage in the process is the assigning of discrete numbers to identify each of the classes on the several levels of generality and the establishment of evenly distributed, unassigned classes to allow for ready assimilation of new subjects or for unanticipated growth in the number of documents in any given subject. The number of assigned and unassigned categories and the number of levels of generality will depend on the magnitude of the literature at the time the classification system is designed and on the rate of growth of the field and its literature.

In the case of the atomic-energy literature, there exist approximately 20,000 documents, and the rate of growth is of the order of 2000 documents a year. To accommodate such a body of literature and such a rate of growth, a total of 3000 subject classes will be utilized. At the outset, 1536 of the available 3000 subject classes will be used, leaving a total of 1464 unassigned for future expansion. These figures are based on an allowance of an average of about 12 to 15 documents per class.

In actual practice, each of the basic subject classes will constitute a fourth level of generality. Therefore, for con-

venience, the basic subject classes are termed subclasses. In this context, there will be a total of 3000 subclasses, 300 classes, 30 superclasses, and three major subject categories. Every fifth superclass, class, and subclass will be unassigned. By this means, allowance is made for expansion throughout the classification and at every level of generality.

Structurally, once the pilot classification process is completed, the system is built from the bottom up. Subclasses will be combined to form classes; classes will be put together to form superclasses; and superclasses will be combined to form major subject categories.

All subclasses, classes, superclasses, and major subject categories will be identified by four-digit numbers. Thus, the classification will take the following form:

1000–major subject category
  1100–superclass
  1200–superclass
  1300–superclass
  1400–superclass (unassigned)
  1500–superclass
  etc.
  1100–superclass
    1110–class
    1120–class
    1130–class
    1140–class (unassigned)
    1150–class
    etc.
    1110–class
      1111–subclass
      1112–subclass
      1113–subclass
      1114–subclass (unassigned)
      1115–subclass
      etc.

Superimposed on the various processes in the design of the classification is a feedback process which permits members of the user-population to edit and otherwise alter the system at every stage in its development. For this purpose, the equivalent of a consumer panel in marketing is utilized. This panel, which is carefully selected to represent the user-body, reviews the system in its various stages to insure that the arrangement of the classes and the language used to describe them truly reflects the user viewpoint. The editing by the user-panel also serves as a further means of establishing the best remedies for redundancy, when it is uncovered in the course of the pilot classification process. It also permits an evaluation of the subclasses and newly combined classes resulting from the pilot classification and literature count.

When the design of the classification is completed, all of the terms used in each of the subclasses, classes, superclasses, and major subject categories are arranged in a single alphabet to form a subject index to the system. Where indicated, cross-references to related terms are included in the index.

A second feature, which is installed at this point, is a set of "keys" to the subclasses, classes, superclasses, and major subject categories as well as a network of cross-indexing among related subjects. By means of the "keys," it becomes possible for the user of the classification to go from the broadest or most generic identification of his subject interest to the most specific identification in four simple stages. Thus, the user is afforded three means of approaching his subject: (i) the subject index, (ii) the "keys," and (iii) going directly to the shelves and browsing.

Advantages of the System

A basic advantage of the tailor-made classification and indexing system is its input and output speed. By actual time studies, the process of locating and writing down appropriate class numbers and indexing subjects for specific documents has been found to consume an average of 2 to 3 minutes. (These time studies were made with librarians who had no prior experience with this method of classification and no particular knowledge in the subject fields of the documents they were classifying.) This is several times faster than the time required for classification and indexing by conventional library procedures.

The speed of the system derives, basically, from the fact that it utilizes only established viewpoints and terminology. As a result, there are fewer subject headings and categories. Those headings and categories that *are* utilized in the system are mutually exclusive and reflect the most likely approach of the user to the literature. Thus, the choices available to the cataloger are sharply limited, and the classification and indexing processes are greatly simplified.

In much the same manner that it increases the input speed, the system increases the output speed. As has already been suggested, the library user, in tracking down a piece of published information, is forced to duplicate, in many respects, the thought processes of the person who cataloged and indexed it. Therefore, the simpler and more consistent the input processes, the simpler and more consistent the output.

Mechanical Searching Codes

Thus far we have been discussing the use of tailor-made classsification systems as a substitute for general classification systems in special libraries. There is another possible application of tailor-made systems which relates to the fact that such systems utilize consistent notations and discrete subject categories. This has to do with the use of such systems as bases for machine-searching codes.

One of the shortcomings of conventional classification systems as bases for such codes has been their inconsistency and the length of their notations. As a rule, in such systems, the more specific the subject category, the longer the subject class number. In punched-card information retrieval systems, where conventional classification systems are used, it is necessary to dedicate as much space on the card as is necessary to accommodate the largest number of digits used to denote any subject category. This results in a loss of valuable space on the card and a serious diminution in the flexibility of the system.

The same problem arises in systems where punched or magnetized-tape memories are used. In the case of devices using tape as a medium of storage, the problem is complicated by the fact that excessively complicated code notations necessitate the use of more code channels. This increases costs and slows the rate of search. Through the use of a constant, short notation, which reflects a carefully limited number of subject categories, the tailor-made classification may minimize these problems and make possible a more efficient use of mechanical searching devices.

Another feature that may make tailor-made systems readily amenable to mechanical searches is the discreteness of their subject categories. This discreteness of categorization makes possible a discreteness in coding which is not possible in systems where there are overlaps in subject categories. Where unique identification does not exist, there is always the danger of missing the target altogether or of getting a good deal of extraneous material along with the material that is actually wanted.

In addition to specific searches on the lowest level of generality, codes based on classification systems also permit mechanical searching devices to perform generic searches on a higher level of generality. This stems from the grouping of related terms in larger, more general categories. In coding systems that are not based on classification, generic searches are extremely difficult, since the relationships among the various concepts are not explicitly defined and are not incorporated in the code.

Areas for Future Consideration

Although the usefulness of tailor-made classification and indexing systems has been demonstrated through the design, installation, and day-to-day use of two such systems, there are still some questions that have to be answered. The first of these questions is that of the relative efficiency of such systems when they are compared in quantitative terms with existing systems. Factors to be considered in such comparisons are input costs, out-

put costs, and the relative effectiveness of the various systems in the performance of subject searches. The question of input costs of tailor-made systems versus conventional library systems has been fairly well settled. But the remaining elements of the overall question still remain unanswered.

As part of the study which is being performed under National Science Foundation auspices, an attempt is being made to find answers to these remaining elements of the question. Carefully designed experiments have been set up to test each of the comparative aspects of the new system against those of existing systems. These experiments have been designed to minimize biases resulting from differences in the mental and physical dexterity and doggedness of the persons using the systems as well as biases resulting from differences in the complexity and intellectual level of the literature that the systems are designed to organize. The tests will also take into account the effect of reference questions of varying levels of difficulty being put to the systems.

For the purposes of the present tests, a comparison will be made between the classification system which is being constructed for atomic energy literature and the indexing system now in use in most Atomic Energy Commission libraries.

Another question that still requires an answer is the comparative usefulness of tailor-made classifications as a basis for codes in mechanical searching systems. Although there is good theoretical evidence that tailor-made systems, their notations, and the mutual exclusiveness of their classes lend themselves to machine-encoding, this evidence has never been tested on a comparative basis in actual machines and real-life situations.

In order to furnish a basis for such tests, we have under construction a small, computerlike device which is capable of sorting and correlating literature references and various types of data. This machine will be amenable to codes based both on indexes and on classification systems. Thus, it will be useful for controlled, unbiased comparisons between tailor-made classifications and most other systems as coding media for mechanical searching devices.

It is very probable that there will develop, in the wake of the answers to the foregoing questions, many new questions that require answers. But, as new questions and new answers arise, they are bound to result in more and more effective means for making information available to the scientist and to other members of the scholarly community.

Science, Imagination, and Art

Norman Robert Campbell

Two criteria [may be used] to determine why a scientific proposition has value and what degree of value it has. . . . First, a proposition is valuable if truly universal assent can be obtained for it; second, it is valuable if its contemplation causes intellectual satisfaction to students of science. These two principles are to some extent contrary, and, if the test provided by each of them is applied to the same proposition, one might sometimes determine that the proposition is valuable and the other that it is not. For a student of science is a student of science in virtue of some difference between his intellectual constitution and that of the rest of mankind; if he finds intellectual satisfaction in a proposition it is almost certain that persons with different training and different interests can be found to whom it will give none; and on the other hand the mere fact that a proposition is approved by everyone, however different their modes of thought, will deprive it for him, not of course of all its value, but of that very special value which is the basis of the second principle. It is necessary therefore to examine the two principles rather more nearly and to determine exactly what part each of them plays in the establishment of scientific propositions. . . .

Everybody recognizes today that what I have called truth is an essential element of a scientific proposition and few, if any, will deny explicitly that what I have called meaning is also important. But it does not seem to me that facts which are universally admitted openly, or their implications, are always remembered when the most general and fundamental questions concerning science are raised. In such discussions attention is apt to be concentrated on the truth and the meaning is apt to be left out of sight.

The tendency is natural. The great advance or, more accurately, the first beginnings of scientific knowledge which took place in the 16th and 17th centuries were a consequence of the recognition of the possibility of scientific truth. To say that science must be based on experiment and observation is simply to say that it must satisfy the first principle of value, for it is only concerning the results of such experiment and observation that universal agreement of the kind which is characteristic of science can be obtained. It is the neglect of truth, the failure to test evidence according to the canons of modern science, the acceptance of well-attested fact, vague rumor, and the product of riotous imagination as equally valuable—it is the attitude of mind to which such things were possible which raises an insurmountable barrier between ourselves and the most enlightened of the ancients. That science should have meaning, they would have agreed readily; it was the doctrine that it should have truth which was strange to them. The ghost of Greek learning still stalks ruins not yet abandoned; it still disturbs timid minds and has still to be exorcised; the weapon of Galileo cannot be allowed to rust in its sheath, and while it has still to be used other dangers may be neglected.

However there is a more cogent reason why truth rather than meaning receives emphasis whenever any question is raised of the value of science or of its relation to other studies. Truth, it has been said, is a quality of which we may hope to convince others; it is a valuable quality because it is appreciated by everyone. And there is actually no doubt that scientific propositions have the kind of truth that is here attributed to them and that this truth has some value. Nobody disputes that truth, if they once agree to use that word in our sense; what they may dispute is whether or not it is misleading to call this quality truth and what is its value in comparison with that

Dr. Campbell is a noted British scientist who has made significant contributions to the philosophy of science. This article is reprinted from his *Physics: the Elements* (Cambridge University Press, London, 1920), chap. 8. By arrangement with Cambridge University Press, Dover Publications, New York, will issue a reprint of this book in May.