

Classification Today— Shadow or Substance

Mortimer Taube

President, Documentation, Inc., Washington

The topic which has been assigned to me, "Classification Today—Shadow or Substance," might more appropriately have come at the end of the Institute, rather than the beginning. If I could convince you that our pursuit of valid classifications was the pursuit of a shadow, there would be no reason to listen to the papers on the remaining part of the program. We could all pack up and go home. Hence, I must conclude that when those who planned this Institute gave me this topic, they assumed that regardless of what I might say about classification, I would certainly be unable to demonstrate its ephemeral or shadowy nature and that I would conclude that classification had substantial value for librarianship and related information activities.

Confronted with this dilemma, it occurred to me that the way out for an erstwhile student of logic like myself might be found in the first instance not in examining the nature of shadows nor the nature of substances, but in examining the meaning of the connective between them, namely, the logical operator "or." Most of us, when we think of the word "or," think of it in the exclusive sense as meaning "either or," that is, the word used in this title, "Shadow or Substance," would ordinarily be interpreted to mean that if classification were substantial it could not be shadowy, or if it were shadowy, it could not be substantial. There is, however, another meaning of "or" which is the usual meaning attributed to it in works of logic, where the "or" is taken as meaning logical disjunction with reference to propositions and logical sum with reference to classes. In this sense "or" means "and/or" rather than "either or." Thus if I say "It will rain tomorrow or I will stay home," both statements could be true; that is, it might rain tomorrow and I could still stay home. Similarly, if I say of an item that it is a member of the class A or B, it could be a member of A, a member of B, or a member of AB, and the general proposition "X is a member of A or B" is true in all three cases. This general proposition is only false when the item is a member of neither A nor B. This logical relation can be illustrated by the truth table for disjunction at the top of the following page.

Now then, if we assume that the "or" in the title is the "or" of logical disjunction, then it is possible for me to take the line that classification in some sense is substantial, in some sense is shadowy, and

A	B	A B
T	T	T
F	T	T
T	F	T
F	F	F

in some sense is both. My text, then, becomes one of indicating the sense in which it is substantial and in warning against the sense in which it is shadowy, so that you will be able to judge subsequent papers in this Institute in this context.

We must admit in the beginning that the concern of librarianship with problems of classification represents one of the oldest and strongest links of librarianship with basic intellectual and theoretical questions. As a first year student in library school many years ago, John Lund and I found that questions of classification constituted an intellectual oasis in a barren waste of learning how many spaces should go between the author and title in descriptive cataloguing, or how one colates a book when the publisher has gotten mixed up in his numbering procedure. Hence, the earliest contribution I attempted to make to the art or science of librarianship was a paper on classification. Some of you may have read it. It was called "A Non-Expansive Classification System" and it appeared in the *Library Quarterly*¹ over twenty years ago. In this paper we took the line that a classification system covering all knowledge for all time was certainly chimerical or, as the title assigned to me has it, "shadowy." Hence, we felt that in order to save classification as an intellectual activity for librarianship, it would be necessary to set up our major classes in terms of time divisions; that is to say, the major classes we recommended, instead of being such things as science, literature, art, etc., were historical epochs. We thought it possible that within these historical epochs one might construct adequate classifications; and by this we meant not classifications of knowledge, but classifications of library material itself. One of the major considerations which led us to this conclusion was something we were taught very early in library school. We were taught that the Dewey system was a theoretical system which attempted to legislate for books and that its pigeonholes were created independently of a concern with the content of the pigeonholes. We were taught at the same time that the great advantage of the Library of Congress classification system was that it was made from the books themselves and based upon an empirical study of the material at the Library of Congress. Hence one could say that the pigeonholes or classes in the Library of Congress classification system were actually designed to contain the material in the Library and therefore one could predict an excellent fit. From such an argument, however, it is a simple matter to draw the conclusion that the Library of Congress classification might have fitted the contents of the Library of Congress at the time it was made, but that for future materials to be received by the Library of Congress, the classification system took on the same theoretical and

predictive character as the Dewey system; that is to say, once the pigeonholes were set up, all new material would have to be fitted into the pre-existing pigeonholes. Hence, with reference to new material, the Library of Congress system differs from the Dewey system only in having different pigeonholes. It was this predictive character that a classification system based on temporal epochs was designed to avoid. We felt that the great virtue of the Library of Congress system, namely, its development from an actual examination of the material to be classified, could always be retained if new classifications were set up as required by the changing pattern of literature.

In the twenty years that have elapsed since this paper, I have seen no reason to weaken its conclusions but I am now convinced that Dr. Lund and I did not go far enough. At that time we did recognize a changing pattern of literature. What we overlooked were the different interests which might exist in the same historical epoch. Now we would say that not only is it necessary to make classifications for different periods of time but that it is necessary to make classifications for different special purposes.

It is not my intention to give you a biographical sketch at this time based upon the various papers that I have written about classification, but Anatole France once described literary criticism as "the adventures of a soul among masterpieces." He meant by this expression that the important thing about literary criticism was not the book criticized, but the nature of the critic. Hence, I feel that I can best carry out my assignment at this Institute by telling you of the various considerations and the steps along the way which have led to my present conclusions about classification in librarianship.

In 1950 I was privileged to give an opening paper at a similar Institute, although the title of the Institute was different. I refer to the Institute on Bibliographical Organization held at the University of Chicago at the Fifteenth Annual Conference of the Graduate Library School. At that Conference I was assigned the topic "The Functional Approach to Bibliographical Organization."² In preparing that paper I felt that my first obligation was to define the concept of function, and following suggestions from the biological sciences, I concluded that a function could only be defined in terms of purposes. Hence, a functional approach to bibliographical organization could only mean an analysis of bibliographical organization in terms of its purposes. I concluded, then, that there were no universal purposes and hence that there could not be any universal bibliographical organizations. This indicated that the *Universal Decimal Classification* is certainly "shadowy" since it assumes that the scientific and intellectual enterprises of all men everywhere could be subsumed under a universal purpose. Certainly we can say that an increase of knowledge is the universal purpose of all scientific and intellectual endeavor, but what is required here is not such a general and vague universal purpose but a universal purpose in terms of which we can design and construct an elaborate system of major classes, sub-classes, sub-sub-classes, etc., into

which we can organize the products of all intellectual endeavor. I did not believe then and I do not believe now that this is even a remotely feasible enterprise. Hence our conclusion at that time was that different individuals or different groups should determine the specific bibliographical organization necessary to their own purposes. I suggested that these special purposes might be related by having each special group make its selection of major class, sub-class, sub-sub-class, etc., from a common vocabulary. Let me say at this time, parenthetically, that at the present time I despair of even such a universal apparatus as a common vocabulary for all sciences.

Aside from my theoretical interest in the problem of classification, I had learned a good deal about its nature and utility from working in libraries, and one of the things I did learn from working in both large and small libraries was that for most such institutions, classification has become a method of shelving books and has ceased to be, if it ever was, a way of organizing the information in such books. This was brought home to me most clearly in my years at the Library of Congress. The Library of Congress, as you know, has closed stacks. The approach to the content of these stacks is through the standard type of dictionary catalogue. Beyond the dictionary catalogue, those who consult its collections use the standard type of printed bibliographies, e.g., *Chemical Abstracts*, *Physical Abstracts*, the publications of H. W. Wilson and Company, *Public Affairs Information Service*, etc. It therefore seemed that the effort expended in setting up and maintaining an elaborate system of close classification is wasteful, since it has no real impact on the users of the Library. Of course, I knew that there remained within the system of American libraries a number of institutions, such as the Crerar Library and the Engineering Society's Library in New York, which utilize classed catalogues, but it still remains true that in general, classification is not a major tool for the use of contents of libraries. Certainly I was also aware of the very great value of classification as a method of arranging books in open shelf libraries, mainly public libraries or small academic libraries; but I felt in this instance that these classifications had a special purpose, namely, making available to the general reader a rough breakdown of books which reflected a similar rough breakdown of the interests of the general reader, i.e., fiction, travel, science, religion, etc.

I should like at this time to refer to just one other previous paper which we did on this subject. In 1953 we prepared a report for the Office of Naval Research on "Machines and Classification in the Organization of Information." This report was published in Volume II, *Studies in Coordinate Indexing*. In this report I raised the following question: Why, in the face of a general decline in interest of problems of classification in regular library organizations, was there such a renewed interest in the problem among documentation people and people who were concerned with machine searching of information?

How then do we account for the renewed interest in classification as a method of information control? Within the last few years, we have

witnessed the birth (and in some cases, the rapid death) of dozens of new classification systems, among which we can name, *The Story of Classification for the Army Technical Reference Service*; the Office of Naval Research *Project Status Classification*; the Research and Development Board *Classification of Research Projects*; the American Society for Metals-Special Libraries Association *Metallurgical Literature Classification*, and the *Standard Aeronautical Indexing System*. There has been a revival of interest in the *Universal Decimal Classification*, in the Patent Office *Classification*, and in Ranganathan's *Colon Classification*. Western Reserve University has labored for several years and is still laboring on the development of "abstraction ladders" and "semantic factoring."

This renewed search for the solution to an unsolvable problem results from a paradox, namely, the promise of machine organization and retrieval of information, and the actual slowness of the machine in the linear searching of an index. Classification becomes one of the methods proposed for dividing an index in order to shorten the time required for a machine search.

Let us suppose we are searching for the name "Baker, Able Charlie" in a village telephone book containing about 1,000 names. To search for this name might take a minute or two, occupied with picking up the book, finding the proper page and column, and scanning the proper column for the name being sought. Now it is quite practical to utilize an IBM machine, or some similar machine, or even a deck of edge-notched cards, to find one name in a random file of a thousand names, in about the same time required for the manual search of an alphabetical file in a minute or two. But suppose we are looking for the name "Baker, Able Charlie" in a list of a million names comparable to the New York telephone book. It might take us a little longer to lift the heavier book, to find the right page and the right column, and to scan by the given names and address as well as the last name. Nevertheless, the time required for a search for one name in an alphabetical list of a million names is of the same order of magnitude as the time required to find one name in an alphabetical list of a thousand names. But a machine search for one name in a random list of a million names will take one thousand times as long as a machine search for one name in a thousand.

It was the more or less vague realization of this fact that led the early advocates of the application of punched-card machines for the organization and retrieval of information to recognize that machine methods could not be applied efficiently to the random searching of large masses of information. No machine search of a large random list can approach the speed with which the mind can jump to the exact position in an ordered list. It would be silly to randomize a list of names in a phone book, or subject headings in an alphabetical index, in order to search for any particular name or heading with punched-card machines. An ordered list when it is over a certain size always enables the mind which recognizes and utilizes the order to beat the

machine. The conclusion to be drawn here is that contrary to popular misconceptions, the larger the number of qualitatively different units in a linear system of information, the less applicable are standard punched-card systems or even magnetic tape systems to the problem of searching; and this conclusion leads, in turn, to a search for (1) ways to cut down the size of indexes; and (2) ways to prefile or classify items of information.

So long as it seemed that machines could only be used for linear search of large files of information, the search for classification systems which could divide such files hierarchically or in any other way, although doomed to defeat, still appeared to be necessary. However, in recent years machine searching of literature has with rare exceptions adopted the method of look-up and coordination, rather than linear scanning, and this means that it is no longer necessary to invent classification systems in order to make machine search of information feasible or economic and practicable.

Some of those who wished to use classification in machine searching systems developed the notion of generic coding. This large mouthful means nothing more than the use of subordinate digits to indicate subordinate topics, which every student of librarianship learns in learning the Dewey Decimal system. For example, 500 is science in general; 510 is mathematics; 520 is astronomy; 511 is arithmetic; 521 is theoretic astronomy; 511.3 is prime numbers; 521.3 is orbits, etc. The advantage of such coding for machine systems lies in the ability to search by a portion of the number rather than the whole number. For example, if I search for everything on 51 — —, I pick up everything on prime numbers, without asking or knowing that anything is in the system on prime numbers. There are some people who feel that this type of generic searching is necessary for machine systems. This is most usually the case in the field of chemistry, where instead of searching for a specific compound I may wish to search for all amines or all chlorides or all purines, etc. It has been felt that the coding for any specific compound which is an amine should also contain the coding for amines as a generic group. Without going too much into detail on this matter, it can be said that this type of generic coding is totally unnecessary in order to make generic searching possible. Furthermore, it is much more expensive than other methods of carrying out generic searches. In a study of the cost of generic coding which we published in 1956³ based upon a study of the number of digits being employed in some of the systems being experimented with by the Patent Office, we determined that generic coding would increase the size of a mechanical store by a factor of three to one, as compared to other and simpler methods of carrying out generic searches. Since that time our conclusions have been reinforced by the attempt made by the National Bureau of Standards to develop a system of generic coding for compounds. The system developed by the Bureau of Standards used so many digits that the computer was in actual fact slower in its look-up procedure than an individual turning over and examining cards in a 3 x 5 drawer.

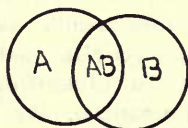
If it is the case that all classification systems so far produced or suggested have shown themselves to be inadequate as instruments of such bibliographical control, and if it is the case that such systems are not necessary for mechanized retrieval of information, why, may we ask, must we continually be faced with the problem of laying the ghost of classification or dissipating its shadows in the clear light of analysis? This has been a problem which has troubled me for some time. The issue seems so clear and yet we have this recurrent interest in and time spent on the theory and problem of classification in librarianship. I found the answer to this question in the "Report of Conclusions and Recommendations" issued by the International Study Conference on Classification for Information Retrieval, held at Beatrice Webb House, Dorking, England, May 13th - 17th, 1957. In a certain sense the classification group which has been started in this country and this Conference itself may be considered reactions to the Dorking Conference. In studying the conclusions and recommendations of this Conference, Paragraph (1), called "The Scope of Classification," gives us our clue:

Traditional classification has been concerned with the construction of hierarchies of terms - chains of classes and co-ordinated arrays. Modern information retrieval techniques also necessitate the combination of terms to express complex subjects. This conference takes the term 'classification' to include the problems raised by both these forms of relation. Some members use the term 'codification' for this field of study.

This is a complicated way of saying what earlier defenders of classification have said, namely, that all intellectual organization is classification and that even such things as alphabetical indexing or numerical arrays are species of classification. It is said that no matter how much we try to get away from classification, we must come back to it. And thus we see the Dorking Conference, which was presumably called to deal with classification as a specific method of organizing information, generalized the term so that classification became the name for any method of organizing information. We wish to do more at this point than quarrel about the meaning of words. Hence, we will admit that there is a sense in which all intellectual activity involves classification. The modern theory of arithmetic involves the notion that all numbers are classes, that is, one is the class of all classes having a single member, two is the class of all doubles, three is the class of all triples, etc., that is to say, a number is a class of classes. Further, it is certainly true that any general term involves the notion of class. Any word which does more than point or indicate this or that, is a word connoting or denoting a class. For example, I can point to a particular color, but I cannot use the term "red" without implying a class of shades, or the term "color" without implying a class of hues. When I use a man's name as the entry in a descriptive catalogue, his

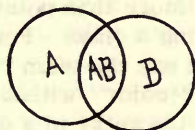
name becomes the class of all items written by him. In an alphabetical catalogue any subject heading is the class of all items which follow it in the catalogue. Certainly in this sense we must admit that all intellectual endeavor involves classification and that if we use the word "classification" in this wide sense, then all particular systems of organizing information are species of or varieties of classification. But on this point there is no quarrel nor really any reason to hold the type of Institute we are now holding. It seems to me that if we have a conference on classification, or if someone is asked to read a paper on whether classification is substantial or shadowy, there must be implied that there are other forms of organization of information, other forms of library organization, to which the term "classification" does not apply. In other words, if we say that a dictionary catalogue is a classed catalogue in just the same sense in which the John Crerar Library catalogue is a classed catalogue, then the question of whether we should have classed catalogues or dictionary catalogues becomes meaningless, sort of like saying that "A includes B and B is not included in A." This is a flat contradiction. What we must look for, then, both at the present time and in the following papers presented at this Institute, is a definition of classification which distinguishes it from other forms of organization and which permits an evaluation of classification as contrasted with an evaluation of other forms of organizing information. Unless we make this distinction, all of our discussion from now on will be shadowy and essentially meaningless. I wish, then, to offer a simple definition of classification as librarians have always used it which distinguishes it from other forms of organization. And here, if you will forgive me, we must utilize some simple logical notions to make this problem clear.

- (1) The product of any two classes is a class, as illustrated by the following diagram:

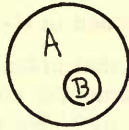


In this diagram A is a class, B is a class, AB is a class.

- (2) The sum of any two classes is a class; that is, "A or B" is a class.

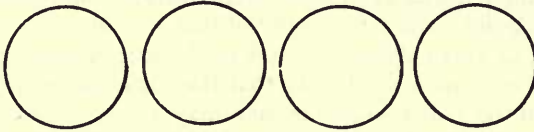


- (3) Given the situation where A includes B, AB is a class, but the class B not A is null; that is, it has no members.

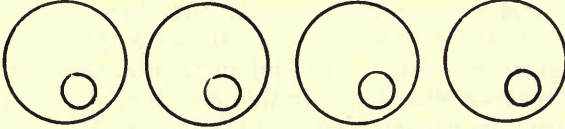


The class B is included in the class A, when all the members of class B are also members of class A.

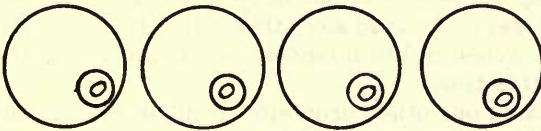
A library classification system like the Dewey system, the L.C. system, the U.D.C. system, etc., may now be defined as follows: There are a set of main classes, illustrated as follows:



All sub-classes are included in only one main class:



And this relation of inclusion continues, no matter how far we carry this subdivision; thus, all sub-sub-classes are included in only one sub-class:



It seems to me that those who defend classification systems are saying that knowledge, books, or the information in books can be organized in this way and that an organization carried out in this manner will serve the interests of scientific research and other intellectual activities. In terms of logic, class inclusion is only a special case of

class intersection. For example, two standard theorems in any logical work are:

(y) (x) $x \cap y \subset x$: The product of x and y is included in x

(y) (x) $x \subset x \cup y$: x is included in the sum of x and y

This is equivalent to saying that class inclusion can be defined in a Boolean system of products, sums and complements.

Where, then, does the issue lie? We have first rejected the notion that classification is a purely general notion and insisted upon its distinction from other types of organization. Now it appears we have insisted on the general character of Boolean relations and have pointed out that hierarchical classification or class inclusion is only a special relationship within Boolean algebra. What issue, then, remains? For myself, I think there isn't any, but historically there have been two issues which may provide substance in addition to shadow during the coming deliberations of this Conference. There have been metaphysicians, philosophers, and even some scientists, notably certain botanists and zoologists, who have insisted that in addition to the mathematical notion of class there do exist in the world real classes or archetypes. These men would say that the class of geraniums is much more real than the class which anyone may set up which has as its members any two flowers, e.g., a geranium and a rose. These men would say that the class of red things is more real than the class of colored things. Following this line, it would be said that scientific investigation will disclose that the universe and all the items in it are organized in a set of real classes and that the business of library classification or any other type of classification is not to make classes but to discover such classes. It is my present feeling that there are no serious scientists who still hold this position, at least not since the development and popularization of the theory of evolution and since the development of Boolean algebra in the middle of the Nineteenth Century. Let me remind you that it is traditional in library literature to recognize that Dewey was very much influenced by Harris, that Harris was an Hegelian, and that Hegelians are a species of unscientific German metaphysicians who believe that all reality is constituted by an hierarchy of classes reaching up to the Prussian State as the class of all classes. I would say further that the emphasis on real classes in this sense in librarianship is a cultural lag which should be eliminated at this time.

There remains one other problem. It might be said that an empirical investigation of how men actually organize knowledge or write books discloses that some classes are better than others and that some classes include other classes and that a good library organization should reflect this empirical fact of how people study, do research, or use libraries. This is a valid point of view and if the empirical facts could be demonstrated, then a library classification based upon such empirical facts would certainly be useful. On the other hand, if the librarians make classifications for themselves based upon theo-

retical considerations and insist that the users of libraries modify their own interests or own groupings in order to fit the librarians' theoretical classifications, such a procedure would have no warrant in either fact or logic.

Notes

1. J.J. Lund and Mortimer Taube, "A Non-Expansive Classification System," *The Library Quarterly*, VII (July, 1937) 373-394.
2. M. Taube, "The Functional Approach to Bibliographical Organization," *Bibliographic Organization*, (Chicago: University of Chicago Press, 1951), pp. 57-71.
3. M. Taube and Associates, *Studies in Coordinate Indexing*, Vol. III, (Washington: Documentation Inc., 1956), pp. 34-57.