

JURIMETRICS: THE METHODOLOGY OF LEGAL INQUIRY

LEE LOEVINGER*

INTRODUCTION

The terms "science" and "law" have both been used for so long by so many writers with such a variety of meanings, clear and unclear, that one who aspires to clarity or rigor of thought or expression might well hesitate to use either one. The lawyers are no more agreed on what constitutes "law"¹ than are the scientists on the meaning of "science."² Further, there have been many who claimed that law is a science, and it is still asserted by eminent scholars that jurisprudence is "the science of law."³ Exhaustive reading is not required to establish that there is neither an authoritative nor a generally agreed definition for any of the terms "jurisprudence," "science" or "law."

Nevertheless, each of these terms does designate an activity that is being conducted by an identifiable group of men. Lawyers and judges are engaged in practicing law and adjudicating. There are physicists, chemists, biologists, anthropologists, psychologists, and a host of others, engaged in activities that are universally recognized as science. And numerous professors, joined by an occasional eccentric lawyer, are engaged in writing articles and books that are either labelled or indexed as "jurisprudence."

Without undertaking either an exhaustive or definitive analysis of the activities of these groups, the general nature of their respective activities is fairly evident. Lawyers and judges generally are engaged in seeking to apply the principles or analogies of cases, statutes, and regulations to new situations. Scientists generally are engaged in collecting experimental and statistical data and in analyzing them mathematically. Writers on jurisprudence are engaged in the philosophical analysis of legal concepts and ideas.⁴

* B.A. 1933, LL.B. 1936, University of Minnesota. Assistant Attorney General, Antitrust Division, U.S. Department of Justice. Author, *THE LAW OF FREE ENTERPRISE* (1949); *AN INTRODUCTION TO LEGAL LOGIC* (1952).

¹ See Loevinger, *Jurimetrics—The Next Step Forward*, 33 MINN. L. REV. 455 (1949).

² Compare KARL PEARSON, *THE GRAMMAR OF SCIENCE* ch. 1 (1892), and JAMES B. CONANT, *ON UNDERSTANDING SCIENCE* 4 *et seq.* (1947); P. W. BRIDGMAN, *THE LOGIC OF MODERN PHYSICS* ch. 1 (1927), and A. S. EDDINGTON, *THE NATURE OF THE PHYSICAL WORLD* 323 *et seq.* (1928); GEORGE A. LUNDBERG, *CAN SCIENCE SAVE US?*, 66-67 (1947), and JAMES B. CONANT, *MODERN SCIENCE AND MODERN MAN* 118 *et seq.* (1952). Also see HANS REICHENBACH, *THE RISE OF SCIENTIFIC PHILOSOPHY* (1951); JAMES R. NEWMAN (Ed.), *WHAT IS SCIENCE?* (1955).

³ ROSCOE POUND, *JURISPRUDENCE* ch. 1 and authorities cited in note 1 (1959).

⁴ The following are taken as reasonably representative of modern jurisprudence: WOLFGANG FRIEDMANN, *LEGAL THEORY* (4th ed. 1960); ROSCOE POUND, *JURISPRUDENCE* (1959); MORRIS R. COHEN & F. S. COHEN (Eds.), *READINGS IN JURISPRUDENCE AND LEGAL PHILOSOPHY* (1951); JULIUS STONE, *THE PROVINCE AND FUNCTION OF LAW* (1950); P. L. SAYRE (Ed.), *INTERPRETATIONS OF MODERN LEGAL PHILOSOPHIES* (1947).

Through the years there have been those who called upon law in the name of jurisprudence to become truly scientific. As early as 1895 Justice Holmes asserted that "An ideal system of law should draw its postulates and its legislative justification from science."⁵ There can be no doubt that this great legal mind thought that at least potentially, the techniques of physical science could solve the most basic problems of law. For the rational study of law, he declared, "the man of the future is the man of statistics";⁶ and he projected the ideal of "an ultimate dependence upon science because it is finally for science to determine, so far as it can, the relative worth of our different social ends"⁷

Following Holmes, there have been others in the present century, although relatively few in relation to the whole profession, who have thought that "scientific methods of inquiry" should be applied to the field of values, including ethics, politics and law,⁸ or that we should have an "experimental jurisprudence" in which scientific data would be used to determine legal issues.⁹ However, little actual work has been done along these lines. The "law-science" symposia that have been published have been devoted to the narrow field of utilizing medical evidence to prove damages in personal injury cases.¹⁰ The resources of science have been used by law enforcement agencies, notably the Federal Bureau of Identification, to identify and convict criminals. There have been numerous studies of criminality from the viewpoint of psychology, psychiatry, and sociology. But nothing that can fairly be called even a beginning toward a "scientific jurisprudence" has yet appeared.

I

JURIMETRICS: A NEW SCIENCE

It would appear that either the problems of jurisprudence are not susceptible of investigation by science, or that there has not been sufficient interest in this possibility to produce any results. These are not necessarily the only alternatives, and it may be that science will yet fulfill the fondest hopes of those who think it capable of

⁵ OLIVER W. HOLMES, *Learning and Science*, speech delivered at a dinner of the Harvard Law School Association, June 25, 1895, reprinted in *COLLECTED LEGAL PAPERS* 138 (1920).

⁶ OLIVER W. HOLMES, *The Path of the Law*, an address delivered in 1897, *id.* at 167.

⁷ OLIVER W. HOLMES, *Law in Science and Science in Law*, an address delivered in 1899, *id.* at 210, 242.

⁸ Walter Wheeler Cook, in *MY PHILOSOPHY OF LAW* 51, 59 (Julius Rosenthal Foundation, Northwestern University, 1941); see also Cook, *Scientific Method and the Law*, 13 *A.B.A.J.* 303 (1927); and references cited in appendix to first citation.

⁹ Beutel, *An Outline of the Nature and Methods of Experimental Jurisprudence*, 51 *COLUM. L. REV.* 415 (1951); FREDERICK K. BEUTEL, *SOME POTENTIALITIES OF EXPERIMENTAL JURISPRUDENCE AS A NEW BRANCH OF SOCIAL SCIENCE* (1957); also see Moore & Callahan, *Law and Learning Theory: A Study in Legal Control*, 53 *YALE L.J.* 1 (1943); Hull, *Moore & Callahan's "Law and Learning Theory": A Psychologist's Impressions*, *id.* at 330; Yntema, "Law and Learning Theory" *Through the Looking Glass of Legal Theory*, *id.* at 338; Cowan, *Legal Pragmatism and Beyond*, in P. L. SAYRE (ED.), *INTERPRETATIONS OF MODERN LEGAL PHILOSOPHIES* 130 (1947); Simpson & Field, *Social Engineering Through Law: The Need for a School of Applied Jurisprudence*, 22 *N.Y.U.L.Q.* 145 (1947); Cowan, *The Relation of Law to Experimental Social Science*, 96 *U. PA. L. REV.* 484 (1948).

¹⁰ *Symposium—Scientific Proof and Relations of Law and Medicine, second series*, 30 *MINN. L. REV.* 409-541 (1946); *Law-Science Symposium*, 31 *TEXAS L. REV.* 625-831 (1953).

producing a blueprint for Utopia. I think this most unlikely. Rather, it seems that the activities of those who practice jurisprudence and those who practice science are so unlike that there is very little communication and no exchange of work between them. Jurisprudence is engaged in asking questions such as: What is the nature of law? What is the end or aim of law? What is property? Why should people perform promises? Why should we punish criminals? Why should a man be held liable for negligence? These are questions that seek ultimate answers, like the "Why" of a curious child.¹¹ These are not questions that can be asked in or answered by any scientific discipline. In so far as there are answers to such questions as these, the answers are those of philosophy, ethics, aesthetics, or theology. In response to such questions man can offer only speculation, preference or faith. The unanswerable questions of life belong to the realm of philosophy,¹² and jurisprudence is the philosophy of law.

On the other hand, in science a question is meaningless unless it is possible to find some operation by which an answer may be given to it.¹³ While form is certainly not determinative, the questions of science, in contrast to those of philosophy, are likely to be of the "How?" variety: How do you know that? How do you do this?¹⁴ The questions of science do not seek ultimate answers, but only immediate answers, subject to further correction and modification as additional questions are formulated.¹⁵ One cannot convert philosophy to science merely by adopting the vocabulary or imitating the methods of science. To conduct a scientific inquiry, one must first ask a scientific question—one that poses a problem that science is capable of investigating. A scientific question must be one that can be answered, at least partially, by doing something and observing the result.

These considerations suggest why we do not have and are not likely ever to have a jurisprudence that is "experimental" or "scientific." Those who are interested in jurisprudence are, naturally enough, interested in the traditional questions and problems of jurisprudence. Although the term and the field of jurisprudence could be changed or extended to designate a new set of problems and operations, there is neither a reason for nor an advantage in doing this. On the contrary, an insistence that new problems and procedures are within the field of jurisprudence will inevitably be provocative of more futile disputes as to the "proper" scope and method of jurisprudence. It seems much more appropriate and profitable to use a different term for a different set of activities. Thus the term "jurimetrics" has been suggested, and is

¹¹ See JEAN PIAGET, *THE LANGUAGE AND THOUGHT OF THE CHILD* ch. 5 (Meridian ed. 1955).

¹² Edwin W. Patterson, in *MY PHILOSOPHY OF LAW* 231 (1941).

¹³ BRIDGMAN, *op. cit. supra* note 2, at 28; REICHENBACH, *op. cit. supra* note 2, ch. 16; WILLIAM S. BECK, *MODERN SCIENCE AND THE NATURE OF LIFE* ch. 3 (1957); KARL POPPER, *THE LOGIC OF SCIENTIFIC DISCOVERY* 40 *et seq.* (1959); *but cf.* BERTRAND RUSSELL, *HUMAN KNOWLEDGE: ITS SCOPE AND LIMITS* 447 *et seq.* (1948).

¹⁴ ANATOL RAPOPORT, *SCIENCE AND THE GOALS OF MAN* 35 *et seq.* (1950).

¹⁵ HANS REICHENBACH, *ATOM AND COSMOS* 291 (1933); JAMES B. CONANT, *ON UNDERSTANDING SCIENCE* 24-25 (1947); JAMES B. CONANT, *MODERN SCIENCE AND MODERN MAN* 106-07 (1952); BRONOWSKI, *Science as Foresight*, in JAMES R. NEWMAN (ED.), *WHAT IS SCIENCE?*, at 433 (1955).

gaining some use, as a designation for the activities involving scientific investigation of legal problems.¹⁶

It is unnecessary, and perhaps impossible, to give a precise definition to the field of jurimetrics. As in any pragmatic discipline, the definition will be given by the activities of its practitioners, and will undoubtedly change and expand as experiment and experience give answers to specific questions. The distinction between jurisprudence and jurimetrics is already evident. Jurisprudence is concerned with such matters as the nature and sources of the law,¹⁷ the formal bases of law,¹⁸ the province and function of law,¹⁹ the ends of law and the analysis of general juristic concepts.²⁰ Jurimetrics is concerned with such matters as the quantitative analysis of judicial behavior, the application of communication and information theory to legal expression, the use of mathematical logic in law, the retrieval of legal data by electronic and mechanical means, and the formulation of a calculus of legal predictability.²¹ Jurisprudence is primarily an undertaking of rationalism; jurimetrics is an effort to utilize the methods of science in the field of law. The conclusions of jurisprudence are merely debatable; the conclusions of jurimetrics are testable. Jurisprudence cogitates essence and ends and values. Jurimetrics investigates methods of inquiry.

The validity and importance of this approach has recently received official recognition. The Behavioral Sciences Subpanel of the President's Science Advisory Committee issued a report on March 21, 1962, with the approval of the White House, in which it recommended action to strengthen the behavioral sciences and improve their use.²² The report stated, *inter alia*:

The general aims and criteria of evidence of the behavioral sciences are the same as they are in other sciences; however, it has so far frequently been necessary to settle for more approximate answers—errors of measurement may be large, and often, where experiments are not yet possible, correlations still substitute for cause-effect relations. The number of variables apparently needed to understand many kinds of human behavior, when combined with random or uncontrolled variations familiar in most of the life sciences, account for imprecision of results. Nevertheless, behavioral scientists are finding ways to develop and test meaningful theories; they have managed to amass a considerable store of tested and useful information.

¹⁶Loevinger, *Jurimetrics—The Next Step Forward*, 33 MINN. L. REV. 455 (1949); Loevinger, *Jurimetrics: Science and Prediction in the Field of Law*, 46 *id.* 255 (1961); Johnson, *Jurimetrics and the Association of American Law Schools*, 14 J. LEGAL ED. 385 (1962); also see 62M MODERN USES OF LOGIC IN LAW 34, 36 (1962) [hereinafter cited as M.U.L.L.].

¹⁷JOHN CHIPMAN GRAY, *THE NATURE AND SOURCES OF LAW* (1st ed. 1909; rev. ed. 1921).

¹⁸GIORGIO DEL VECCHIO, *THE FORMAL BASES OF LAW* (1st ed. 1914).

¹⁹JULIUS STONE, *THE PROVINCE AND FUNCTION OF LAW* (1950).

²⁰ROSCOE POUND, *JURISPRUDENCE* (pt. 2 and pt. 7) (1959).

²¹In view of the present symposium it seems supererogatory to offer a bibliography of recent work in the field of jurimetrics. However, in addition to the present symposium, attention might be called to *LAW AND ELECTRONICS: THE CHALLENGE OF A NEW ERA* (Jones ed. 1962), and to M.U.L.L., the quarterly newsletter of the American Bar Association Special Committee on Electronic Data Retrieval, published quarterly in collaboration with Yale Law School. Each issue of M.U.L.L. contains a wealth of material within the field of jurimetrics.

²²*Strengthening the Behavioral Sciences*, 7 BEHAVIORAL SCIENCE 275 (1962); 136 SCIENCE 233 (1962); also see comment by John Lear in *The Saturday Review*, May 5, 1962, p. 35 *et seq.*

The impact of the behavioral sciences on our society is far greater than most people realize. At one level they are providing technical solutions for important human problems. But at a deeper level they are changing the conception of human nature—our fundamental ideas about human desire and human possibilities. When such conceptions change, society changes.

.....

Progress in behavioral science has come about by using the scientific processes of observing, experimenting, and extensively following up and correcting working hypotheses. Indeed, all the general attitudes and strategies of physical and biological science have found a place in behavioral science.

.....

Unsolved behavioral-science problems that are clearly solvable, and for which methods of attack are already identified, are no longer minor and trivial. Instead, both their scope and their scientific importance are substantial and steadily increasing.

As the report points out, recent developments in many areas of the behavioral science have been so rapid and diverse that it is difficult for anyone to get a complete picture of their scope and depth. This can only be suggested by examining illustrative examples of current research. One of the most interesting and seminal areas of work now being done in the field of jurimetrics involves the utilization of electronic computers for the storage and retrieval of legal data.

II

ELECTRONIC DATA RETRIEVAL

The task of data retrieval is one of the most basic, pervasive, and important of all the functions performed by lawyers and judges. This includes the activity which lawyers commonly refer to as "legal research," but also considerably more. It is important to note that when lawyers use the term "legal research" they mean library-searching, whereas scientists use the term "research" to mean laboratory experimentation. For the sake of both clarity and generality the term "data retrieval" is more useful in the present context.

One of the principal aspects of data retrieval in the law is that of finding applicable, analogous, or relevant precedential authority in the reported cases for determination of some current question.²³ Indeed, a large part of the formal professional education of the lawyer consists of training and exercise in the analysis of problems, the use of a legal vocabulary, and the use of legal index systems in order to perform this task. But this is by no means the extent of the data retrieval problem confronting the lawyer. The problem exists, in varying degrees of difficulty with respect to at least these categories:

- 1) statutes
- 2) court decisions
- 3) administrative regulations

²³ On the general considerations applicable to this problem, see Loevinger, *An Introduction to Legal Logic*, 27 *IND. L.J.* 471 (1952).

- 4) administrative decisions and orders
- 5) title records
- 6) mortgages, liens, and similar recorded instruments
- 7) cases and judgments in courts of record
- 8) patents
- 9) trademarks
- 10) legislative history
- 11) legal periodicals and other literature
- 12) files, records, and evidence in big cases

Consideration of the growing mass of material in nearly all of these categories has pointed to the need for improved techniques of handling the material.²⁴ Indeed it is not too much to suggest that unless some more efficient means is found for dealing with the constantly increasing volume of case material alone, the common law system itself will be in some peril. It is becoming increasingly difficult to search for and find the relevant authority in the enlarging bulk of all reported cases; and as this difficulty increases, so does impatience and dissatisfaction with the system.

A. The Pittsburgh "Key Words in Combination" Project

In response to the twin stimuli of this need plus the emergent development of computers developed for other purposes, a number of experimental projects have been undertaken in the field of electronic retrieval of legal data. One of the first projects in this field, and probably the most extensive and sophisticated one undertaken to date, is that of the Health Law Center at the University of Pittsburgh. Under the direction of John F. Harty, a research staff at the University began a study of hospital law in 1956 and undertook the writing of a manual on the subject.²⁵ The staff found that much of the material from the several states was statutory, but that there was little uniformity in indexing from state to state. Therefore research could not proceed in reliance on available indexes. As the interest moved into broader areas of "health law," the difficulties increased. To meet such practical problems, the staff undertook the storage of relevant statutory material on magnetic tapes for use in electronic computers.

In order to avoid problems of abstracting, indexing, or pre-coding, the full text of all statutes within the universe selected is recorded on tape. As an indication of the order of magnitude involved, the full text of the entire body of Pennsylvania statutes, composed of about 31,000 sections and comprising some 6,230,000 words, is recorded on four reels of tape.²⁶ As each statutory section is recorded it is given a document

²⁴LAYMAN E. ALLEN, ROBIN B. S. BROOKS & PATRICIA A. JAMES, *AUTOMATIC RETRIEVAL OF LEGAL LITERATURE: WHY AND HOW* (1962). This volume contains a most useful bibliography of the subject.

²⁵Harty, *Experience With the Application of Electronic Data Processing Systems in General Law*, 60D M.U.L.L. 158 (1960).

²⁶Harty, *The "Key Words in Combination" Approach*, 62M M.U.L.L. 54, 58 (1962).

number in serial order. When the statutes have been recorded, a vocabulary of the words used in the statutes is prepared by computer. Common words, such as "the," "a," "an," "therefore," "however," "by," and various other pronouns, conjunctions, and articles are eliminated as of no significance for retrieval purposes. All other words are recorded on separate reels of vocabulary tape which also list the document number of each section in which each word appears. The four reels containing the full text of the Pennsylvania statutes require five reels of vocabulary tape.²⁷

When material is sought from the recorded statutes in the Pittsburgh system, the question is formulated and then phrased in terms that appear in the vocabulary of the statutes (the terms appearing in the statutory vocabulary can readily be determined as an alphabetical print-out is obtained from the vocabulary reels). This question is then "programmed," or arranged in a series of steps each involving an operation that the machine can perform. In effect, the machine is instructed to find and report all documents containing a certain word, or, preferably, certain words appearing in specified combinations. Both the words, the combinations, and the relationships (including order and spacing of words) are contained in the instructions given the machine. The appropriate words to use in searching for statutes relating to a subject of interest are determined either by examination of the vocabulary list or by consulting a thesaurus of legally-related (rather than strictly synonymous) terms. The search instructions are then formulated to secure all sections containing the key words in the specified combinations. These may call for all sections in which two words (or any other number) appear regardless of order; they may require the words to appear in a certain order (as in the phrase "good faith"); they may require the words to appear in the same sentence or within a certain number of words of each other.

For example, suppose the question put is: What are the rights of illegitimate children and the duties owed to them by their parents under Pennsylvania statutes? The steps that are to be programmed, expressed in ordinary language, will be these.

First, find and store the numbers of all documents which contain one of the following words:

baby
babies
child
children
foundling
infant
minor
offspring
Label this list—A

²⁷ *Ibid.* For a more complete technical description of the KWIC system, see Kehl, Harty, Bacon & Mitchell, *An Information Retrieval Language for Legal Studies*, 4 COMMUNICATIONS OF THE ACM (Association for Computing Machinery) 380 (1961).

Second, find and store the numbers of all documents which contain one of the following words:

parent
 parents
 unwed (or) unmarried (in the same sentence and within one word
 of one of the following)
 father
 fathers
 man
 men
 mother
 mothers
 woman
 women

Label this list—B

Third, set up and store a list of the numbers of those documents which contain both a term on list A and a term on list B. Label this resulting list—C.

Fourth, find and store the numbers of all documents which contain one of the following words:

bastard
 bastards
 born (in the same sentence and separated by two words from the
 word) wedlock
 illegitimate (where it does not occur with the word "purpose" in
 the same sentence)
 natural (occurring within one word of any of the terms in list
 A)

Label this list—D

Fifth, merge the numbers in lists C and D and print the full texts of the documents whose numbers appear on the resulting list—E.

It is possible to give the computer three forms of output commands: "list," "cite," or "print." The command "list" causes the computer to list the document numbers of all sections meeting the criteria of the search. The command "cite" causes the machine to print the section numbers resulting from the search in the form of conventional legal citations. The command "print" produces the full text of all sections (or documents) resulting from the search.

This system of storage and retrieval has been called the "Key Words in Combination" approach by the project director, John Harty.²⁸ It has certain obvious advantages and disadvantages. The full text recording avoids the problem of conventional abstracting, which is not only a problem of analyzing the legal significance of a document, but also requires foresight as to all possible applications and implications of the document that may arise in the future. Of course, such foresight can never be complete, and to that extent every abstract is likely to be inadequate.

²⁸ *Ibid.* KWIC is also used to refer to the similar Keyword-In-Context index developed by H. P. Luhn of International Business Machines Corporation.

Abstracting involves the further practical problem that it is most difficult to employ sufficient first-rate talent to do all such work that needs to be done in the law. Thus, every lawyer searching for material that is indexed by conventional abstracts, regardless of his own ability or diligence, is limited by the ability and performance of the abstracter. The system of full text recording is completely open-ended in the sense that it permits search for novel relationships, and, indeed, imposes no limitation or requirements upon the search except that it be in the vocabulary of the text. This, of course, may be a substantial limitation but there are available techniques to deal with it.

Probably the principal merit of this system is its efficiency in affording a high degree of probability of retrieving relevant data. A test of the comparative efficiency of the system was made by running six searches (including one dealing with the rights of illegitimate children, used as an example above) both by computer and manually (which is the current idiomatic euphemism for mental effort in this field). The computer searches produced more than twice as many references (177) deemed relevant by the researchers as the manual searches (72), and the manual searches produced a minimal number of references missed by the machine (2).²⁹

There are several disadvantages of the system of full text storage. It may very well be impractical for handling data that are more voluminous and less structured than statutes, since this would require recording an impractical mass of material. It is estimated that it would require 100 reels of tape to record the reported decisions of Pennsylvania, which is twenty-five times as many reels as those containing the statutes of that state. Presumably the vocabulary reels for the cases would be in proportion. Further, the mass of case material will increase at a much greater rate than the mass of statutes. The cost of recording this mass of material would be high. More important, a single search of this number of tapes, using present equipment and techniques, would be uneconomic by any standard. Further, full text recording of documents of substantial length, such as judicial decisions, vastly increases the difficulty of developing sophisticated and discriminating search systems.

B. The Western Reserve "Semantic Coded Abstract" Project

A system particularly keyed to handling voluminous and complex material has been developed at Western Reserve University and has been called the "Semantic Coded Abstract" approach.³⁰ Western Reserve University established a Center for Documentation and Communication Research in 1955, which has been engaged since that time in the investigation of methods of searching literature by electronic machines.³¹ The principal work done by the WRU Center has been in the field

²⁹ *Id.* at 60. Also see Swanson, *Searching Natural Language Text by Computer*, 132 SCIENCE 1099 (1960).

³⁰ Melton, *The "Semantic Coded Abstract" Approach*, 62M M.U.L.L. 48 (1962).

³¹ Melton & Bensing, *Searching Legal Literature Electronically: Results of a Test Program*, 45 MINN. L. REV. 299 (1960); see also Kent, *A Machine That Does Research*, Harper's Magazine, April 1959, p. 67 *et seq.*; Kent & Barhydt, *Compromise—A Key to Documentation*, Datamation, Oct. 1961, p. 26 *et seq.*

of metallurgy, but it has also done substantial work in other fields of science, engineering, and medicine. In 1959 and 1960, some work was done in the legal field on an experimental and pilot project basis, using the sales provisions of the Uniform Commercial Code and cases interpreting this part of the code. The same methods of abstracting, indexing, storage, and retrieval were applied to the legal material as to the non-legal material.

In the semantic coded abstract system, the documents to be indexed are first abstracted manually in some detail.³² The abstract is recorded in natural language stating the significant index terms, role-indicators assigned by the abstracter to show syntactical relations between the terms, and punctuation symbols performing a largely conventional function. This abstract is then copied on a key-punch unit and the punch card is put through equipment which automatically retrieves semantic-code equivalents from the file for the abstract terms, role indicators, and punctuation, and transcribes the abstract onto magnetic tape in the semantic code. The semantic code is the heart of the WRU system. It consists of a series of four-unit combinations of letters and numbers which symbolize the logical elements of the terms to be encoded, and of a highly simplified symbolic grammar which consists of three-unit combinations of letters called "role indicators" which are used to translate the variety of English syntactical expressions into a uniform and consistent coded pattern.

The four-unit combinations of letters are called "semantic factors" and are constructed by using three letters to symbolize a particular concept while varying the fourth letter, called an "infix," to indicate variations of that concept. For example, the basic concept *information* is symbolized by the semantic factor D()CM. The infix "A" symbolized the idea *a form of*. Thus the semantic factor DACM symbolizes the logical concept *a form of information*. A four-digit numerical factor added to the four letter semantic factors symbolizing a term serves to distinguish synonyms or close approximations from the logical elements of the specific term. For example, the concept of *swimming* is analyzed into the logical elements of *motion* and *water*. The forms of the term are differentiated by various infixes symbolizing different logical correlates, thus:

swim: MATN HVDR 2001 (*i.e.*, a form of motion acting upon water)

swimming: MUTN HVDR 2001 (*i.e.*, used for motion acting upon water)

swam: NWTN HVDR 2001 (*i.e.*, acted upon by motion acting upon water)

Each of the foregoing terms has essentially the same basic logical concept and therefore has the same exterior numerical factor. Examples of related concepts that are distinguished by the numerical factor are *blueprint* and *specifications*. In encoding the term *blueprint* it is analyzed into the logical elements of (a) a form of information, (b) used to regulate, and (c) acting upon construction. These ideas are symbolized by the semantic factors of:

³² The description of the semantic coded abstract system and the illustrations given in the text are taken from the references cited in notes 30 and 31, *supra*.

DACM—a form of information

RUGL—used to regulate

CVNS—acting upon construction

However, the logically similar term *specifications* analyzes into the same elements and is represented by the same semantic factors. Therefore an exterior numerical factor is used to distinguish these terms, which are represented in this fashion:

blueprint: CVNS DACM RUGL 3002

specifications: CVNS DACM RUGL 3001

It is obvious that the analysis of terms into logical elements may be a difficult task. Therefore the system utilizes a code dictionary in which terms and phrases are correlated to generic concepts and which also acts as a thesaurus of terms. The code dictionary is on punch cards so that abstracts and questions can be quickly and automatically checked against the dictionary.

When a request for search is made in the semantic coded abstract system, the question is analyzed in the same manner in which a document is analyzed for abstracting. When the logical elements of the question have been determined by manual analysis, they are checked against the code dictionary and encoded in the semantic code. The coded question is then put on punched paper tape, which is run through the computer to secure documents that fit the specifications of the question. The output of the machine consists of the serial numbers of relevant abstracts, which can then be examined in their uncoded version to determine the citations of original documents that are desired.

The semantic coded abstract system is by no means a complete departure from traditional indexing. It depends upon persons skilled in the relevant field to select and record significant aspects of the subject matter for future retrieval, and also to construct search specifications, leaving only the mechanical process of comparison and correlation of search specification with index to the machine operation. Furthermore, the use of coded semantic factors to represent conceptual relationships involves the creation of indexing "cluster points" which are similar to the "Key Numbers" of the West Publishing Company's digest system. The WRU system does avoid the rigidity inherent in the hierarchical topical arrangement of the conventional digest system and permits coordinate indexing and searching. It certainly has a flexibility that exceeds any conventional index; permits searches for specific words, generic concepts, or both simultaneously; and provides for specificity or generality at almost any level of abstraction.

The normalizing of language and use of a machine thesaurus permits the retrieval of documents that may be relevant although they do not contain any of the terms used in the search. This characteristic would be of particular value in problems that might involve comparative or foreign law. However, the WRU system not only does not avoid the problems and limitations of manual abstracting but imposes

some additional functions on the abstracter, thereby exacerbating the difficulties of this process. Further, the system imposes its own logic on both the library of stored material and on each question. It remains to be demonstrated by further work both that this can be done successfully with a large body of legal material and that, if done, it will not unduly constrain the use of such material.

There are also "computer languages" that have been developed and used successfully in particular applications, such as COBOL.³³ However, the primary purpose of these is to facilitate programming and secure compatibility among various computer systems, rather than to establish the conceptual or logical framework within which data may be manipulated. The law possesses its own conceptual framework, and it seems dubious that it will prove useful to attempt to impose an additional logical structure on legal material for purposes of an automatic retrieval system.

C. The Oklahoma State "Points of Law" Approach

The existing conceptual framework of the law itself is used for storage and retrieval in the "Point of Law" approach used at Oklahoma State University.³⁴ This system involves the analysis of each document (case, legal periodical article, or whatever) by a lawyer to determine the legal issues decided or covered. Each point of law is given a code number, which, of course, represents the particular concept involved in that point of law. For example, the legal concept *res ipsa loquitur* is represented by a single code number. When a field of law has been completely analyzed, a directory is prepared, listing all points of law in alphabetical order and showing the code numbers corresponding to each legal concept. In storing data, legal material is recorded in its conventional form on magnetic tape except that it is preceded by the appropriate code number or numbers.

To secure material in this system, the lawyer must analyze his problem to determine the point (or points) of law involved and then check against the system directory to insure that he has chosen a point that is included in the system. If he finds his point of law listed, he ascertains the corresponding code number and requests the machine to search for all material indexed under that code number. The machine will print out the legal citation of material identified by the code number, the citation plus head-notes or the citation, head-notes and complete text. A search can be made for a number of points simultaneously.

This system appears to be essentially a matter of putting the familiar West Key Number system, or something very much like it, on computer tape. It has the advantage of utilizing concepts and terms that are already well known to lawyers and thus working with an idiom that lawyers are willing to accept. However, it seems to be limited in both its adaptability to new concepts and its ability to absorb

³³ COBOL—1961, REVISED SPECIFICATIONS FOR A COMMON BUSINESS ORIENTED LANGUAGE (1961).

³⁴ Morgan, *The "Point of Law" Approach*, 62M M.U.L.L. 44 (1962). The work at Oklahoma State University in this field was being conducted by Professor Robert T. Morgan, whose untimely death may handicap full experimentation with this approach.

large additional quantities of data while still retaining the advantages of fast and economical searching. The system obviously has the limitations and disadvantages of all systems that rely on manual abstracting, and, in addition, depends upon the coincidence of analysis by the abstracter and the questioner, possibly even to a greater extent than present conventional digest systems. This is the kind of a system that the lawyer first contemplating automatic data retrieval imagines the future will provide.³⁵ However, the scientist in the field says, "Direct mechanization of traditional library classification is like building locomotives to run with legs."³⁶

D. The "Association Factor" Method

While some have been seeking to use the new technology of electronic data machines to perform conventional tasks faster and more easily mechanically, others have been developing genuinely new methods of problem solving. One of the most interesting techniques in the data retrieval field, of particular significance to law, is the development of the "Association Factor" method of retrieval.³⁷ This is a system in which the terms of a question are automatically expanded to retrieve relevant documents that may be indexed under other terms, and the search is so processed by computing correlation factors electronically that the documents resulting from the search are delivered with a rank order of "weight" indicating probable relevance to the inquiry.

In the association factor system the collection of documents is initially indexed by the uniterm coordinate index system. This is indexing by terms drawn from the text of the documents themselves without any topical or hierarchical arrangement. In coordinate indexing the index terms are independent of each other and data is retrieved by manipulation of the index terms, usually (although not necessarily) to secure a "match" or "intersection" of two or more terms that will designate an intersection, or other logical function, of the classes represented by the index terms and that is likely to contain documents desired.³⁸ In the coordinate indexing used in reported applications of the association factor system the effort was made to eliminate all synonyms by cross referencing them to a single term in a thesaurus used as part of the system.³⁹ Two files are prepared, as in usual coordinate term indexing. The first is a document-term file showing the index terms assigned to the respective documents. This information is recorded on punch cards in this format:

Document 1: Term a, Term c, Term d, Term x, etc.

Document 2: Term b, Term d, Term e, Term y, etc.

Document 3: Term a, Term b, Term f, Term z, etc.

³⁵ See, e.g., Loevinger, *The Industrial Revolution in Law*, 60 J. M.U.L.L. 56 (1960).

³⁶ Fairthorne, *Automatic Retrieval of Recorded Information*, 1 COMPUT. J. 1 (1958).

³⁷ Stiles, *The Association Factor in Information Retrieval*, 8 J. ASS'N FOR COMPUTING MACHINERY 271 (1961).

³⁸ See JOSEPHINE J. JASTER, BARBARA R. MURRAY & MORTIMER TAUB, *THE STATE OF THE ART OF COORDINATE INDEXING* (prepared for the Office of Science Information of the National Science Foundation by Documentation, Inc., 1962). This contains an excellent bibliography of the subject.

³⁹ Stiles, *supra* note 37.

The cards are then processed by machine to produce an inverted term file on magnetic tape in the following format:

Term a: Doc. 1, Doc. 3, Doc. 5, etc.

Term b: Doc. 2, Doc. 4, Doc. 6, etc.

Term c: Doc. 1, Doc. 4, Doc. 8, etc.

This gives a list of all index terms, together with the serial or accession number of each document indexed by each term stored on magnetic tape available for subsequent computer manipulation. The actual indexing in the applications of this system that have been reported to date has been done manually, but it could just as well be done automatically when satisfactory methods of automatic indexing have been developed.

In the use of this system, when a search question has been formulated it is fed into the computer, which then derives the association factor expressing the numerical value of the degree of association between each term of the request and each index term by an established formula.⁴⁰ In essence, this shows the index terms that occur in the same document as any of the request terms, and the degree of coincidence between them as measured by the number of documents in which both occur more frequently than might be expected if there were no logical association. Terms are dropped if they have an association factor of less than 1.00, or some other specified value, with terms of the request. The remaining terms provide a list of index terms associated with each term of the request. This list of associated terms is called a "term profile."

It should be noted that the term profiles contain terms that are statistically, but not necessarily semantically, related. The second step, after preparing a term profile for each term of the request, is to compare the term profiles of all the request terms and then select those index terms which appear in all or a given number of profiles. These selected terms are called "first generation" terms. In selecting the first generation terms it is possible to select terms that represent the logical product or logical sum or other logical function of request term profiles. Synonyms or near synonyms are not likely to be found in these first generation terms.

The third step in this method is to take the list of first generation terms and to use these as request terms repeating the process of the first two steps. The new terms secured at this stage are called "second generation" terms. These second generation terms are likely to include synonyms, various grammatical forms, and

⁴⁰ The formula used by Stiles, *supra* note 37, for computing the association factor is:

$$\log_{10} \frac{\left(|fN - AB| - \frac{N}{2}\right)^2 N}{AB(N-A)(N-B)} = \text{ASSOCIATION FACTOR},$$

where A is the number of documents indexed by one term; B is the number of documents indexed by a second term; f is the number of documents indexed by the combination of both terms; and N is the total number of documents in the collection. If AB is greater than fN the association is negative.

even variations in spelling of the request terms. The fourth step is the preparation of an expanded list comprising the original request terms, the first generation terms, and the second generation terms. All of the terms on this list will now have an association factor greater than the minimum specified as acceptable for the search. The association factor of each term in this list to every other term in the list is computed and added. The sum of these association factors for each term is then divided by the total number of terms on this expanded list. This gives a weight which indicates the probable relevance of each term on the expanded list to the original request.

The fifth step in this system is to search the document collection by matching terms in the expanded list with the index terms of each document in the collection. Whenever any terms match, the weight of the request term on the expanded list is assigned to the corresponding document. The sum of these weights for each document is the document relevance number. This indicates the degree of fit between the request and the contents of the document. The documents thus retrieved are then arranged according to these weights and the search thus produces a list of documents arranged in order of probable relevance to the initial inquiry, the higher numerical value indicating greater probable relevance. It is notable that this system will retrieve documents that may be relevant to a request even though the document has not been indexed by any terms in the initial request.

The association factor system is extremely flexible since it develops the inter-relationship of terms at the time of each search, rather than fixing it in advance according to any predetermined system. The system may be used to obtain a response in as much or as little depth as is desired by varying the weight of probable relevance that is acceptable in retrieving documents. The association factor system has been successfully used in experimental application to legal material, including both statutory and case material, although the results of this experiment have not yet been reported in a publication.⁴¹ The initial use of an automatic thesaurus to expand the terms of a request, coupled with association factor manipulation, promises to provide one of the most powerful tools for data searching and retrieval that has yet been devised.

E. Probabilistic Indexing

The association factor system operates upon the search procedure to provide retrieval based upon comparative relevance. There are other systems that secure a similar result by modification of the indexing. The basic concept seems to have

⁴¹ An experiment in utilization of the association factor system was undertaken with a body of legal material relating to antitrust law at the Graduate School of Public Law of George Washington University, Washington, D.C., in 1961 and 1962. The project was supported jointly by the school and the Datatrol Corporation, and the work was in large part supervised and conducted by John Lyons. A paper giving a summary report of the experiment was prepared in August 1962 and distributed at the American Bar Association convention in San Francisco; however, the project has not yet been fully reported in any publication. Mr. John Lyons has graciously made the summary report available to the author of the present article and has assisted the author by providing additional information.

been suggested by Maron and Kuhns in proposing a system of "probabilistic indexing."⁴² They point out that there is no strict one-to-one correspondence between the information content of a document and its set of indexes. All indexing is based upon an estimated probability that if a user wants information of the kind contained in the document he will formulate a request using the assigned index term. However, since there is a degree of uncertainty inherent in the assignment of all index tags, it is more reasonable and realistic to make the assignment of index tags on the basis of a probabilistic judgment. This is done by weighting index terms assigned to documents with relevance numbers which indicate the probable relevance of a document for anyone framing a request including that index term. This technique is called "probabilistic indexing."

In the procedure actually adopted for testing this technique, a collection of documents was read, and keywords were selected and listed by the readers. The keywords were then sorted into categories on the basis of a subjective judgment as to their meanings. A keyword could belong to more than one category. The categories of keywords were then assigned names or meaningful labels or tags, which were thereafter used as index terms. The index terms, or category names, were then assigned to documents by giving each document the index term of every category containing a keyword that appeared in that document. Each document was then re-read and the reader, or indexer, assigned a weight from an eight point scale indicating the degree of relevance of the index term to the contents of the document. This, of course, is a probabilistic estimate, or an indication of probable relevance to a request for information.

Retrieval in this system is accomplished by presenting a request in terms taken from the index vocabulary, together with the value of a relevance number below which documents are considered not to be of sufficient importance to be useful. The computer will then locate documents indexed by request terms, rejecting those that have less than the threshold relevance value, and will produce an ordered list of documents which satisfy the request, ranked according to their probable relevance. In experimental usage, it was found that the system efficiently retrieved relevant documents and required the reading of substantially fewer documents than conventional systems. However, the experiments were of limited scope, and it appears in any event that the high degree of dependence of this system upon human reviewers greatly reduces its efficiency.

A proposal to secure the advantages of probabilistic indexing while making the assignment of relevance values automatically by mathematical formula, rather than by subjective estimate, is involved in "latent class" analysis and indexing.⁴³ Latent

⁴² Maron & Kuhns, *On Relevance, Probabilistic Indexing and Information Retrieval*, 7 J. ASS'N FOR COMPUTING MACHINERY 216 (1960).

⁴³ Baker, *Information Retrieval Based Upon Latent Class Analysis*, 9 *id.* 512 (1962).

class indexing depends upon the use of keyword index terms, but it eliminates the step of reducing the keywords to categories and assigning relevance numbers to the connection between the data and the category names. Latent class analysis forms large categories (called latent classes) of documents, which deal with a common topic, upon the basis of the relationships among the keywords in the total collection of documents.

First, the categories, or latent classes, are formed by determining the degree of association among the keywords of the whole document collection in a manner roughly similar to the determination of the association factor mentioned above, but utilizing different mathematical procedures. Next, a "response pattern" is determined for each document. This is simply a pattern of concurrence of index keywords found in each document. Then the documents are assigned to the latent classes on the basis of correspondence between the response (or keyword) patterns in the documents and the keyword organization of the latent classes, and an ordering ratio is computed for each document which represents the probability that the keyword pattern found in that document will be generated by (or be found in) a document from that latent class. In other words, the ordering ratio represents a statistical probability relationship between the keyword pattern that might be expected from a document that fits the keyword pattern of a latent class perfectly, and the keyword pattern actually occurring in each document assigned to the latent class. Thus the ordering ratio is a practical indicator of probable relevance of each document to the keyword pattern of each latent class.

Documents are then to be stored by latent classes, with a specified minimum level of relevance as indicated by the ordering number of each document for inclusion in each class. A single document may be stored in more than one class, provided that its ordering number for each class is greater than the minimum specified for inclusion in that class. Searching is done by putting a request in terms occurring in the keyword vocabulary used for latent class analysis, and by specifying a minimum ordering number or degree of relevance desired in retrieved documents. The computer uses the pattern of keywords to locate the latent class which has the highest probability of containing documents that may yield the pattern of keywords in the request. The search will then secure the documents stored in the identified latent class and having ordering numbers higher than the minimum specified, arranged in rank order according to probable relevance to the request. The latent class indexing system has not yet been used in information retrieval, and therefore its utility is speculative. It is apparent, however, that it would require recomputation and reorganization of the latent classes and of the documents assigned to the latent classes as the total collection of documents grows significantly in size. While not an insuperable problem, this may be an important practical objection to the use of this system.

III

LEX—THE ANTITRUST DIVISION'S SYSTEM

An attempt to use machines for the storage and retrieval of legal material that is more practical than experimental has been undertaken recently in the Antitrust Division of the Department of Justice.⁴⁴ Since this attempt was made for utilitarian rather than scientific reasons, the system may be less interesting and original than some of those mentioned above that are seeking to develop or test new techniques. However, because it is designed and used for practical purposes, the DoJ system may have its own particular interest for lawyers.

In designing and constructing a data retrieval system for the Antitrust Division, there were certain requirements that constituted minimum specifications. First, the system had to be capable of handling a large body of data of various kinds including such material as briefs, judicial opinions, and statutes. Second, the system had to be open ended so that material could be added continuously without undue difficulty. Third, the system had to be operable with the relatively limited equipment presently owned by the Department of Justice. Fourth, the system had to be adaptable to, and capable of taking advantage of, more advanced equipment when that becomes available. Fifth, the system must be capable of being designed and operated with a relatively small staff of specially trained professionals (there are just two lawyers with training in computer theory and technology on the staff at the present time), although some additional lawyers without technical training in the field and a small staff of technicians (machine operators etc.) might be available to assist. Sixth, the system must be readily explicable, as well as comprehensible to and usable by a large staff of lawyers with a high order of legal skills but no special training or interest in the new techniques of data retrieval. Seventh, the system must be compatible with existing library and legal methods, which will necessarily continue to be used by the staff for all matters not covered by the system. Eighth, the system must be more efficient than existing available indexes and methods of indexing. Ninth, the system must be economical.

The system which has been designed and put into partial operation toward the end of 1962 appears to meet the requirements specified, although no firm conclusions can be hazarded until there has been a substantial period of experience in using it. In any event, it is anticipated that there will be significant modifications made as experience is gained. The DoJ Antitrust Division system is designed as a legal index and accordingly has been named "LEX" by its designers. LEX will be in four parts. Part one will index court decisions relating to antitrust law. Part two will index

⁴⁴A Legal Reference and Data Retrieval Unit has been established within the Antitrust Division to deal with problems of storing and retrieving various kinds of data. This unit has been in charge of Mr. John Lyons and Mr. Michael Duggan who have devised and put into operation the system described in the text. This system has not previously been described in any publication. The description in the text is based upon the author's personal knowledge of the system. Credit and thanks must be given to Mr. Lyons and Mr. Duggan both for their achievement in designing and constructing the system and for their assistance to the author in providing information for this article.

material originating in the Antitrust Division, such as memoranda, briefs, and official statements. Part three will index the legislative history of antitrust and related statutes. Part four will index legal periodical material related to antitrust law. At the time of writing, parts one, three and four are still in preparation and only part two has been substantially completed and put in use. However, the system is essentially the same for all four parts, so a description of part two will serve adequately to characterize the other parts.

The urgent need for part two of LEX arose out of the fact that the Antitrust Division has had several hundred lawyers continuously engaged in practicing antitrust law for many years, and that these lawyers have been busily engaged in producing briefs, memoranda, and similar material on many facets of antitrust law throughout that period. Nevertheless, there has been no systematic indexing of this material previously. Consequently there was no way of knowing whether a given problem had been exhaustively analyzed and considered in some previous memorandum or was being encountered for the first time, other than the memory of older staff members and such information as might be obtained by inquiry among other staff members. Even these sources were not very readily available to the approximately one hundred lawyers stationed in the six field offices of the Division. Thus the prompt provision of a workable means of retrieving this material was a matter of some importance.

Like most other present indexing systems, LEX uses a manually created abstract of the underlying document. For material generated within the Antitrust Division, it is thought that the author can do the best job of abstracting. Accordingly, abstract sheets have been mimeographed and distributed to all lawyers—with instructions to fill out one abstract sheet and furnish it, together with two copies of the document, to the Legal Reference Unit for every memorandum, brief, or similar document drafted. Of course, it is not essential that the abstracting be done by the author of the original document, and there is a small staff of lawyers available to the Legal Reference Unit for abstracting. Each abstract sheet shows the title and date of the document abstracted, an abstract in enough length and detail to indicate fully the contents of the document, the name of the author of the document, the title of the case to which the document relates, and a few other items that are thought to be of possible interest, such as the commodity involved, names of parties, court, and various file numbers.

The abstract is reviewed in the Legal Reference Unit by one and sometimes two lawyers with some experience in abstracting. When thought appropriate, reference is made to the original document to correct or expand the abstract. The reviewer in the Legal Reference Unit then assigns descriptors to the document to serve as index terms. The descriptors are simply terms that are thought likely to be associated in the mind of a lawyer with the contents of the document. So far we have devised no system for selecting descriptors objectively, although we recognize this as desirable.

It is hoped that by having descriptors assigned to all documents by a small staff, some uniformity may be achieved. However, experience casts some doubt on this, as documents that have accidentally been processed twice with an intervening period of time have not been assigned identical descriptors. There is no limit to the number of descriptors that may be assigned to a single document, and most documents are expected to be indexed by ten or more descriptors. Each document is then assigned a serial or accession number.⁴⁵ Punch cards are then prepared for each document, containing the document accession number, the descriptors indexing the document, the abstract, and the other information contained on the abstract sheet. These cards are then processed by machine to produce nine different lists which comprise the basic index used. These lists are printed out, reproduced by reduced photo-offset, and made available to the staff as a single index volume in nine sections.

Section one is a thesaurus of terms used. This is to provide cross-references among descriptors and from synonyms or near-synonyms to descriptors. This section can and will be expanded to afford reference from any word likely to be used by a lawyer to some related descriptor.

Section two is a descriptor index. This is a list of descriptors alphabetically arranged, and provides a reference to the number of each document that is indexed under each descriptor.

Section three is a compilation of abstracts arranged in serial order by document number. It is intended that anyone using the index will refer from some other section to this section in order to determine from the abstract whether the documents he has found cited are relevant to his search.

Section four is a statute and rule index which lists antitrust statutes and federal court rules in order by conventional methods of citation. Under each citation is a reference to the numbers of all documents containing significant internal references to the statute or rule cited. In view of the relatively limited number of statutes and rules that are of importance in the antitrust field and the high clustering of references to a few statutes, it is doubtful that this section will prove to be of much value. If it does not prove useful, it will be dropped from future editions of the index.

Section five is a subject matter index. This is an arrangement of topics and sub-topics in conventional hierarchical classification with document numbers under each entry. The arrangement of this section is similar to a conventional legal digest system, except that the LEX system has been constructed for the antitrust field and therefore is more detailed with respect to the law in this field than any other presently available index. This section offers a conventional alternative to the descriptor index as a means of seeking relevant references on the basis of subject matter. Whether such an alternative will continue to remain desirable when more advanced equipment and techniques are available remains to be determined. For the present,

⁴⁵ In the first edition of part 2 of LEX all documents were assigned both a serial number and a micro-film accession number. However this appeared to be unnecessary, and all documents are now assigned a single serial-accession number which serves both to identify and locate the document.

the availability of these two differing index approaches offers the opportunity for comparison of their efficiency in use.

Sections six, seven, eight, and nine are more specialized methods of finding potentially relevant documents from particular known information. Section six is an index based on "Blue Book" numbers. These numbers are assigned in serial order chronologically to all antitrust cases brought by the Department of Justice. The "Blue Book" numbers are available both from internal and from published sources. Section seven is a product classification index. Commodities are listed in numerical order according to the Federal Supply Classification system. Section eight is an index based upon the names of defendants arranged in alphabetical order. Section nine is an index based upon the names of the authors of documents arranged in alphabetical order. In each of these last four sections the index shows the numbers of all documents generated by the case, product, defendant named (*i.e.*, case) or author listed in the respective index section.

The full text of all documents listed is stored on microfilm. The microfilm is contained in magazines identified by letters. The first part of each document accession number consists of two letters that identify the microfilm magazine which contains the text of that document. To read a document, the appropriate magazine is inserted into the microfilm reading machine.⁴⁶ This automatically illuminates the screen and starts the film drive motor. An image appears on the screen which identifies the number of the document positioned for reading. A lever on the machine causes the film reel to be rotated until the desired document image is reached, when the motor is stopped and the document is read. If a copy of a page or of an entire document is desired, a button is pushed and a print of the desired page or document is made automatically.

It is self-evident that in its present form LEX is a relatively rudimentary and unsophisticated apparatus for retrieval of legal data. Although the assignment of descriptors permits coordinate indexing and retrieval, the descriptors will probably not be utilized for the discriminating selection of which they may be capable. However, there are several important advantages of this system.

First, it is the only one now in actual use by a group of practicing lawyers who have no special training in the use of such systems. It is thought that whatever system is ultimately developed will eventually have to be adapted to use by lawyers who are not specially trained in its use. Experimentation with a system that, whatever its other shortcomings, is in actual use by lawyers is likely to contribute greatly to the development of a better system to meet the needs of lawyers.

Secondly, it is believed that the basic structure of the LEX system, involving the storage of full text of documents on micro records with index terms and abstracts recorded on punch cards, tape discs, or similar means for manipulation, is the pattern that will ultimately be found most practical and economical.

⁴⁶ The machines presently in use are Recordak Lodestars.

Thirdly, LEX provides the opportunity for feedback in use. In actual searching for legal reference material, there is a constant interchange of information and instructions between the searcher and the body of data. As information is secured, the search is modified; finding a particular case, citation, or dictum may wholly change the direction of the search. It is thought essential to provide the means of maintaining such feedback in any system that is to be found practical by lawyers. An elaborate and sophisticated search system that is of such complexity that a substantial period of time must elapse between asking one question, securing a response, and asking a second question is likely to be less useful than a less complicated and less sophisticated system that furnishes a rapid response to questions and permits the asking of successive questions quickly.

Fourth, LEX is flexible and can be easily adapted to more refined techniques and more advanced equipment when that becomes available. The index terms used by LEX provide coordinate indexing that is adapted to automatic searching, and can also be used in a system of association factor searching. Since documents are also independently indexed by their own abstracts, LEX can utilize natural language questions in automatic full text searching of abstracts. While this is not quite the equivalent of full text searching of original documents, it may serve as a practical substitute.

Finally, automatic indexing and abstracting may be utilized when and if the techniques and equipment for performing this function satisfactorily become available. Since the full texts of all documents are stored on micro records, automatic abstracting and indexing may be applied to documents already recorded, as well as to new accessions.

IV

MICRO-IMAGE STORAGE OF DOCUMENTS

A. From Circulating to Duplicating Libraries

The structure of LEX is dictated in substantial part by an appraisal of the potentiality of equipment. No one can predict what may be possible in the future in recording, manipulating, and retrieving data by electronic devices. Conceivably, it may become possible to record all the decisions in a field of law, or perhaps in the entire law, in such fashion that their full text can be searched electronically in microseconds. But no device capable of doing this is available now or foreseeable in the near future. On the contrary, the mass of material to be stored in even a limited field such as antitrust law is so great that putting it all on computer tape is not only unduly expensive but is almost prohibitive to search. On the other hand, techniques and devices now available will permit full text storage and retrieval by micro records of virtually unlimited quantities of material.

For example, a recently reported technique for the storage and dissemination of microdocuments has made high density document storage feasible at a linear reduc-

tion of 200:1, representing an area reduction of 40,000:1. This is known as the photochromic micro-image technique, or PCMI.⁴⁷ Using this technique, it is possible to record a 300-page book within one square inch of film. A 3 x 5 inch photochromic plate can contain 2,625 micro-images, and by reproduction on a micro-image card, the entire contents of eight to ten average size books can be recorded on a 3 x 5 inch card. A file of one million document pages can be stored in micro-image form on less than 400 3 x 5 inch cards, which are a stack about four inches high. It is calculated that the total volume of all reported judicial decisions is slightly less than 5,000,000 pages, and that all statutes constitute about another 1,000,000 pages.⁴⁸ Accordingly, by this method all reported judicial decisions and all statutes can be recorded in full text on 3 x 5 inch cards in a file approximately 24 inches in depth. In this system, any desired document page can be retrieved and read or printed out in a hard, full size copy quickly and on demand.

Another system recently advertised, and presently in commercial production, can store 30,000,000 documents in micro-image in the space of an ordinary file cabinet.⁴⁹ Within one minute, this system can locate and produce a black and white hard copy of any page in the file.

The storage of documents in micro-image form has many advantages besides saving space. The documents are preserved intact and the wear and tear of handling and use that is inevitable in an ordinary library is avoided. Perhaps most important, the documents are not lost, stolen, or misplaced, but remain constantly available. The ability to reproduce desired portions of the documents easily eliminates any substantial disadvantage, and may offer a significant advantage, at least for reference material. It seems likely that the duplicating, rather than the circulating, library is the library of the future.⁵⁰ The storage, retrieval, and reproduction techniques now available for use with microdocuments indicate that the storage problem has been, or can readily be, solved. However, this manner of storage precludes any practical method of direct searching of the text and therefore requires retrieval by accession number, necessitating an independent indexing system. So it appears that indexing or abstracting, or both, will continue to be an integral part of future systems for handling large volumes of data.

B. Indexing and Abstracting Problems

The techniques for manipulating index terms and matching request terms to index terms are certainly capable of further development and refinement. Nevertheless, the work that has already been done suggests that there are known techniques which will produce reasonably satisfactory results, and that, in any event, the

⁴⁷ *Photochromic Micro-Images*, Reproduction Methods, November 1962, p. 48 *et seq.*

⁴⁸ Harty, *The "Key Words in Combination" Approach*, 62M M.U.L.L. 54, 56 (1962).

⁴⁹ See Magnavox advertisement for MEDIA system in Wall Street Journal, Nov. 9, 1962, p. 11.

⁵⁰ See Heilprin, *Communication Engineering Approach to Microforms*, American Documentation, July 1961, p. 213 *et seq.*; Heilprin, *On the Information Problem Ahead*, *id.*, Jan. 1961, p. 6 *et seq.*

capabilities of present equipment permit modification and improvement of these processes within an established and operating system. The greatest problem now apparent in the field is that of establishing adequate clues to document contents by indexing or abstracting. These procedures are still done manually in all present operating systems, although some provocative experiments have been performed with automatic indexing and abstracting.

On first impression the task of indexing or abstracting may appear to be one requiring the exercise of human intelligence and therefore beyond the range of automation. However, communication and information have been subject to mathematical analysis for more than a decade,⁵¹ and there are some techniques for mathematical "content analysis" of meaning.⁵² All such objective techniques arise from the fact that "meaning per se is not negotiable for purposes of communication, except by means of arbitrary physical tokens such as the spoken or the written word."⁵³ It is quite possible, although difficult, to manipulate and analyze objectively the objective manifestations of meaning—both the discrete symbols constituting words and the complex patterns constituting syntax and even style.⁵⁴ It has been remarked that the semantic aspects of communication are irrelevant to the engineering aspects, but this does not mean that engineering aspects are necessarily irrelevant to the semantic aspects.⁵⁵

To illustrate by a simple example, the meaning of the messages that are to be carried is of no significance to the problem of constructing an efficient telephone system. On the other hand, the efficiency of the telephone system is indispensable to the communication of meaningful messages by telephone, and an inefficient system or a "bad" connection can result in a garbled message or wholly prevent communication. Similarly, automatic indexing experiments are based upon the thesis that statistics on the kind, frequency, location, order, and arrangement of words are adequate to make reasonably good predictions about the subject matter (*i.e.*, appropriate classification for library purposes) of documents containing those words.⁵⁶ It is not necessary to know the significance of an individual word to a writer or reader in order to establish its value in conveying information or acting as a clue for document retrieval. Thus it may reasonably be assumed that a writer will use terms bearing on his special topic with greater frequency than those which do not.

⁵¹ See C. E. SHANNON & W. WEAVER, *THE MATHEMATICAL THEORY OF COMMUNICATION* (1949).

⁵² See, *e.g.*, HAROLD D. LASSWELL, NATHAN LEITES & ASSOCIATES, *LANGUAGE OF POLITICS* (1949); CHARLES E. OSGOOD, GEORGE J. SUCI & PERCY H. TANNENBAUM, *THE MEASUREMENT OF MEANING* (1957).

⁵³ HANS P. LUHN, *AUTOMATED INTELLIGENCE SYSTEMS—SOME BASIC PROBLEMS AND PREREQUISITES FOR THEIR SOLUTION* (IBM, April 1, 1960).

⁵⁴ See Yngve, *Computer Programs for Translation*, *Scientific American*, June 1962, p. 68.

⁵⁵ SHANNON & WEAVER, *op. cit.* *supra* note 51, at 99-100.

⁵⁶ MARON, *Automatic Indexing: An Experimental Inquiry*, 8 J. ASS'N FOR COMPUTING MACHINERY 404 (1961); but cf. EDMUNDSON & WYLLYS, *Automatic Abstracting and Indexing—Survey and Recommendations*, 4 COMMUNICATIONS OF THE ACM 226 (1961); also cf. SALTON, *Manipulation of Trees in Information Retrieval*, 5 COMMUNICATIONS OF THE ACM 103 (1962).

However, mere frequency of use within a document is not a valid indication of content, since the most frequently used words are general terms (like "important") and logical connectives (like "and," "if," "but," "all," "the," and "an").⁵⁷ In fact, a substantial part of any message may be disregarded for purposes of information theory analysis. Only about fifty per cent of the words in a message are determined by the free choice of the sender, the other fifty per cent being determined by the accepted rules governing the use of symbols or the structure of language; so that one-half of every ordinary message is redundant in the sense of being unnecessary to complete the sense of the message.⁵⁸ By general information theory, the amount of information conveyed by a word should vary inversely, rather than directly, with its frequency of usage, since infrequency evidences greater selectivity in its choice.⁵⁹ It is the rare, special, or technical word that will indicate most strongly the subject of an author's discussion. However, the frequency referred to here is frequency in general usage, not within a particular document.

This has led to the formulation of the "relative frequency technique" as a method of automatic indexing.⁶⁰ This uses the ratio between the frequency of a word's occurrence in a particular document and the frequency of its general use to measure the significance of the word as an indicator of the contents of the document. The matrix for the determination of general frequency may be the literature of the language generally, a broad field of interest, such as law, or a special field of interest, such as antitrust law. Not enough work has been done to establish the utility of the several alternative matrices or ratio measures, but subjective judgment suggests that the use of a special field matrix will be both more practical and more informative. Refinements of this technique may involve attaching weights to the position of occurrence of words; for example, giving more weight to an occurrence in the title, first paragraph, or summary and conclusion than in some other part of an article. Further, this technique is capable of establishing a statistically weighted connection between a given word and the contents of a document, in order to provide the foundation for probabilistic, or comparative relevance, indexing.

The methods of automatic abstracting are essentially similar to those of automatic indexing. Some relatively simple expedients have been suggested, such as extracting the first and last paragraphs of a document, or taking the first sentence of each paragraph or of every other paragraph. These do not appear to be satisfactory, however, and more refined and sophisticated techniques are undoubtedly required. When such simple and relatively mechanical techniques are rejected, the problem of automatic abstracting becomes essentially the same as the problem of automatic indexing. Once it becomes possible to determine statistically the significant words in a document, an abstract can be similarly constructed statistically by determining

⁵⁷ Maron, *supra* note 56; Edmundson & Wyllys, *supra* note 56.

⁵⁸ SHANNON & WEAVER, *op. cit.* *supra* note 51, at 104.

⁵⁹ SHANNON & WEAVER, *op. cit.* *supra* note 51, at 99-106; Edmundson & Wyllys, *supra* note 56.

⁶⁰ Edmundson & Wyllys, *supra* note 56.

the relative significance of sentences. Sentences may be considered significant if they contain single words having a specified significance value or if they contain a plurality of words having a specified aggregate significance value. Then such sentences, together with so much of the context as is found necessary to make them usefully readable (as, perhaps, the sentence preceding and the sentence succeeding each significant sentence), may be extracted from the document to form the abstract. To those who are concerned about the intellectual quality of an abstract produced automatically, the words of working scientists in the field may be instructive.⁶¹

Any technique for automatic abstracting will, ultimately, succeed or fail according as the author of the document being abstracted succeeds or fails in expressing his thoughts clearly in the document. One cannot put more into an automatic abstract than the author of the document provides. This dependence may actually turn out to be an advantage, for automatic abstracts will reveal the poverty or richness of a document without either disguise or embellishment; whereas conventional abstracts sometimes make a document appear more valuable than it really is and are liable to subconscious biases and misunderstanding on the part of the human abstracter. One may even hope that when, inevitably, automatic abstracting becomes widely used, it will tend to induce authors to set forth more clearly and explicitly the main points of their articles.

V

OUTLOOK

There is much more that can and will be said on the subject of legal data retrieval. In an article such as this it is possible only to suggest the character and scope of the work being done in the field and its possibilities. While major interest naturally centers on problems of indexing, abstracting, storing, and retrieving textual matter such as court decisions, briefs, memoranda and similar material, this is by no means the limit of application. On the contrary, the techniques of automatic data processing have been and are being successfully applied to such legal problems as the filing of annual corporation reports;⁶² the filing and retrieval of chemical patents;⁶³ trademark searches;⁶⁴ the manipulation of large accumulations of evidence in a "big case";⁶⁵ storage and retrieval of real estate title records;⁶⁶ record keeping by the Internal Revenue Service, including setting up and posting taxpayers' accounts, computing taxes and refunds due, searching for failures to file returns and

⁶¹ *Id.* at 232.

⁶² Russo, *Mechanized Processing for Annual Corporation Reports in Connecticut*, 60D M.U.L.L. 133 (1960).

⁶³ Andrews, *Experience with Electronic Searching of United States Patents*, 60D M.U.L.L. 168 (1960); Newman, *Information Retrieval Research in the U.S. Patent Office*, 60J M.U.L.L. 45 (1960).

⁶⁴ Trademark International of Detroit, Mich., a commercial organization, is now offering an automated international trademark searching service.

⁶⁵ Freed, *Machine Data Processing Systems for the Trial Lawyer*, The Practical Lawyer, April 1960, p. 73. Individual lawyers in the Antitrust Division of the Department of Justice have also used simple punch card systems for handling evidentiary material in an antitrust case.

⁶⁶ *Computer Storage of Title Records Offers Solution to Paperwork Explosion*, 47 IOWA L. REV. 382 (1962); 12 LAW REV. DIG. 1 (1962).

selecting returns for detailed audit;⁶⁷ and the compilation and analysis of identical bidding reports by the Department of Justice.⁶⁸ These techniques are more widely used, both in and out of government, for the handling of technical information not directly related to law.⁶⁹ Such technical information does, of course, become relevant to law when legal investigation or litigation touches technical fields. Accordingly, the Department of Justice has established liaison with both ASTIA (the Armed Services Technical Information Agency) and NASA so that abstracts and bibliographies on a wide variety of technical subjects can readily be secured by lawyers concerned with those subjects.⁷⁰

The advantages of automated data retrieval are so great that there is a tendency to view these developments principally in terms of convenience, speed, economy, and efficiency,⁷¹ although thought may also be given to some of the legal problems involved in widespread computer employment, such as the effect upon traditional notions of negligence and the shopbook rule.⁷² Certainly these are important considerations. But they should not wholly obscure the broader implications of the new technology. There are at least two fundamental differences between the electronic computers now available and devices (such as calculating machines) which preceded them. First and foremost, the new computers can be internally programmed to duplicate or simulate such important psychological functions as remembering, forgetting, analyzing, differentiating, deciding and pursuing other goal-directed activities.⁷³ Secondly, the new computers have the ability to manipulate non-numerical information, including words and the abstract symbols of mathematical logic.⁷⁴ The computers have just one inescapable theoretical limitation: every term and operation must be made explicit and nothing can be presumed, assumed, implied, or based on intuition.⁷⁵ The computers can handle all symbols and programs that

⁶⁷ Smith, *Tax Practice Under Automatic Data Processing*, 62M M.U.L.L. 7 (1962).

⁶⁸ The program under which identical bidding reports are received and its genesis are described in IDENTICAL BIDDING IN PUBLIC PROCUREMENT, REPORT OF THE ATTORNEY GENERAL UNDER EXECUTIVE ORDER 10936 (1962). No sophisticated techniques for analysis of the data have yet been utilized, but automatic manipulation has been found indispensable to an orderly arrangement and presentation of the mass of data involved with the limited staff available. Pages 112 to 404 of the printed report are simply reduced photo copies of the actual print out of the machine from the data cards.

⁶⁹ See Senate Comm. on Government Operations, *Documentation, Indexing, and Retrieval of Scientific Information, A Study of Federal and Non-Federal Science Information Processing and Retrieval Programs*, S. Doc. No. 113, 86th Cong., 2d Sess. (1960), and *Addendum to Senate Document No. 113 of the Eighty-sixth Congress*, S. Doc. No. 15, 87th Cong., 1st Sess. (1961).

⁷⁰ The extensive libraries of ASTIA and NASA are automated. See references cited in preceding footnote.

⁷¹ Freed, *Prepare Now for Machine-Assisted Legal Research*, 47 A.B.A.J. 764 (1961).

⁷² Freed, *A Lawyer's Guide Through the Computer Maze*, The Practical Lawyer, Nov. 1960, p. 15; Brown, *Electronic Brains and the Legal Mind: Computing the Data Computer's Collision with Law*, 71 YALE L.J. 239 (1961); also cf. Freed, *Some Legal Aspects of Computer Use in Business and Industry*, [1961] J. INDUSTRIAL ENGINEERING 289.

⁷³ Lingoes, *Information Processing in Psychological Research*, 7 BEHAVIORAL SCIENCE 412 (1962).

⁷⁴ *Ibid.* Also see the pioneer article in the field, Bush, *As We May Think*, Atlantic Monthly, July 1945, p. 101.

⁷⁵ Stone, Bales, Namenwirth & Ogilvie, *The General Inquirer: A Computer System for Content Analysis and Retrieval Based on the Sentence as a Unit of Information*, 7 BEHAVIORAL SCIENCE 484 (1962).

are completely explicated. In other words, the computers can do anything we tell them to do; their only absolute limitation is our ability to provide instructions.

Even this limitation has uses, for it permits us to test the clarity and consistency of our own thought and expression. This can be done for large numbers of similar cases by computer programming. It can be done more efficiently for individual cases by using the intellectual model or analogue of a computer, symbolic logic. This is a far more flexible and powerful intellectual tool than syllogistic logic, and is adapted to detecting ambiguities and inconsistencies and determining validity or invalidity in complex problems of logic, expression and communication.⁷⁶ Any operation that can be performed on a computer can be expressed in modern logic; and any operation that can be expressed in modern logic can be performed on a computer. The basic elements that these procedures have in common are explication, model-making (or operational organization or programming), and the production of a testable conclusion. These elements carry us beyond the subject matter that either computers or symbolic logic ordinarily work with and into the broad field of scientific research generally on social problems.

The yearning for a "scientific" answer to legal problems, as noted at the beginning, is an old one. The great Cardozo cried for a "table of logarithms" to produce "the formula of justice."⁷⁷ What we are learning is that the logarithms and other mathematical formulae are available—not to produce a "formula of justice," but only more modestly to answer specific questions by programmed operations. If questions are properly put to science by the law, sensible answers can be secured. The field of behavioral science is far too broad today to review briefly, but examples of particular pertinence to law abound.

Four examples of behavioral science research methods have recently been reported from the Jury Project of the University of Chicago Law School.⁷⁸ One question asked was: What, if any, difference is there between the way a judge and a jury decide the same case? The method adopted was to ask the presiding judge in a large number of jury cases to answer a questionnaire before the jury verdict was returned, stating how he would have decided the case. The judges' answers and the jury verdicts were then compared. The results of this inquiry showed that judges and juries agreed in eighty per cent of the cases, that in eighteen per cent the jury acquitted where the judge would have convicted, and that in only two per cent did the jury convict where the judge would have acquitted.⁷⁹

⁷⁶ See Allen, *Some Uses of Symbolic Logic in Law Practice*, *The Practical Lawyer*, May 1962, p. 51; Allen, "Beyond Document Retrieval Toward Information Retrieval: Federal Estate Tax," a paper presented at the Conference on Tax Administrative and Tax Policy Implications of Electronic Data Processing, Harvard University, April 13-15, 1961; Stern, *Syntactical Ambiguity in the Clayton Act, Section 5(a)*, 60D M.U.L.L. 129 (1960); Vonneuman, *But How Do We Apply Modern Logic to Law?*, 60D M.U.L.L. 138 (1960).

⁷⁷ BENJAMIN N. CARDOZO, *THE PARADOXES OF LEGAL SCIENCE* 1 (1928).

⁷⁸ Zeisel, *Social Research on the Law*, in WILLIAM M. EVAN (Ed.), *LAW AND SOCIOLOGY* 124 *et seq.* (1962). The discussion in the text following this note is taken from the essay of Zeisel.

⁷⁹ Judge Gustavus Loevinger, who sat on the district court of Ramsey County, Minnesota, from

Another question was whether there was a greater tendency for people to file claims in some cities of the United States than in others. The investigation was made by a statistical study of claims filed with a cross section group of insurance companies arising out of automobile accidents and accidents in department and chain stores. The analysis produced a fairly consistent rank order of cities with Philadelphia emerging as the most "claim conscious" and Detroit as the least. A related question was whether juries in some regions or cities give higher awards in comparable tort cases than in other regions or cities. After analyzing a number of difficulties in establishing comparability, the experimenters secured detailed written descriptions of five personal injury cases. These were then presented to a large number of experienced insurance adjusters, each of whom was asked to estimate the jury verdict that would be expected from a jury in his locality for that case. The results indicated reliable regional variations, with estimated jury awards being about eighty per cent of the national average in the rural south and midwest and about 120 per cent of the national average in the metropolitan east and west coasts.

Another experiment was designed to determine what effect the formulation of the legal rule as to the defense of insanity in criminal cases actually has on jury verdicts. A mock trial based on the real Durham case was tape recorded, with all parts of the trial included from the judge's opening statement to the final instructions. Three recorded versions were produced that were identical except for the instructions to the jury. In one version the instructions were based on the M'Naghten rule, in a second version on the Durham rule, and in the third version no instructions on the subject were given. Juries were selected by lot from the actual jury pools of three metropolitan areas, and were presented with the recorded trial, followed by the same opportunity to deliberate and reach a verdict as in a real trial. The results of this experiment are yet to be published. The point of significance here is not that any particular investigation was made, or results secured, but rather that the problems of law are subject to investigation by the construction and operation of experimental models.⁸⁰

Computer techniques and applications seem to have attracted more attention, both from the technicians and the general public, than scientific theory and investigation of other kinds in the behavioral field. One of the reasons obviously is that the problem of data retrieval is so pervasive that automated data retrieval systems will have very wide utility. However, there are other problems of law that are almost, if not quite, as general as that of data retrieval. One of these is that of determining the reliability and credibility of testimony. Attention was first called to

1931 until 1955, kept a similar record of his own agreement or disagreement with jury verdicts over a period of more than ten years. His results were in substantial agreement with those of the Chicago experiment.

⁸⁰ For another example of interesting model design in the construction of scientific experiments for the investigation of forensic problems, see CARL I. HOVLAND (ED.), *THE ORDER OF PRESENTATION IN PERSUASION* (1957).

this problem and to the possibility of its scientific investigation by the classic statement of Münsterberg more than half a century ago.⁸¹ Some scientific work has been done in the field since that time, principally focused on the detection of deception,⁸² but the basic problem still remains and awaits interdisciplinary and scientific investigation.⁸³

Another problem of great generality and deep interest among lawyers is that of predicting judicial decisions. Recent work (some of it reported in other articles in the present symposium) has made it evident that this problem too, if presented in properly posed questions, is amenable to scientific investigation.⁸⁴ Two aspects certainly seem beyond question. First, as the methods of securing, indexing and retrieving legal data improve, it will become more likely that lawyers and judges faced with a given case will at least have the same legal precedents and principles at hand from which to begin their consideration. This will tend to give the lawyer somewhat greater assurance in forecasting the judicial decision than if he must take account of the possibility that the judge has discovered a different set of precedents than the lawyer was aware of when presenting the case. In the second place, if lawyers are to deal intelligently and effectively with scientific data—regarding judicial behavior or anything else—they must have at least some understanding of the intellectual framework of science, particularly mathematical expression, statistical measures and probability.⁸⁵ All of this is merely to say that jurimetrics is now practical, and that in the near future, its study will probably become essential for the individual lawyer.

Certainly the approach that has been characterized as jurimetrics does not offer any social panaceas. Essentially, it involves putting a series of questions that are capable of investigation to the test of investigation. It seeks not sudden revelations or universal laws but the slow accretion of tested information. It seeks to apply to legal problems "the same humble, honest objective approach that has characterized the development of science" in other fields.⁸⁶ Jurimetrics does not seek to oust

⁸¹ HUGO MÜNSTERBERG, *ON THE WITNESS STAND* (1908).

⁸² See Loevinger, *Jurimetrics—The Next Step Forward*, 33 MINN. L. REV. 455 nn.78 and 79 (1949); Rowell, *Admissibility of Evidence Obtained by Scientific Devices and Analysis*, 6 ARK. L. REV. 181 (1952); Symposium—*The Polygraphic Truth Test*, 22 TENN. L. REV. 711 *et seq.* (1953); Skolnick, *Scientific Theory and Scientific Evidence: An Analysis of Lie-Detection*, 70 YALE L.J. 694 (1961).

⁸³ See Kubie, *Implications for Legal Procedure of the Fallibility of Human Memory*, 108 U. PA. L. REV. 59 (1959).

⁸⁴ See GLENDON A. SCHUBERT, *QUANTITATIVE ANALYSIS OF JUDICIAL BEHAVIOR* (1959); Loevinger, *The Elements of Predictability in Judicial Decision Making* (delivered in 1960), in *LAW AND ELECTRONICS: THE CHALLENGE OF A NEW ERA* 249 *et seq.* (Jones ed. 1962); Loevinger, *Jurimetrics: Science and Prediction in the Field of Law*, 46 MINN. L. REV. 255 (1961); Kort, *Quantitative Analysis of Judicial Behavior*, 60 D M.U.L.L. 143 (1960); Nagel, *Using Simple Calculations to Predict Judicial Decisions*, *The Practical Lawyer*, March 1961, p. 68; Schubert, *Psychometric Research in Judicial Behavior*, 62 M.U.L.L. 9 (1962).

⁸⁵ See Loevinger, *The Elements of Predictability in Judicial Decision Making*, in *LAW AND ELECTRONICS: THE CHALLENGE OF A NEW ERA* 249 *et seq.* (Jones ed. 1962); Loevinger, *Jurimetrics: Science and Prediction in the Field of Law*, 46 MINN. L. REV. 255 (1961).

⁸⁶ The quoted phrase is from Szent-Gyorgyi, *Science, Man, and Politics*, *Saturday Review*, Oct. 20, 1962, p. 24.

jurisprudence, philosophy, or faith from men's lives. These, too, have their place. There will always be assumptions and choices to be made by the free spirit of a man, and no scientific operation or test can ever properly make or constrain such choices. Fears for the dangers of a "mechanized jurisprudence" are both quixotic and uncomprehending. Jurimetrics is not concerned with a debate as to whether the metaphorical life of the law has been logic or experience. Jurimetrics is concerned only with investigating the structure and dimensions of all experience that is relevant to the law.