

The Empirical Study of Reference

The first section of this essay reconstructs four empirical studies of reference, which then serve as a basis for the discussions which follow. Section II shows that the principal notions on which these studies are based are deficient. It is observed that their deficiencies do not permit reliable results. Section III discusses the conventional concept of reference, through which reference is perceived as performed only by reference librarians; it concludes that the empirical interpretation of the conventional concept is fruitless as a framework for the development of the empirical study of reference. Section IV considers reference outside of librarianship. It suggests that reference is a special case of problem solving. On the basis of this consideration, Section V proposes a new framework for the empirical study of reference.

SECTION I

FOUR EMPIRICAL STUDIES of reference are briefly reconstructed in this section. Of the studies of reference with which the author is familiar, they appear to contain the most insight. They are presented in chronological order. By chance the first two studies were written by librarians; the second two, by non-librarians. The reconstructions serve as a basis for the discussions of the following sections.

Cole understood reference to be the "queries which people bring to the desk of the reference librarian" and the librarian's response to the queries in terms of "materials best suited to the needs of the reader."¹ Her objective was "to identify reference questions . . . and

¹ Dorothy Ethlyn Cole, "An Analysis of Adult Reference Work in Libraries" (unpublished master's thesis, Graduate Library School, University of Chicago, 1943), p. 1.

describe . . . the sources of reference information"² in public, academic, and special libraries.

The data of her study were the result of a questionnaire. For each reference, as interpreted under Cole's definition, the questionnaire requested the following returns:

1. What the patron wanted.
2. Exact statement of the question.
3. Occupation of the patron.
4. Specific books and other tools used in finding the answer.
5. Other functions performed.
6. Was the question answered satisfactorily. If not, indicate reason.³

Fourteen libraries reacted to the questionnaire. Four of them were public, four were academic, and six were special libraries. The data were first manipulated for readings on reference questions, then for readings on reference sources.

² *Ibid.*, p. 2.

³ *Ibid.*, p. 24.

Mr. Gardiner is Director of Libraries in Normal State University, Normal, Illinois.

The first set of manipulations included these. The returns of No. 2, above, were classified by subject. That is, each statement of a question was mapped into one of the divisions of the Decimal Classification. The subject of a question was then characterized as a function of its origin, *i.e.*, public, academic, or special library. The returns of No. 3 were classed into such occupational categories as housewives, unskilled laborers, and skilled laborers. The returns of No. 2 were classified by time. The temporal classification consisted of the following historical periods: -499, 500-1899, 1900-1940, 1941, 1942- . The subject of a question was characterized as a function of its time. The returns of No. 2 were classified by complexity. This classification, in outline, included: fact, how-to-do, supporting evidence, and general information categories. The returns of No. 1 were classed by specificity. The returns of No. 1, that is, were compared with the respective returns of No. 2. If "what the patron wanted" was the same as "the exact statement of the question," then that question was classed as specific; otherwise it was classed as non-specific. Specificity was then characterized as a function of both origin and occupation.

The second set of manipulations required the following operations. The returns of No. 2 were classified by complexity. The returns of No. 4 were classified by form, *i.e.*, such as reference, non-reference, periodical, and document. The form of a source was then characterized as a function of the complexity of a question.

Unlike Cole, Breed⁴ did not indicate his concept of reference. Since his study was largely a follow-up of Cole's study, it was entirely unnecessary that he should. Indeed, as the following recon-

struction will show, it is reasonable to assume his entertainment of a concept of reference more similar to Cole's than to any alternative. Breed undertook the study of difficult reference questions in a large university library.

To Breed a difficult reference question was one for which "the ordinary reference sources failed to produce an answer."⁵ Breed purposed "to discover something about the persons asking the questions . . . the kinds of questions asked" and "to compare the extent to which certain kinds of knowledge . . . were called upon by the reference librarian in the search process."⁶ Breed's knowledge typology included knowledge associated with a general liberal arts background, subject specialization, personal knowledge, knowledge gained in the process of search, and library school and library experience. These knowledge types were intended to explain each step of the reference librarian's search. A reference step was defined as the reference librarian's consultation of any source or entry in a source.

The data of Breed's study was the result of a questionnaire. The questionnaire provided for the following returns on each reference:

1. A precise statement of the information wanted.
2. The name and/or occupation of the person or organization making the request, and the manner in which it was made (telephone, letter, or in person).
3. The time required to answer the question.
4. A step-by-step account of the method employed to deal with the question.
5. Whether the question was answered, partially answered, or not answered.⁷

The reference staff of Harper Library, University of Chicago, acted upon the questionnaire. After difficult questions were isolated from other questions, the data were manipulated for readings on those questions and for readings on the

⁴ Paul Francis Breed, "An Analysis of Reference Procedures in a Large University Library" (unpublished master's thesis, Graduate Library School, University of Chicago, 1955).

⁵ *Ibid.*, p. 19.

⁶ *Ibid.*, p. 1-2.

⁷ *Ibid.*, p. 4.

processes by which those questions were answered.

For readings on the difficult questions the following manipulations were made. The returns of No. 1, above, were classed by subject, *i.e.*, social, biological, or physical sciences, and the humanities. The returns of No. 2 were classified by origin, campus or non-campus. The questions falling into the former category, "were examined for deviations from a known subject specialization."⁸ The returns of No. 1 were classified as being specific or general. Generality was indicated if a question "called for broad treatment of a subject or . . . involved a process of selection and evaluation of material."⁹ There were very few general questions; they were not given further consideration in Breed's study. Those questions characterized by specificity were dichotomously classed as bibliographical or factual. Bibliographical questions were first classed by form, *i.e.*, questions about books, periodicals, learned society publications, or U.S. government documents, then classed by inaccuracies of personal or corporate authorship, title, and date of publication. The inaccuracy of a question was then characterized as a function of its form. Factual questions were classified into numerous categories, often individual categories.

The second set of manipulations were undertaken for readings on the librarian's search process. The returns of No. 4 were categorized by knowledge, as indicated above. The subject of a question was then characterized as a function of the knowledge required to answer it. Final steps used in answering questions were isolated. A final step was defined as "the step which resulted in a solution."¹⁰ Such steps were not necessarily final in the sense of last steps. The subject of a question was

then characterized as a function of the knowledge utilized by the librarian in taking final steps.

To the HERNERS, reference consisted of "all the problems within organizations or among groups of workers which require searches for information."¹¹ Though they condemned a study of reference "which leads or permits the information seeker . . . to place the task of getting the information he needs in the hands of a person or group outside of himself" as "a study of a certain, narrow type of information requirement,"¹² they produced such a study. And, though their concept of reference differed from Cole's, the concept of reference which guided their study did not. Their objective was to define "those information requirements which reference librarians and other information specialists are likely to be called upon to meet."¹³

The data of the HERNERS' study also resulted from a questionnaire. It requested the name of the organization receiving the question and the question itself. Fourteen atomic energy research and reference organizations forwarded to the HERNERS all questions received from the fall of 1956 through the fall of 1957. The data which the HERNERS received were manipulated only for readings on the questions.

After technical questions, those "involving one or more of the natural or engineering sciences,"¹⁴ were isolated; technical as well as non-technical questions were separately classified by subject. Non-technical questions were then disregarded. Technical questions, however, were further classified by number of discrete concepts and the logical re-

¹¹ Saul Herner and Mary Herner, "Determining Requirements for Atomic Energy Information from Reference Questions," *Preprints of Papers for the International Conference on Scientific Information, Washington, D.C., November 16-21, 1958* (Washington: National Academy of Sciences-National Research Council, 1958), p. 172.

¹² *Ibid.*

¹³ *Ibid.*

¹⁴ *Ibid.*, p. 173.

⁸ *Ibid.*, p. 22.

⁹ *Ibid.*, p. 24.

¹⁰ *Ibid.*, p. 52.

relationships among the discrete concepts.

A discrete concept was defined as a significant concept which could not be subdivided without changing its essential meaning. For example, "the question 'Give me information on engineering in nuclear reactors,' was taken to contain two concepts, 'engineering' and 'nuclear reactors.'"¹⁵ Only questions containing two or more discrete concepts were classed by logical relationships. The logical relationships were defined as logical sums, "where the requestor would settle for information about concept A or concept B"; logical products "where the requestor had to have information about concept A and concept B"; and logical differences, "where the requestor was interested in concept A, but not concept B."¹⁶

Carlson described reference as "human searching behavior."¹⁷ His study emerged from the problem of determining an optimum search procedure of a very large file. Reference librarians were chosen as subjects for his study for they are "the most trained in search procedures."¹⁸ Like the HERNERS, though his concept of reference differed from Cole's, the concept of reference which guided his study did not. The objective of Carlson's study was to describe search procedures used by reference librarians in order:

1. To illustrate that human search behavior can be precisely described. . . .
2. To improve the present search procedures used by humans. . . .
3. To develop new training procedures for librarians.
4. To make it possible to develop computer routines that could assist the human in making complex searches of a file.¹⁹

Three persons served as subjects.

¹⁵ *Ibid.*, p. 174.

¹⁶ *Ibid.*, p. 175.

¹⁷ G. Carlson, *Search Strategy by Reference Librarians* (Part 3 of the Final Report on the Organization of Large Files [Sherman Oaks, California: Advance Systems, Hughes Dynamics, 1964]), p. ii.

¹⁸ *Ibid.*, p. 2.

¹⁹ *Ibid.*, p. 1.

They were the reference librarians in a university medical library. The file searched was the library collection with which the subjects were familiar. "The basic procedure was to present a search problem to a reference librarian and then record in detail his search techniques."²⁰ Protocols, verbatim records, and descriptions of all that subjects say and do during a study, were obtained from the observations of two persons. The observers recorded everything they could see or hear during the searches. Presumably the subjects were requested to "think aloud" during the searches. After several protocols were gathered, they were analyzed for consistent search routines. The resultant routines were then tested against another protocol. The process was repeated. The product of these operations was a flow chart. The initial flow chart was compared with and modified by the original and additional protocols.

SECTION II

In this section it is argued that the major notions on which the reconstructed studies of reference are built are productive of results which are unreliable. For various, often similar reasons, these notions are in need of serious explication. In several cases it is noted that such explication does not seem forthcoming.

That the notion, the subject of a question, warrants discussion is by no means obvious. Cole was unable to correlate the subject of a question with either its origin or its historical period. Indeed, the HERNERS observed that "the results of a study such as the present one, based on questions from workers in a field other than atomic energy, would produce results quite different from those obtained in the present one . . . conducted at some future time, would also produce results at a variance with the present ones."²¹

²⁰ *Ibid.*, p. 3.

²¹ Hermer and Hermer, *op. cit.*, p. 176.

To Cole the significance of the subject of a question was its conjectured power to predict subject interests of reference patrons. Its significance to the HERNERS presumably was quite similar. It is questionable, however, that either Cole's or the HERNERS' results characterized the subject interests of patrons in the simple one-to-one correspondence supposed.

Cole defined the subject of a question by those operations which specify the class number of a book by means of the Decimal Classification. The subject of a question, then, was the subject as interpreted under that scheme. The interpretation of questions under that scheme, however has several obvious disadvantages.

Under interpretation of the Decimal Classification, for example, it is obvious that two or more questions about the same subject are not necessarily mapped into the same class. Though the pair of questions "Who chopped down the cherry tree?" and "Who was the first president of the U.S.?" both have the same subject, the latter is decimally classifiable under history but the former is classifiable under the social sciences. Or, given the pair of questions "What are the properties of an irregular cube?" and "What are the basic tenets of Lawsonomy?" the latter is unclassifiable until or unless it is known what indeed Lawsonomy is and the former, though it has no subject, is classifiable in the pure sciences.

To circumscribe the artificiality of the Decimal and similar classification schemes, consider criteria for a more natural scheme:²² (1) the subject of a question is independent of its various formulations and (2) it is about the extension of its subject term(s). Though these criteria dispel some of the difficulties just observed in connection with

the Decimal scheme, they create difficulties of their own. By condition 1, question (a) "Is Henry Miller the author of *Little Women*?" is by subject equivalent to question (b) "Is the author of *Little Women*, Henry Miller?" By condition 2, however, question (a) is about Henry Miller and question (b) is about Louisa May Alcott; therefore, questions (a) and (b) do not have the same subject and criteria 1 and 2 are contradictory.

Indeed, besides the absence of a simple one-to-one correspondence between the subject of a question and its interpretation under a classification scheme, as well as the lack of a formalization of our concept "about," there is experimental evidence that subject classification is far from the objective process assumed by either Cole or the HERNERS.²³ Appropriately recast, the subject of a question undoubtedly has something important to tell us of reference behavior. Under its present formulations, however, it has almost nothing to tell us.

Breed's notion of the difficulty of a question is similar to Cole's notion of the specificity of a question in several respects. Rather than descriptive of the patron, as she supposed, Cole's notion of the specificity of a question was equally descriptive of the librarian. Cole defined specificity as the absence of a difference between the patron's question and the librarian's interpretation of that question. To exemplify non-specificity Cole wrote of the student who asked for the other author of *The Taming of the Shrew* and the librarian who gave Henry James as the answer. If the librarian of Cole's example arrived at his answer by comparing the formulation of the question and the formulation of the solution at hand, then it is ob-

²² For this example credit is due Hilary Putnam, "Formalization of the Concept 'About,'" *Philosophy of Science*, XXV (April 1958), 125; cf. Nelson Goodman, "About," *Mind*, LXX (January 1961).

²³ See Mary Cuddy St. Laurent, "A Review of the Literature of Indexer Consistency" (unpublished master's thesis, Graduate Library School, University of Chicago, 1966).

vicious that a librarian who held "the Taming of a Shrew," or possibly *Turn of the Screw* as well, in his cognitive structure, undoubtedly would have answered the student of Cole's example in quite another way. Indeed, rather than exemplifying non-specificity, the question "Who besides Shakespeare wrote *The Taming of the Shrew*?"²⁴ would have exemplified specificity.

Breed observed that, when they were received, difficult reference questions "appeared on the surface to be unexceptional questions which the usual reference books would answer."²⁵ For that reason he defined difficult questions as those for which "the ordinary reference sources failed to produce an answer."²⁶ Like Cole's notion of specificity, Breed's notion of difficulty is indeterminately characteristic of either the patron or the librarian or, indeed, the environment in which the question is asked. Under Breed's definition the difficulty of a question is an empty notion. Under Breed's definition, that is, a question is difficult only if a question is difficult. Breed's manipulations on the set of difficult questions were productive, then, of a miscellany of tenuously related results.

Let us probe this matter somewhat further. Given Reitman's description of a problem as having an initial state and a terminal state, a problem solution may be described as being a set of operations by which the former is convertible into the latter.²⁷ On the basis of this description it is possible, for example, to express the question "How is a sow's ear changed into a silk purse?" by the following rule:

Description—contains sublist 1.
 sublist 1—sublist 2, sublist 3, sublist 4.

sublist 2—is, initial state; contains sublist 5.
 sublist 3—is terminal state; contains sublist 6.
 sublist 4—process; input, sublist 5; output sublist 6.
 sublist 5—is, sow's ear.
 sublist 6—is, silk purse.²⁸

If a person has an element in his cognitive structure which satisfies this rule, then that person has a solution to this question. If, however, that person does not have an element in his cognitive structure which satisfies this rule, then he cannot achieve a solution to the question unless he is successful in setting up substitutive or subordinate questions to the one at hand. For example, it is possible to achieve a solution to the question given above by replacing the implicit physical constraint by a value constraint such that the value of the sale of a sow's ear will equal the value, or partial value, of the purchase of a silk purse. In any case, short of knowing precisely what elements are held by a given cognitive structure and its environment and, indeed, of knowing what counts as a satisfactory solution to a given question, the task of explicating the notion "the difficulty of a question" will prove a formidable task.

At various points in his study Breed observed deficiencies in his knowledge typology. For example, he observed that liberal arts and personal knowledge "were too subtle to be grasped in a study of this kind,"²⁹ that knowledge gained in the process of search and library specialized knowledge were not readily distinguishable "where final successful steps were concerned,"³⁰ and that "the separation of kinds of knowledge is slightly unrealistic; the reference librarian combines knowledge of many differ-

²⁴ Cole, *op. cit.*, p. 40.

²⁵ Breed, *op. cit.*, p. 19.

²⁶ *Ibid.*

²⁷ Walter R. Reitman, *Cognition and Thought, An Information-Processing Approach* (New York: John Wiley & Sons, 1965), p. 133.

²⁸ *Ibid.*, p. 134.

²⁹ Breed, *op. cit.*, p. 52.

³⁰ *Ibid.*, p. 56-57.

ent kinds in his exercise of judgement and his interpretation of the data at hand.³¹ Although almost nothing is known about one's personal knowledge, what is reasonably conjectured when related to Breed's self-criticism, makes that criticism seem largely understated. Newell, Shaw, and Simon, for example, have conjectured that cognitive domains are structured by

1. A control system consisting of a number of memories, which contain symbolized information and are interconnected by various ordering relations. . . . 2. A number of primitive information processes, which operate on the information in the memories. . . . 3. A perfectly definite set of rules for combining these processes. . . .³²

If this conjecture is correct or even nearly correct, then one's knowledge consists not only of hierarchies of information but also hierarchies of rules for manipulating and modifying the information as well as the rules themselves. In that case, each sort of one's knowledge would be so intimately bound up with every other sort that to speak of identifying the sort of knowledge which led to a specific action would be misleading; to suppose the reality of such identification would be ridiculous.

Breed defined a reference step as the librarian's consultation of any source or entry in a source. Use of the reference step as the basic unit in the process of answering reference questions, however, has the same obvious disadvantages as the use of the human pulse beat to measure time. Like the human pulse beat, the reference step would differ for different people and would differ for the same person at different times, at different places, and under different conditions of mental and physical health.

Gagné has suggested several variables of individual difference which point up the problems inherent in such notions as the specificity or difficulty of a question, the knowledge used in answering questions and the reference step as defined by Breed.³³ One person, for example, may possess more information and more rules for applying that information than another; be more facile in the recall of stored information; more able to select and maintain conceptual distinctions in the face of conflicting cues; more fluent than another in formulating hypotheses; more steady in the retention of a solution model; more rapid in verifying a solution. Implicit in Gagné's comments in relation to such notions is that variables of individual difference must be given experimental control or that such notions must be given independent study and explication. Simon and Kotovsky, for example, have proposed a theory which "predicts rather successfully which problems, from a set of letter series completion test items, will be the more difficult for human S's."³⁴ In Section VI Hayes's experimental control over variables of individual difference is observed in a study of problems characterized by sequences of well-defined steps.

Though the HERNERS interpreted their results on the conceptual structure of reference questions into information retrieval, intuitively at least, these results would seem to have implications for reference. It is intuitively reasonable to suspect that in a spoken question-answering system, the number of significant concepts per question will not exceed the span of immediate memory, *i.e.*, the number of symbols which a person can repeat, orally or manually, in

³¹ *Ibid.*, p. 58.

³² Allen Newell, J. C. Shaw, and Herbert A. Simon, "Elements of a Theory of Human Problem Solving," *Psychological Review*, LXV (May 1958), 151.

³³ Robert M. Gagné, "Human Problem Solving: Internal and External Events," Benjamin Kleinmuntz (ed.), *Problem Solving: Research, Method, and Theory* (New York: John Wiley & Sons, 1965), p. 143-46.

³⁴ Herbert A. Simon and Kenneth Kotovsky, "Human Acquisition of Concepts for Sequential Patterns," *Psychological Review*, LXX (November 1963), 534.

correct sequence and without error after a single presentation. For example, as Miller has observed, "a person who can repeat nine binary digits will have a span of about eight decimal digits, seven letters of the alphabet, or five monosyllabic English words."³⁵ Also, it seems reasonable to expect a human encoding device to produce logical products more frequently than logical sums and to produce logical sums more frequently than logical differences. Several studies have suggested that humans categorize conjunctively more readily than disjunctively and prefer, in categorizing, positive over negative instances.³⁶ Quite why and under what conditions these results obtain is barely speculated upon as, in Bruner's words, "one eventually begins to wonder whether Nature herself does not abhor disjunctive concepts."³⁷ The HERNERS' results on the conceptual structure of reference questions reinforce these suspicions.

Recall that a discrete concept was defined by the HERNERS as a concept which is both significant to the information desired and indivisible with respect to its essential meaning. Essential meanings, however, presuppose criteria, determinancy, and uniformity of usage, which common language does not meet.³⁸ At the present time it can, at best, be said that a person "has a concept if he has a disposition on the basis of which he can make nominal classificatory statements or responses ('this is X that is not X'); it is assumed that the disposition is

learned from a number of instances which vary among themselves; it is assumed that the responses also be made to instances other than those contained in the set on which the concept was learned; it is taken that the classificatory response should not be the only possible one."³⁹ Indeed, the essential meaning of little from the essential subject of a question in Cole's study. In Cole's study it was the Decimal Classification which seemed to provide intersubjectively synonymous interpretations of subjects. In the HERNERS' study it undoubtedly was "a special classification scheme for librarians in the atomic energy field,"⁴⁰ a question in the HERNERS' study differs rather than the questions alone, which provided the results shown in Table 1 on the number of discrete concepts per question.

Logical products were defined by the HERNERS as information about A and B; logical sums, as information about A or B; and logical differences, as information about A and not B. Under this interpretation of the conceptual structure of reference questions the HERNERS reported the results shown in Table 2.

What is of interest to reference about these results is that they need not have turned out as they did. Assuming the psychological reality of discrete concepts, for the sake of argument, it is possible to interpret the logical relationships among the discrete concepts of the question, "Who besides Shakespeare wrote *The Taming of the Shrew*?" in any one of the following ways: (author and *The Taming of the Shrew*), (Henry James and *The Taming of the Shrew*) or (Henry James and *Turn of the Screw*), (*The Taming of the Shrew* and not Shakespeare) and so forth. Indeed, it is entirely clear that such re-

³⁵ George A. Miller, "Human Memory and the Storage of Information," *IRE Transactions on Information Theory*, IT-2, 1956 Symposium on Information Theory, Held at Massachusetts Institute of Technology, Cambridge, Massachusetts, September 10-12, 1956 (Cambridge, Massachusetts: M.I.T. Press, 1956), p. 131.

³⁶ Jerome S. Bruner, Jacqueline J. Goodnow and George A. Austin, *A Study of Thinking* (New York: John Wiley & Sons, 1956).

³⁷ *Ibid.*, p. 162.

³⁸ Carl G. Hempel, *Fundamentals of Concept Formation in Empirical Science*, Vol. II, No. 7 of the *International Encyclopedia of Unified Science*, ed. Otto Neurath, Rudolph Carnap and Charles Morris (Chicago: University of Chicago Press, 1952), p. 9-10.

³⁹ John P. Van De Geer and Joseph M. F. Jaspers, "Cognitive Functions," *Annual Review of Psychology*, XVII (Palo Alto, California: Annual Reviews, 1966), 149.

⁴⁰ HERNER and HERNER, *op. cit.*, p. 171.

TABLE 1
NUMBER OF DISCRETE CONCEPTS PER QUESTION⁴¹

	1	2	3	4	5	6
Number of concepts	1	2	3	4	5	6
Number of questions	466	1818	1167	327	73	0
Per cent of questions	12.1	47.2	30.3	8.5	1.9	0

TABLE 2
LOGICAL RELATIONSHIPS AMONG DISCRETE CONCEPTS⁴¹

	Logical products	Logical sums	Logical differences	Totals
Number of questions (with more than one concept)	3773	45	33	3851
Per cent of questions (with more than one concept)	98.0	1.2	0.8	100

sults as those given in Table 2 are largely dependent on the experimenter. For a more general example, from an arbitrary array of four attributes of two values each, as Bruner has observed, "one can define the *same* subset of instances with different types of concepts . . . the way in which a person will categorize *new instances encountered* will depend drastically upon the type of concept he has constructed . . . when one learns to categorize a subset of events in a certain way, one is doing more than learning to recognize new instances. . . . One is also learning a rule that may be applied to new instances."⁴²

The major notions around which the results of Carlson's study were developed are what he called regular human behavior and inconsistent human behavior. Since what he intended by either regular or inconsistent human behavior is not at all clear, the following paragraphs attempt to clarify the meanings of these notions and to point up their shortcomings.

One of Carlson's "most encouraging findings" was that "human search behavior is really quite regular."⁴³ By this

he meant, as he explained two sentences below, that "there is a great deal of common behavior"⁴⁴ in human search behavior. If regular behavior is that behavior which is common among searchers, then it is not unreasonable to expect inconsistent behavior to be that behavior which is not common among searchers. As Carlson was at pains to point out by numerous examples, however, inconsistent behavior was consistently practiced by the subjects of his study. In other words, it was also found to be quite common. Indeed, he explained that inconsistent behavior is that behavior which is detrimental to a successful search, *e.g.*, "human searchers are often influenced by the physical environment . . . he scans mostly those documents at a convenient eye level . . . he has a strong tendency to select clean, brightly colored books."⁴⁵ If, then, inconsistent behavior is that behavior which is both common and detrimental to a successful search, it does not seem unreasonable to surmise that regular behavior, in contrast, is that behavior which is both common and characteristic of successful searches. If these are the meanings which Carlson intended for these notions, then we shall see in

⁴¹ *Ibid.*, p. 175.

⁴² Bruner, Goodnow, and Austin, *op. cit.*, p. 44-45.

⁴³ Carlson, *op. cit.*, p. 28.

⁴⁴ *Ibid.*

⁴⁵ *Ibid.*, p. 131-32.

considering his flow charts of regular behavior and his discussion of inconsistent behavior that he stipulated unfortunate meanings for both of these terms.

If Carlson has indeed flow-charted human search behavior, then his flow charts must tend to meet at least one criterion. They must tend to specify a set of operations which is sufficient to achieve an answer to a question much as a reference librarian would. However, since a description of human search behavior must account for both that behavior which is successful and that behavior which is not successful in a search, Carlson's elimination of unsuccessful search behavior from regular behavior, indeed from the entire study, precludes that possibility. One is led to suspect, in fact, that instead of preparing a flow chart of human search behavior Carlson prepared a flow chart of information retrieval.

Considering Carlson's Generalized Flow Chart, reproduced on the following page, first from back to front, it is evident, on inspecting the second level of detail, that his "Exit" intends the printout of a bibliography. Unless Carlson intends something more general than what is ordinarily intended by "bibliography," then one must conclude that either Carlson equated the conclusion of all successful searches with listings as outputs or that Carlson has, indeed, not flow-charted human search behavior at all. That is, certainly, successful searches conclude with answers other than bibliographical listings, as answers which are simply affirmatives or negatives.

Considering this Generalized Flow Chart from front to back, again on the second level of detail, his "Select 2-4 keywords" intends: If terms of input not previously encountered, generate synonyms from a general dictionary; otherwise, generate synonyms from memory; then, if synonyms are not available from either a general dictionary or from memory, generate synonyms from a medical

dictionary. "Enter synonyms into synonym list; enter adjectives into search list, and rank."⁴⁶ Delete common nouns. Finally, place the remaining nouns on the search list and rank them. Surely, however, "Select 2-4 keywords," as just one detailed example from the Generalized Flow Chart, is not representative of successful or unsuccessful human search behavior, not even in Carlson's own terms since he observed that the reference librarians of his study *never* checked "the accuracy of their spellings, pronunciation, and meaning before the search starts."⁴⁷ Indeed, if in an act of communication, the decoding process of the listener is an approximate inverse of the encoding process of the speaker, then Carlson's "Select 2-4 keywords" taken inversely will surely not elicit a question from any speaker. The suspicion that Carlson's Generalized Flow Chart is a model of information retrieval rather than of human search behavior, as advertised, is more than warranted.

One of Carlson's "most significant findings" of the inconsistent sort was that "there is no feedback of the results of their search to the librarian."⁴⁸ On the basis of this finding Carlson inferred that "the reference librarian has no consistent way of improving the quality of his search."⁴⁹ He concluded that "any searches completed for a requestor should only be given out on the strict condition that the librarian be told about the usefulness or inadequacies of the result."⁵⁰

The inference that there is no feedback to the reference librarian was based on evidence illustrated by the following example: During the recording of protocols one of the librarians

⁴⁶ *Ibid.*, p. 8.

⁴⁷ *Ibid.*, p. 29.

⁴⁸ *Ibid.*, p. 35.

⁴⁹ *Ibid.*, p. 35-36.

⁵⁰ *Ibid.*, p. 36.

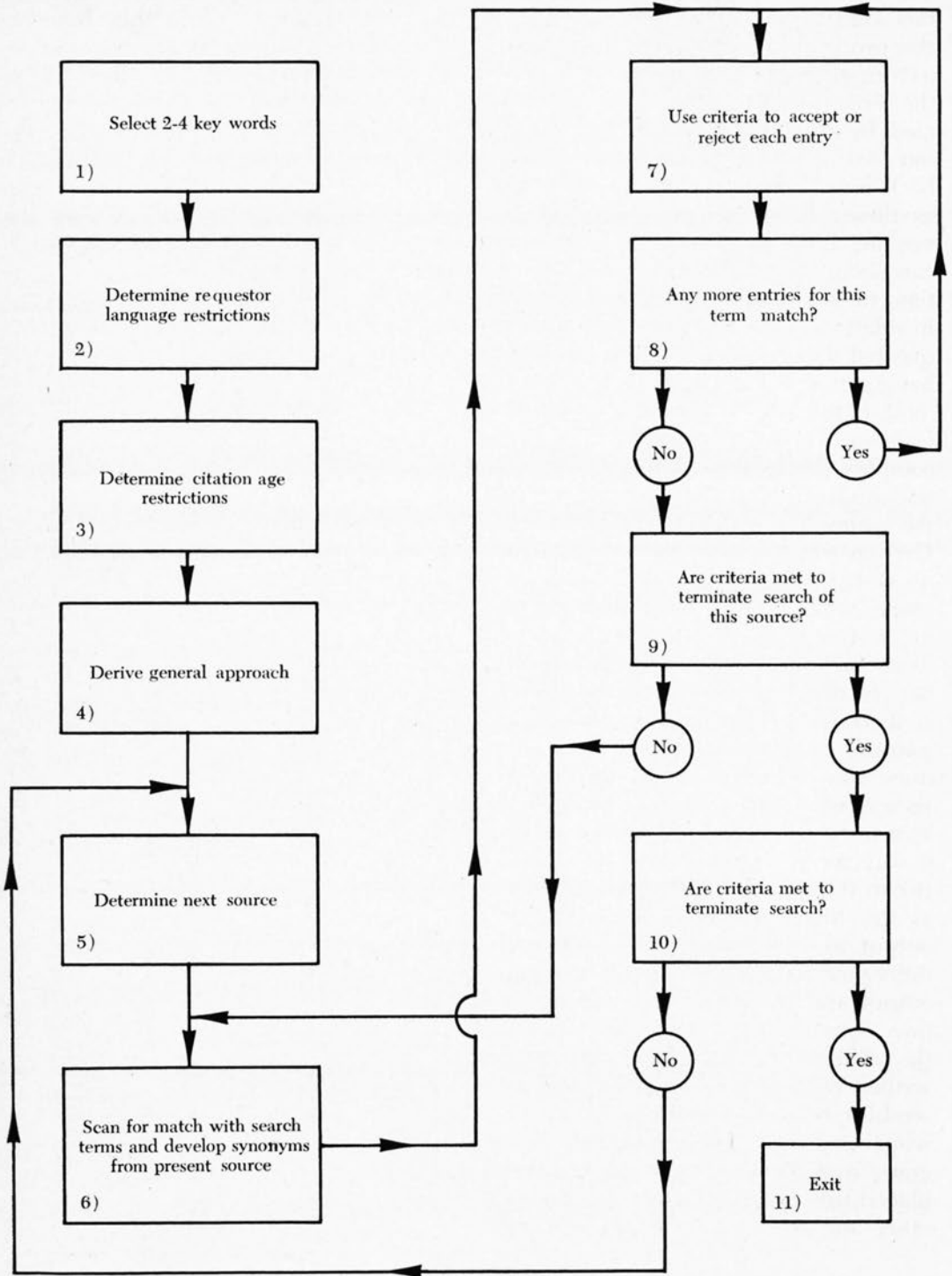


FIG. 1.—Generalized Flow Chart of Reference Librarians Search Behavior. G. Carlson, *Search Strategy by Reference Librarians*, Part 3 of the Final Report on the Organization of Large Files (Sherman Oaks, Calif.: Advance Information Systems, Hughes Dynamics, 1964), p. 7.

stated, "I've never had a reader along on this kind of search."⁵¹ On the basis of this protocol, it seems reasonable to conjecture that on other kinds of searches the librarian was sometimes accompanied by the reader. According to Carlson's construal of feedback, then, feedback must have occurred in such cases as these. It seems more reasonable to assume, however, that patron feedback can occur only after the patron has had time to evaluate the work done for him in relation to the work for which he requested the librarian's help. In that case, having the reader along would for the most part be irrelevant. But what does Carlson intend by feedback? Surely in a complex question-answer system as the one under discussion, feedback is not the simple or obvious activity which Carlson describes as "He receives a request from a person; makes the search, often without the requestor present; gives the requestor the results; and hears no more."⁵² In the simplest manner feedback should have to be described as the librarian's response to the patron's output in such a way as to reduce the difference between it and a prescribed input and the patron's response to the librarian's output in such a way as to reduce the difference between it and a prescribed input; as well as the librarian's response to his own output in such a way as to reduce the difference between it and a prescribed output. "The picture roughly resembles two rings (but rings each of which is composed of a series of ongoing and intervening events) that are placed in planes at an angle to each other and touching each other at com-

mon points in their circumferences."⁵³ To describe feedback in this, the simplest of ways, makes it absurd to assert the absence of feedback on the basis of data derived from protocol analysis of the librarian alone. Indeed, to refuse the reader the desired information until he feeds back to the librarian in an obvious manner surely makes this the most absurd of Carlson's findings, especially since Carlson's construal of inconsistent behavior implies that it, like regular behavior, results in a successful search.

SECTION III

This section of the essay discusses the conventional concept of reference and its relationship to the empirical study of reference.

Derived from the preparation for and the practice of reference, the conventional concept is made explicit in numerous definitions in the literature of librarianship. In these definitions it is clearly asserted that only reference librarians perform reference. As focal points for our discussion, let us consider several definitions of reference; as our point of departure, Pierce Butler's counter to the conventional concept.

To Pierce Butler such definitions of reference as "the assistance given by the librarian to readers in acquainting them with the intricacies of the catalogue, in answering questions, and, in short of doing anything and everything . . . to facilitate access to the resources of the library,"⁵⁴ were invalid. To Butler such definitions were invalid in that they tended to define reference in the trivial circle as "what goes on in the reference room" and the reference room as "the

⁵¹ *Ibid.*, p. 35.

⁵² *Ibid.*

⁵³ Floyd H. Allport, *Theories of Perception and the Concept of Structure, A Review and Critical Analysis with an Introduction to a Dynamic-Structural Theory of Behavior* (New York: John Wiley & Sons, 1955), p. 526-27.

⁵⁴ William B. Child, "Reference Work at the Columbia College Library," *Library Journal*, XVI (October 1891), 298.

place where reference work is carried on."⁵⁵ Pierce Butler proposed that reference is "that process by which civilized man is able to obtain specific information at will by use of books which have been organized into a library."⁵⁶ Unlike other definitions of reference, Butler suggested that reference is not performed solely by librarians and is not performed solely in reference rooms. Butler's counter to the traditional concept of reference, however, has made little impression.

In her empirical study of reference, Dorothy E. Cole observed no difference between Butler's concept of reference and her own, as we have noted, the "queries which people bring to the desk of the reference librarian" and the reference librarian's response to the queries in terms of "materials best suited to the needs of the reader."⁵⁷ Indeed, Cole felt Butler's to be supportive of her concept of reference. In another empirical study of reference a difference was observed. With unshaken confidence in the traditional concept, however, Lois Fern⁵⁸ rejected Butler's counter. Fern, that is, first observed that Butler's concept was inconvenient, *i.e.*, she would have had to consider "those hundreds of instances in which the patron engaged in 'reference work' without the librarian's assistance"; she then observed that Butler's concept was absurd, *i.e.*, she would have had to entitle her study "Use of Library Catalogs by Reference Librarians Engaging in Reference Work."⁵⁹

As students of reference, librarians either have not perceived Fern's "hun-

dreds of instances" or perceived them, like Fern, as something different from reference, perhaps something less than reference. They have perceived reference through the conventional concept. They have perceived only what experience in the preparation for and practice of reference has led them to expect.

There is evidence that "when such expectations are violated by the environment, the perceiver's behavior can be described as resistance to the recognition of the unexpected or incongruous."⁶⁰ In the familiar Bruner and Postman experiment "On the Perception of Incongruity," subjects were shown by tachistoscope a series of five playing cards at controlled and gradually increasing levels of exposure. Some of the playing cards were normal, *i.e.*, expected, as the five of hearts; others were incongruous, *i.e.*, unexpected, as the red six of clubs. "At each exposure the subject was asked to report everything he saw or thought he saw."⁶¹ A stimulus series was completed by three successive correct responses to each of the five cards at various exposure levels from ten up to one thousand milliseconds.

A few of the subjects perceived the normal cards erroneously. One subject, for example, mistook the black five of spades for the red five of hearts. The vast majority of subjects, however, perceived the incongruous cards erroneously. Subjects identified the red six of spades, in a number of cases, as the six of spades or as the six of hearts. The subjects reported such perceptions with considerable assurance. Without uncertainty, that is, they fitted what they presumably saw into a normal form or color pattern that they expected to see. Often subjects compromised on what

⁵⁵ Pierce Butler, "Survey of the Reference Field," Pierce Butler, ed., *The Reference Function of the Library, Papers Presented before the Library Institute at the University of Chicago, June 29 to July 10, 1942* (Chicago: University of Chicago Press, 1943), p. 4.

⁵⁶ *Ibid.*, p. 11.

⁵⁷ Cole, *op. cit.*, p. 1.

⁵⁸ Lois Fern, "Library Catalogs as Reference Sources" (unpublished master's thesis, Graduate Library School, University of Chicago, 1962).

⁵⁹ *Ibid.*, p. 8.

⁶⁰ Jerome S. Bruner and Leo Postman, "On the Perception of Incongruity: A Paradigm," *Journal of Personality*, XVIII (December 1949), 222.

⁶¹ *Ibid.*, 210.

they saw. With the red spade and club cards, for example, subjects reported seeing such colors as black with red lights, lighter than black but blacker than red, olive drab, and so on. A few of the subjects came suddenly to recognize the incongruity of the trick cards. Other subjects found something wrong with them and came to recognize the incongruity gradually. "A subject viewing a red spade may start by reporting a red tint which gradually becomes redder on succeeding trials until he finally asserts that the card is a red spade."⁶² Indeed, there were several subjects who never managed to unmask the incongruity of certain of the playing cards.

Analogically, the Bruner-Postman experiment provides a suggestive schema for viewing the librarian's perception of reference; fitting, that is, the incongruity of reality to one of the neat conceptual categories prepared by prior training and experience, training and experience which do not and never need to indicate that anyone other than librarians perform reference. Let us consider now some of the implications of the librarian's perception of reference.

If the student of reference perceives reference as performed only by reference librarians, then in his research he is constrained into conceptually setting reference librarians apart from other human beings. In setting reference librarians apart from other persons, it has seemed manifest to the student of reference that what differentiates the reference librarian from other persons is his preparation for and practice of reference librarianship. Under the conventional concept of reference, then, the empirical study of reference poses two pertinent questions to the student of reference: How efficient is education for reference librarianship? How successful is the practice of reference librarian-

ship? In the former case the empirical study of reference has become an instrument for investigating some aspect of that preparation and, in the process, for suggesting some modification of it; in the latter case, an instrument for investigating and modifying some aspect of that practice. Indeed, the results of the reference studies reconstructed above were interpreted in just that manner. Cole, for example, observed that classifying reference questions should aid the library administrator in evaluating "the amount and kind of work done by the reference department" and aid the library training agencies in determining the nature and scope of the reference courses."⁶³ Breed wrote, "The findings concerning the reference process related to difficult questions have implications for library schools and the material presented in courses on reference work. More attention might well be paid to the analysis of the reference process itself—and problems on the advanced level might include more examples of the kind of incomplete and inaccurate data with which the reference librarian must so frequently work."⁶⁴ The HERNERS said, "It is evident that useful data on the information requirements of a body of users can be obtained from collecting and analyzing statistical quantities of their reference questions."⁶⁵ By *useful* the HERNERS presumably meant helpful in specifying "those information requirements which librarians . . . are likely to be called upon to meet."⁶⁶ Carlson declared that "In many cases, human behavior is inconsistent to the detriment of successful search. . . . These [inconsistencies] could be used as guidelines by librarians to improve their search procedures. Perhaps some of these inconsistencies or omissions could

⁶² *Ibid.*, 222.

⁶³ Cole, *op. cit.*, p. 58.

⁶⁴ Breed, *op. cit.*, p. 65.

⁶⁵ HERNER and HERNER, *op. cit.*, p. 176.

⁶⁶ *Ibid.*, p. 172.

be generalized and provide guidelines for the training program of reference librarians."⁶⁷

This viewpoint, of course, does not at all invalidate the results of the empirical studies of reference. It does, however, make the process by which they are derived circular. It does, therefore, profoundly affect the manner in which they are perceived. Given that the preparation or practice of reference is both point of departure and point of conclusion of the empirical studies of reference, the results of these studies must often seem, then, to be little more than reflective formulations of the very observations which reference librarians encounter in their education or the very observations which reference librarians make in their work. Consider now several of the results derived from these empirical studies of reference. Cole, again as an initial example: Reference patrons most frequently ask questions in the social sciences, useful arts, and history. Breed: Difficult questions are characteristically specific rather than general and involve bibliographical references rather than facts. Bibliographical questions are difficult to answer because of omissions or inaccuracies in the patron's citation of author, titles, and dates of publication. The HERNERS: Patrons generally ask non-technical questions about descriptions of processes or methods of procedure. Carlson: "Human searchers should be very careful to write down their search terms and check the accuracy of their spellings, pronunciation, and meaning before the search starts, and whenever they adopt a new search term as the search progresses. . . . A human searcher would be more consistent if he would follow more rigid rules in checking every document or source. . . . The reference librarian has no consistent way of improving the quality of his own search. Any searches

completed for a requester should only be given out on the strict condition that the librarian be told about the usefulness or inadequacies of the results."⁶⁸

If the results of the empirical studies of reference seem to be little more than reflective formulations of everyday observations, observations for which there are hundreds of counter-examples ready at hand, then it is hardly surprising that Ennis, for example, has characterized the endeavor which has produced such results as "among the weakest in all library research."⁶⁹

To the non-librarian, the librarian's commitment to the conventional concept of reference in his empirical research perhaps seems strange. To the librarian, for the non-librarian to condemn and then use that concept to guide his empirical study of reference must likewise seem strange. Consider, then, two cases: the one illustrated by what psychologists have called functional fixedness, the other illustrated by what they have called distortion produced by popular opinion.

First consider the case for librarians, as Cole and Breed. "The student attends an educational institution on the presumption that the training he receives there will enable him to solve problems in later life more adequately and efficiently than would be the case had he not received this training. . . . It is also possible, however, that the training which the student receives may often, alas perhaps more often than we are aware of, make the student less able to solve certain new problems."⁷⁰ The classical description of functional fixedness is Maier's pendulum problem.⁷¹

⁶⁸ *Ibid.*, p. 29-36.

⁶⁹ Philip H. Ennis, "The Study of the Use and Users of Recorded Knowledge," *Library Quarterly*, XXXIV (October 1964), 309.

⁷⁰ Rudolph W. Schulz, "Problem Solving Behavior and Transfer," *Harvard Educational Review*, XXX (Winter 1961), 61.

⁷¹ Norman R. F. Maier, "Reasoning in Humans: II, the Solution of a Problem and Its Appearance in Consciousness," *Journal of Comparative Psychology*, XI (August 1931), 181-94.

⁶⁷ Carlson, *op. cit.*, p. 28.

Subjects were introduced into a room where two pieces of cord were suspended from the ceiling. The room contained, besides the suspended cords, "many objects such as poles, ringstands, clamps, pliers, extension cords, tables and chairs."⁷² The subjects were told to tie the two pieces of cord together. The cords, however, were too far apart for a single individual to hold both ends simultaneously. Though several solutions to the problem were possible, one seemed especially difficult for the subjects to grasp. It was the solution achieved by weighting one of the cords with pliers, setting the cord in motion and catching it after the other cord had been seized. Most of the subjects did not perceive the cord as a potential pendulum and the pliers as a potential weight. In general terms the subjects of Maier's study demonstrated an inability to displace old, learned concepts with new ones required by a novel task. Similarly, the subjects who are our concern in this essay have demonstrated an inability to displace that concept of reference learned in the preparation for its practice and in the practice itself by concepts required for the task, generally novel to librarians, of the study of reference.

Consider now the case for non-librarians as the HERNERS and CARLSON. Recall that the HERNERS condemned and that CARLSON called the conventional concept of reference into question. Recall that unlike COLE and BREED, who were committed to that concept by prior training and experience, the HERNERS and CARLSON were not. Recall that though their concepts of reference differed from that of COLE and BREED, the concept of reference which guided their studies did not. The evidence here suggests that "the interfering influence of the familiar extends even to situations where the subject is presumably confined to rigid rules

of thought that should exclude it."⁷³

In a series of tests on "The Distortion of Syllogistic Reasoning Produced by Personal Convictions," MORGAN and MORTON showed that subjects tend to reach conclusions which, though logically invalid, are consistent with popular opinion.⁷⁴ The subjects of these tests were required to draw conclusions from two premises by the process of immediate inference, as

If all men are mortal (major premise),
and

If John is a man (minor premise); then
John is mortal (conclusion).⁷⁵

The subjects of this study were given two sets of fifteen syllogisms in parallel form. In the first set of syllogisms letter symbols as X, Y, and Z were used as terms of the premises; in the second set, "vital issues which were being currently discussed in the papers and over the air were"⁷⁶ incorporated as the terms. Given the invalid syllogism, for example,

Some ruthless men deserve a violent death; since one of the most ruthless of men was Heydrich, the Nazi hangman:

1. Heydrich, the Nazi hangman, deserved a violent death.
2. Heydrich, the Nazi hangman, may have deserved a violent death.
3. Heydrich, the Nazi hangman, did not deserve a violent death.
4. Heydrich, the Nazi hangman, may not have deserved a violent death.
5. None of the given conclusions seems to follow logically.⁷⁷

in letter symbols, subjects tended to choose the second conclusion presumably under the influence of the particu-

⁷³ Bernard Berelson and Gary A. Steiner, *Human Behavior, An Inventory of Findings* (New York: Harcourt, Brace & World, 1964), p. 205.

⁷⁴ John J. B. Morgan and James T. Morton, "The Distortion of Syllogistic Reasoning Produced by Personal Convictions," *Journal of Social Psychology*, XX (August 1944), 39-59.

⁷⁵ *Ibid.*, 40.

⁷⁶ *Ibid.*, 45.

⁷⁷ *Ibid.*, 48.

lar-affirmative atmosphere of this syllogism. Given the same syllogism as a popular issue, however, the same subjects tended to shift from the second to the first conclusion, an opinion which was "quite strong at the time that this test was given."⁷⁸ In general terms the subjects of the Morgan-Morton study believed that they were guided in their choice of conclusions by logical thought, though, in fact, when current issues were embodied in the syllogisms, they were being guided by popular opinion. Similarly, it is not implausible that the other subjects who are our concern in this essay were influenced by a point of view traditional among librarians even though they faced a situation requiring rigid rules of thought which should have eliminated that point of view from consideration.

Whether it has been librarians or non-librarians who have used the conventional concept of reference to guide their empirical studies is largely unimportant. What is important is this. Under the conventional concept it has been assumed in the empirical study of reference that only reference librarians perform reference, that the librarian's proficiency in the performance of reference is the result of his professional training and experience, that the goal of the empirical study of reference is the modification of the librarian's education or of his practice, that observations of the librarian's performance are a sufficient basis on which to propose those modifications which will improve the librarian's proficiency. Since the empirical study of reference under these assumptions is capable only of augmenting the rules of thumb which underlie the librarian's professional training and practice, it is clear under these assumptions

that no amount of study will permit an explanation of how the performance of reference takes place except in the trivial sense that a manual of reliable reference guidelines, e.g., Downs's *How to Do Library Research*,⁷⁸ explains the performance of reference. Though the conventional concept of reference serves a useful purpose, the basis for an *esprit de corps*, in the training and molding of reference librarians, it can hardly be surprising that this section of the essay concludes that empirical interpretation of the conventional concept is fruitless as a framework for the development of the empirical study of reference.

SECTION IV

In a previous section of the essay we observed that the notions basic to four empirical studies of reference were deficient. They did not permit results which are reliable. In the last section, we observed that the empirical interpretation of the conventional concept of reference is fruitless as a framework for the development of the empirical study of reference. For these reasons, this section of the essay considers reference outside of librarianship. It suggests, intuitively, that reference is a special case of problem solving.

First, several definitions of problem-solving behavior are discussed, then several approaches to the study of problem solving are described, first on the theoretical, then on the empirical level. The studies of problem solving which are described below were selected from those known to the author as the most suggestive for reference. In the next section, these studies are briefly compared with the studies of reference previously reconstructed.

To Duncker "a problem arises when a living creature has a goal but does not know how this goal is to be reached. Whenever one cannot go from the given situation to the desired situation simply

⁷⁸ Robert B. Downs, assisted by Elizabeth C. Downs, *How to Do Library Research* (Urbana: The University of Illinois Press, 1966).

by action, then there has to be recourse to thinking."⁷⁹ Duncker, as have many psychologists, differentiated between routine and non-routine problem solving. In the terms of Dollard and Miller, that is between problems solved by automatic habits and trains of thought.⁸⁰ Surely, however, this distinction, used to differentiate illegitimate from legitimate problems, is a fruitless one. What is reflexive to one may well be cognitive to another. To Reitman "a system has a problem when it has or has been given a description of something but does not yet have anything that satisfies the description."⁸¹ But in emphasizing what he calls "degrees of problematality,"⁸² Reitman must tend to ignore a question implicit in Duncker's definition, why may a given task be reflexive to one but cognitive to another person. Indeed, Gagné suggests that "one of the fundamental criteria of problem solving is that a kind of performance which could not be exhibited before the 'problem' was solved *can* be exhibited after the 'problem' is solved."⁸³ To Gagné problem solving is "an inferred change in human capability that results in the acquisition of a generalizable rule which is novel to the individual, which cannot have been established by direct recall, and which can manifest itself in applicability to the solution of a class of problems."⁸⁴

The simulation of cognition is one theoretical approach towards an understanding of problem-solving behavior. A computer is programmed to respond as a human would respond within a speci-

fied task domain. The computer program, then, is a model or set of hypotheses about the internal processes underlying human performance of a given task. "No physiological or neurological assumptions are made . . . these models conceive of the brain as an information processor."⁸⁵ A program is run. The run generates predictions of the model. "In order to provide an adequate test, the behavior with which the computer output is compared must, of course, be different from the behavior which served as a basis for the model's construction."⁸⁶ The results of the run are compared with relevant human results. That is the way predictions of the model are tested and, potentially, improved.

Presumably a model of artificial intelligence may also model human behavior. The difference between models of cognitive simulation and artificial intelligence is this: "An artificial intelligence researcher interested in programming a computer to play chess would be happy only if his program played good chess, preferably better chess than the best human player. However, the researcher interested in simulating the chess-playing behavior of a given individual . . . wants his program to make the same moves as the human player, regardless of whether these moves are good, bad, or indifferent."⁸⁷ Several of the following paragraphs describe Paige and Simon's test of Bobrow's artificial "Question-Answering System for High School Algebra Word Problems"⁸⁸ as a first approximation towards a model of "Cognitive Proc-

⁷⁹ Karl Duncker "On Problem Solving," *Psychological Monographs*, LVIII (1945), 1.

⁸⁰ John Dollard and Neal E. Miller, *Personality and Psychotherapy, An Analysis in Terms of Learning, Thinking, and Culture* (New York: McGraw-Hill Book Co., 1950), p. 97-105.

⁸¹ Reitman, *op. cit.*, p. 126.

⁸² *Ibid.*, p. 130.

⁸³ Gagné, *op. cit.*, p. 130.

⁸⁴ *Ibid.*, p. 132.

⁸⁵ Edward A. Feigenbaum, "The Simulation of Verbal Learning Behavior," Edward A. Feigenbaum and Julian Feldman (eds.), *Computers and Thought* (New York: McGraw-Hill Book Co., 1963), p. 297.

⁸⁶ K. R. Laughery and L. W. Gregg, "Simulation of Human Problem-Solving Behavior," *Psychometrika*, XXVII (September 1962), 279-80.

⁸⁷ Feigenbaum and Feldman, *op. cit.*, p. 269.

⁸⁸ Daniel B. Bobrow, "A Question-Answering System for High School Algebra Word Problems," *AFIPS Conference Proceedings*, XXVI, 1964 Fall Joint Computer Conference (Baltimore: Spartan Books, 1964).

esses in Solving Algebra Word Problems."⁸⁹ The question posed was what cognitive processes are necessary in solving algebra word problems; what processes are sufficient?

The model is STUDENT. STUDENT communicates with human beings over a limited range of natural language. Its task environment is algebra word problems as

(A NUMBER IS MULTIPLIED BY 6. THIS PRODUCT IS INCREASED BY 44. THIS RESULT IS 68. FIND THE NUMBER.)⁹⁰

STUDENT reads a problem statement, rewrites it as a set of simpler sentences, transforms these sentences into equations, and attempts to solve the resultant set of equations. The student model understands a problem statement by "transforming it into an equivalent (in meaning) sequence of simpler kernel sentences. A kernel sentence is one which the listener can understand directly; that is, one for which he knows a transformation into his information store."⁹¹

Paige and Simon compared STUDENT's methods of transforming with human methods of translating algebra word problems into algebraic equations. The humans in this comparison were largely students. The students' processes were established on the basis of protocols; STUDENT's processes, on the basis of Bobrow's descriptions of his computer program.

STUDENT's transform of the problem statement given above reads:

(THE EQUATIONS TO BE SOLVED ARE)
(EQUAL $X00001$ (NUMBER))
(EQUAL (PLUS (TIMES (NUMBER) 6) 44)
68)⁹²

⁸⁹ Jeffrey M. Paige and Herbert A. Simon, "Cognitive Processes in Solving Algebra Word Problems," Benjamin Kleinmuntz (ed.), *Problem Solving: Research, Method, and Theory* (New York: John Wiley & Sons, 1966).

⁹⁰ Bobrow, *op. cit.*, p. 604.

⁹¹ *Ibid.*, p. 593.

⁹² *Ibid.*, p. 604.

Some of the students translated the same problem statement in a similar manner, as

The unknown is a 'certain number,' which would be x . Multiply x by 6. Write down '6x' please. 'Increased by' means add, so you put a plus 44. 'The result is'—indicate equals—write please—'68.'⁹³

Paige and Simon labeled this method direct translation, "a step-by-step substitution of algebraic symbols and expressions for the English words and phrases of the original problem statements."⁹⁴

One of the sub-processes of translating English prose into algebraic equations, observed in both students and STUDENT, is the labeling of variables. The labels are of two sorts, conventional, *i.e.*, x may be used as the abbreviatory name of some unknown, and relational, *i.e.*, $6x$ or $6x + 44 = 68$ may be used as the relational names between the same unknown and some known quantity. Unlike the students, however, STUDENT "introduces relational names only in the special case . . . where it is cued by the occurrence of the word 'this.'"⁹⁵

A sub-problem of naming, again observed in both STUDENT and students, is ambiguity of names. Two similar but distinct common language terms, for example, may refer to the same object, though two distinct but similar common language terms may refer to different objects. To the problem

(THE NUMBER OF SOLDIERS THE RUSSIANS HAVE IS ONE HALF OF THE NUMBER OF GUNS THEY HAVE. THE NUMBER OF GUNS THEY HAVE IS 7000. WHAT IS THE NUMBER OF SOLDIERS THEY HAVE Q.)⁹⁶

a problem in which two different terms, "Russian soldiers" and "they," refer to the same object, Russian soldiers, STUDENT responds:

⁹³ Paige and Simon, *op. cit.*, p. 70.

⁹⁴ *Ibid.*, p. 82-83.

⁹⁵ *Ibid.*, p. 78.

⁹⁶ Bobrow, *op. cit.*, p. 605.

THE EQUATIONS WERE INSUFFICIENT TO
FIND A SOLUTION
(ASSUMING THAT)
((NUMBER OF SOLDIERS (THEY/PRO)
(HAVE/VERB)) IS EQUAL TO
(NUMBER OF SOLDIERS RUSSIANS (HAVE/
VERB))
(THE NUMBER OF SOLDIERS THEY HAVE
IS 3500)⁹⁷

To the problem

The number of quarters a man has is seven times the number of dimes he has. The value of the dimes exceeds the value of the quarters by two dollars and fifty cents. How many has he of each coin?⁹⁸

a problem in which the expressions number and value are synonymous in certain contexts, a student responded in this fashion:

Let's represent--the fir-- --x as the dimes. Write down x please. The number of quarters is--is seven times the dimes. Therefore, the number of quarters is represented by 7x. Put '7x.' 'The value of the dimes exceeds the value of the quarters by two-fifty'; therefore, x . . . Since it really is a money problem we will check this and discover that we really should multiply to make this--ah--to make it equal in value. Since we have 250 pennies we have to multiply the dimes by 10. So that should be 10x and a plus. Since it's quarters it would be 7 times 25 times x, on the other side. That is the correct solution.⁹⁹

In the former, the soldier problem, the cue through which STUDENT achieved a sufficient equation was a grammatical cue. The relationship, that is, between "Russian soldiers" and "they" is a grammatical one. In the latter, the coin problem, the cue was a substantive one. The student, that is, had to "know that the value of a quantity of coins equals the number of coins times the value per coin."¹⁰⁰ Given the problem

(IF 1 SPAN EQUALS 9 INCHES, AND 1 FATHOM EQUALS 6 FEET, HOW MANY SPANS EQUAL 1 FATHOM Q.)¹⁰⁰

like a student, STUDENT "is able to draw on a body of facts, picking out relevant ones, and use them to obtain a solution"¹⁰¹ from its store of global information. To that problem STUDENT responds:

THE EQUATIONS WERE INSUFFICIENT TO
FIND A SOLUTION
(USING THE FOLLOWING KNOWN RELATIONSHIPS)
(EQUAL (TIMES 1 (FEET)) TIMES 12
(INCHES))
(1 FATHOM IS 8 SPANS)¹⁰²

It became evident, however, that the students were using more powerful processes than those exhibited by STUDENT's use of definitional relationships. In translating the coin problem, one of the students

. . . found incongruity . . . if a man has seven times as many quarters as he has dimes and the value of the dimes exceeds the value of the quarters by two-fifty, the quarters must really not be worth too much, because if he has 7 times as many quarters as he has dimes, the number of--the value of the quarters must exceed the dimes by $7 * 2.5 - x$, or what not. . . negative quarters--no such thing exists.¹⁰³

Presumably, STUDENT would not have perceived this incongruity. Indeed, given the problem

A car radiator contains exactly one liter of a 90 per cent alcohol-water mixture. What quantity of water will change the liter to an 80 per cent alcohol mixture?¹⁰⁴

which is anomalous in the sense that it cannot be translated solely on the basis of the information given, it became obvious that those students who achieved a successful translation of the problem statement did so by creating a physical representation of the problem which en-

⁹⁷ *Ibid.*

⁹⁸ Paige and Simon, *op. cit.*, p. 79.

⁹⁹ *Ibid.*, p. 80-81.

¹⁰⁰ Bobrow, *op. cit.*, p. 604.

¹⁰¹ *Ibid.*, p. 605.

¹⁰² *Ibid.*, p. 604.

¹⁰³ Paige and Simon, *op. cit.*, p. 86.

¹⁰⁴ *Ibid.*, p. 90.

abled them to read off certain conservation assumptions as though they were stated as part of the problem. As illustrated by De Soto, humans translate such problems as "Alice is taller than Mary; Elsie is shorter than Mary: Is Elsie taller than Alice? . . . into an 'up-down' image and the answer is 'read off' the image."¹⁰⁵ On this level, the iconic, STUDENT exhibits no processes similar to those of humans. Presumably, such processes "will have to be added to that program if it is to provide a satisfactory theory of the human behavior observed in handling these problems."¹⁰⁶

The theoretical approach to problem solving behavior has experimental analogs. In one a human is programmed to respond like a computer. "A list containing objects and relations between objects is stored in [human] memory, problems are presented to the human which require discovery of implications of the relations by successively operating on objects, the operations defined by the stored relations."¹⁰⁷

Hayes, for example, has been concerned with problems solved by "a sequence of linked phases or steps which form a chain or path connecting the initial conditions of the problem with its goal."¹⁰⁸ These problems are characterized by homogeneity, information control, and modularity. That is, each step is of the same kind and of the same difficulty, the information necessary for solution is under the control of the prob-

lem solver, and each problem can be constructed to any length with any number of blind alleys. Hayes has referred to these as spy problems.

In a spy problem a subject was to imagine that he was running a spy ring. For security reasons some spies could communicate with each other, other spies could not. Prior to a run, subjects were required to commit to memory connection lists as

SHOWER	—————→	CLERK
DROUGHT	—————→	HILL
LARYNX	—————→	BETH
ADJECTIVE	—————→	SHOWER
HILL	—————→	HORSE
BEEF	—————→	LARYNX
ADJECTIVE	—————→	PARCHESI
DROUGHT	—————→	KEVIN
SHOWER	—————→	BEEF
LARYNX	—————→	DROUGHT
BEEF	—————→	TAFT ¹⁰⁹

"to the criterion of three successive error-free trials."¹¹⁰ The subjects were then set problems to solve as "Get a message from HILL to ADJECTIVE." During the problem solutions subjects were requested to think aloud.

Almost all of the subjects attempted to solve the problems by starting either at the initial state or the terminal state. Given the problem noted above, that would be either at HILL or at ADJECTIVE. Twenty-five per cent of all steps were executed in a backward direction. In general, however, subjects attempted backward solutions only when forward solutions failed. In some cases subjects employed list searching, a link-by-link search from the beginning to a desired point in the list under consideration. Like backward searching, list searching was also infrequent. It was frequently

¹⁰⁵ C. De Soto, "Reasoning and Spatial Representations," Paper read at the Center for Cognitive Studies Colloquium, Harvard University, 1965, quoted in Jerome S. Brunner and Others, *Studies in Cognitive Growth, A Collaboration at the Center for Cognitive Studies* (New York: John Wiley & Sons, 1966), p. 9.

¹⁰⁶ Paige and Simon, *op. cit.*, p. 118.

¹⁰⁷ Garlie A. Forehand, "Epilogue: Constructs and Strategies for Problem-Solving Research," Benjamin Kleinmuntz (ed.), *Problem Solving: Research, Method, and Theory* (New York: John Wiley & Sons, 1966), p. 369.

¹⁰⁸ John R. Hayes, "Problem Topology and the Solution Process," *Journal of Verbal Learning and Verbal Behavior*, IV (October 1965), 371.

¹⁰⁹ John R. Hayes, "Memory, Goals, and Problem Solving," Benjamin Kleinmuntz, ed., *Problem Solving: Research, Method, and Theory* (New York: John Wiley & Sons, 1966), p. 150.

¹¹⁰ Hayes, "Problem Topology and the Solution Process," *loc. cit.*, 372.

observed, however, that as subjects approached the solution to a problem, their rate of progress increased. A second experiment, with a second set of subjects, was designed to probe this observation.

The second experiment bore out the observation of the first, that as subjects approached solution their rate of progress accelerated. A second observation was made. When steps to go were taken into account, progress time did not vary with problem length. The rate of progress, that is, was being determined, not by the number of steps completed in a solution, but by the number of steps to be completed. Presumably subjects were covertly planning steps to be taken.

Hayes inferred two sorts of strategies to account for these observations, local and remote planning. If, that is in the latter case, a subject was working through the ordered sequence A to F, for example, and noted "connections in the neighborhood of F, specifically 'E-F,'" then, "when S gets from A to E, the step 'E to F' occurs rapidly because it has been planned."¹¹¹ Indeed, it seems reasonable that a subject would not enter a blind alley at E, if he knew that the link E to F leads to the desired solution. If, in the former case, a subject was exploring a few steps ahead "of his present position to a depth of one or perhaps more steps," then he is unlikely to fall into blind alleys especially when "with fewer steps to go, S is more likely to discover a solution path in his local covert exploration."¹¹²

In another experimental analog to the theoretical approach to problem solving, human responses are compared with postulated computer responses. That is, given a problem to solve a human, in his response, would tend towards an ideal strategy of solution; a computer, on the other hand, presumably would be programmed to employ an ideal strate-

gy. Thus, according to Bruner, Goodnow, and Austin, an ideal strategy is "an analytic device used as a yardstick against which to compare the performance of human operators in the situations we set them."¹¹³

A Study of Thinking, by Bruner and his associates, concerned itself with problems of concept attainment, the task in which a "human being seeks to sort the environment into significant classes of events so that he may end by treating discriminably different things as equivalents."¹¹⁴ Presumably, when a human organism has attained a concept he has done so by discovering cues by which he can infer the category membership of the objects and events he encounters. Presumably, that is, the organism has learned a set of relevant attributes and the conceptual rules by which the appropriate values are combined to form the concept. For example, the concept spades, associated with playing cards, is cued by the relevant attributes, color and form, and is satisfied just by the joint presence of the values, black and spade-shaped. Presumably an organism attains a concept through his cognitive commerce with the environment by testing a series of exemplars and non-exemplars of the concept. When the organism is free to choose instances to test, presumably he employs selection strategies. For example, "a neurologist is interested in six cortical areas and their bearing on pattern vision. He knows that with all six areas intact, pattern vision is unimpaired. With all six areas destroyed, pattern vision is absent. In planning his research, how shall he proceed? Destroy one area at a time? All but one at a time? In what order shall he do his successive experiments?"¹¹⁵ One of the numerous empirical questions asked in this study of thinking was

¹¹¹ *Ibid.*, 377.

¹¹² *Ibid.*, 377-78.

¹¹³ Bruner, Goodnow and Austin, *op. cit.*, p. 241.

¹¹⁴ *Ibid.*, p. viii.

¹¹⁵ *Ibid.*, p. 81.

what effect does a disordered environment have upon human selection strategies.

Two groups of fifteen subjects each were set in the following task environments. An array of instances each exhibiting one of two possible values of six attributes was presented each subject. Each subject was shown a card exemplifying a concept to be attained. The subject was to attain the concept by selecting instances from the board for testing. Each time he selected a card the subject was told that the choice either satisfied or did not satisfy the concept. Once per choice, whenever he wished, a subject could venture the identity of the concept. "The only difference between the two groups of fifteen subjects was in the arrangement of the cards on the board in front of them."¹¹⁶ For one group the environment was made orderly; for the other, it was made random or disordered.

Presumably both sets of subjects would attempt to insure each instance selected for testing, informative. The subjects working in the ordered environment, however, clearly had an advantage over their colleagues working in the disordered environment. "The Random Group," that is, "was faced with an array of instances whose arrangement gave little or no visual support to the subjects in the task of sorting out positive from negative instances in order to attain the correct concept."¹¹⁷ Presumably, then, both sets of subjects would attempt to insure each instance selected informative within the constraints imposed by the environment in which they worked.

Under ideal conditions a concept could be attained in six choices. It was predicted that the Ordered Group would tend toward a conservative-focussing strategy, *i.e.*, "finding a positive instance

to use as a focus, then making a sequence of choices each of which alters but one attribute value of the first focus card and testing to see whether the change yields a positive or a negative instance."¹¹⁸ It was predicted that the Random Group would tend toward a successive scanning strategy, *i.e.*, testing a single hypothesis at a time until the correct one is found.

The results of the experiment generally confirmed the predictions. The Ordered Group required an average of 6.1 choices per problem to attain the correct concepts; the Random Group required 10.4 choices. The former group employed 1.1 attribute values which distinguished a first choice and the first positive card given, made an average of 1 redundant choice per problem and offered an average of 0.5 incorrect hypotheses per problem; the latter group, respectively, 1.8 attribute values, 4.1 redundant choices and 1.4 incorrect hypotheses. Presumably the Ordered Group was successful in selecting informative instances, largely focusing on the first card given, whereas the Random Group was unsuccessful, selecting redundant and inconsistent instances.

SECTION V

This, the final section of the essay, compares briefly and generally the studies of reference reconstructed in Section I with the studies of problem solving described in Section IV. It purposes to indicate several ways in which the study of reference may be clarified when interpreted as a special case of the study of problem solving.

The studies of problem solving either inferred human competence, *i.e.*, what can reasonably be hypothesized about the internal processes underlying the performance of a task, from performance, *i.e.*, what a person does in the performance of a task, or inferred per-

¹¹⁶ *Ibid.*, p. 97.

¹¹⁷ *Ibid.*, p. 98.

¹¹⁸ *Ibid.*, p. 87.

formance from competence. In Bruner's words:

To infer a person's representation of the world, if we are uncanny experimenters, we design tasks that permit us to infer how he does these things. We ask him to tell us the fifty states of the Union. If he 'reads out' in this order, 'Maine, New Hampshire, Vermont . . .,' we can guess that the supporting representation for his recital is spatial. If the order is 'Alabama, Arkansas, California . . .,' the support is inferred to be more list-like, ordered by an alphabetic rule.¹¹⁹

The studies of reference, on the other hand, inferred performance from particular performance. Generated by the inference of performance from particular performance, the results of the studies of reference, unlike the results of the studies of problem solving, are devoid of any generality whatever. Devoid of generality, these results cannot reasonably be used to satisfy the objectives of the study of reference, the modification either of the preparation for the practice of reference or of the practice of reference itself.

What the empirical studies of reference have to tell us about reference was told by a direct confrontation with reality, a reality, we observed in Section IV, which has been perceived imperfectly. The studies of problem solving have eschewed a confrontation with reality. They have thereby been freed to invent a modeled reality.

Though other models have been de-

vised,¹²⁰ the basic models used in the studies of problem solving were information-processing models. These models were variously formulated as a computer program, as a human programmed like a computer, or as a human with processes analogous to the routines of a computer program. In all cases, these models were set to operate over models of well-defined task domains. By virtue of the information processing models, it was possible for the studies of problem solving to specify and verify complex human processes as focusing and scanning strategies, local and remote planning, and routines for translating algebra word problems into algebraic equations. Indeed, given the information-processing models of human competence, it becomes clear that what differentiates the reference librarian from other human beings is not simply his preparation for and practice of reference, but the way in which he has learned to encode the task domain over which he works.

The concepts which structured the studies of problem solving differed drastically in precision and predictive content from the deficient notions which structured the studies of reference. This was especially noticeable of the constructs governing the task domains. Paige and Simon, for example, were able to identify nine distinct substantive quantities entailed by the physical circumstances described in the mixture problem:

INITIAL SITUATION	CHANGES	FINAL SITUATION
T ₁ —total quantity of original mixture	ΔT—quantity of liquid added	T ₂ —total quantity of final mixture
A ₁ —quantity of alcohol in original mixture	ΔA—quantity of alcohol added	A ₂ —quantity of alcohol in final mixture
W ₁ —quantity of water in original mixture	ΔW—quantity of water added	W ₂ —quantity of water in final mixture ¹²¹

¹¹⁹ Bruner and Others, *Studies in Cognitive Growth*, p. 7.

¹²⁰ For examples see David Harrah, *Communication: A Logical Model* (Cambridge, Massachusetts: M.I.T. Press, 1963); and Mihajlo D. Mesarović, "Toward a

Formal Theory of Problem Solving," Margo A. Sass and William D. Wilkinson, eds., *Symposium on Computer Augmentation of Human Reasoning* (Washington: Spartan Books, 1965).

¹²¹ Paige and Simon, *op. cit.*, p. 100.

As a result, they were able to infer just those conservation assumptions which must have been used by the subjects who solved this problem.

The ubiquitous construct of the information processing models was the strategy. Though referred to by different terms, simply as routines and sub-routines in the Paige and Simon paper and as planning in Hayes's papers, a strategy is generally a sequence of decisions, some conscious and others not, which governs the order and placement of transformations and storage holds through which a set of problem data is manipulated in an attempted problem solution.

The basic method of probing strategies in these studies was protocol analysis, a method of tracing human data processing by means of a stream of conscious verbal behavior. Though superior to the questionnaire, protocol analysis has several disadvantages which we shall note. Nevertheless, in the hands of Paige and Simon it was cleverly used to its fullest advantage. By setting their subjects incongruous problems to solve, Paige and Simon were able to elicit the employment of conservation assumptions and the visual representations from which they emerged. Presumably these are human processes which are not consciously used by sophisticated subjects in handling routine problems.

Studies of problem solving are not, of course, without internal difficulties. Protocol analysis is often cited as being intersubjectively variable. When sophisticated subjects are used, the extent of previous processing of problem data prior to the emission of protocol statements is indeterminant. As a construct, the strategy has not been made entirely clear. What occurs, for example, when a step in a strategy is executed? What unit of information undergoes transformation in a single step, a word, a phrase, or is it even a linguistic unit? Current studies seem to imply different strategies for different tasks, an inelegant implication

at best. Indeed, the information-processing models have several obvious drawbacks. They have proven useful only over well specified task domains as algebra word problems and stripped down concept attainment tasks. Statements of artificial intelligence and simulation of cognition are limited by the exigencies of computer programming and model, generally, a particular individual, possibly the programmer himself.

It has been observed that to state an empirical problem "which will not be either completely trivial or hopelessly beyond the range of present-day understanding and technique,"¹²² is extremely difficult. The studies of reference are surely a case in point. A ubiquitous concern of these studies, what is a reference question, has clearly preceded the delimitation of a necessarily prior concern, what is a question.

Indeed, what is a question? As Hamblin has put it by these examples,

- (a) Rhetorical question: 'I ask you, gentlemen of the jury, can such a man be innocent?' is really not a question but a statement.
- (b) An indicative sentence with a characteristic tone of voice: 'This is the Canberra train I'm on?'—not really a statement but a question.
- (c) 'Tell me how many fingers I'm holding up.' a command, but at least almost equally a question.¹²³

a question is not a simple grammatical distinction. Sprotsy has identified questions as psychological phenomena, as phases in problem solving:

A question follows the detection of a gap in what can be termed . . . one's cognitive map of an area. But it seems likely that as soon as this gap is identified the thinker bridges it with some concept or idea based on the knowledge he has at hand. Further-

¹²² Noam Chomsky, "Review of Verbal Behavior," by B. F. Skinner, *Language*, XXXV (January-March 1959), 55.

¹²³ C. L. Hamblin, "Questions," *Australasian Journal of Philosophy*, XXXVI (July-September 1957), 388.

more, the extent to which he can communicate the nature of this 'bridge' varies. . . . Questions, then, are not simply requests for information. The information that a subject seemingly requests and receives, actually confirms or refutes . . . the cognitive bridge he has already erected to close a cognitive gap.¹²⁴

In transformational generative grammar, it has been posited that questions are both a linguistic and psychological concern. Transformation grammar has hypothesized three sets of rules, phrase structure, transformation, and morphophonemic, through which sentences of the language are generated. Under this view of grammatical structure, the sentence "*What are you looking for?*" can be described as a what-question transform of *You are looking for it*, and *What are you running for?* as a why-question transform from *You are running.*¹²⁵

There is some evidence for the psychological reality of such transformations. Savin and Perchonock, for example, assuming that "immediate memory has a small fixed capacity and that sentences are encoded in the way that a transformational grammar suggests,"¹²⁶

had human subjects encode such transformations as negatives, passives, questions, into immediate memory. If, Savin and Perchonock hypothesized, transformations required more space in immediate memory than kernels, short active indicative assertions, then spatial differentials were calculable on the basis of the number of words a subject could recall after the encoding of sentence plus word string. The results of this and similar studies have nicely supported the theoretical predictions that transformations, as negatives and questions, are encoded in human memory independent of other characteristics of a sentence. Though the evidence is still questionable, *i.e.*, the results of the Savin-Perchonock study, may, as the authors suggest, be explicable on such grounds as sentence length or retrieval interference, it is clear that the transform hypothesis will have to be taken into account in any serious formulation of our concept of questions.

In part the basic question posed here, what is the disposition of the empirical study of reference, has been answered. But only a very small part of that question was asked in this essay. From many points of view the specific questions asked were probably not pertinent to the basic question. Nevertheless, a posit of criticism is that criticism begets criticism. Hopefully, then, other points of view will be heard from. Hopefully, some will be given expression as empirical studies, others as critical studies. ■■

¹²⁴ Paul Sprotsy, "The Use of Questions in the Problem Solving Process" (unpublished PhD dissertation, Department of Psychology, Western Reserve University, 1962), p. 18-19.

¹²⁵ Robert B. Lees, "Review of *Syntactic Structures*" by Noam Chomsky, *Language*, XXXIII (July-September 1957), 388.

¹²⁶ Harris B. Savin and Ellen Perchonock, "Grammatical Structure and the Immediate Recall of English Sentences," *Journal of Verbal Learning and Verbal Behavior*, IV (October 1965), 349.

