Research Report

RC-5866
(#25388)
2/18/76
Computer Science
Psychology
20 pages

Quantifiers and Question-Asking

John C. Thomas

Behavioral Sciences Group
Computer Science Department
IBM Thomas J. Watson Research Center
Yorktown Heights, New York 10598

18 Feb 76

34p.

This research was supported in part by the Engineering Psychology Programs, Office of Naval Research, Contract Number N00014-72-C-0419, Work Unit Number NR-197-020.

IBM Research Division
San Jose · Yorktown · Zurich

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER RC5866 ✓	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Quantifiers and Question-Asking		5. TYPE OF REPORT & PERIOD COVERED Interim Technical Report
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) John C. Thomas		8. CONTRACT OR GRANT NUMBER(s) N00014-72-C-0419
9. PERFORMING ORGANIZATION NAME AND ADDRESS		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR-197-020
11. CONTROLLING OFFICE NAME AND ADDRESS International Business Machines T. J. Watson Research Center PO Box 218 Yorktown Heights, New York 10598		12. REPORT DATE 2/18/76
		13. NUMBER OF PAGES 29
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) Office of Naval Research Code 455 Arlington, Virginia		15. SECURITY CLASS. (of this report)
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

quantifier query human logic cognition reasoning question-asking man-computer

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

ABSTRACT: Data concerning the use of universal quantifers in question-asking is presented. These data were collected in variously translated English questions into a query language, generated their own English questions, translated Venn diagrams into English or vice versa, gave judgements about the consistency of two English statements, or manually looked up answers to questions. Subjects showed considerable difficulty with the logician's notations of set relations (except disjunction) on all tasks.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

over

The interpretations given quantified sentences varied between subjects on a given task and even within a subject, between tasks. Generally speaking, subjects gave interpretations consistent with quantified natural language questions or Venn diagrams but not equivalent to them. Subjects used explicit set specifications rarely in spontaneous English.

Tentative suggestions are made for the design of formal and natural-language question-answer interfaces.

RC 5866
(#25388)
2/18/76
Computer
Science

Psychology

20 pages

Quantifiers and Question-Asking

John C. Thomas

Behavioral Sciences Group
Computer Science Department
IBM Thomas J. Watson Research Center
Yorktown Heights, New York 10598

Typed by Linda Callahan on CMC

Abstract: Data concerning the use of universal quantifiers in question-asking is presented. These data were collected in a variety of procedures using nonprogrammers. These non-programmers variously translated English questions into a query language, generated their own English questions, translated Venn diagrams into English or vice versa, gave judgements about the consistency of two English statements, or manually looked up answers to questions. Subjects showed considerable difficulty with the logician's notations of set relations (except disjunction) on all tasks. The interpretations given quantified sentences varied between subjects on a given task and even within a subject, between tasks. Generally speaking, subjects gave interpretations consistent with quantified natural language questions or Venn diagrams but not equivalent to them. Subjects used explicit set specifications rarely in spontaneous English.

Tentative suggestions are made for the design of formal and natural-language question-answer interfaces.

LIMITED DISTRIBUTION NOTICE

This report has been submitted for publication elsewhere and has been issued as a Research Report for early dissemination of its contents. As a courtesy to the intended publisher, it should not be widely distributed until after the date of outside publication.

Copies may be requested from:
IBM Thomas J. Watson Research Center
Post Office Box 218
Yorktown Heights, New York 10598

The purpose of this paper is to present some observations on the use of quantifiers (e.g., "all", "some", "no") in question-asking. These observations were not the planned results of laboratory experiments expressly designed to test some particular hypothesis about how people use quantification. Rather these were incidental observations of studies conceived of with other purposes in mind. For this reason, the observations and recommendations in this paper must be considered as preliminary. The results provide suggestive leads for those interested in the scientific investigation of quantification or question-asking behavior. For those interested in designing query systems, this report is not meant as a set of pat answers to complex questions, but a stimulus to the research that should be done for the particular users and particular tasks for which a particular system will be designed.

I. INTRODUCTION

It appears that the use of quantifiers like "all" might be a major stumbling block to the use of computers by non-programmers. On the one hand, there are a number of places which require quantifiers, conditionals (e.g., "If X, then Y"), or other logical expressions (e.g., "not X", "A or B"). For example, the designers of powerful artifical query languages have typically tried to include all logical and set operators in their systems----universal quantification, set union, negation or set complement and so on. Examples of such systems include SEQUEL and SQUARE, (Chamberlin & Boyce, 1974), HQL, (Fehder,1973) and Query By Example (Zloof, 1974, 1975). Investigators attempting to develop "natural language" query systems are also aware of the uses and possible ambiguities that arise from the use of quantification, (e.g., Thompson, Lockemann, Dostert, and Deverill, 1969). In addition, there are a number of other important areas where a better understanding of quantification could be useful. One major problem in the data processing (DP) industry is human-human communication. Programming managers themselves, according to one recent survey (Scott & Simmons, 1974) feel that communications problems are the major difficulty in programming. One cause of communication difficulties may be that quantifiers are used by DP professionals in a manner consistent with mathematical and logical usage. In constrast, quantifiers will probably be interpreted by businessmen according to the norms of conventional English.

Another example of the use of quantifiers is in the Application Customizing Service (ACS) questionnaires used by IBM to allow small businessmen to describe their particular business practices by filling out multiple choice forms. Later, sections of program code are selected for the businessman's software system on the basis of his choices. Many of these questions are of the form "Are all your X's Y's?" Thus, quantifiers are important in several areas of human interaction within the DP industry as well as in computer query systems.

Review of quantification difficulties.

While it seems clear that quantifiers like "all" and "some" appear often in data processing, evidence from psychological studies indicates that most people have a fair amount of difficulty using these quantifiers "correctly", that is, according to their definitions in symbolic logic. (For a review of recent research in this area see Neimark and Santa 1975.) It is well-known that college students often make or accept erroneous inferences in syllogistic reasoning. (Roberge, 1970). For example, from the premises, "All B are C" and "All B are A", many people are willing to conclude that all A are C. In fact, in Roberge's study, over twice as many students chose this incorrect inference as the correct inference that some A are C. (It is generally assumed in this paper that these sets are nonempty.) There are a variety of explanations for this finding, but one apparent difficulty for people is that they misinterepret premises, or, to put it more fairly, they interpret premises in a different manner from logicians and those of us that have been trained in logic. (See, for example Pezzoli, 1970, Ceraso and Provitera, 1971, and Revis, 1975). In contrast to logician's, other people tend to limit their interpretation of a quantified statement to a single specific set relation.

Other psychological studies indicate that much of the difficulty that subjects appear to have in interpreting quantifiers (and other logical operators) largely disappears when they are applied in a meaningful context and in a manner consistent with the person's expectations about the language. See, for example, Staudenmeyer (1975) and Wason & Johnson-Laird, (1972). However, errors of interpretation and inferrence do not seem to totally disappear even with meaningful text (Griggs, 1974).

In addition to what are considered to be actual errors that subjects make, there are also accepted problems of true ambiguity in the English use of quantifiers. (e.g., p.163, Anderson and Bower, 1973.) For example, does the sentence, "All research reports are written by some authors" mean that every research report that is written is written by one or more authors (none written by computers or dogs or ghosts) or does it mean that there is a particular set of authors (perhaps a proper subset, perhaps not) who ghost-write every report? However, logical ambiguity is not the same as behavioral ambiguity. It may be that though native speakers would agree that either interpretation of the above sentence is possible, one of the interpretations is both intended and interpreted in the vast majority of cases. Johnson-Laird (1969) found that subjects tended to be more consistent than chance in their interpretations of sentences like "All research reports about machines are written by some authors." But people were still far from unanimous in their interpretations.

To summarize, it appears on the one hand that there are a number of DP-oriented actitivities that involve the correct (i.e. logical) use of quantifiers and yet, on the other hand, that many people, including college students have considerable difficulty interpreting quantification statements. It would seem to be of some interest then to obtain more data on how great a problem this is likely to be for non-programmers.

II. OBSERVATIONS FROM A STUDY OF A QUERY LANGUAGE.

Preliminary Results

In preparation for an extensive psychological study of a new research query language 17 pilot subjects were taught this query system individually and then asked to write a number of queries employing it. This query language permits expression of queries essentially by means of examples, and an evolved form of this language, known as Query by Example, has been of recent interest (cf. Zloof, 1974, 1975; Thomas and Gould, 1974). The present discussion concerns performance with a previous version of the quantification syntax of this language which is referred to here simply as the query language. The reader may get some feeling for the style of this query language by looking at Table I and reading the explanation below. In the query language, 'p.' (for print) means that specified items from that column are to be printed as output. A non-underlined word in a column means a constant. An underlined word is a variable; i.e., an example. Thus in the first query in Table I, the 'p.' indicates that names and salaries are to be printed out. 'Jones' and '23K' are arbitrary examples. Sports' is a constant and restricts the names and salaries to be printed. In the second query, 'Toys' serves as a linking variable. The query essentially means 'find the department Riley works in and use that department as a constraint on the names to be printed out.' The construction illustrated in queries 4 and 5 by 'all (pen)' means to treat the specified items as a set. A set specified by '≥ all (pen)' is a superset of 'all (pen)'. A set specified by '< all (pen)' is a proper subset of the set specified by 'all (pen)'. Since people seemed to have difficulty during training expressing the notion of universal quantification, a number of different surface level syntaxes were tried. The author's opinion, after attempting to teach a variety of surface structure syntaxes, was that syntax was not the basic source of the difficulty.

Method and results of main QBE experiment

After the pilot data were analyzed a more thorough experiment with the query language was carried out. The method and results are reported in Thomas and Gould, 1974, 1975.

Briefly, 39 high school and college students were trained to use the query language and given several pencil and paper tests of their ability to translate English questions into its syntax. None of the 24 high school students had had any previous experience with computers or programming. Four of the college students had had one programming course. None of these subjects could be considered as programmers. Data concerning the IQ and class rank of the high school students showed them to be somewhat above average on these measures, but not exceptionally so. Nearly half of the 3 hour training time was spent dealing with the use of universal quantification and subsetting since earlier subjects had shown so much difficulty on these points. (In contrast, most other features of the language was learned easily.)

Despite the emphasis in training, those questions that necessitated the use of universal quantification and subsetting resulted in a disproportionate number of errors. The overall accuracy of translated queries was 67 per cent. The queries that involved quantification are shown in Table 2 along with the perentage of correctly written queries. It should be noted that the questions requiring quantification were simpler than average on other dimensions of complexity such as number of variables, number of constants, number of relations.

A further analysis of the types of errors that students made with quantification questions was performed. One observation from this analysis was that four of the students made no errors on the five questions that involved universal quantification. In addition, eight of the 39 students made no errors with the quantification syntax (though they made other errors on these problems). Thus, the concepts and syntax involved were not impossible to learn, about a third of the students having been able to do five problems without any errors on the syntax of quantification. A chi-square test confirms that it was unlikely that the individual differences in the number of errors were due to chance (Chi-square (df=3) = 15.42, $p < .005$).

Two potential sources of students' major errors were 1) in determing when to use the syntax for subsetting or 2) in how to use it. In those problems that should have used the "all (pen) --- ≥ all (pen)" construction, 27 errors were because the subject used the wrong directionality on the '≥' sign. Five of the errors were due to the subject having used the "All the pens" construction. In fourteen cases, the subjects did not use any kind of universal construction at all. In the problem that required the "all the pens" construction, 12 of the subjects used "all (pen) --- > all (pen)" instead. Thus, the errors that subjects made were of several types. One apparent difficulty was that subjects were confused about what was a subset of what. Perhaps a better syntax may help solve this problem. But what about the subjects who simply ignored the required universal quantification? It is doubtful that an improvement in syntax alone would help much in those cases. In summary, subjects learned most of this query language easily but had a fairly difficult time translating questions explicitly involving universal quantification. These difficulties could not be traced to a single source.

Self-generated queries

After students were asked to translate 40 questions from English into the query language, another five "problems" were presented to 35 of the students. The student was presented with a problem situation having to do with a high school, and a set of ten tables that contained attributes pertaining to the teachers, students, and classes at the high school. Each subject was asked to write a question in English that met two criteria 1) the question could be answered by reference to information that would be in the tables and 2) finding the answer to the question would be relevant to solving the problem. Each subject was also asked to translate each of his questions into the query language. The five problems are presented in Table III.

Several of the problems could have reasonably given rise to questions involving universal quantification, e.g., "Are all the student rankings for Jones's physics course below the average of all courses," "Are all the faculty salaries of those under 35 years lower than any of the salaries of those over 50?", "Does Thornstein have all the prerequisites for all his courses?". In fact, not a single one of the 160 English questions used explicit universal quantification. In addition, none of the query language translations used the syntax for quantification correctly, though there was one case where a student mistakenly used the word "all" to mean "print a list containing every element of the set that..." It should be noted in passing that although many of the students did quite well after brief training in translating the questions provided by the experimenters into the query language, only a few consistently used reasonable questions in the course of problem solving. Many students asked questions in English that were too vague to be answered in the data base given. Others asked "how" or "why" questions which could not possibly have been directly answerable by retrieval from a relational data base. As noted above, in fact, several students apparently were not sure quite what a question was and gave potential solutions or attempted to provide mythical data for the tables.

Retention of quantification syntax

Two weeks after these students had originally been taught the query language syntax, the eleven college students who were working for a temporary employment agency were asked to return. These subjects were not told when they were originally trained that they would later be tested for retention. Six of these students were able to return and were given immediate retests. Of fifteen total cases requiring the use of quantification, there were no instances in which the subjects used this syntax correctly. This was despite the fact, that overall, the use of the query system was retained quite well by the subjects, with the average percentage correct being 53. After an hour refresher course, the subjects were given another test in which they translated questions from English into the query language. Out of fifteen questions requiring the use of the universal quantifier, eight were written incorrectly. In addition to these errors, there were two cases wherein elements of the syntax for universal quantification were used when unnecessary. Among the errors that were committed when the universal quantification structure was required, six were because the subject used syntax appropriate to the other kind of "all" construction. In five cases, the subject simply ommitted any symbols for universal quantification. In another four cases, the subject used 'greater than' instead of 'greater than or equal to.' Thus, not only did subjects have difficulty originally learning the syntax for universal quantification, they also easily forgot what they learned. Again, their errors were not all of one type.

III. STUDIES OF QUANTIFICATION

In this section are described the results of several pilot experiments dealing with the way in which adults who have not had formal training in logic interpret quantificational statements. Since these studies were all pilot studies, the exact findings must be generalized with care. However, the overall pattern of results seems clearcut.

Interpreting Venn diagrams

In one experiment, the concern was to see how people would spontaneously describe, in English, various set relations illustrated by Venn diagrams. Ten subjects were individually given a sheet containing fifteen Venn diagrams. Each Venn diagram showed the relation between two sets. Some diagrams were labelled with abstract terms and some with concrete terms. Three diagrams showed partial overlap of two sets (see Figure 1-IV). Six of the diagrams depicted proper subsets (see Figure 1-I + 1-II). Four of the diagrams showed two disjoint sets (1-V). Two of the diagrams were of two sets that were identical (1-III). Subjects were told that for each Venn diagram they were to describe as exactly as possible the relationship shown. The ten subjects in the experiment were all IBM Research personnel with at least some college education. English was the first language for each subject.

There were a wide variety of expressions used for the set relations, even among ten people with relatively similar educational backgrounds. For diagrams that showed partial overlap, there were ten basically different expressions used for the thirty occassions. The most common expression was of the form "Some B is A and vice versa." The complete list of expressions is given in the Appendix. These may be of interest to those who wish to allow natural language input to a natural languge query system and the interpretation of set relations from this input. For those diagrams which showed a proper subset relationship, there were twenty basically different expressions among the sixty occassions. The most common expression was of the form "All A are B, but not (vice versa | all B are A)" The next most common expression with 6 cases was "All A are B; some B are A." For those diagrams that showed equivalence, there were seven different expressions. The most common was of the form "All A are B and vice versa." For the diagrams that showed A and B as disjoint sets, there were thirteen different expressions. The most common ones were "No A are B and vice versa", "A and B are disjoint (sets)", "A are not B, and vice versa", "A not related to B". It is interesting to note the diversity of expressions capable of describing these set relations that were actually used. Naturally, it has been clear to linguists that a large number of English expression could be used, but this gives no clue as to how many actually would be used by real people. It should be noted that the numbers of different expressions used above referred only to basically different descriptions, and ignored minor differences such as the use of synonyms.

Responses can also be characterized in terms of the accuracy of each description. Three categories are of interest here. The description can be "Inconsistent" with the Venn diagram. The descriptions could be "Consistent" with the Venn diagrams but not describe it uniquely. That is, a given description might describe the set relationship shown, but one might not be able to map unambiguously back from that English description to the Venn diagram. Finally, the description can be "Exact" in the sense that it is not only consistent with the set relations depicted but could reasonably refer only to such a relationship.

Every one of the English descriptions of the disjoint sets was exact. All reasonable interpretaions of the English sentences mean that the two mentioned sets were disjoint. When subjects described equivalent sets, they were also generally accurate. However, one subject merely responded with a question mark to the equivalent sets. Only one of the other descriptions was "Consistent" rather than "Exact".

Subjects were more inexact with their descriptions of the superset-subset relation. In this case, there were nine cases in which the description was judged by the experimenter to be only "Consistent" with the diagram rather than "Exact". For partially overlapping sets, eight of the thirty English statements were only consistent with the diagrams. Sixteen were "Exact". Six of the statements were best described as "Anomolous". They were certainly not "Exact" nor "Inconsistent" but yet not really "Consistent" in the usual sense either. One subject used the form "Some A's might be B's or some B's might be A's." on one occassion. This subject used this basic form another time but used "and" rather than "or" and still another time used a period between the two phrases. Another subject consistently described partially overlapping sets with the form "All A's and all B's."

Another point of interest was that there was a considerable bias to mention first in the English statement whichever set label was on the left in the diagrams for disjoint and partially overlapping sets. (For subset and equivalence relations the two terms were above each other.) This bias to mention the terms in the order they appear in the diagram could be of importance to some experimental designs. One might also want to take this bias into account in display systems. For both the partially overlapping sets and the disjoint sets, the term on the leftmost set was mentioned first in the subject's English statement about 80 per cent of the time.

Summarizing, subjects with college training can generally describe the set relations expressed by Venn diagrams fairly well. They do this spontaneously in a wide variety of ways, however. Subjects exhibited a bias to mention first in the sentence the term that labelled a leftmost set. Subjects were very accurate in describing disjoint sets unambiguously and also did very well in describing equivalence relations. However, some subjects were not "Exact" in their descriptions of partially overlapping sets or of proper subsets. There were no cases however in any of the 150

English descriptions that were collected that could be considered "Inconsistent". This is a rather interesting finding in itself. The finding suggests that wherever possible, one should attempt to reduce the human's task to that of giving consistent descriptions of set relations. Also of interest is the observation that it seems quite easy (at least in English) to describe set disjunction unambiguously, while subsets and partial overlap cause people difficulty. Even when subjects gave exact descriptions of partial overlap, they often sounded forced or elaborate, e.g. "Venn diagram of two sets, A and B with some but not all elements of each in common."

English language consistency judgements.

In this task, English statements involving quantification were given to each of four subjects. Each subject was an adult female with some college training--though none had a technical background.

For each statement in Table IV, the following procedure was employed. The subject was shown a command statement in Table IV. Then the subject was shown descriptions like those in Table V (with appropriate qualifiers). For each description, the subject was asked whether all the departments specified by that description should be given as answers in response to the command statement from Table IV.

The only sentence form that illustrated absolute consistency both within and between subjects was the negative expression. Subjects always said departments such as those lablelled 'e' in table V would be printed out in response to question 5, but to no other question. On the other hand, there was considerable inconsistency with regard to the meanings of statements 6,7 and 8 and f,g, and h. The other forms also showed some inconsistency between subjects. Note that forms a, b, c, d, and e were exact forms. For each statement, there was only one possible set relation referred to, whereas the last three statement forms, used the more common quantifiers "all" and "some" each of which admits to several possibilities. The results of this study are highly tentative since there were only four subjects. However, the same pattern of results was found here as for several other studies done in our laboratory as well as earlier experiments reported in the psychological literature. People have difficulty and differences of opinion concerning the interpretations of statements using "some" and "all". People seem to have little or no trouble interpreting the statement "no A are B."

Venn diagram generation task.

In this task subjects generated Venn diagrams in response to English sentences. They were given brief pretraining that illustrated the possible types of set relations that might hold between two sets and how to indicate these with a Venn diagram. Then the six female subjects, ranging in age from 26 to 57, were given a series of statements that used terms denoting set relations. Each subject was at least a high school graduate. None of the subjects were programmers or had technical college degrees. Venn diagrams that corresponded with all the possible relations between sets that were consistent with the statement. The instructions stressed that more than one diagram could be drawn for each statement. Statements were of five forms : 1. All A are B, 2. Some A are B, 3. All A are not B, 4. Some A are not B, 5. No A are B. Of these statement forms, note that many are consistent with several specific set relations. For example, statement form 1 is consistent with A being a subset of B and is also consistent with A and B being identical (Figure 1-II and III).

Statement form 2 is consistent with either A or B being a subset of the other or with A and B overlapping (partially or completely). Statements 3 and 4 are also consistent with several set relations. Only statement 5 is really unambiguous, consistent only with the relation that A and B are disjoint sets. The statement forms listed above (1-5) were expressed in several sentences each, some of these being abstract and some concrete. The concrete sentences were either consistent with world knowledge or inconsistent with world knowledge. An abstract example of statement form 1 was "All Y are Z". A concrete example consistent with (one interpretation) of form 1 was "All men are mammals" and an inconsistent example was "All cats are dogs". Note that the concrete examples that were consistent with world knowledge, were, in all cases, only consistent with one of the several potential (logically possible) interpretations. It was of interest here to see whether people would only pick the interpretation consistent with that world knowledge, particularly when this interpretation was generally less preferred according to the data of Pezzoli (1970).

Again the results of this experiment indicated a wide range of individual differences, even with the small number of subjects employed. One can consider the performance of these subjects relative to the logician's answers or relative to each other. Consider first comparing the subject's answers to the answers that logicians use. None of the subjects consistently gave all possible interpretations that were logically admissable. Every subject always agreed with the logician's interpretation of "No A are B." (By drawing two disjoint sets). However for all other forms, there was considerable divergence from the logician's interpretations. However, it should be noted that there was only one case wherein a subject drew a diagram that was inconsistent with the English statement. The "errors" that subjects made were overwhelmingly errors of ommission. However, it is interesting to note, that when one considers the responses of the subjects as a group, the set of these responses is exactly the set of possibilities that are logically possible. (Excepting the one case noted above).

Responses of subjects for the various statements were also compared. Combining all responses for statements of the universal "All A are B", the most common form of diagram shown was that of A as a subset of B. This preference coincides with the results of Pezolli (1970) who used a multiple choice paradigm. There was no noticeable tendency here for subjects to normalize their drawings to make them more consistent with world knowledge. With statements of the particular form "Some A are B" the most common response seemed to depend upon knowledge of the world. When abstract terms were used, and when the real world relation between the two sets was that B was a proper subset of A, then showing B as a subset of A was the most common response. The other response was to show A and B as partially overlapping. However, when the real world situation reversed the terms -- A was a subset of B, -- the most common response was to picture A and B as partially overlapping, although other responses were also given. Overall, subjects tended to agree with the logicians interpretations of universals much more often than for particulars. (Cf. Niemark and Chapman, 1975). For statements of the form "All A are not B", the most common case was to show A and B as disjoint sets. However, partial overlap and B as a subset of A were also given by some subjects. For statements of the form "Some A are not B", the most common response was to show A and B as partially overlapping sets. Occassionally however, A and B were shown as disjoint or as B with a subset of A. As mentioned earlier, every subject drew "No A are B" as disjoint sets. The high level of accuracy in interpreting "No A are B" is replicated by Niemark and Chapman (1975).

Comparisons have been made between the interpretations of these subjects and those of formal logic. It is also of interest to consider how well the subjects would have communicated with one another. One primitive measure of this is to see whether there are cases in which every interpretation drawn by a given subject falls entirely outside the range of possibilities considered by other subjects. First of all, note that this never happened with the form "No A are B". With the form "All A are B" there were some cases in which the set of interpretations drawn by one subject was disjoint from the set of interpretations drawn by other subjects. The same was true for statements of the form "Some A are B", "All A are not B", and "Some A are not B." In other words, even with so small a sample size as six individuals, there is no single interpretation that one could pick that would be sure to include a response from every individual tested. Thus, in addition to the logical ambiguity in quantificational statements pointed out by linguists, there is a real behavioral ambiguity as well. (At least when people are forced to deal with statements in isolation).

Manual table look-up task.

In another task, subjects were given questions stated in English. For each question, they needed to find the answer manually. The subjects did this by looking at the data tables shown in Table VI. They spoke aloud the answers as they found them. The subjects were allowed to make notations on the tables themselves or on scratch paper. The experimenter sometimes asked questions or made comments in a manner similar to the clinical experimental method espoused by Piaget (Flavell, 1963). Subjects were asked several questions, the exact number depending upon their ability to answer progressively more complex questions. Several illustrative examples are given below.

Consider the question "Print (find) the departments whose entire line of items is supplied by a single company." This was given to five subjects. First note that there are two possible interpretations of this question. One interpretation is that any department that gets part of its supply of all its items from a single company should be printed out. Another interpretation is that one should only print out companies that get their entire supply of all their items from a single company. According to the first interpretation, the correct answers are Cosmetics, Toy and Hardware. According to the second interpretation, only Cosmetics should be printed out. The first subject's initial impression of the meaning of the question was that it meant that a single department sold all the articles. (?). The experimenter re-explained the question. The subject finally said "Hardware. I think its Hardware. Don't they sell all these articles? (gesturing to the entire set of items in the Sales table). This subject took six and a half minutes and apparently interpreted the question as being equivalent to "Which departments sell all the articles?" This would indeed seem a strange interpretation of the question, but the second subject had a similar interpretation. After 37 seconds, SB said "There is none." Upon question it was clear that she was looking for a single company that supplied all the articles. A third subject said "Stationary. No excuse me, Parker doesn't supply dishes. Cosmetics." This took a minute and a half. Apparently, this subject "understood" the question and used an appropriate table look-up procedure. A fourth subject said, "Cosmetics, Toy, Stationary, and Hardware." This took only fifty seconds. It was clear from the subjects checking (she held her place with her fingers) between tables that she either did not interpret the question correctly or was unable to produce an appropriate algorithm for checking. A fifth subject said "I can't do that. Question doesn't make sense." after 34 seconds. She kept looking at it and finally gave up. Thus, this question was interpreted incorrectly by at least two of five subjects, who in fact seemed unable to perceive either of the "correct" meanings even with coaching from the experimenter.

A similar question is "Print out any departments that sell every article that some company makes." Again, there is an ambiguity as to whether departments that sell every KIND of article made by a given company should be printed out (regardless of whether other departments sell some of that kind of article too) or whether "every" means "every one of every type". According to the first interpretation, the answers are Toy (for Parker), Stationary (for Parker and for Bic and for Dupont) Household (for Dupont). According to the second interpretation, there are no departments which should be printed out. (It should be noted in passing that much of the ambiguity in this and the preceeding question could be avoided by an appropriately organized relational data base. In Table VI, it is not possible to tell whether the dishes that are sold by stationary are the dishes sold by Dupont, Bic, or a mixture.)

The first subject, BV, started by saying "How can you do that? It doesn't tell you that. You only have..." Upon questioning, it became clear that she interpreted the question to refer to every article a company made, regardless of whether it was listed in the table. The experimenter eliminated that ambiguity and the subject started again. This subject's algorithm was then to find the set of items sold by a department and then see whether that entire set was sold by a single company. This algorithm is appropriate to the previous question (...departments whose entire line of items...) but not to this one (viz....departments that sell every article that...). Apparently, this subject interpreted the question as equivalent to "Print the list of Departments all of whose items are sold by one company." A second subject, MK, took the first "correct" interpretation though she failed through a clerical error to find all three departments. A third subject, SB, apparently picked a "correct" interpretation but was unsure whether the question meant "Print out departments that sell exactly the set of items supplied by a company." Thus, looking up data to answer questions involving universal quantification proved difficult for these subjects.

In the Venn diagram and consistency tasks, subjects did fairly well in interpreting and producing statements of the form "No A are B". The manual look-up task included several questions that combined negation with set relations. These questions, which at first seem fairly simple caused the subjects considerable difficulty. In some cases, there is real ambiguity of meaning. Consider the question "List all the departments that sell articles that do not come in green." Of four subjects, one was completely unable to interpret this question. One subject gave the names of departments that sold any item that came in a color other than green. (Even if the same item also came in green). Two subjects however, proceeded as follows: They first located the color green in the TYPE table. They then proceeded to check off every article that came in green. They then moved to the SALES table and elminated every department that sold any one of these items that sometimes came in green. Note that this interpretation is doubly more strict than that of the second subject. The sentence "List all the departments that sell articles that do not come in green." appears clear. In reality, there are a number of unspecified quantifiers. Even with the results of four subjects, it seems clear that different people interpret this sentence differently.

Next consider the similar question "List departments that don't sell items supplied by Bic." Four subjects interpretted this question to mean that if a department sells any item that Bic supplies, this eliminates that department. One subject interpreted the question to mean that in order to be printed out the necessary condition is only that the department sell at least one item that is supplied by a company other than Bic (even if that item is also supplied by Bic). Another subject also followed this procedure but then rechecked her list of departments and eliminated any that did not fit with the interpretation that selling any Bic item eliminated the department.

As a final example, consider the question "List all departments except those that sell articles that don't come in red." None of the seven subjects given this Nor did they agree with each other. A look at the answers to these questions indicates that different subjects interpreted these questions that use quantification differently. A comparison of interpretations across various tasks by the same subjects reveals that even a given subject may interpret a sentence involving quantification differently depending upon what the task is.

General conclusions from quantification studies.

Taken as a whole, the results of the preceeding experiments illustrate several points. First, subjects are not terribly accurate at extracting the "logical" meaning of various statements. This was true whether their understanding was tested by having them decide which specified sets of data were consistent with various questions, having them draw Venn diagrams to represent all the meanings of English sentences. or having them look up answers to questions in a relational data base. Second, subjects are not consistent among themselves and sometimes even a given subject is inconsistent across tasks (but a given subject is generally consistent within a given task). So, behaviorally, there is more ambiguity to many quantificational statements than even the logician would have us believe. (Not less, as might have been hoped.) Third, when presented with Venn diagrams that show various relationships, people produce a wide variety of English statements to express these relationships.

IV. OBSERVATIONS OF DIALOGUES

Analysis of Performance with the query language, section II, illustrated a point consistent with the existing psychological literature, viz., at least for non-programmers, using quantifiers in the way that logicians do is difficult. The quantification studies discussed in Section III strengthened this conclusion. On the other hand, the results with questions that the students themselves generated suggested that the need for universal quantification may be rather minor. This suggestion was somewhat strengthened by some pilot experiments done by John Gould in which five college students were given a variety of problems and some associated tabular (relational) data bases. These students, for each problem, were asked to generate a series of questions that would be appropriate to solving these problems. Again, out of 185 questions, only 7 seemed to involve the explicit use of universal quantification and, several of these showed some apparent confusion. This was despite the fact that some of these problems had been specifically designed to elicit the use of quantification.

In addition, two of the students recorded every question that they heard in the course of a day. Out of 100 questions so recorded none of them involved the use of quantification in the logician's sense. All this suggests that perhaps the logician's use of quantification was less universal than may have been supposed. These observations do not seem to be limited to the dialogues collected here. The dialogue presented as an example by Mann, Moore, Levin and Carlisle (1975) does not really contain an example of quantification in the logican's sense among the 129 lines that are presented.

Despite the fact that many query languages provide facilities for the use of universal quantification in the logician's sense, these facilities may not be vital for some applications. It was the subjective impression of the data base managers for two large data base retreival systems that complex questions involving quantifiers were seldom, if ever used. Indeed, some query languages, such as Interactive Query Facility do not provide the capability for using universal quantification (Gould & Ascher, 1975).

Also relevant are some natural language dialogues collected by the author and analyzed for the use of quantification. These dialogues were between a semi-automated dialogue system (see Thomas, 1975) and subjects who were attempting to find out how a particular computerized order-handling and invoicing system worked. The subject in the experiments typed questions about the system and received answers on an IBM-3277 display. These messages were sent to the subject by a person knowledgeable about the application. The subject continued to ask questions until he felt that he understood the order-handling and billing system in the sense that given any order as input to the system, he would be able to produce the same invoice that the system would. Of 117 questions, only three cases could be construed as involving universal quantification. In addition, there were another three messages that used explicit quantificational statements. (These were not questions.) There were of course, many cases in which statements could arguably have involved "hidden" quantification. For instance, when a subject asked "What does 'acrec' stand for?" one might argue that he REALLY means, "For all cases of 'acrec', what does it stand for?". This seems rather forced, and, in addition, introduces numerous difficulties of interpreting what the universe of discourse is with respect to which "all" is meant. Some of these difficulties will be described below. For now at least, one can conclude that the occassions of explicit use of quantifiers were quite rare. Recall that in the query language study, subjects were required to translate questions from English into the query language. There were several questions in that experiment similar to "List the departments all of whose items are supplied by a single company." None of the uses of quantication in the natural language dialogues that were collected achieved this level of complexity. The first two questions are as follows: (This subject had had no experience with business terms or procedures and seemed overwhelmed by the complexity of the system.)

USER: Help.
SYSTEM: Which notion is causing difficulty.
USER: All notions.
SYSTEM: Like what?
USER: Like everything you've sent me.

Although some might rephrase the user's first comment in terms of symbolic logic "Given anything which is a notion, that notion is causing difficulty." it seems clear that a computer system ought not interpret this literally or logically but rather "realize" that this is merely a way of expressing an emotional difficulty rather than a logical statement. Similarly, the subject's second statement cannot be accurately paraphrased as "Given anything you've sent me, that thing is causing difficulty." Again, there is a certain sense in which the subject feels everything that has been sent caused difficulty, but is the subject referring to every element of the set of things (notions, words, messages?) or to the entire body of information? Indeed, its quite possible here that each and every individual message was quite understandable and the difficulties of the subject were caused by attempting to remember and integrate the totality of the information. Even if the subject literally meant every single message caused difficulty, could that really mean EVERY message including the innocent "hello" at the beginning of the dialogue? It seems clear then that even if one were to make the rather dubious claim that this subject were referring to every element of a set, how is this set defined? How could any system know? What the subject's statement really seems to say, if one assumes it refers to every element of a set, rather than to the Gestalt is that every message that causes difficulty is causing difficulty, or in other words, that there are some notions that are causing difficulty. As the subsequent questions and answers of the dialogue indicated, the subject was not really able to identify the locus of difficulty. This example illustrates three potential difficulties for a natural language system that attempts to translate an expression involving "all X are Y" too readily into a logical expression. First of all, the person's statement may simply be an expression of an emotional state and not a literal statement at all. Second, the statement may not refer universally to the elements of a set but to their totality or even configural properties. Third, the set to which the statement refers may not be clear from context or may even violate the set explicitly mentioned in the statement.

Now consider the following example of a question from another dialogue. This subject also had no familiarity with business but as a systems programmer was quite familiar with understanding complex systems.

USER: Apparently something called a control listing is printed each day.

Now can this person honestly think that a control listing is printed EVERY day? Even if the computer breaks down? Even if this company goes out of business? What is meant here is not the universal quantifier but something like "On every reasonable day, a control listing is printed". Note that even the relevant dimensions of what constitutes reasonableness are not explicit here. For this reason, one cannot deal with this difficulty merely by introducing the notion of fuzzy sets (Zadeh, 1974).

Another subject in the experiment used what might appear to be universal quantification three times. These messages are shown below. Interestingly, none of these occassions was during the phase of the experiment when system understanding was the goal, but occurred during the shorter phase of diagnostic problem solving that followed. A consideration of the pragmatics of these two tasks makes it quite plausible that a "true" use of quantification is much more likely to appear when one is attempting to debug a system than when one is attempting to understand it.

USER: Are errors being made at all levels of taxation (state, local, and federal) or only one or two of these levels?

Here is an instance of "all" followed by a listing of the elements that are involved. Here the use of "all" is fairly clear, though understanding it is not really necessary for an appropriate system response. All the system need do is answer the question "Which kinds of taxes have errors." Understanding the "all" is really unnecessary. In fact, the user doesn't want the literal, logical structure of his question answered. (That is, he would not be happy with a "yes" or "no"). Another example occurs slightly farther in the dialogue.

SYSTEM: Only on the state tax level.

USER: I presume that the customer files contains a txcd2 column with all correct entries, right?

This seems to be an actual use of the logician's use of universal quantification wherein the set that is referred to is well-defined. However, again to give a satisfactory reply to the subject, the system only needs to "realize" that the user is asking about the correctness of the codes in the txcd2 column. The same system response could equally well serve for "What about the correctness of txcd2?" "Are most of the codes in txcd2 correct?" "Are any of the codes in txcd2 correct?" A little later in the dialogue, the following question appears.

USER: Does the system consult txcd2 before calculating tax on each item in the invoice?

The use of "each" seems quite close to the logical meaning of univeral quanfication. Again, however, it would not really be necessary for the system to understand that meaning in order to give an appropriate response.

In summary, only three users even appeared to use explicit universal quantification. Five of the eight users who interacted with the semi-automated dialogue system in order to understand a computerized order-handling and invoicing system never used explicit universal quantification at all. Furthermore, we see that there are number of difficulties for any natural language system that attempts to interpret a user's messages as involving the logician's use of quantification, at least if these dialogues are even vaguely representative. In many cases, something which seems to involve quantification on the surface, actually doesn't. In other cases, in which the subject really does mean universal quantification it isn't really necessary for the system to "know" this in order to provide an appropriate response. For some applications the logician's quantifers are probably desirable. For a number of applications though, one can avoid many difficulties by presenting appropriate information sufficient for the questioner's needs. In very many cases this would not require mind reading but simply listing the attributes and values relevant to the question.

Quantification in system description dialogues

Dialogues were also collected in the following situation. Subjects described a particular type of order-handling and invoicing to the semi-automated dialogue system. The semi-automated system attempted to find out about how the subject's particular order-handling and invoicing system worked by asking the questions on the Application Customizer Service questionnaire (1971). A number of the questions on the questionnaire contain items like "Are all your X's, Y's?" "Do you ever H your Z's?". When answering such questions, some subjects in our experimental dialogues consistenly repeated back the qualifying phrases in the question along with a quantifier. Apparently, this was a technique to insure that the system and the subject both had the same understanding of the quantificational aspects of the question and answer. For example, when the system asked "Are sales ever made for cash or COD?" the user chose to respond "Yes, in some cases, ..." rather than the shorter "Yes, ..." When the system asked "Good. Do all your customers enjoy extended price discount within..." the user responded with "They all have the option but not all take advantage." As noted, the ACS questionnaire for order-handling and invoicing contains many statements of the form "Are all X also Y?" The use of questions of this form is probably not as effective in eliciting rare exceptions from the user as the alternative form of the question "Are any X, not Y?" Since recall is generally harder than recognition, a few examples of when X's might not be Y's, might also serve as a reminder. Although no experiments have actually been done with real customers using the actual ACS questionaire, a number of studies in the psychological laboratory indicate that existential questions are generally easier to answer than universal questions. (Meyer, 1970, Just, 1974). However, there may be some exceptions to this rule (See p. 405, Anderson and Bower). As is generally the case with psychological effects, these effects need to be verified in the actual situation of interest. One can observe the following example (taken from the system description experiment) of an exception almost missed.

SYSTEM: Are all items and special charges subject to tax?

USER: Yes. However, freight is a special charge that is nontaxable.

In a real use of the ACS questionaire, if the user had not thought of this until the program was installed, the resulting costs could have been considerable.

Strategies that people use for dealing with quantification

In the sequences of questions collected by John Gould as well as those collected by the author, the subjects used strategies that enabled them to gain information without having to resort to complex uses of quantification, subsetting and negation. In some applications, it might be possible to design a system that would allow the user to use these strategies rather than forcing him to formulate complex queries. The first of the subjects' strategies was to ignore quantifiers and allow the feedback from the question to provide some quantificational disambiguation. This is the technique that is used in the game of twenty questions. One says "are X's Y's?" If X and Y refer to identical sets or if X is a subset of Y, then the answerer is expected to say "Yes". If Y is a subset of X or if X and Y are partially overlapping sets, the answerer is expected to say "partly" and if X and Y are disjoint, the answerer is expected to say "no".

The second common strategy that avoided specifying complex set relations was to use a sequence of questions like the following: "Let's call students who have characteristics a,b, and c U-students." "Let's call students who have characteristics d or e, V-students." "Let's call teachers who teach any of the U-students, U-teachers." "Let's call teachers who teach any of the V-students, V-teachers." "Print out a list of V-teachers who are also U-teachers."

A third strategy subjects used was simply to ask for two or more sets of data to be printed out. Presumably, in these cases the subjects themselves would have attempted to judge important set relations among the sets of data. Still another alternative to using constructions like "Are all the X's Y's?" was to ask a question that involved a negative plus an existential. This seemed particularly common when there was an existing word in English for the negative of the attribute-value the subject was interested in. For example, rather than asking whether all the students were happy with the class, a person could ask whether there were any students who were unhappy with the class.

As pointed out earlier, the statement "All A are B" is consistent with two set relations: A is equivalent to B, and A is a proper subset of B. Subjects who attempted to solve problems by asking a sequence of questions appeared to be "homing in" on the possibility of an equivalance relation between A and B by asking a sequence of questions about the elements of A and B until there could be little doubt that they were equivalent sets.

In addition to these multi-question strategies, it should be noted that often people use qualificational statements rather than quantificational or conditional statements. For example, a person might find the expression "Put the red blocks in the box." quite natural in contrast to the conditional statement "If a block is red, then put it in the box." or the quantificational expression "Given anything which has the property red, and has the property of being a block, that thing also has the property that it belongs in the box." The apparent preference for people to make qualificational statements was pointed out by Miller and Becker (1974).

V. RECOMMENDATIONS

Naturally, since so little research has been done on these matters, these recommendations, except for the first one, should be viewed as tentative. They are meant as a catalyst to comment and criticism and a starting point for research, not as absolute design criteria.

1. Studies should be undertaken concerning the usability of a query system with the particular users and tasks that the system is designed for.

2. Unless one has a logically sophisticated population of users, one should make it possible for users to gather information in ways that are consistent with their natural strategies. Some of the strategies observed above may be fairly universal. The safest course, though, would be to see what strategies particular users may want for a particular system.

3. If, for some reason, a system must use the logician's quantifiers, then a high proportion of errors should be expected and the system designed accordingly. (Intelligible error messages, recovery procedures, etc.)

4. Whenever practical, the human's quantification tasks should be limited to producing or choosing descriptions that are consistent with his needs rather than forcing him to unambiguously specify his needs.

5. Whenever practical, communicate with the user in terms of set identities and set disjunctions. (Obviously, in some cases, there is no choice.)

6. A natural language query system should generally not attempt to answer exactly the user's precise question when that question involves quantification. Two users even in the same context

may well have in mind by the same string of English words two different set relationships. A more modest and workable strategy ---which humans themselves seem to use in communicating with each other---is to provide information relevant to the query and satisfying to the user. Note that this strategy does not require that the question answering system induce from the user's question a deep structure corresponding with the user's.

ACKNOWLEDGMENTS

I wish to thank John Gould for access to his data on question-asking. I also wish to thank Lance Miller and John Gould for their many helpful suggestions on this manuscript as well as Linda Callahan for her excellent secretarial services.

REFERENCES

Anderson J. R. and Bower G. H., Human Associative Memory, New York: John Wiley, 1973.

Application Customizer Service System/3 (Disk) Sales and Distribution. Questionnaire Publication No. S320-1165-1) IBM Corporation (September 1971).

Ceraso and Provitera, 1971. Sources of Error in Syllogistic Reasoning Cognitive Psychology, 1971, *2*, 400, 410.

D. D. Chamberlin, R. F. Boyce, "Sequel: A structured English Query language, IBM Technical Report, RJ 1394.

Falmagne, R. J. (Ed.), Reasoning: Representation and Process in Children and Adults. New York: Wiley, 1975.

Flavell, J. H., The Developmental Psychology of Jean Piaget, Princeton: Van Nostrand 1963.

Gould, J. D. and Asher, R. N., Use of an IQF-like Query Language by Non-programmers. IBM Technical Report, RC 5279, February, 1975.

Griggs, R. A., Logical Errors in Comprehending Set Inclusion Relations in Meaningful Text. Indiana Mathematical Psychology Program Report No. 74-7.

Fehder, P. L., The Hierarchic Query Language IBM Technical Report, RJ 1307, November 20, 1973.

Interactive Query Facility (IQF) for IMS/360, Publication No. GH 20-1074, IBM Corporation, While Plains, New York.

Johnson-Laird, P. N., On Understanding Logically Complex Sentence, Quarterly Journal of Experimental Psychology, 1969, *21*, 1-13.

Just, M. A., Comprehending Quantified Sentences. The Relation between Sentence-Picture and Semantic Memory Verification. Cognitive Psychology, 1974, *6*, 216-236.

Mann. U. C., Moore, J. A., Levin, J. A., and Carlisle, J. H., Observation Methods for Human Dialogue, 151 Research Report 75-33, (June, 1975).

Meyer, D. E., on the Representation and Retrieval of Stored Semantic Information, Cognitive Psychology, 1970, *1*, 242-300.

Miller, L., and Becker, C. Programming in Natural English, IBM Technical Report RC 5137, November 1974.

Niemark, E. D. and Chapman, R. H., Development of the comprehension of logical quantifiers. In Falmagne, R. J. (Ed.), Reasoning: Representation and Process in Children and Adults. New York: Wiley, 1975.

Niemark, E. D. and Santa, J. L., Thinking and concept attainment. In M. R. Rosenzweig and L.

W. Porter (Eds.) Annual Review of Psychology (26). Palo Alto: Annual Reviews, 1975.

Revlis, R., Syllogistic reasoning: logical decisions from a complex data base. In Falmagne, R. J. (Ed.), Reasoning: Representation and Process in Children and Adults. New York: Wiley, 1975.

Roberge, J., A re-examination of the interpretations of errors in formal syllogistic reasoning. Psychonomic Science, 1970, *19* (6), 331-333.

Scott, R. F. and Simmons, D. B., Programmer productivity and the Delphi technique, Datamation 1974, *20* (5), 71-73.

Staudenmeyer, H., Understanding Reasoning, Unpublished Ph.D. Thesis, University of Colorado 1973.

Staudenmeyer, H., Understanding conditional reasoning with meaningful propositions. In Falmagne, R. J. (Ed.), Reasoning: Representation and Process in Children and Adults. New York: Wiley, 1975.

Thomas, J. C., and Gould, J. D., A psychological study of Query by Example, IBM Technical Report RC 5124, November, 1974.

Thomas J. C. and Gould J. D., A psychological study of Query by Example., IN AFIPS Conference Proceeding Volume 44, Montvale, New Jersey: AFIPS Press, 1975.

Thomas J. C., A method for studying natural language communication. IBM Technical Report (in preparation).

Thompson, F. B., Lockemann, P. C., Dostert, B., and Deverill, R. S., A Rapidly Extensible Language System. Proceedings of the 24th National Conference ACM, Publication 1969, pp. 399-417.

Wason, P. C., and Johnson-Laird, P. D., Psychology of Reasoning: Structure and Content Cambridge Massachusetts: Harvard University Press.

Zadeh, L. A., The Concept of a Linguistic Variable and its Application to Approximate Reasoning, IBM Technical Report, RJ 1355, February, 1974.

Zloof, M. M., Query by Example, IBM Technical Report, RC 4917, July 1974.

Zloof, M. M., 1975, The invocation and definitions of tables and forms, IBM Technical Report, RC 5115, February, 1974.

TABLE I

SAMPLE QUESTIONS IN THE QUERY LANGUAGE

PERSONNEL FILE

	NAME	SALARY	AGE	MANAGER	DEPARTMENT
1.	p. Jones	p. 23K			Sports
2.	p. Smith Riley				Toys Toys
3.	p. Scholz				Computers

SUPPLY FILE

	COMPANY	ITEM	NO. RECEIVED
3.	IBM	360	
4.	p. ABC	all (pen)	
5.	p. ABC	≥ all (pen)	

SALES FILE

DEPARTMENT	ITEM	NO. SOLD
3. Computers	360	
4. DP	≥ all (pen)	
5. DP	all (pen)	

SUPPLIER DATA FILE

	COMPANY	LOCATION	SIZE	PRESIDENT
3.	IBM	Mass.		

English

1. Print the names and salaries of people who work in sports.
2. Print the names of people who work in the same Department as Riley.
3. Print the names of people who work in a department that sells an item supplied by a company located in Massachusetts.
4. Print the names of companies all of whose items are sold in the DP department.
5. Print the names of companies that supply all the items that the DP department sells.

TABLE II

This is a list of the questions that involved universal quantification. The numbers in the parentheses refer, respectively, to the percent of subjects who wrote correct queries and the mean time in minutes to write the query.

Print out any departments that sell every item that some company makes. (38.5 percent, 1.3 min.)

Print the names of anyone who works in a department that gets all its items from a single company. (48.7 percent, 1.5 min.)

Print the names of company presidents whose companies supply every item to the department Jones works in. (48.7 percent, 1.8 min.)

Print the departments whose entire line of items is supplied by a single company. (51.3 percent, 1.2 min.)

Find out whether any departments sell every item that is supplied to our company. (51.3 percent, 1.5 min.)

TABLE III

201. Student ratings for Jones' physics course are really low.

202. Some students have complained about the stiff requirements for a math major.

203. Thornstein is flunking out.

204. The younger faculty claim that the older faculty are getting paid more than their share.

205. How could we help a student pick a major?

TABLE IV

SAMPLE QUESTIONS FOR THE CONSISTENCY TEST

1. Print out departments that sell some but not all of the large articles; and that also sell articles in other sizes.

2. Print out departments that sell all the blue articles and, in addition, sell articles of other colors.

3. Print out any department that sells only ACME articles, but only if that department sells less than all of the articles that ACME supplies to us.

4. Print out departments that sell all the articles that BIC supplies and that sell no articles supplied by any other company.

5. Print out departments that sell no yellow articles.

6. Print out departments that sell all the articles that ABC supplies.

7. Print out departments that sell small articles.

8. Print out departments that sell some red articles.

TABLE V

ANSWERS FOR.CONSISTENCY JUDGEMENTS

a. Departments that sell some but not the large articles; and that also sell articles in other sizes.

b. Departments that sell all the large articles and, in addition, sell articles in other sizes.

c. Any departments that sell only large articles, but only if that department sells less than all of the large articles that are supplied to us.

d. Departments that sell all the large articles and that sell no articles in any other size.

e. Departments that sell no large articles.

f. Departments that sell all the large articles.

g. Departments that sell only large articles.

h. Departments that sell some large articles.

TABLE VI

DATA TABLE FOR QUERY LANGUAGE MANUAL LOOK-UP

EMP	NAME	SALARY	MGR	DEPT
	JONES	8K	SMITH	HOUSEHOLD
	ANDERSON	6K	MURPHY	TOY
	MORGAN	10K	LEE	COSMETICS
	LEWIS	12K	LONG	STATIONERY
	NELSON	6K	MURPHY	TOY
	HOFFMAN	16K	MORGAN	COSMETICS
	LONG	7K	MORGAN	COSMETICS
	MURPHY	8K	SMITH	HOUSEHOLD
	SMITH	12K	HOFFMAN	STATIONERY
	HENRY	9K	SMITH	TOY

SALES	DEPARTMENT	ITEM
	STATIONERY	DISH
	HOUSEHOLD	PEN
	STATIONERY	PENCIL
	COSMETICS	LIPSTICK
	TOY	PEN
	TOY	PENCIL
	TOY	INK
	COSMETICS	PERFUME
	STATIONERY	INK
	HOUSEHOLD	DISH
	STATIONERY	PEN
	HARDWARE	INK

SUPPLY	ITEM	SUPPLIER
	PEN	PARKER
	PENCIL	BIC
	PARKER	INK
	PERFUME	REVLON
	INK	BIC
	DISH	DUPONT
	LIPSTICK	REVLON
	DISH	BICK
	PEN	REVLON
	PENCIL	PARKER

TYPE	ITEM	COLOR	SIZE
	DISH	WHITE	M
	LIPSTICK	RED	L
	PERFUME	WHITE	L
	PEN	GREEN	S
	PENCIL	BLUE	M
	INK	GREEN	L
	INK	BLUE	S
	PENCIL	RED	L
	PENCIL	BLUE	L

APPENDIX A

Expressions used by native speakers to refer to various relations between two sets. In the expressions below exact references to particular sets have be substituted by the capital letters 'A' and 'B'. Parentheses indicate alternatives that some subjects used. 'e' is used for the null string. The number in parthentheses is the number of cases each form was used.

Relation I A B are partially overlapping.

Some B is A and (vice versa/some A is B).	(7)
Some but not all A are B and (vice versa/some but not all B are A).	(4)
Some B are A.	(4)
All A's and all B's.	(3)
There are things which are A & not B, B & not A, A & B, and not A & not B.	(3)
Some A's might be B's or some B's might be A's.	(3)
Part of B is included in A and (vice versa/part of A is included in B).	(1)
Some B are A but not all A are B.	(1)
Some A are B; not all A are B; not all B are A.	(1)
A, B, and A and B.	(1)
Not all A are B & not all B are A.	(1)
A & B have some but not all elements in common.	(1)

Relation II A and B are equivalent.

All A are B and (vice versa/all B are A).	(8)
A are B and (vice versa/B are A/conversely).	(3)
A=B.	(2)
A and B are equivalent labels.	(2)
? (Apparently this subject did not understand this relation.)	(2)
A # B	(1)
All A are B.	(1)
A and B are the same.	(1)

Relation III A & B are disjoint sets.

No A are B (and/e) (vice versa/no B are A).	(10)
A are not B and (conversely/vice versa/B are not A).	(5)
A and B are disjoint (sets/e).	(5)
A not related to B.	(4)
A, B, not A and not B.	(4)
A and B are mutually (exclusive/distinct).	(3)
The intersection of A and B is the empty set.	(3)
B is not an A.	(1)
There are no A who are B.	(1)
A & B are separate.	(1)
All B are not A.	(1)
A and B does not exist.	(1)
A are not B.	(1)

Relations IV & V A is shown as a proper subset of B.

All A are B (but/e) not (vice versa/all B are A).	(16)
All A are B. Some B are A.	(6)
There are things which are: A which are B. B which are not A. Something that is not B.	(6)
Some B are A.	(5)
A's are (special/particular) B's.	(4)
(The set/e) A is a (proper/e) subset of (the set of/e) B.	(3)
A is contained in B.	(2)
All A are B; some B are not A.	(2)
A's are B's.	(2)
If an A exists it is also a B; however a B can exist without being an A.	(2)
All A (e/that exist) are B but all B's are not necessarily A.	(2)
All A are B.	(2)
All A's are B's.	(1)
A thing can be a B without being an A but if a thing is an A then it is a B.	(1)
A are B & there are other B's too.	(1)
A are a kind of B, but not all B are A.	(1)
Some B are A and all A are B.	(1)
A is a subset of B – i.e. some B's are A.	(1)
All A are B but there might be some B that are not A.	(1)
Not all B are A but all A are B.	(1)

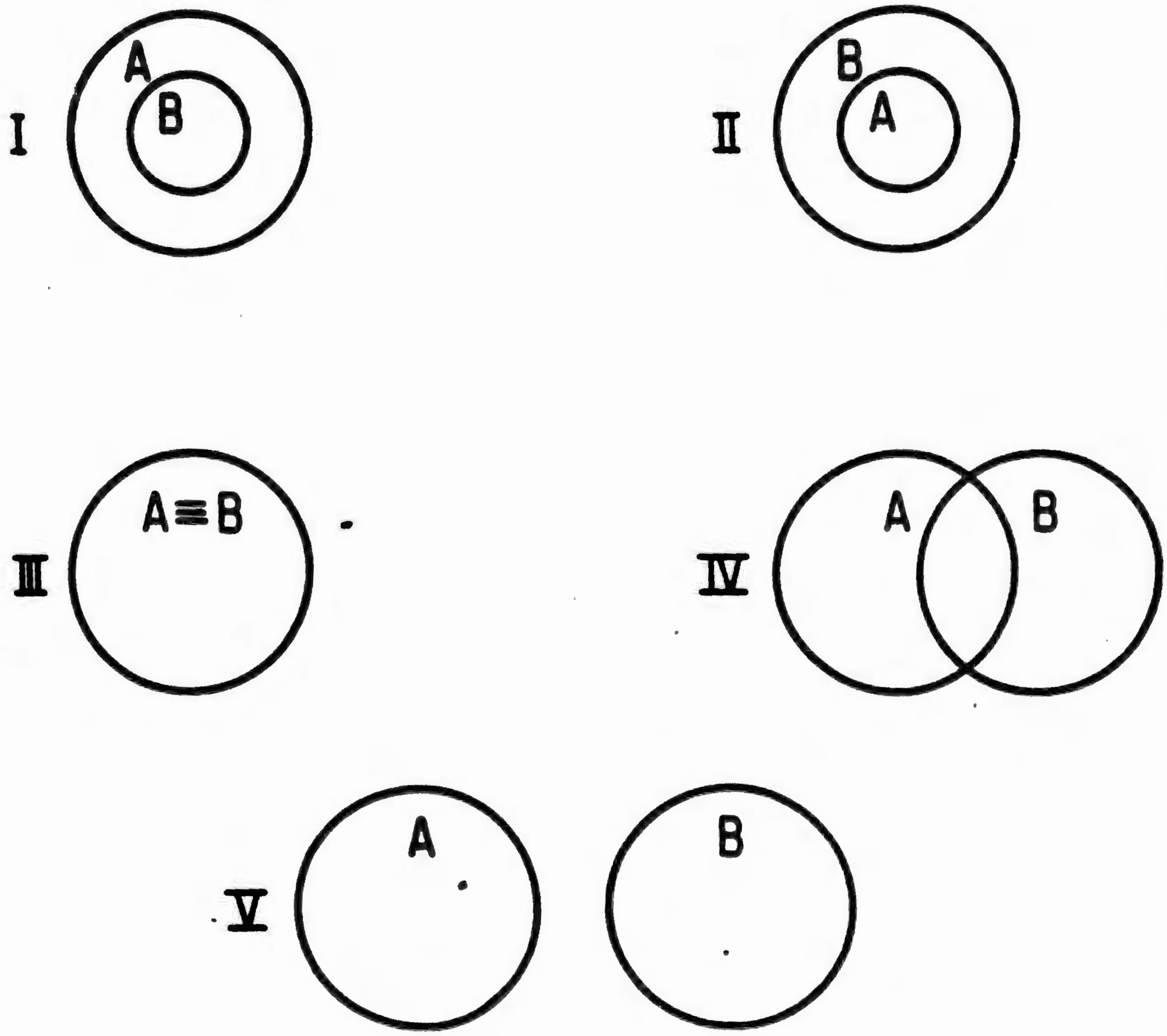

Figure 1. Venn diagrams of possible set relations between A & B.

TECHNICAL REPORTS DISTRIBUTION LIST

CODE 455

Director, Engineering Psychology (5 cys)
Programs, Code 455
Office of Naval Research
800 North Quincy Street
Arlington, Virginia 22217

Defense Documentation Center (12 cys)
Cameron Station
Alexandria, Virginia 22314

Director, ONR Branch Office
ATTN: Dr. J. Lester
495 Summer Street
Boston, Massachusetts 02210

Director, ONR Branch Office
ATTN: Dr. M. Bertin
536 S. Clark Street
Chicago, Illinois 60605

Director, ONR Branch Office
ATTN: Dr. E. Gloye
1030 East Green Street
Pasadena, California 91106

Director, ONR Branch Office
ATTN: Mr. R. Lawson
1030 East Green Street
Pasadena, California 91106

Dir., Naval Research Laboratory (6 cys)
Technical Information Division
Code 2027
Washington, D. C. 20375

Dir., Naval Research Laboratory (6 cys)
ATTN: Library, Code 2029 (ONRL)
Washington, D. C. 20375

Mr. John Hill
Naval Research Laboratory
Code 5634
Washington, D. C. 20375

Office of Naval Research
Information Systems Program, Code 437
800 North Quincy Street
Arlington, Virginia 22217

Office of Naval Research
Operations Research Program, Code 434
800 North Quincy Street
Arlington, Virginia 22217

Office of Naval Research
Naval Analysis Programs, Code 431
800 North Quincy Street
Arlington, Virginia 22217

Office of Naval Research
ATTN: Dr. K. T. Wallenius, Code 436
800 North Quincy Street
Arlington, Virginia 22217

Office of the Chief of Naval
Operations, Op-987E
Department of the Navy
Washington, D. C. 20350

CDR H. J. Connery
Office of the Chief of Naval Operations,
Op-987M4
Department of the Navy
Washington, D. C. 20350

Dr. A. L. Slafkosky
Scientific Advisor
Commandant of the Marine Corps
Code AX
Washington, D. C. 20380

Dr. Heber G. Moore
Hqs. Naval Material Command
Code 0331
Department of the Navy
Washington, D. C 20360

Mr. Arnold Rubinstein
Naval Material Command, NAVMAT 03424
Department of the Navy
Washington, D. C. 20360

Commander, Naval Electronics
Systems Command
Command and Control Div., Code 530
Washington, D. C. 20360

Naval Electronics Systems Command
Human Factors Engineering Branch
Code 4701
Washington, D. C. 20360

Bureau of Medicine and Surgery
Human Effectiveness Branch, Code 713
Department of the Navy
Washington, D. C. 20372

CDR Robert Wherry
Human Factors Engineering Branch
Crew Systems Department
Naval Air Development Center
Johnsville
Warminster, Pennsylvania 18974

LCDR Robert Kennedy
Human Factors Engineering Br., Code 5342
U.S. Naval Missile Center
Point Mugu, California 93042

Lt. Col., Henry L. Taylor, USAF
OAD*E&LS) ODDR&E
Pentagon, Rm. 3D129
Washington, D. C. 20301

Mr. Richard Coburn
Head, Human Factors Division
Naval Electronics Laboratory Center
San Diego, California 92152

Dean of Research Administration
Naval Postgraduate School
Monterey, California 93940

Navy Personnel Research and
Development Center (Code 10) (5 cys)
San Diego, California 92152

Mr. James L. Long
Weapons Systems Research (N-332)
Naval Education and Training Command
Naval Air Station
Pensacola, Florida 32408

Human Factors Dept., Code N215
Naval Training Equipment Center
Orlando, Florida 32813

Dr. George Moeller
Head, Human Factors Engineering Branch
Submarine Medical Research Laboratory
Naval Submarine Base
Groton, Connecticut 06340

Lt. Col. Austin W. Kibler
Director, Human Resources Office
Advanced Research Projects Agency
1400 Wilson Blvd.
Arlington, Virginia 22209

U.S. Air Force Office of Scientific
Research
Life Sciences Directorate, NL
1400 Wilson Blvd.
Arlington, Virginia 22209

Dr. J. M. Cristensen
Chief, Human Engineering Division
Aerospace Medical Research Laboratory
Wright-Patterson AFB, OH 45433

Dr. J. E. Uhlaner
Dir., U.S. Army Research Institute
for the Social & Behavioral Sciences
1300 Wilson Blvd.
Arlington, Virginia 22209

Chief of Research and Development
Human Factors Branch
Behavioral Science Division
Department of the Army
ATTN: Mr. J. Barber
Washington, D. C. 20310

Dr. Joseph Zeidner
Dir., Organization and Systems
Research Laboratory
U.S. Army Research Institute for
the Behavioral & Social Sciences
1300 Wilson Blvd.
Arlington, Virginia 22209

Dr. Stanley Deutsch
Chief, Man-Systems Integration
OART, Hqs., NASA
600 Independence Avenue
Washington, D. C. 20546

Dr. Jesse Orlansky
Institute for Defense Analyses
400 Army-Navy Drive
Arlington, Virginia 22202

Dr. Edgar M. Johnson
Organizations & Systems Research Lab.
U.S. Army Research Institute for the
Behavioral and Social Sciences
1300 Wilson Blvd.
Arlington, Virginia 22209

Dr. James Parker
BioTechnology, Inc.
3027 Rosemary Lane
Falls Church, Virginia 22042

Dr. Edwin A. Fleishman
Foxhall Square
3301 New Mexico Avenue, N.W.
Washington, D. C. 20016

American Institutes for Research Library
135 N. Bellefield Avenue
Pittsburgh, Pennsylvania 15213

Psychological Abstracts
American Psychological Association
1200 17th Street, N.W.
Washington, D. C. 20036

Dr. A. I. Siegal
Applied Psychological Services
404 E. Lancaster Street
Wayne, Pennsylvania 19087

Dr. Joseph Wulfeck
Dunlap and Associates, Inc.
115 South Oak Street
Inglewood, California 90301

Dr. Robert R. Mackie
Human Factors Research, Inc.
Santa Barbara Research Park
6780 Cortona Drive
Goleta, California 93017

Mr. Wes Woodson
Man Factors, Inc.
4433 Convoy Street, Suite D
San Diego, California 92111

Dr. C. H. Baker
Director, Human Factors Wing
Defense & Civil Institute of
Environmental Medicine
P. O. Box 2000
Downsville, Toronto, Ontario
Canada

Journal Supplement Abstract Service
American Psychological Association
1200 17th Street, N.W.
Washington, D. C. 20036

Dr. Bruce M. Ross
Department of Psychology
Catholic University
Washington, D. C. 20017

Mr. Harry Chipman
WR Systems, Inc.
2531 S. Jefferson Davis Highway
Arlington, Virginia 22202

Dr. David Meister
U.S. Army Research Institute
1300 Wilson Blvd.
Arlington, Virginia 22209

Mr. George Graine
Naval Ship Systems Command
(SHIPS 047C12)
Department of the Navy
Washington, D. C. 20362

Lt. Col. John F. Ahlborn
Headquarters, AFSC-DLSE
Andrews AFB
Washington, D. C. 20334

Dr. H. H. Wolff
Technical Director (Code N-2)
Naval Training Equipment Center
Orlando, Florida 32813

Dr. Donald A. Topmiller
Chief, Systems Effect. Branch
Human Engineering Division, USAF
Wright Patterson AFB, Ohio 45433

Col. Robert O. Viterna
DA/OCRD
Headquarters, Dept. of the Army
DARD-ARS-B
Washington, C. C. 20310

Dr. Anthony Debons
IDIS
University of Pittsburgh
135 N. Bellefield Avenue
Pittsburgh, Pennsylvania 15260

Dr. Alfred F. Smode
Training Analysis & Evaluation Group
Code N-OOT
Orlando, Florida 32813