DOCUMENT RESUME

ED 085 648		CG 008 597

AUTHOR	Kochen, Manfred; Badre, Albert N.
TITLE	Recognizing and Formulating Problems: Learning to Comprehend and Organizing Knowledge into Structures.
INSTITUTION	Michigan Univ., Ann Arbor. Mental Health Research Inst.
SPONS AGENCY	National Center for Educational Research and Development (DHEW/OE), Washington, D.C. Regional Research Program.
BUREAU NO	BR-2-E-057
PUB DATE	Aug 73
GRANT	OEG-5-72-0050(509)
NOTE	44p.
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	Algorithms; *Cognitive Processes; Decision Making Skills; Deductive Methods; *Hypothesis Testing; *Inquiry Training; Learning; Performance Criteria; *Problem Solving, Questioning Techniques; *Research Projects

ABSTRACT

Comprehension of a problem or task that is generated in the real world rather than presented as a well-defined problem-statement of the kind encountered in textbooks or psychological laboratories was related to the ability of recognizing, selecting and formulating problems. The process of acquiring and utilizing this ability was conceptualized with the help of flow diagrams for algorithms. This resulted in the furthering of a new and fruitful theory of cognitive learning which stresses the formation and use of hypothesis and how to represent them. New experimental techniques were developed for measuring performance and quality of questions, on problems requiring shifts in representation. These were applied to investigate the effect of experience in learning to formulate such problems in fifth graders and in college students. New procedures for exposing learners to such experiences were also derived and tested. Results suggest that children learn problem recognition and formulation if they are exposed to inquiry-provoking situations where they have to form hypotheses. College students with experience in having to shift representations perform better on tasks requiring such shifts than those who don't. Question-quality was demonstrated to be correlated with problem-solving performance. (Author)

ED 085648

FINAL REPORT

Project No. 2-E-057
Grant No. OEG-5-72-0050(509)

RECOGNIZING AND FORMULATING PROBLEMS
LEARNING TO COMPREHEND AND ORGANIZING KNOWLEDGE INTO STRUCTURES

Manfred Kochen and Albert N. Badre

Mental Health Research Institute
The University of Michigan
Ann Arbor, Michigan

August, 1973

U.S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE

Office of Education
National Center for Educational Research and Development

RECOGNIZING AND FORMULATING PROBLEMS
LEARNING TO COMPREHEND AND ORGANIZING KNOWLEDGE INTO STRUCTURES

1. Introduction

The main goal of this project was to specify both conceptually and operationally some of the criteria necessary for the attainment of comprehension in learning. It became clear that the best novel approach to detecting and measuring the level of comprehension in learning was to let the subject experience the process of formulating and solving problems.

The subjects in all of our experiments were given minimal instructions and confronted with an initially ill-defined problem-situation. In all cases, the task of the subjects was to formulate and solve the problem. The data collected and analyzed were the verbal protocols of questions asked and actions taken. The questions posed by the subjects were analyzed for their degree of comprehension. It was hypothesized that a three-stage intellectual process was involved.

This process reflected in shifts of representations is as follows: First, a person becomes aware of a problem-situation which stimulates him to generate a problem-statement. This may be in writing, expressed orally, or merely thought and evidenced by other behavior. This statement is based on (a) making assumptions about a newly encountered environment (problem) on the basis of previous learning, and (b) formulating new assumptions on the basis of the newly perceived environment. Secondly, he transforms the formulated problem-statement from a statement of belief to one of knowledge. Operationally this involves testing, verifying, and reformulating such a statement. Thirdly, organizing the knowns and givens about the problem into a final statement which we call the stage of comprehension.

The above described approach stems from a well-established line of research on learning that emphasizes the concepts of cognitive and step-wise (sequential) structuring. Among adherents of this kind of research are: Gagné (1970); Ausubel (1960, 1963); Miller, Galanter, Pribram (1960); Estes (1959); Minsky (1970); Suppes (1964); and Bruner (1960). More specifically, Gagné and associates (1962, 1970a, 1970b) have hypothesized and shown that step-wise, hierarchical organization is necessary to the mastery of the terminal task in learning. Similarly, Ausubel (1960, 1963), assuming the hierarchy hypothesis, goes on to specify it further by demonstrating that effective and meaningful learning occurs when material is introduced to the learner, at the highest levels of the hierarchy, in its most abstract and universal form (advance organizers), to be followed subsequently and step-wise by the more detailed and concrete tasks.

Likewise, Miller, Galanter, and Pribram (1960) have emphasized the "cybernetic hypothesis", and drawn on it to generalize the TOTE pattern, which describes a hierarchical organization underlying behavior. Hovland (1960) in his studies on human thinking and computer simulations, and Newell, Simon, and Shaw (1958), in their design of the "Logic Theorist", have emphasized and demonstrated the need to specify not only the prior information a subject possesses, but also the structural sequence of steps (algorithm) by which he uses the attained information in order to solve a problem.

Bruner (1963) and more recently Ebel (1969), have emphasized the notion of structure in the recall of meaningful knowledge. Ebel has theorized that the essence of achievement (mastery) is the command of a structure of knowledge. Minsky (1970), Suppes (1964), and Kochen (1970) have argued that helping students learn means helping them build cognitive models (structures) of their encountered environment. The process of building these models involves step-wise heuristic procedures.

More recently the study of comprehension was approached by us (Kochen, 1970; Kochen and Badre, 1973a; Kochen, Badre and Badre, 1973; Kochen and Badre, 1973b; Badre, 1973) by evaluating the generality of questions posed by human learners. This was originally conceived in the context of a novel approach to learning that stresses the formation, revision, and use of internal representations in the learning process (Kochen, 1971). A representation is akin to a model. It is a set of interpreted sentences or hypotheses in an internal language which enables a learner to recognize, formulate, and cope with an ever increasing variety of traps and opportunities in his environment.

In contrast with models in which hypotheses are selected from a fixed set according to a Markov process (Trabasso and Bower, 1968), this approach stresses the formation of and shifts in a set of logically connected hypotheses. A hypothesis is a proposition expressed by a well-formed English sentence together with an associated "strength of belief" and a degree of saliency. When a set of highly salient hypothesis is inconsistent, contains glaring gaps, or is of low weight, in order to remove these defects, the subject will be motivated to inquire by forming and using hypotheses. The answers should help him cope with the task.

Another line of research that has contributed to question of comprehension in learning has been that concerned with problem-solving. Gestalt and organization theories (Tulving & Donaldson, 1972; Kohler, 1926) are concerned with how, for example, a chimpanzee acquires the "insight" to join 2 poles for reaching a banana that is beyond the reach of one pole. Psycholinguistic and information processing theories (Carroll and Freedle, 1972) on the other hand try to account for how people obey verbally stated commands, such as "Invert the match-stick sketch of the cocktail glass so that the olive is outside by moving just two sticks" or "substitute numerals for letters in SAM + JIM = BILL". Organizational theories dealt primarily with episodic memories (Tulving, 1972) which receive and store information about dated episodes and temporal relations between them. Psycholinguistic theories deal mainly with semantic memories, such as thesauri, which are necessary for the use of language. To our knowledge there has been no extension of these theories beyond concern with memory to processing, and to synthesize episodic and semantic approaches.

There are at least two assumptions that seem common to the above-cited research: (1) that learning-behavior is most efficient when it is step-wise and hierarchical; (2) that there is a cognitive structuring, mental modeling of transmitted knowledge as it is assimilated by the learner. It is in this context of sequential and cognitive structuring that we perceive the role of the proposed research. For, our research hypothesis states that the process of comprehension involves a sequence of necessary steps, the last of which entails organizing knowledge into a structure.

This study differs from the above-cited researches in two respects: (1) It aims at specifying the assumptions of sequential and cognitive structuring as behavioral criteria for comprehension. Neither structural

learning nor educational-research literature on comprehension (Carroll, 1969; Otto, 1969) deals specifically with comprehension-behavior; (2) It focusses on question-asking rather than question-answering behavior as the medium of observation.

In sum, the key question we asked here was how people recognize and formulate problems, and how their ability to do this relates to their problem-solving performance. This question is important because it leads to results of practical value in meeting a great need in American schools.

The need is this. Americans are exemplary problem-solvers. Technologists have often developed "solutions" and search for the problems. An important cause of our collective failure to recognize and cope with many of the well-known real problems we have recently begun to sense (and also to become preoccupied with problems that may not correspond to real ones) may be that we were educated, from the first grade through graduate school, to solve problems someone else formulated for us rather than to recognize and formulate problems by and for ourselves. Eighth-graders, for example, become proficient at solving a problem like "How old is Joe who is twice as old as Jim and whose age added to Jim's gives 30?", or a trickier one like "A wrapped gift costing $1.10 costs a dollar more than the wrapping. How much is the wrapping?". The first example typifies what is commonly found in the texts, and students might justifiably ask where in real life such a problem - or even one like it - would ever occur. But do they systematically learn that the two examples correspond to mathematically very similar problems, and could they recognize a real problem as similar? The second example could actually be transformed into a scenario, a PLS, by motivating the student to want to know the price of the wrapper alone. He might then be much more motivated to learn algebra in the classroom or the text and see it quite differently. Moreover, he could actually use his school-learning in life-situations.

2. Methods, Procedures, and Results

The design of all our experiments was such that subjects were brought into the experimental situation lacking in all of the pertinent information they needed in order to solve the problem. In every case the problem was ill-structured and the problem-statement was ill-defined. The subject could learn or gain new information only if he asked the proper YES-NO questions. Thus the medium for acquiring information was partially fixed. This enabled the experimenter to observe and analyze the "search" and "thinking" strategies, (formulation and verification of assumptions, as well as, the logical inter-relation of verified assumptions in the process of comprehension), independent from an analysis of the medium (in this case, the question-asking method), in which the search strategy occurs. This means that we infer the subject's thinking process - the steps involved in the process of comprehension - by looking at the sequential structure of questions asked; i.e., the logical relation of one question to the next. The following is a report of each of the experiments, procedure and results.

EXPERIMENT I

In this experiment we measured comprehension levels on the basis of question-quality.

Problem-Solving Representation and Question-Quality

Problem-Formulation

In our experiment, a human subject enters a room where he sees an array of inverted cups on a table. He is given next to no instructions but he knows that he is to be paid for serving as a subject in a psychological experiment during this designated hour. Subject senses that he is, in this environment, in some problem-state, but he still has a very diffuse, vague "image" of his need. His initial internal representation might be a sentence he said or thought to himself like: "Here is a room with a rectangular array of upside-down cups, with two people who expect me to do something like, perhaps, turn cups over, play a game, or ask questions". With such a representation, the subject may ask questions or turn over cups, and we call such behavior coping with the situation or problem-state, rather than problem-solving.

If the subject copes successfully, he asks questions or turns over cups which sharpen his representation of the problem-situation. He might, after some exploration and conversation say or think to himself: "So they want to see how I choose the cups that are likely to hide dimes". If he does not cope successfully, his hypothesis about the nature of his task may not be any more precise or close to the mark than it was at first. The subject has probably coped even more successfully by the time he forms a hypothesis like "I think the dimes are all in the far lower right corner of the array".

At such a point the subject can formulate a well-defined problem-statement, such as "How can I determine the spatial pattern according to which the experimenter has distributed dimes under the cups by asking questions or looking under one cup at a time?". When the subject has reached this point, we say that he has completed the problem-formulation stage. Completion of the problem-formulation phase is ascertained when subject asks a question or a sequence of questions containing a well-defined problem statement of the relevant problem.

Problem Solving

Subject now enters a problem-solving stage. This may have two sub-phases. The first is unplanned information-gathering or random sampling. He may, for example, pick seven cups at random and ask if they conceal dimes. The second is more systematic question-asking or search, based on a specific hypothesis about how the dimes are arranged. Thus, if four of the seven cups he investigated in the first sub-phase contained dimes and all four were in the same row, he might, as his first question of the second sub-phase investigate a cup in that same row.

A representation is useful in both the problem-formulation and the problem-solving stage. It enables us to interpret incoming information as knowledge, to structure this knowledge, and to analyze a problem-statement into sub-problem statements and relate these to one another. Successful problem-solving behavior (second stage) is exemplified, in our view, by two properties:

(a) The problem-statement is structured into a logically connected sequence of other problem-statements so that the solution is a consequence of the solution to the problems in this sequence.
(b) A problem-statement is regarded as a special instance of a more general class of problem-statements to which a unifying pattern of finding solutions applies. (Polya, 1962; Gagné, 1970).

Representations

We view a representation to be specified by an internal language and a structured set of sentences within it. To specify an internal language is to specify a vocabulary which denotes constants, variables, predicates, quantifiers, functions, and rules for forming well-formed propositions as in predicate calculus; this specifies a (generally infinite) set of possible well-formed strings we have informally called "images". They can represent events, states, laws of the environment, to a greater or lesser degree. To specify a structured set of sentences is to specify certain of these well-formed strings as axioms, others as hypotheses, and theorems; to specify also: special rules of inference; certain well-formed strings as questions and a set of logical connections among the questions. Altogether, such a structure is not only a set of formal strings but an associated system of interpretation, in the sense of model theory. Thus, to each well-formed string is associated an interpretation in a universe of discourse which can be compared with a corresponding state in the external environment.

Question-Quality in Problem Formulation

How a learner (L) represents this task-environment to himself, is, we believe, revealed by the questions he asks. If L is uncomfortable with the irrelevance, or imprecision of his representation, he will tend to ask "groping" questions, such as "Is there money under some cup?". During the later problem-solving stage, when L has a more relevant and precise representation, he will tend to ask specific and generic, yet precise and relevant questions.

Three aspects of a representation, as revealed by corresponding three qualities of a question, are of interest for this study: degrees of relevance, precision, specificity. A question is highly specific if it yields information about a single noun-object such that the information cannot be generalized to any other noun-object or element in a class of noun-objects, then it is highly unspecific or generic. A question is relevant if it reveals information about the experimenter's problem-state and irrelevant if not. If the predicate of a question can be sharply defined, it is a precise question, otherwise it is fuzzy.

During the problem formulation phase, greater priority is given to degree of relevance than to degree of precision. An irrelevant but precise question, like "Must all chairs in this room stay fixed?" is likely to elicit less information than a relevant but imprecise question, like "Is there only one dime towards the right end of the bottom row?". Degree of precision is given higher weight than degree of specificity. A precise but specific question is informationally more useful than a generic but imprecise one because it does not leave the subject undecided about the exact subset of information he may use.

Of two questions which are equally precise and specific, the one which is more relevant to the representation of the problem that the experimenter has in mind is of higher quality. Without achieving relevancy, problem-formulation could never succeed. Of two questions which are both relevant and specific, the one which is more precise is of higher quality. The interpretation of the answer to a precise question is more useful than that for an imprecise one because it is less ambiguous, more unique. But which is the better of two questions that are equally relevant and precise but differ in degree of specificity? The quality of a generic question should be greater because it has greater potential for reducing uncertainty.

Suppose we encode the quality q of a question as a three-bit number, (s, p, r). Here s denotes degree of specificity, which is 1 if the question is generic, 0 if specific; p denotes degree of precision with 1 if the question is precise, 0 if not; and r=1 if the question is relevant, 0 if not. This encoding partitions the set of all questions into eight possible quality-classes, ranked in Figure 1 (next page).

Question-Quality in Problem Solving

In the problem-solving stage, the criteria for question-quality are different. Though this is not germane to the main point of this study, it relates to the notion of specificity. We are dealing here only with "pattern-specificity", or questions such as "Are dimes distributed according to a letter of the alphabet?".

As soon as a subject imagines possible dime-distribution patterns, he will select a representation that uses a conceptual repertoire corresponding to terms like "rows", "columns", "under every other cup in a row", et cetera. With 40 cups, there are 2^{40} possible patterns; even if each pattern could "flash" through the subject's mind in one nano-second (more than the speed of a computer) and if he could in that time decide whether or not to entertain questions based on that pattern, it would take him about 30 hours to go through them all. Of course, these patterns are aggregated, classified into a few major classes, each characterized by certain predicates chosen from the conceptual repertoire in a system of representation. Even with a given vocabulary of such properties, the 2^{40} patterns could be classified in many, many different ways, some of much greater value for "efficient" problem-solving than others. Furthermore, changing the vocabulary of predicates - modifying the entire system of representation - can have a dramatic effect.

The quality of questions depends on the context of other questions the subject is asking, and on the representation on which they are based. Suppose that a particular representation admits of n_o possible hypotheses about the pattern for distributing dimes. A first question, Q_1, if answered "Yes", eliminates n_{1e}^{Y} of these n_o hypotheses, and n_{1e}^{N} if it is answered "No". Let $N_1 = n_o - wn_{1e}^{Y} - w^1 n_{1e}^{N}$. Here w and w^1 are weights, like ½. A second question, Q_2, would eliminate n_{2e}^{Y}, n_{2e}^{N} for Yes and No answers, respectively. Again, $N_2 = N_1 - wn_{2e}^{Y} - w^1 n_{1e}^{N}$ measures the number of remaining hypotheses. We repeat to get N_k as a measure of the number of remaining hypotheses.

1 1 1	Best: generic, precise, relevant
0 1 1	specific, precise, relevant
1 0 1	generic, imprecise, relevant
0 0 1	specific, imprecise, relevant
1 1 0	generic, precise, irrelevant
0 1 0	specific, precise, irrelevant
1 0 0	generic, imprecise, irrelevant
0 0 0	Worst: specific, imprecise, irrelevant

Figure 1

A rank ordering for three-bit encoding of question-quality

We now assume:

(1) If a hypothesis corresponding to the actual pattern k is in the representation, then it is among the N_k that are not eliminated.

(2) If not, and the subject eventually learns the pattern, then a shift to another representation which includes the corresponding hypothesis must have occurred.

A certain question may fail to eliminate any hypothesis in a given representation, no matter what the answer, because it does not apply to this representation. It may, however, eliminate the entire representation. Ideally, it eliminates all but one of a set of representations from which we (the observers) consider the subject capable of choosing. This would be a perfect question early in the sequence. We would recognize it as such only later, after the subject has asked more questions that reflect how he eliminated all but one representation. A question which elicits a contradiction as its answer is good because it eliminates a representation. Likewise, a good question is one which brings out the incompleteness of a representation.

Once the subject appears to be locked into a representation - at least for some time - the perfect question at the end of a partial question-sequence of k questions is one that makes $N_k - N_{k+1}$ as large as possible; ideally, it reduces N_{k+1} to one, with one hypothesis corresponding to the correct pattern. Thus, a question, the answer to which implies the answer to numerous other questions is good because it will make N_{k+1} very close to one.

A question is good, at the problem-solving stage, to the extent to which it comes close to the above ideal questions. We indicate how to specify question-quality operationally at the problem-formulation stage, in the next section.

METHOD

Subjects and Procedure

Eighteen University of Michigan freshmen were chosen from among paid volunteers. Of the 18 only 14 were used for the reported experiment.

The other four were given a slightly different task where the object was to discover whether they would act to maximize their earnings on the basis of knowing distribution probabilities. A total of 64 cups were arranged in 32 columns. Each of the four subjects was told that the following information was true: Distributed randomly,

		First row	Second row
4 columns	=	no dime	no dime
4 columns	=	dime	dime
8 columns	=	dime	no dime
16 columns	=	no dime	dime

Subject was then told to pick any eight of the 32 columns and by asking Yes and No questions attempt to maximize his earnings. The outcome showed that none of the subjects utilized the information they were given.

Also because of the pre-specified nature of this task, the category of specific questions predominated the subjects' protocols. When later they were given the main experiment, they continued to ask specific rather than groping or generic questions.

Each of the remaining 14 subjects entered the experiment to be faced with five separate arrays of inverted, opaque cups. In four of these arrays dimes were placed under cups to form a regular pattern. In three of these four arrays, dimes were distributed (different in each array) according to rows and columns. Under the fourth array, coins were regularly distributed about the perimeter. The fifth array constituted a random distribution of coins.

The systematic change in the nature of pertinent information was tied to the hypothesis that changes in the environment will cause shifts in representation which will correspond to shifts in questions asked. At first, questions would reflect a model of the previously learned environment which no longer held. But as the representation of the new environment improved so would the question-quality and learning rate. Three of the arrays required similar representations, but different from the other two. Counterbalancing was used in the order of presentation within the three arrays.

The instructions given each subject were:

(a) You may ask questions to get information.
(b) The only allowable question is one to which a Yes or No answer can be given.
(c) If you lift a cup it will cost you a nickel.

If L discovered a dime under some cup he found it to be his. The use of the money was intended as a motivating factor. L would soon set himself the goal of finding all and only the dime-hiding cups. This in turn would induce him to "imagine" the possible patterns of distributing dimes under cups, one of which the experimenter might have picked. This mental image of the possible patterns is a representation of the kind in which we are interested.

A run is a period during which the subject can ask questions, look under cups concealing dimes in a fixed pattern until he has collected all the dimes or spent five minutes, whichever occurs first. The pattern by which dimes are distributed constitutes a problem-state or task. In the next run, the subject is faced with another task of the same kind. Each subject underwent five runs during his experimental hour.

Data Collected

The exact protocols of questions asked were recorded for all five runs for each of the 14 subjects. The experimenter and an observer were alone with the subject during the entire hour. When an ambiguous question was asked the experimenter said that this could not be answered Yes or No.

The observer tape-recorded all questions asked by the subjects and all answers given by the experimenter. All actions, money transactions and the time these took were also recorded. Protocols of each session were later transcribed and typed. A protocol for a sample is given in Table I (see next page).

TABLE I

A sample question-sequence protocol and coding

Question-Sequence

S By asking questions am I supposed to come to conclusion about something?

E Yes

S Do I earn money by coming to the right conclusion?

E Yes

S Does the conclusion concern the things under the cups?

E Yes

S I will uncover a cup....there is a dime under it.

S Is the object of the experiment to discover dimes under cups?

E Yes

S Is there a dime under this cup? (Points to a specific cup.)

E Yes

S Are the dimes distributed in a regular pattern under every other cup?

E Yes

TIME

Earnings = 10 (# finds) + 40 - # questions + 5 (# lifts)

10 (1) + 40 - 6 + 5 = <u>39¢</u> (Note that is the amount for first run. To get total earnings, this must be added to earnings in other 4 runs.

Question-Sequence Quality for the Formulation Phase =

$$\frac{\boxed{10100} + \boxed{00100} + \boxed{11001} + \boxed{01111} + \boxed{00011} + \boxed{11111}}{6} = 10$$

Data Analysis

The data were then coded as follows. Instead of encoding q as a three-bit number, we used a five-bit number, allowing three bits for specificity. Conceptually, q is an n-bit number. However, an inspection of subjects' protocols for this task revealed that the number of noun-objects per question reached a maximum of three; hence, the three bit code for specificity. This required us to extend the ranking scheme of Figure 1 into the scale of Figure 2 (next page).

Given an example-question such as "Must all chairs in this room stay towards the middle east half of the room?", how could one assign a five-bit code to it. On the relevancy dimension, a coder looks for words and strings in the question that are necessary in describing the task-environment. For this particular task, words such as cup, dime, pattern, and array would satisfy the criterion for relevancy. This makes the example-question irrelevant as it does not contain any of these words. On the precision dimension, we look for the well-definability of predicates. If the predicates are well-defined and clear, such as in "are the number of cups greater than twice the number of dimes?", then the question is precise. If a predicate does not define a sharp boundary such as "are the number of cups much greater than the number of dimes?", then the question is imprecise. In the example-question, the predicate "towards the middle east" renders the question imprecise. On the specificity dimension, a coder looks for the inflections and referents of noun-objects. If for instance an s, indicating number, is affixed to the noun-object or if the noun-object has more than one referent, then it is generic; otherwise, it is specific. In the example question, the noun-object, "chairs" is generic; "room" in both instances is specific. Thus the example-question can readily be assigned the code $\boxed{10100}$. This encoding corresponds to rank number three in Figure 2.

If there are less than three noun-objects in a question, then the empty noun-object cells are relegated to an encoding which places them in the lowest possible rank. The reason is that the greater the number of noun-objects associated with a given question, the greater the amount of information elicited.

Finally, how do we compute the quality of the sequence of questions associated with the problem-formulation phase for a single subject? We determined that it takes an average of ten questions to reach the problem-solving stage - the point at which the problem statement is formulated by the subject and discernable in his vocabulary. Thus the quality of the sequence of the first ten questions = $\frac{\Sigma q}{10}$.

Problem-solving performance for each run was measured by L's total earnings at the end of that run. The total earnings are in cents, 40, the subject's initial capital, plus 5 x number of cups lifted (net gain of 5¢), less the number of questions (1¢ each).

A correlation measure was computed for sequence-quality of the problem formulation phase and total performance. We chose this measure as we were merely interested in an estimate of the degree of closeness of the two variables. The elementary nature of the data and the task did not warrant a strong technique such as regression analysis. Thus, the analysis used says nothing about either the shape of the curve or the predictive power of either variable.

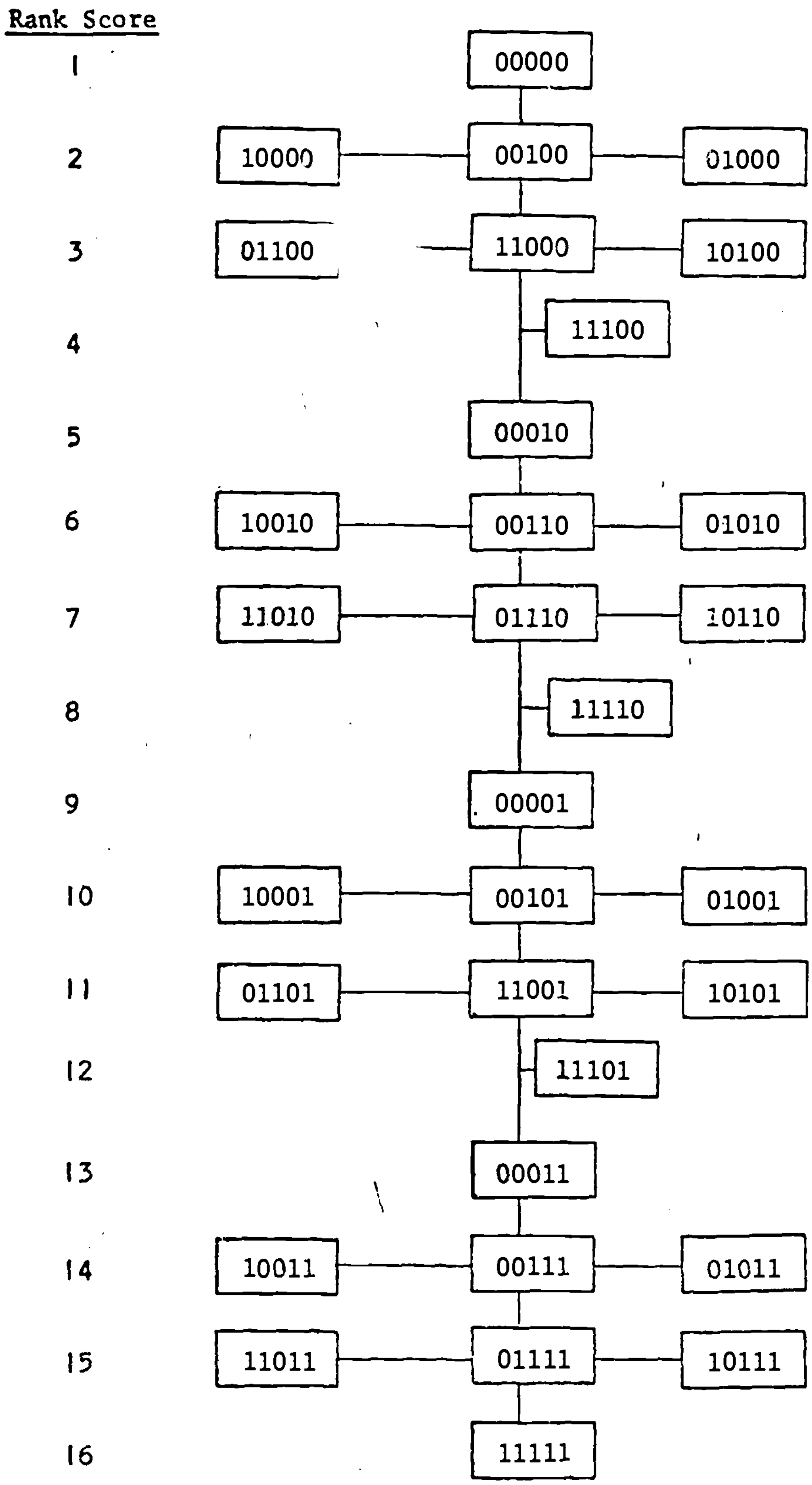


FIGURE 2

A scale for measuring question-quality during problem-formulation phase.

RESULTS AND DISCUSSION

The main result is embodied in Figure 3 (see next page). This shows performance as a function of question-quality. The correlation coefficient is .74. The relationship between the two variables is significant at t(12d.f.) = 3.81, $p < .01$. The result indicates that there is a relation between asking good questions during the problem-formulation stage and subsequent problem-solving performance. We cannot, of course, infer from a correlation that this relation is causal. Nor are we saying that improvement in question-quality during the problem-solving stage did not contribute.

The finding that improvement in question-quality during problem formulation is accompanied by improved performance is in line with demonstrations in artificial intelligence research (Amarel, 1971) that the way a problem is formulated is highly related to the efficiency with which the problem will be solved. That is, the process of finding a solution depends on the choice of an appropriate representation during the initial part of the problem coping process.

This finding suggests that it may be possible to predict systematically problem solving performance from a problem solver's formulation vocabulary. However, a more comprehensive experimental undertaking, using different problems is necessary for a good test of this hypothesis.

The results in Figure 4 (page 15) show that, on the whole, problem-solving performance improves from one problem to the next. Improvement in runs one, two, and three is constant. But as we move from the third to the fourth problem, a slight decrement appears in the graph. The most likely explanation for such a trend is that while the first three problems were similar, the fourth and fifth introduced new elements into the situation which required shifts in representation. As he began on the fourth problem, the subject had not yet experienced and therefore learned to expect changes in the problem-state which required shifts in his representation of the problem. This caused a delayed shift and therefore a slight decrement in performance. The delay in shifting was quite evident in the subject's vocabulary. The words he used at the start of the fourth problem pointed towards a representation of the previous problem. But as soon as L became aware of changes in the specifics of the problem, a shift in vocabulary indicating a shift in representation occurred. The trend from the fourth to fifth problem indicates that L may have begun to anticipate changes in the problem-state and meet them with needed shifts in representation.

If we look at the data in terms of the partitioning of total time for the five runs on the basis of the three problem-coping phases of formulation, sampling, and solving, we observe further support for the findings of Figure 4. Figure 5 (page 16) shows, as expected, that at the beginning of the experiment, the formulation of the problem occupies most of the subject's time. But as he moves into the second run, formulation decreases and gives way to sampling and solution times. During the third run, formulation phase disappears to be replaced by minimal sampling and solving predominance. However, when he moves into what appears to be the same but in fact a new problem in the fourth run, the subject begins to sample, but because this gets him nowhere, he reverts to the formulation phase as indicated by his vocabulary. This reversal causes a delay which accounts for the decrement in problem-solving performance between the third and fourth run. His vocabulary pattern

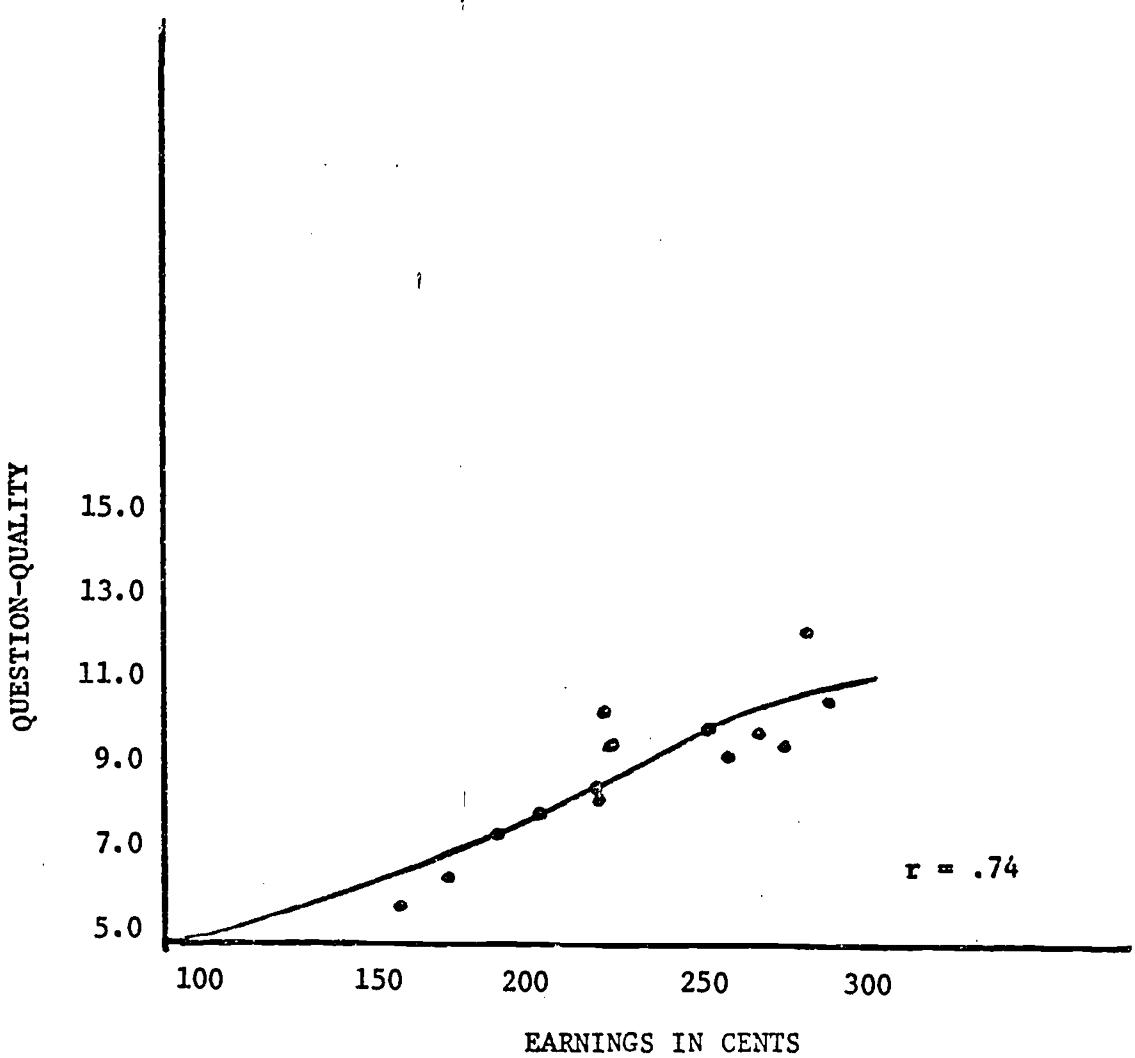


FIGURE 3

Correlation of problem-solving performance as measured by earnings with question-quality in problem-formulation phase.

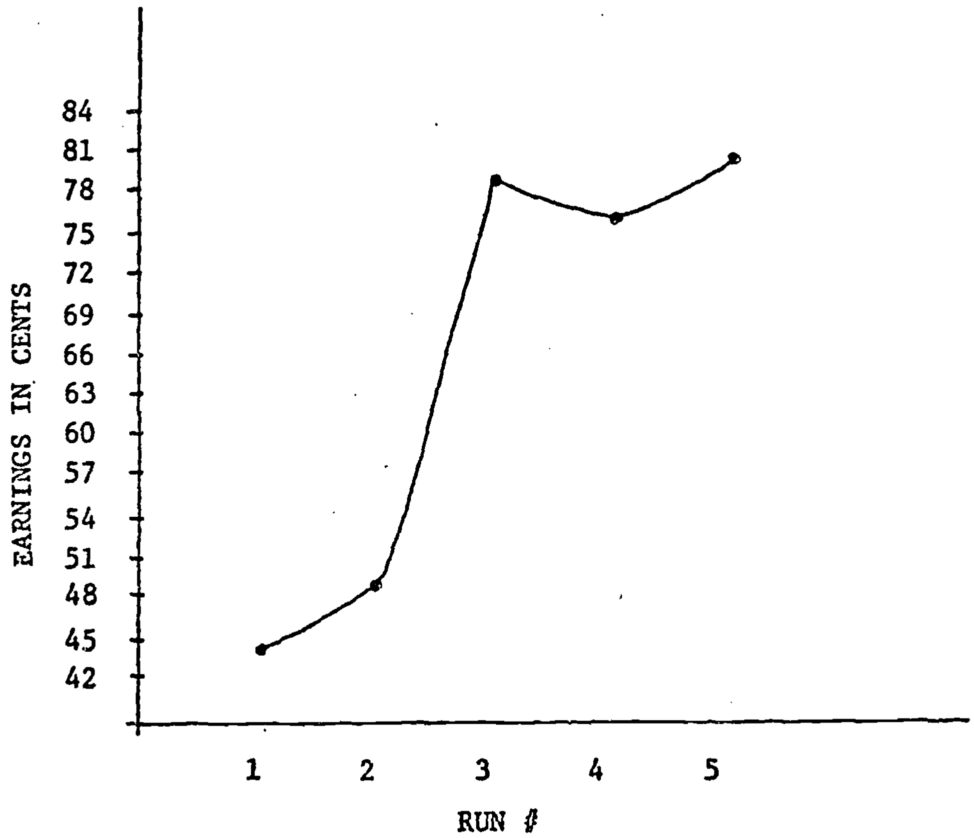


FIGURE 4

Trend in problem-solving performance in runs one through five.

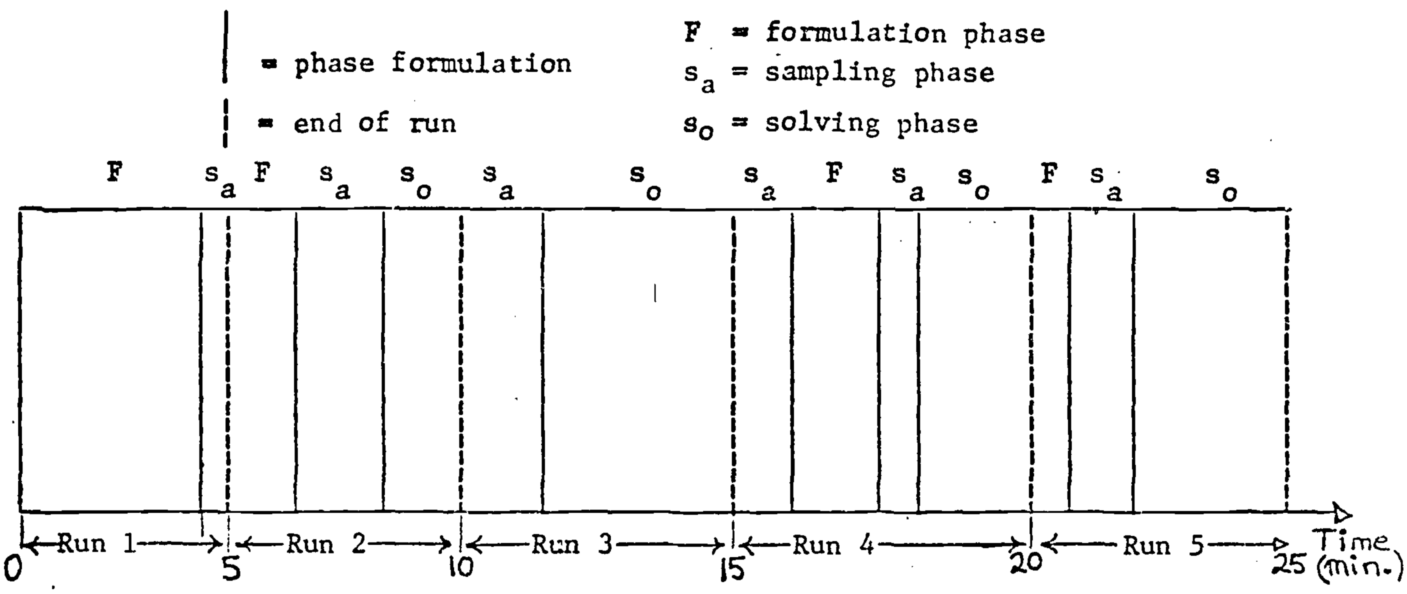


FIGURE 5

Phase occurrence on the basis of question-vocabulary against time by minutes from zero to twenty-five. (Note: The experiment was designed so that each lasted five minutes. The time of the phases indicated in this figure are estimates based on the average number of questions asked.)

in the fifth run indicates that learning of representational shifting occurred because of the fourth run experience.

Had we found good problem-solving without good questions in the formulation stage, this could have been due to: (1) shifts of representation just the same, but not expressed as verbal (questioning) behavior as might be the case for chimpanzees; (2) problem-solving performance on our task being governed predominantly by perceptual or rote memory processes but not by cognitive maps or internal representations, which might also be the case for chimpanzees or people with a lot of experience with tasks of this kind; (3) defects in our method of measuring question-quality.

Had we found good questions without good problem-solving, this might have been due to: (1) inability to utilize good representations; (2) inability to register, maintain, or retrieve relevant memories long enough, if memory plays an important role; (3) inability to form coherent questions, as in aphasia or other disturbances of linguistic performance; and (4) the above possible defects of our method.

It is therefore not trivial or obvious that question-quality in the formulation-stage is correlated with problem-solving quality, because this lends credence to the psychological reality of "internal representations" that we take for granted in fellow humans.

Our experimental results lead us to suggest that question-asking behavior at the formulation stage is a good indicator of the overall problem-solving performance; that certain question-types occur more frequently at given stages of the problem-solving experience than others. At the start, a problem-solver's questions are of the groping and irrelevant type. As he progresses, his questions become more generic and more precise. As questions get better, so does problem-solving performance.

EXPERIMENT II

In this experiment, the main question we raised was: How does a fifth grade child achieve a sophisticated level of mathematical comprehension, and how do we detect, measure, and improve such achievement? Formulating a mathematical story problem is more difficult than solving one already formulated mathematically. A mathematical story problem is a verbal problem such as: Drove 3 hours. Average speed 65 mph. Then drove 3 more hours. Average speed 55 mph. Traveled how far? (Eichholz and D'affer, 1964). Formulation of a problem involves a degree of structural organization of certain "knowns" and their relations; the ability to achieve such structural organization is what we call comprehension. The best indication of whether one has achieved comprehension of the problem is if he could formulate it when initially given no information.

In this experiment we pose the central point of how to achieve comprehension through the problem-formulation question in terms of problems requiring mathematics. We report an experimental technique of testing or assessing whether a problem has been recognized and formulated; it uses questions asked by a subject as the basic data (Kochen and Badre, 1973). We extend to theory to suggest how a computer program could generate questions in a problem-formulation environment. We also report a technique for improving the performance of children in grades 4 and 5

on tasks requiring them to recognize and formulate problems; to achieve a state of comprehension; it resembles the game "Twenty questions".

The central point is an experimental verification of the hypothesis that a large population of children can be taught to improve in recognizing, describing, and understanding some real situations as ones requiring problem-statements which resemble story-problems in arithmetic texts used in grades 4 and 5. In other words, there exists an environment that stimulates the formation of internal problem-statements (hypotheses) which manifests itself as observable questions.

Question-Generation

Improved inquiry modes can improve problem-recognition. To make this more precise, we ask how we would program a computer to recognize problems and to ask questions. Complete rigor would demand a very lengthy exposition. Hence we only sketch some central ideas.

To start, we have to specify the input to L, the program. This input is to mirror, for example, the physical stimuli which would motivate a given traveler in Houston to be concerned about whether he could drive to New Orleans in 6 hours; they are also answers to questions. Then we must specify the output, which is mainly questions such as "How do I drive from Houston to New Orleans?" and actions such as driving. We must also sketch what L has in storage prior to input, and the general outline of the algorithm according to which it processes the inputs and generates outputs.

Input to L: This is a state s of L's environment. Suppose it to be a string of several variables, s_1, s_2,..., each of which ranges over some dimension of state-space, and varies with time t, measured in hours. For simplicity of exposition, suppose that $s_1(t)$ is the name of a town on the route from Houston to New Orleans (or ϕ to denote no town) where L might be t hours after L after Houston. That is, s_1 ranges over all the town-names along the route. Initially, $s_1(0)$ = Houston. The state $s_1(t)$ = New Orleans with $t \leq 6$ is the only reward state. Let $s_2(t)$ be an answer to the last question L asked prior to t. Let $s_3(t)$ be an extraneous instruction, verbal stimulus or datum, a question to be imitated. This ranges over a specified set of sentences.

Output of L: This is an action a, from L to the environment. Suppose it to be a string of several variables, a_1, a_2,... . In this case let $a_1(t)$ denote the imagined speed (mph), say -90 to 90, where a negative number means heading back to Houston. Another output variable is the decision:

$$a_2(t) = \begin{array}{l} \text{drive to New Orleans} \\ \text{don't drive} \\ \text{defer decision.} \end{array}$$

Yet another variable is $a_3(t)$ which ranges over the set of possible questions L could ask.

In Storage Prior to Input: This includes a production system for questions and answers. Formally, this is specified by a terminal vocabulary, V_T e.g. {Houston, how, far, from}, a non-terminal vocabulary V_N, 2 special symbols used to start generation of questions (Q) and answers (A), a set of rewrite rules R. L also has, in storage, a list of rules for recognizing (parsing) answer sentences and for translating them into

an internal representation (Kochen, 1969). Most importantly, L has in storage a set of hypotheses. These are statements in an internal representation exemplified by: H1 = "For all t, if $s_1(t)$ = Houston and $a_1(0)$ = 40 mph, then $s_1(t+1)$ = Austin; weight .8, saliency 1", and H2 = "If $s_1(t)$ = New Orleans and t = time the Mardi Gras in New Orleans starts, value is high; weight 1, saliency 1".

Some hypotheses, such as "If I go faster than 90 mph, I am likely to cause an accident or receive a fine, either of which I dislike more than I like speeding. Weight = 1, Saliency = 0," are stored in long-term memory. Other hypotheses, such as H1 and H2 may be in L's short-term memory for the few seconds or minutes in which he is recognizing the problem and making a decision. All the hypotheses in short-term memory (STM) have high saliency.

Algorithm: The main function of L is to select outputs which maximize the expected value of a future state. First L registers the input by parsing and translating it if it is a sentence, classifying it if it is not. The initial input in the above example might be: $s_1(0)$ = Houston, $s_3(0)$ = "The Mardi Gras starts in New Orleans at t = 6". This input is classified as an opportunity-state by matching the phrases "Mardi Gras starts" and "New Orleans" in a stored hypothesis such as H2, which has a high value. If L could not parse an input sentence or if the sentence has a word not in V_T, L's generates a stylized question: "What does ___ mean?". It processes the answer by forming new hypotheses and adding them to the store.

Secondly, L searches its store (a program for this has been implemented in SNOBOL4) for useful hypotheses. A useful hypothesis is one that helps L choose and attain a valued "goal"-state. It selects these from short-term memory with a probability proportional to the weights of the hypotheses in STM. Both H1 and H2 might be retrieved in the above example because H1 shares with the input the term "Houston" and H2 shares "New Orleans". Ideally, L would like to find, besides H2, an hypothesis like "If $s_1(0)$ = Houston and $a_1(0)$ = 80, then s(t) = New Orleans for some $t \leq 6$". If that is present, the output is: the decision, $a_2(0)$ = "drive to New Orleans"; $a_1(0)$ = 80 mph; and $a_3(0)$ = no further questions. The environment now responds and the interaction continues. During the short time interval, $(\Delta t, 0)$ that decision $a_2(0)$ is made, a "within-representation, high saliency shift" (Badre, 1973) may have occurred in that the weight of a hypothesis containing a_2 = "don't drive" has increased while the weight of an hypothesis containing a_2 = "drive to New Orleans" has decreased.

If such an hypothesis is not there, L forms an hypothesis of the form: "(At)(AT)(Ay)(Ax)(Av), If $s_1(t) = x$ and $a_1(t) = v$ then $s_1(t+T) = y$, where the distance from x to y is $v \cdot T$." Once L has formed this hypothesis - particularly the underlined phrase - he has recognized and formulated the mathematical problem which must be posed and solved for L to make a rational decision. This indicates a state of comprehension. We must now sketch how L might generate evidence of this by asking questions. The implied questions are: "What towns are between Houston and New Orleans?" (formally, what is x such that, for $0 \leq t \leq 6$, $s_1(t) = x$).

What is the maximum speed between towns x and y? (What is v such that $a_1(t) \leq v$?) What is the distance from x to y? et cetera. When enough such questions are posed, L should be able to synthesize them into a decision $a_2(0)$. After observing all these questions as output, we infer that L has formulated the problem.

But how can L form such an hypothesis involving a product (and perhaps a sum, $v_1 \cdot T_1 + v_2 \cdot T_2 + \ldots$)? We assume that multiplication ($\cdot$) and addition (+) is in V_T, and that there are in storage general hypotheses of the form: "(Av)(An). If 1 unit of a property 1 is associated with v units of property 2, and n units of property 1 are chosen, then the n units are associated with n $\cdot$ v units of property 2". Such a general hypothesis is specialized, with the help of hypotheses that constitute a thesaurus, which has entries such as "Time is a property", "Distance is a property", "Hour is a unit", "Mile is a unit". The specialized hypothesis now is: "(Av)(An). If 1 hour of time is associated with v miles of distance, and n hours are chosen, then the n miles are associated with n $\cdot$ v miles of distance".

Where does the general hypothesis come from? Like all other hypotheses, it may be direct verbal input that is simply recorded; or it may be formed by imitating types of questions asked by another L which reflected the use of such hypotheses. It may also be the result of induction and generalization from other hypotheses in memory that is the heart of the algorithm in representation theory.

<u>It follows that an environment which provides inputs, such as questions reflecting hypothesis-formation processes to be imitated, can produce in L the formation of general hypotheses, and, from these, the formation of hypotheses that indicate recognition and partial formulation of a problem. A structured version of "Twenty Questions" may be such an environment. It is this hypothesis we test with a controlled experiment.</u>

<u>Hypothesis</u>

The first question of interest to us was: does the technique we specify for improving problem-recognition and formulation behavior work? We chose a simple experimental design to test this technique. We selected a random group of subjects, exposed half of them to our procedure and let the other half continue their exposure to the ongoing classroom methods of learning mathematical problem-formulation and then compared the difference.

More precisely, let T (for treatment) denote the set of subjects who were exposed to our procedure and C (for control) that set of subjects who were not. Let X_T and X_C denote the corresponding test scores for randomly chosen subjects from T and C. The null hypothesis is that the expected values, EX_C and EX_T are equal.

Let H be the time it takes a subject to form useful hypotheses when called for by a problem-situation either because he formed a general hypothesis, an algorithm that forms hypotheses, or because such hypotheses (or programs to generate them) were previously formed and stored for rapid retrieval. Let Q be the time it takes a subject to pose questions, the answers to which are necessary in coping. It is also plausible to assume that $H_T = H_C$ implies $Q_T = Q_C$, other factors being

the same. Finally, we assume the implication: $Q_T = Q_C \Rightarrow X_T = X_C$. Therefore, $X_T \neq X_C \Rightarrow Q_T \neq Q_C \Rightarrow H_T \neq H_C \Rightarrow E_T \neq E_C$. If the subject exposed to T gets a lower test score (faster problem-recognition and formulation) X_T than does an otherwise equal subject not so exposed then it takes the subject exposed to T less time to form useful hypotheses than the subject not so exposed, according to the above assumptions.

Subjects

The population about which we wish to generalize consists of children in grades 4 and 5 of upper middle class, predominantly white families in an American university city. From a group of thirty fourth and fifth graders, two equal groups of 10 each were randomly selected.

Improvement Method

The experimental "treatment" group underwent six days of training sessions. Each session was one hour long. The main thrust of these sessions was to get children to formulate mathematical story problems similar to the ones they encounter in their mathematics texts (e.g. Eichholz and D'affer). The children were specifically told that no attempt must be made to solve formulated problems. The trainer considered her objective to have been met when each child had achieved the formulation of six such problems.

In order to get the children into an "inquiry" and "problem-asking and quizzing" frame of mind, the first session was devoted to playing Twenty Questions, using various topics, e.g. cryptograms, hidden objects, guessing numbers, et cetera. The next session began with 20 questions about story-problems in the text book. Next each child was asked to formulate "for himself" a story problem similar to a specific one in the text book. The rest of the children were to guess it by playing twenty questions.

The next session involved using concrete objects, to stimulate children how to formulate verbal problems. An example of this was the use of a scale and two cars being weighed. The trainer formulated the first problem: "If the weight of the big car + the weight of the small car is equal to 94 grams, and the weight of the big car is equal to the weight of the small car + 24 grams, what is the weight of the big car?". Then children were asked to formulate two different problems each using the same or different objects.

The rest of the sessions were conducted similarly. Real-life objects and situations such as "customer and shopkeeper", "calculation", and "rate problems" were used. A pocket-size electronic calculator was used to do arithmetic at the children's request. This procedure continued until each child had formulated six story problems.

Testing

All 20 randomly selected children were tested 10 days after the training sessions started. Each child was tested individually for about 30 minutes. Testing took 2 days. Like the training sessions, the tests took place on the premises of the school to which all the children went. The testing procedure is detailed next.

Assessment and Test Construction

Before we can test the hypothesis that the ability to recognize and formulate certain problems improved, we must have a way of assessing that ability. To this end, we devised a three-way test, covering algebra, geometry, and arithmetic analysis, which are the traditional main divisions of mathematics taught in grades 4 and 5. In each task, we were testing performance ability for recognizing a situation as one requiring certain mathematical operations.

A description of the three-tasks test follows.

Set-Up

Subject entered the test room to find 3 tables, D1, D2, D3, and 6 chairs. Each table had associated with it 2 chairs facing each other on either side of the table. The experimenter, E, sat in one chair facing S (this corresponds to what we called L - for Learner - earlier), who was sitting in the other chair. In a different part of the room, an observer-coder sat with a pen, a paper, and a stop-watch.

D1 had on it: (a) a cardboard sheet 26" x 23"; (b) three cardboard houses on (a) labeled MacDonalds, School, and Bank; (c) the cardboard houses were placed on corners of (a) at three different intersections of three main roads (drawn on (a)); (d) 3 signs placed at the three different roads: Sign 1 read: "Speed limit 2 seconds per inch, distance to MacDonalds is 18 inches"; Sign 2 read: "Speed limit 1 second per inch, distance to School is 12 inches"; Sign 3 read: "Speed limit is 3 seconds per inch"; (e) a car placed at upper right corner of board.

D2 had on it: (a) 5 boxes that ranged in volume from 260 to 630 cubic inches; (b) 360 1" polystyrene cubes.

D3 had on it: (a) 3 spools of orange, white and black wire; (b) price tags - "White wire is 11¢ per inch", "Orange wire is 13¢ per inch", "Green wire is 7¢ per inch".

Tasks

The items on D1 were associated with Task 1, T1; D2 with T2, and D3 with T3. There was a sign on each table which read: "Keep asking questions until you know what to do". E told S that "this is a game that requires the use of some mathematics". Then, he gave S instructions that varied with each task. The instructions in every task began: "I would like you to make up questions for me to answer. The answer should make it possible for us...

Task 1 - to figure out how long it takes a car traveling at maximum speed to get from where it is now to the Bank.

Task 2 - to choose one of those boxes that will exactly fit those cubes as they are placed near and on top of each other in the box.

Task 3 - to sell me some of these wires. Now I am the customer and I want to order from you 10 inches of white wire, 12 inches of orange wire, and 7 inches of green wire.

After S was seated, E told S this was a game and that he had in mind three tasks involving the objects on the three desks. He instructed S to ask E any questions, which E promised to answer truthfully and which were to help S guess what task E had in mind. E then proceeded to answer

the questions asked by S, responding to questions like "What am I supposed to do?" with "That is what you are to figure out", or to "In which box will all the cubes just fit?" with "I can't tell you directly, but will answer another question that might help you find out". This continued until either one-half hour was up or S had asked questions indicating that he had figured out the 3 tasks in a way that was equivalent to the following three statements:

1. The time (in seconds) for the car to go from the start to the Bank is the speed allowed on the Start-MacDonalds stretch, in inches/second times the distance (in inches) of that stretch plus the speed allowed on the MacDonald-School stretch times the length of that stretch plus the speed allowed on the School-Bank stretch times the length of that.
2. The box I should pick if E gives me all his cubes and I want to just fill the box is one whose volume is equal to the number of cubes, and the volume (cubic inches) is the product of the length, width and height of a box (all in inches).
3. The amount of money I should get for delivering the order is the price of the white wire, in cents/inch, times the length of white wire I sold (in inches), plus the price of the orange wire times the number of inches of orange wire, plus the price of the green wire times the amount of that.

Data Collection

The observer, O, recorded the time, to the nearest second, between the termination of E's instruction or response to a question and the onset of S's next question for each question asked or comment made by S. E also recorded, for each question, whether it contained words on a checklist. For Task 1, for example, the checklist contained such words as "time", "speed", "times or multiplication", "length or distance", "plus or addition", et cetera. Near-synonyms were also checked. E also judged when S seemed to have asked a sequence of questions that, in their totality, indicated that S had recognized and formulated a problem equivalent to statements 1 - 3.

Data was recorded on two coding sheets for each subject, one for the time, one for the coding of the questions. In addition, careful records of actual behavior and special questions, both during the training and the test sessions were kept.

Scoring

The score for a randomly chosen subject on the first task was a random variable we called X_1 which was the sum of all the recorded inter-question intervals for that subject on that task. Let X_2, X_3, and X_4 denote corresponding random variables for tasks 2, 3, and 4. The total score on the test was intended to be $X_1 + X_2 + X_3 + X_4$, though only $X = X_1 + X_2 + X_3$ was used because none of the 20 subjects were able to formulate Task 4 as we intended it.

Results and Discussion

In order to test the nul hypothesis, $EX_C = EX_T$, a one-way analysis of variance was computed. The null hypothesis was rejected at the .01 level. We obtained an $F_{.99}(1, 18) = 11.95$. This means that our improvement technique had a significant effect. While working with the children, we formed the "clinical" impression that those of superior intelligence, energy, aptitude, from both groups T and C would do equally well and better than those of lesser "mathematical abilities". Some of the children rated lowest in "mathematical ability" by their teachers, however, did surprisingly well on the test. It is these children for whom the improvement method appears to have made the greatest difference.

For our test to be a good assessment instrument, it should have high reliability. To measure its reliability requires a far larger sample than the 20 children tested here. This has yet to be done. This experiment was intended primarily as a pilot, to guide our conceptualization and give us experience in designing a test and improvement technique. It has served this purpose by supporting the claim that "hypotheses" have psychological reality and that problem-recognition and formulation can be learned by exposing children to inquiry-provoking situations where they have to form hypotheses.

The experimental subjects who were exposed to our improvement procedure did significantly better on the test than the subjects in the control group primarily because the improvement procedure provided exposure to opportunities for original inquiry. This stimulated the subjects to form general hypotheses. These led them to ask questions. The answers led to changes in weight, saliency, and to the formation of new hypotheses. This learned ability to form, pick, and use general hypotheses and specialize them to specific cases may have transferred to the test situation. It is very unlikely that memory alone can account for the higher score of the experimental subjects, because the tasks on the test differed considerably from the tasks in the training sessions.

Some additional findings emerged from our data. A simple test for association indicated that X_1, X_2 and X_3 were not statistically independent. That is, the conditional probability that a subject does well on Task 1 (arithmetic on rate x distance) given that he did well on Task 3 (arithmetic on price x quantity) is higher than it is, given that he did poorly on Task 3. We expected that for the 10 trained children, X_1 would be correlated with X_3, if not also X_2, though for the control group we expected lower correlation between X_1 and X_3, because Tasks 1 and 3 were formally identical. The correlations among X_{1T}, X_{2T}, and X_{3T} gives additional support to the claim that general hypotheses, which can be specialized to both Task 1 and Task 3, for example, were formed.

The hypotheses $EX_{1T} = EX_{2T} = EX_{3T}$ and $EX_{1C} = EX_{2C} = EX_{3C}$ were accepted at the .01 level by means of an analysis of variance. This indicates that the 3 items on the test were approximately equivalent.

Summary of Conclusions

We conceptualized the process of recognizing and formulating real problems as mathematical story problem-statements. This is based on "representation theory", which holds that learners form, select, and use general hypotheses. To test an aspect of this theory, we developed a technique to elicit inquiry behavior in fourth and fifth graders. By exposure to question and hypothesis-formation, such as a variant of "twenty questions", we expected the children to form general hypotheses on their own. This was tested by the speed with which they asked questions indicative of such hypotheses. A controlled experiment with 20 children showed that 10 who were exposed to our technique aimed at improving problem-recognition and formulation did significantly better than the 10 children who were not exposed to this.

This finding shows that problem formulation can be learned. This is important because if offers a feasible remedy for the situation where people are far better at solving problems that were preformulated for them than they are at recognizing and formulating problems on their own.

EXPERIMENT III

In the previous experiments, we coded and evaluated the quality of verbal questions posed by a subject during the process of comprehension attainment in a problem-formulation task. One of the key aspects of the questions reflecting the degree of comprehension is the "precision" with which adjectives and other modifiers are used. When the adjective is imprecise, the question is difficult to answer. This also seems to reflect the degree of comprehension as measured in the other experiments. The difficulty of giving a precise and accurate answer increases with imprecision of the adjective in the question. For example, in the question "Is object x far from object y?", the adjective "far" is imprecise. But, what makes us say that it is imprecise, and how can we determine its degree of imprecision? One way of doing this is to connect the notion of an "imprecise adjective" to fuzzy set theory (Zadeh, 1965). If the connection between fuzzy set theory and the precision of a phrase can be made, psychology, linguistics, and psycholinguistics might be enriched by this body of potentially applicable theorems. Fuzzy set theory in turn might benefit by becoming a behavioral science, its assumptions validated and its problems and results stimulated by empirical findings.

In this experiment, we demonstrate (a) a novel procedure for measuring the imprecision of a given adjective in a sentence, as reflecting the degree of comprehension, and (b) the use of this measure for comparing the precision of adjectives as well as the consistency of such comparisons over trials.

Theoretical Background

The simplest idea for explicating the precision of a phrase is that of an interval. The phrase "Between 3500 and 4500 miles" as an answer to "How large is the earth's diameter?" is more precise than "Several thousand miles". When the question refers to a random variable, such as

the diameter of a randomly chosen planet, interval estimates are widely accepted. Another idea for explicating the precision of a phrase, which has been proposed to capture the response of an ordinary person better than does an interval estimate is to regard an imprecise predicate phrase as denoting a fuzzy set (Zadeh, 1965). An ordinary set, such as E, the set of all even numbers, can be specified by its characteristic function, f_E. This maps the natural numbers, $N = \{0, 1, 2, \ldots\}$ into the set $\{0,1\}$: $f_E(x) = \begin{cases} 1 \text{ if } x \notin E \\ 0 \text{ if } x \notin E \end{cases}$ for all $x \in N$. A fuzzy set is an extension of this idea in which the all-or-none nature of the characteristic function is replaced by a grade of membership, any real number in the interval $[0,1]$. Thus, if L is the set of large numbers, $f_L(0)$, $f_L(1)$, $f_L(2)$ would all be close to 0, while $f_L(10^6)$, $f_L(10^{10})$, et cetera would all be closer to 1. A set, like L, is fuzzy if $f_L(x) \neq 0$, or 1 for some x. The mapping f_L depends on who judges grade of membership, and the purposes and conditions under which he makes this judgment.

More generally and more realistically, f_L maps N, R_1 the set of reals, or an arbitrary ordered set into a finite lattice rather than only into $[0,1]$. Regarding $f_L(x)$ as a real-valued function, which is a (non-fuzzy) set ordered pairs, $\{x, f_L(x)\}$, is contrary to what motivated the invention of "fuzzy sets". It is more consistent with the spirit of smoothing the sharp boundaries of a class to replace $f_L(x)$ itself by a fuzzy set. It is the fuzzy set denoted by the certainty with which a judge believes that $x \in L$. This certainty itself is yet another fuzzy set: how certain he is about his certainty, et cetera. It may be plausible to assume that for different values of x, a judge is always more certain or less certain about a proposition like "The certainty of my belief in P is ___" than he is about P, for all P. Under additional conditions such as continuity and boundedness, a limiting "characteristic fuzzy set" may exist. People, with their limited information processing capacities, can probably not judge the certainty of more than embeddings.

Since its founding, in less than a decade, fuzzy set theory has developed vigorously in the hands of mathematicians, computer scientists, and engineers (Bellman, Kalaba, and Zadeh, 1966; Chang, 1968; Goguen, 1967; Mizumoto, Toyoda, and Tanaka, 1969; Zadeh, 1971). It is now a rather sophisticated discipline with promising applications. Its importance is not only in its potential for solving engineering problems, such as designing a robot to park a car. The concepts and methods of the theory may be of potential value for developing more adequate models of human information processing and for the design of systems to help people with the storage, organization, and use of knowledge.

Even the simple-minded and unrealistic notion of f_L as a mapping of R into $[0,1]$ can help us conceptualize more clearly the difference between phrases like "large" and "very large". If we can suppose that f_L describes how a particular person maps R into $[0,1]$ in response to the instructions "How strongly do you believe that r is a <u>large</u> number?" for a sample of real numbers $r \in R$, then we can compare f_L with f_{VL}, the corresponding function with "<u>large</u>" replaced by "<u>very large</u>". Suppose that f_F is continuous and differentiable for any fuzzy set, and that it has the shape of an S for polar adjectives like "large". (For an adjective like "medium-sized", it would have a bell-shaped curve.)

Suppose further that the derivative, $f_L'(x)$, is jointly proportional to $f_L(x)$ and $1 - f_L(x)$. This means that the marginal increase in the judge's strength of belief that $x \in L$ and the strength of belief that $x \notin L$, assumed to be $1 - f_L(x)$.

The logistic curve, $f_L(x) = \dfrac{1}{1 + e^{a-bx}}$ satisfies the differential equation $f_L'(x) = bf_L(x)[1 - f_L(x)]$ expressed by the above assumptions. This has the S-shape we expect, with the property that $\lim_{x\to\infty} f_L(x) = 1$ and $\lim_{x\to-\infty} f_L(x) = 0$. It has an inflection point at $x = a/b$. To see this, note that $f_L''(x) = b[f_L'(x) - 2f_L(x)\, f_L'(x)]$. The value of x for which $f_L''(x) = 0$ must satisfy $f_L(x) = \frac{1}{2}$, and $1 = e^{a-bx}$, or $a - bx = 0$. The maximum steepness of the curve $f_L(x)$ is a possible measure of the precision of the adjective "large" denoting L. That is the value of $f_L'(\frac{a}{b})$, which is just $b/4$. Another plausible measure of precision is the "transition range" of $f_L(x)$: the difference $d = x_1 - x_0$, where $f_L(x_1) = 1 - \epsilon$ and $f_L(x_0) = \epsilon$ for some ϵ, $0 < \epsilon < \frac{1}{2}$. From $a - bx = \ln \dfrac{1-f_L(x)}{f_L(x)}$, it follows easily that $b(x_1 - x_0) = \ln \dfrac{1-\epsilon}{\epsilon} - \ln \dfrac{\epsilon}{1-\epsilon}$ and

$$d = \frac{2}{b} \ln \frac{1-\epsilon}{\epsilon}$$

The more precise of less fuzzy L, the larger b and the smaller d. In comparing f_L with f_{VL}, it is plausible to hypothesize that $b_{VL} > b_L$. The subscript denoting the fuzzy set, L, VL, et cetera should be added to both parameters a and b. (It was omitted for simplicity.) The parameter a helps to indicate where the inflection point occurs.

Though developments in formal analysis of fuzzy sets have taken place during the past eight years, and analytic questions relating fuzzy sets to linguistics (Lakoff, 1972) and logic (Goguen, 1967) have been raised, there have been few attempts to approach the assumptions and questions raised by fuzzy set theory from a psychological and experimental viewpoint. If one is interested in behavior, then the question: "Is 'very far' more precise than 'far'?" might be reposed as: "In the context of a response to a given question is 'very far' more precise than 'far' when the subject is told to respond on such and such a scale?".

The subject's responses and the judgments of precision become exceedingly sensitive to the experimental situation. It makes a big difference if in answering the question, "Is 10^9 greater than 5?" than the question "Is 6 much greater than 5?". Criterial anchoring is of proven importance in the psychology of judgments (John, 1971).

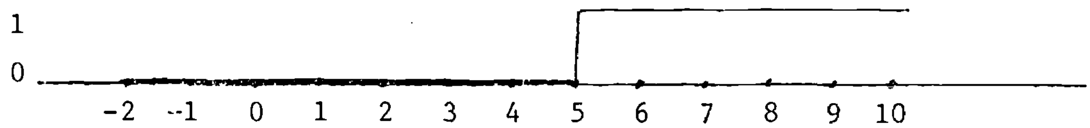
It also makes a difference on how one presents the instructions and the questions to the subject as well as how the subject is allowed to scale his answers. For instance, the answer to the question, "How far

am I from the curb?", asked by a driver trying to park his car, is captured neither by using an interval estimate nor a grade of membership for a scale. "Close", "very close", "somewhat close", are more spontaneous, consistent over time, and possibly more useful responses.

The precision of the answer should sometimes match the precision of the question. We thus hypothesize that if subjects are allowed to be fuzzy in their response to an imprecise question, they will show greater degree of consistency over trials than if they were forced to be precise in their response to the same question.

Experimental Rationale

In this study, we asked human subjects, by various techniques, to assign grades of membership in fuzzy sets to samples of objects. Let I be the set of integers $\{\ldots -2,-1,0,1,2,3,\ldots\}$. The predicate "___ is greater than ___" denotes a two-place relation, or a subset of I x I: $\{(1,0), (2,1), (2,0)\ldots\}$. It can be used to form a one-place relation by filling in one of the 2 slots, as in "___ is greater than 5", which denotes $\{6,7,8,\ldots\}$. If we asked a human subject in a psychological experiment to assign grades of membership in the set of integers greater than 5 to the integers -2,-1,0,1,2,...10, we might expect to get:



We could qualify this one-place predicate by transforming it into "___ is greater than 5 *by 3*" or "___ is greater than 5 *by a factor of 2*" (Kochen, 1969). This would make the predicate more *specific* rather than more *precise*, because it restricts the denotation, in both those cases to one-element sets.

If we presented the sample of integers (8,2,9,29,10,105) and asked the subject to assign a grade of membership in the set of even numbers, we should get (1,1,0,0,1,0). If we did not observe that, there is an interesting finding to be explained.

We would not, however, expect the subject to confine his assignment to 0 and 1 if we modified the one-place predicate to read "___ is much greater than 5". Consider the corresponding two-place predicate, "___ is much less than ___". We could now adapt the method of paired comparisons, and present the subject with two numbers, say (8,10), and ask him to select the one that is much greater than the other if *he judges this to be the case*. We might expect the subject to be inconsistent in his judgment in that he might select (5,90) (5,94) (5,99) (5,100) (5,102) (5,103) (5,104)... but fail to select (5,91) (5,92) et cetera.

It is important to distinguish between the subject's judgment of *how much* greater than 5 he considers x and the subject's judgment about the strength *of his belief* that x is much greater than 5. Insofar as many fuzzy sets are described by predicates that a subject can scale, these two notions are often connected. The ability to classify seems to depend on using predicates which make sentences either true or false; thus, classifying integers into even and odd does not admit of a "degree of evenness", though this might be defined.

If we asked a subject to mark a cross on a line of fixed length with a 5 shown at one point, for each of several numbers, like 3, 7, 10, 17, 1000, we would be measuring something about the way he scales these numbers in this constrained task. But this judgment differs from that of the strength of his belief that 17 is much greater than 5.

A traditional method for measuring strength of belief is to ask the subject to indicate on a scale based on semantic differential (Osgood, 1961) how strongly he agrees or disagrees with a given statement. In this case the statements are all of the generic form "(Stimulus x) is a member of set of (Name of fuzzy set)". (Stimulus x) is replaced by stimulus, such as a card with a number, and (Name of fuzzy set) is replaced by a phrase like "all large numbers" or "all numbers much larger than 5" or "all numbers very much larger than 5". On being presented with

(1) the card,
(2) the statement, and
(3) a scale such as

Agree Strongly	Disagree Strongly

the subject's response is to place a cross-mark along the above scale. We in turn translate the position of that mark into a number between 0 and 1, and plot this number against the corresp nding value of the physical stimulus variable to get a characteristic curve $f_n(x)$ for that subject and phrase.

For the same subject we now compare $f_n(x)$ for phrase n with $f_m(x)$ for another phrase; for example, n = "all points far from *", and m = "all points very far from *". We expect both curves to be S-shaped. We take the slope of the inflection point to be a plausible measure of the precision with which the subject uses the phrase n or m. Thus, we expect the curve $f_m(x)$ in the above example to be closer to a step-function than the curve for $f_n(x)$, which may be a more widely spread S. In addition, $f_m(x)$ should be shifted to the right of $f_n(x)$. In this way, we can quantitatively assess a given subject's interpretation of certain phrases in a given context.

Several research problems are raised by these considerations.
1. How reliable an instrument for measuring a person's conceptualization of phrases is this technique? Is there consistency?
2. How context-sensitive is it, and how can the context be controlled for?
3. What is the variation in conceptualization of phrases over subjects?
4. In what sense can we generalize that "greater than" denotes a more precise concept than does "much greater than"?
5. What is the relation between the assessment of a person's strength of belief that an object belongs to a set specified by an adjective, like "heavy" and that person's judgment of the magnitude of the stimulus to which that adjective applies? In other words, how do scales of strength of belief about membership of a stimulus in a class relate to scales for psychophysical or psycholinguistic judgments?

In this experiment we propose to answer only questions 4, 5, and 1 leaving 2 and 3 for future experimentation, in that order of priority.

Our aim is to establish the psychological reality of fuzzy sets; to test the assumption that when faced with a situation that calls for an imprecise judgment, people would utilize a grade of membership. This is important in the content analysis of questions and other verbal behavior which is to reflect cognitive states.

PART I

Hypothesis

The hypothesis being tested by this experiment may be introduced by the following example: If given the sentence, "Identify all numbers that are (wd) than 5, where (wd) can be replaced by "greater", "very much greater", and "much greater", then the three sentences obtained by replacing (wd) with the three phrases in the order shown decrease in precision, as determined by the characteristic curve of a fuzzy set.

Procedure

The subjects were ten University of Michigan students. Each of them was given three pieces of paper, P_1, P_2, and P_3 with the same seven numbers, 0, 20, 100, 750, 500,000, 1,000,000, 1,000,000,000 written on each. A scale labeled 0 to 1 was drawn under each number. The instructions given to a subject were "By marking a cross-mark on the scale, indicate your strength of belief that the number directly above the scale is greater than 5" for P_1; "much greater than 5" for P_2; and "very much greater than 5" for P_3. P_1, P_2, and P_3 were given to each subject in a random order and one at a time.

Results and Discussion

A plot of the strength of belief that the number x was (wd) than 5 vs x was drawn. Table II gives the strength of belief, $f_{wd}(x)$, averaged over 10 subjects, expressed by them that x is in the set of numbers which are (wd) than 5. (See Table II on next page.)

If C(x) is plotted against log x, a function of the form $1-e^{-k \log x}$ with $k_{MG} \doteq .025$ and $k_{VMG} \doteq .043$. The subscript MG and VMG on k refers to "much greater" and "very much greater", respectively. The effect on $1-f_{MG}(x)$ of preferring "much larger than" with "very", is in this case to raise $1 - f_{MG}(x)$ to the power $\frac{.043}{.025} \doteq 1.8$.

One measure of how well $1 - x^{-.025}$ fits the data is given by the sum of the squares of the deviations, which is 1.46. This is not a very good fit. If we estimated k so as to minimize this figure of merit, it might still be a poor fit. But when the numbers were presented randomly to subjects, a least square regression analysis gave a good fit, with $F(1, 19) = 6.72$, $p < .05$.

It is more plausible for $f_{MG}(x)$ to have the form of the logistic curve, discussed earlier.

We estimated a and b with the help of a computer program to get:

	x	$\log x$	$f_{MG}(x)$	$f_{VMG}(x)$
x_1 =	1	0	$y_1 = 0$	0
x_2	20	1.3	$y_2 = .49$	.30
x_3	100	2.0	y_3 .60	.48
x_4	750	2.9	.68	.63
x_5	5×10^5	3.7	.80	.79
x_6	10^6	6.0	.84	.88
x_7	10^9	9.0	$y_7 = .89$	.91

TABLE II

A Logarithmic Transformation of Responses for "much greater" and "very Much greater".

$a = -.83$, and $b = .13 \times 10^{-8}$. The F statistic is 1.99 at a significance level of .23. This, too, is a very poor fit.

Note that $f_{MG}(x)$ crosses $f_{VMG}(x)$ just above $x = 5 \times 10^5$. The value of d_{VMG} is less than that for d_{MG}. The value of b for $f_{VMG}(x)$ is $.17 \times 10^{-8}$, which is greater than that for $f_{MG}(x)$. This supports the notion that "very much greater" is more precise than "much greater". This is the main result we wanted to establish.

PART II

Hypothesis

Using the characteristic curve of a fuzzy set as the measure of precision, subjects who use an anchor in situations calling for imprecise judgments would show greater confidence and degree of precision in judgment than those who are not allowed to use an anchor.

Procedure

Ten University of Michigan undergraduates were used in this experiment. They were each given seven boxes of different weights one at a time. Then each subject was asked to hold each box in his right hand and and make a decision on whether it was heavy or not, in comparison with a constant. Then they were asked how strongly did they believe that such a box belonged to the set of boxes that were heavier than the comparison box, by marking a scale between 0 and 1. Then they were given the same boxes again, but this time without the comparison and they were asked to decide whether each of the boxes was heavy or not. Then they were asked to rate on a scale between 0 and 1 their strength of belief in this judgment.

Results and Discussion

The characteristic curve of the strength of belief on judgments (See Figures 6 and 7) where no comparison was employed showed a degree of fuzziness greater than that reflected by the curve of the first judgment, where the weight was compared with a constant.

The degree of precision in this case was based on the strength of confidence with which the subject viewed his judgments of which weights. The assumption was that the greater the confidence, the greater was degree of precision as postulated earlier by the characteristic curve. An analysis of variance showed a significant difference in degree of precision as interpreted by the strength of confidence, with $F = 15.1$, $df = 1/8$, $p < .05$. The non-arbitrariness of the finding is strengthened by the other finding that there is no significant correlation between the subject's confidence in his judgment about the weight and the weight itself.

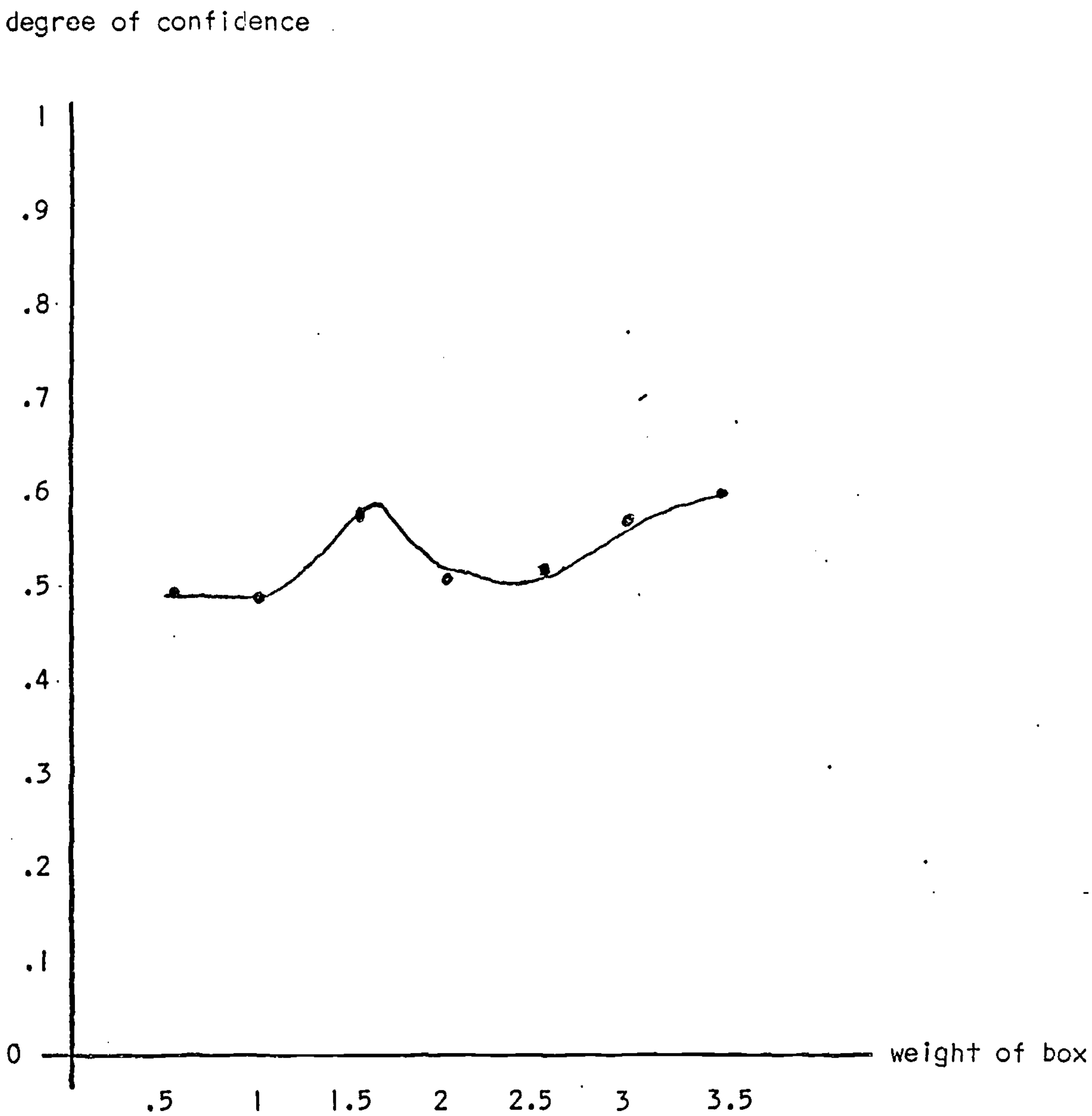


Figure 6

Comparison with Constant

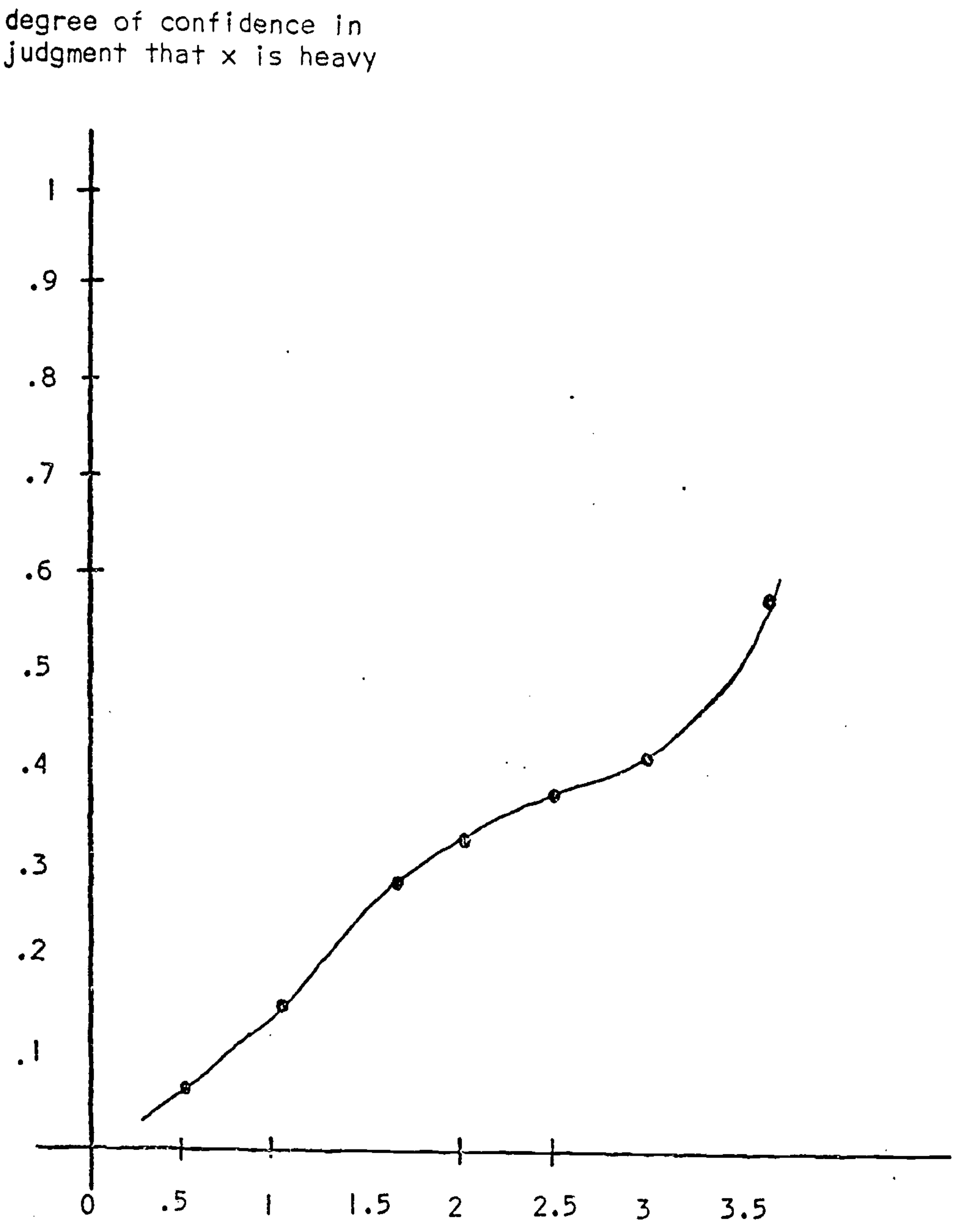


FIGURE 7

No comparison

PART III

Hypothesis

A higher degree of response consistency over trials would occur if subject is allowed to give a verbal imprecise response to a question about a fuzzy set than if he were forced to give a precise answer.

Procedure

Seven adults were used in this experiment. Each of them underwent four trials during which they were asked: "How strongly do you believe that x is much greater than 5?", such that x stands for one of the numbers in the first column of Table III. The presentation of the numbers was in a regular order but identical for every subject and on every trial. What differed was the scaling technique. On trials 1 and 3, the subject was asked to indicate his strength of belief by responding with a number between 0 and 10 where 0 meant "completely disbelieve it", and 10 meant "completely believing it". On trials 2 and 4, the subject was asked to respond with one of the seven verbal categories from "perfectly certain it is" to "perfectly certain it is not", as indicated in Table III. Each of these responses is itself fuzzy. The subject underwent trials 1 and 2, then 24-hours later 3 and 4. The 24-hour span was used in order to minimize the effects of memory.

Results and Discussion

As Table III (page 36) shows, there is very low consistency between subjects as well as within the same subject over the two trials when the subject is asked to respond in terms of numerical grading. When however one compares the responses on the seven verbal categories scale, consistency prevails. In fact, if we were to reinterpret the numerical responses in terms of the verbal categories, by looking at the numerical range rather than the exact chosen numbers on the scale, as indicative of the strength of belief, then consistency goes up between trials 1 and 3 (see Table III on next page). This result may be interpreted as due to instructional sensitivity. This sensitivity to instructions may be more common to many psychological experiments than is commonly granted. Indeed, a great deal of psychological experimentation may be eliciting inappropriately precise responses. The ordering of the seven imprecise forced-choice responses in this experiment elicits higher degree of consistency than a line along which a subject marks or a scale from 0 to 10. It is still not exactly what we need because: (a) of the forced choice process; and (b) of too much sensitivity to details of verbal presentation.

Conclusion

The findings of this paper are perhaps of greater significance for the new questions they raise than for the questions they settle. The new questions raised are readily amenable to experimental analysis. The import of this investigation is, therefore, primarily to open for experimental investigation a new direction of fruitful, convergent

TABLE II: Subjects Responses According to the Numerical (T_1 and T_3) and verbal (T_2 and

Responses

	S_1				S_2				S_3				S_4				S_5				S	
x	T_1	T_3	T_2	T_4	T_1	T_3	T_2	T_4	T_1	T_3	T_2	T_4	T_1	T_3	T_2	T_4	T_1	T_3	T_2	T_4	T_1	T_3
0	0	0	7	7	0	0	7	7	0	0	7	7	0	0	7	7	0	0	7	7	0	0
4	0	0	7	7	0	0	7	7	0	0	7	7	0	0	7	7	0	0	7	7	0	0
5.4	.04	.01	7	7	.09	3	7	7	0	.9	7	7	.01	.5	7	7	.3	.8	7	7	1.2	1.9
8.2	.01	.1	7	7	.1	.13	7	7	.01	1.2	7	7	.4	1	7	7	.5	.15	7	7	.9	1.4
9	.03	.03	7	7	.02	.01	7	7	.04	1	7	7	.45	.9	7	7	.8	.2	7	7	1	2
11	.06	.06	7	7	.11	.4	7	7	2	1	7	7	1	1	7	7	1	2.5	7	7	1.5	4
17	1.2	1.8	7	7	4	2	6	6	1	3	5	5	1	1.5	7	6	2	2.5	7	7	2	4
25	2	1.87	7	7	3.8	3	6	6	2	1.5	6	5	1.3	1	6	6	3	2.5	6	6	2.9	2
32	2.7	3.2	6	4	4.2	4.2	4	6	3	2.1	5	5	2	2	6	6	3	4	6	6	3	2.8
46	4.1	4.1	6	6	4.1	2	6	6	2	3	5	6	2	2.3	4	6	4	3.2	6	6	4.3	4.7
50	4.3	4.7	6	6	4.4	4.0	6	6	4.3	5	4	5	3.3	2.1	6	6	3.9	3.7	4	6	4.2	4.9
100	5.2	5.8	5	5	3.6	3.9	5	5	5.6	6.2	5	5	3.2	4.5	5	5	4.2	4.9	6	6	5	5
700	7.0	5.0	5	5	4.5	4.8	5	5	6.3	7.1	5	5	4.2	3.2	5	5	4.9	3.8	5	5	6.5	6.7
5000	7.5	4.5	2	2	5.2	5.0	2	2	5.9	7.2	2	3	7	6.5	2	3	5.2	5.9	2	2	7.3	8
40,000	8	6.5	2	2	5.9	5.2	2	2	7.3	7.5	2	2	6.1	7	2	1	6.3	6.2	2	2	8.1	8
200,000	8.2	5.0	2	2	6.0	6.0	2	2	5.7	6.3	2	2	7	7	1	1	6.5	7.1	1	1	9.5	9.6
700,000	8.7	6.3	2	1	6.0	6.5	2	2	5.2	7.6	2	2	8	7	1	1	7.3	8.2	1	2	10	10
10^6	9.0	7.7	2	1	8.5	7.0	1	1	8.8	7.6	1	1	9	9	1	1	8.7	8.0	1	1	10	10
5×10^6	10	10	1	1	9.1	9.5	1	1	9.2	9.8	1	1	10	10	1	1	9.2	9.2	1	1	10	10
400×10^6	10	10	1	1	10	10	1	1	10	10	1	1	10	10	1	1	9.5	9.8	1	1	10	10
10^9	10	10	1	1	10	10	1	1	10	10	1	1	10	10	1	1	9.1	9.3	1	1	10	10

*In T_2 and T_4 the numbers 1 through 7 mean:

1 = Perfectly certain - Yes
2 = fairly certain
3 = I think it is
4 = I don't know
5 = I think it is not
6 = Fairly certain it is not
7 = Perfectly certain it is not

II: Subjects Responses According to the Numerical (T_1 and T_3) and verbal (T_2 and T_4) Responses

	S_2				S_3				S_4				S_5				S_6				S_7		
T_4	T_1	T_3	T_2	T_4	T_1	T_3	T_2	T_4	T_1	T_3	T_2	T_4	T_1	T_3	T_2	T_4	T_1	T_3	T_2	T_4	T_1	T_3	T_2
7	0	0	7	7	0	0	7	7	0	0	7	7	0	0	7	7	0	0	7	7	0	0	7
7	0	0	7	7	0	0	7	7	0	0	7	7	0	0	7	7	0	0	7	7	0	0	7
7	.09	3	7	7	0	.9	7	7	.01	.5	7	7	.3	.8	7	7	1.2	1.9	7	7	.06	.03	7
7	.1	.13	7	7	.01	1.2	7	7	.4	1	7	7	.5	.15	7	7	.9	1.4	7	7	0	.3	7
7	.02	.01	7	7	.04	1	7	7	.45	.9	7	7	.8	.2	7	7	1	2	7	7	0	0	7
7	.11	.4	7	7	2	1	7	7	1	1	7	7	1	2.5	7	7	1.5	4	7	7	.12	.4	7
7	4	2	6	6	1	3	5	5	1	1.5	7	6	2	2.5	7	7	2	4	6	6	.9	1.3	6
7	3.8	3	6	6	2	1.5	6	5	1.3	1	6	6	3	2.5	6	6	2.9	2	6	6	1.4	2.0	6
4	4.2	4.2	4	6	3	2.1	5	5	2	2	6	6	3	4	6	6	3	2.8	6	6	2	3	6
6	4.1	2	6	6	2	3	5	6	2	2.3	4	6	4	3.2	6	6	4.3	4.7	6	6	2.9	4.8	4
6	4.4	4.0	6	6	4.3	5	4	5	3.3	2.1	6	6	3.9	3.7	4	6	4.2	4.9	6	6	3.3	3.7	6
5	3.6	3.9	5	5	5.6	6.2	5	5	3.2	4.5	5	5	4.2	4.9	6	6	5	5	6	5	4.7	4.5	5
5	4.5	4.8	5	5	6.3	7.1	5	5	4.2	3.2	5	5	4.9	3.8	5	5	6.5	6.7	5	5	4.5	5.1	5
2	5.2	5.0	2	2	5.9	7.2	2	3	7	6.5	2	3	5.2	5.9	2	2	7.3	8	3	3	5.3	6.2	2
2	5.9	5.2	2	2	7.3	7.5	2	2	6.1	7	2	1	6.3	6.2	2	2	8.1	8	2	2	5.8	7.5	2
2	6.0	6.0	2	2	5.7	6.3	2	2	7	7	1	1	6.5	7.1	1	1	9.5	9.6	2	2	6.9	7.5	2
1	6.0	6.5	2	2	5.2	7.0	2	2	8	7	1	1	7.3	8.2	1	2	10	10	2	2	7.5	7.4	2
1	8.5	7.0	1	1	8.8	7.6	1	1	9	9	1	1	8.7	8.0	1	1	10	10	2	2	8.0	8.4	1
1	9.1	9.5	1	1	9.2	9.8	1	1	10	10	1	1	9.2	9.2	1	1	10	10	1	1	9	9	1
1	10	10	1	1	10	10	1	1	10	10	1	1	9.5	9.8	1	1	10	10	1	1	9.3	9.2	1
1	10	10	1	1	10	10	1	1	10	10	1	1	9.1	9.3	1	1	10	10	1	1	10	9	1

ers 1 through 7 mean:
1 = Perfectly certain - Yes
2 = fairly certain
3 = I think it is
4 = I don't know
5 = I think it is not
6 = Fairly certain it is not
7 = Perfectly certain it is not

research. The results are most likely to build a strong bridge between linguistics, psychology and fuzzy set theory which is itself a bridge between mathematics, computer science and electrical engineering. If the results are strong, they will shed important new light on fundamental problems in all these fields.

We have shown that "very much greater than 5", as used by people in assigning a grade of membership to a number is less fuzzy than "much greater than 5". We have also shown that anchoring has the effect of making a fuzzy adjective less fuzzy. Also we found that response consistency prevails when the subject is allowed to be fuzzy in his scaled answer to an imprecise question.

We have yet to find a reliable method for establishing the characteristic curve of a given subject for a specific phrase. This depends on: (a) the order in which the stimuli are presented, e.g. 20, 100, 750,... vs 100, 20, 750,...; (b) the range over which stimuli are presented e.g. 20 to 10^9 vs 10^{-9} to 100; (c) the number of stimuli presented; (d) the speed with which stimuli are presented; whether they are displayed simultaneously, one at a time with long pauses in between; whether there was a distracting task between presentations; (e) the units attached to the stimuli, e.g. 20 feet, 100 ft., vs 20", 100"; (f) context of the instructions which specify the fuzzy set. This is the subject of another study.

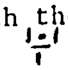
The fourth major activity which was partially supported by this grant was Albert N. Badre's Ph.D. Dissertation on "Hypotheses and Representational Shifting in Ill-Defined Problem Situations". Because of the length and intricacy of this work, and the severe budget constraints on this project, this can only be summarized here. (See Appendix I) The entire 130-page dissertation will be made available to anyone who can reimburse reproduction costs.

CONCLUSIONS

Cognitive learning theories as well as educational practices have stressed the behavior of people on solving problems that were formulated for them and presented to them as well-defined problem-statements. There was a serious gap in our conceptualization of how people recognize and formulate real problems which they must formulate by themselves. There is a corresponding practical need to educate people at all levels to recognize, select, and formulate the real problems they encounter in life.

This work contributed significantly to lessening this gap and meeting the practical needs. The contributions were both on the theoretical and the experimental side. On the theoretical side, how problem formulation is learned was conceptualized by specifying an algorithm that asks questions in response to presentations of staged tasks.

Cognitive theories of learning (e.g. Tolman, Kohler, Koffka, Wertheimer, Lewin) coincide with stimulus-response theories in the view that problem-solving requires "structuring of the problem". This is intended to mean that the learner is able to use experiences that resemble "elements of the problem" or "aspects of the situation". While S-R theorists stress the learner's history of past experiences, cognitive theorists stress "insight", or "current understanding of essential

relations". Operationally, problem-solving tasks given humans in experiments on higher learning are usually presented in verbal instructions such as "Build a hat-rack with these materials" or "Invert the match-stick sketch of a cocktail glass  with the olive outside by moving just two matches", or "substitute numerals for the letters in SAM+JIM=BILL". The vague term "problem" in the above phrases with quotation marks apparently means a verbal problem-statement, quite similar to the story problems that school children solve in arithmetic and algebra.

Animals, such as Kohler's chimpanzees who had the "insight" to join two sticks to reach a banana beyond the reach of one stick, recognize and solve numerous problems all the time. What are the limits to the problems they can learn to solve? One class of problems they cannot solve, we hypothesize, contains those the learner must formulate for himself, linguistically or graphically. There seems to be a major theoretical gap in the development of cognitive learning theories about problem-solving between concern with tasks given chimpanzees such as the above, and tasks given humans, which resemble story-problems. The gap is the lack of attention to the question of how real problems (problem-situations) are recognized and formulated.

A new and practical technique for observing and measuring how people formulate problems was developed. It was found to be useful as a test and assessment instrument, both in the laboratory and in the school. An intervention method to help children improve in recognizing and formulating problems was also developed. It was found to improve problem-recognition performance significantly. College students, too, were found to perform significantly better in solving a task requiring problem-formulation if they had prior exposure to problem-formulation experience than when they had no such exposure. In sum, problem-formulation can be learned. It is far more important for people to learn how to recognize and formulate problems by themselves than to solve problems someone else formulated for them. The schools have not been giving sufficient priority to helping students improve their problem-formulation activities. It is urgently vital that they begin to do so.

The notion of "comprehension" or "understanding" has been explicated in terms of the ability to recognize, select and formulate problems. By a problem we mean a state of the world from which another state, that is greatly preferred, could be reached by the appropriate action. For example, man carries the sickle-cell anemia trait, and his wife carries it too, he has a problem. He may be quite unaware of it. Even if he were aware of it, he may not pay more attention to it or give it higher priority than any of a dozen other problems. Even if he did pay attention, he may not be able to articulate or describe it with any clarity. This notion of problem differs radically from what is usually studied in problem-solving, which are really well-defined problem-statements.

A person understands a problem when he is aware that he needs to know something he does not know and could find out by asking appropriate questions. This awareness arises from the formation and use of hypotheses, which we believe to be the basic units of thought. Awareness can be explicated in terms of hypotheses which refer to the learner's ability to form hypotheses. Operationally, we recognize when a person

understands a problem by evaluating his questions. If a teacher can expose children to problem-generating environments in which they are reinforced for asking questions indicative of comprehension, then he can instil increased comprehension in them. We have shown that this can be done, and devised a way of testing, of recognizing questions indicative of what can be done.

Of course, this is just a beginning. But it provides a strong base on which to build. It makes clear what the specific next steps should be. It is strongly recommended that NIE continue supporting further work in this promising and practically useful and urgently needed direction.

APPENDIX I

ABSTRACT

ON HYPOTHESES AND REPRESENTATIONAL SHIFTING
IN ILL-DEFINED PROBLEM-SITUATIONS

by

Albert Nasib Badre

Chairman: Manfred Kochen

The purpose of this thesis was to investigate the conditions under which people learn to cope in ill-defined problem-situations. It was hypothesized that practice in representational shifting improves coping. Shifting of representations refers to the formation of new hypotheses in the solving of already formulated problems or the formulation of new problems. An ill-defined problem is one that lacks specification of a set of solutions, solution-properties, and solution-methods.

An experiment was designed in which subjects were told to ask questions to help them formulate and solve a problem. Certain words and actions were prespecified, but not shown to subject, and interpreted as indicative of shifts in representation. The time it took the subject to use these words and actions was measured.

The results show that if a problem-solver practices with tasks requiring shifts of representation, he is likely to perform better in solving an ill-defined problem than one who has no prior practice or one who has prior practice with well-defined problems not requiring representational shifting. There is no significant difference in performance between a no-practice group and a group that gets practice with a well-defined problem. Practice with tasks which require shifting of hypotheses has the greatest positive effect on solving problems which are initially ill-defined.

Amarel, S. "Representations and modeling of problems of program formation", in B. Meltzer and Michie (eds.), Machine Intelligence, Edinburgh: Edinburgh University Press, 1971.
Ausubel, D.P. "The use of advance organizers in the learning and retention of meaningful verbal material", J. Educ. Psychol., 51, 1960, 267-272.
Ausubel, D.P. The Psychology of Meaningful Verbal Learning. New York: Grune and Stratton, 1963.
Badre, A.N. "On hypotheses and representational shifting in ill-defined problem-situations", Ph.D. dissertation, The University of Michigan, Ann Arbor, July 1973.
Bellman, R.E., Kalaba, R., and Zadeh, L.A. "Abstraction and pattern classification", J. Math. Anal. Appl., 13, 1966, 1-7.
Bruner, J.S. The Process of Education. First Vintage Edition, 1963.
Bruner, J.S. Toward a Theory of Instruction. Cambridge, Mass: Harvard University Press, 1966.
Carroll, J.B. and Freedle, R.O. (eds.) Language Comprehension and the Acquisition of Knowledge. New York: Wiley, 1972.
Chang, C.L. "Fuzzy topological spaces", J. Math. Anal. Appl., 24, 1968, 182-190.
Ebel, R. "Knowledge vs. ability in achievement testing", Invitational Conference on Testing Problems, Princeton: Educational Testing Service, 1969, 66-76.
Eichholz, R.E. and D'Affer, P.G. Elementary School Mathematics, Books 4 and 5. Palo Alto, Calif: Addison-Wesley Publishing Co., 1964.
Estes, W.K. "Component and pattern models with Markovian interpretations", in R.R. Bush and W.K. Estes (eds.), Studies in Mathematical Learning Theory. Stanford: Stanford University Press, 1959, 9-52.
Gagné, R.M. The Conditions of Learning. New York: Holt, Rinehart & Winston, Inc., 1970.
Gagné, R.M. Some views of learning and instruction", Phi Delta Kappan, May 1970, 468-472.
Gagné, R.M. "The acquisition of knowledge", Psych. Rev., 69, 1962, 355-365.
Goguen, J.A. "The logic of inexact concepts", Synthese, 19, 1971, 325-373.
Hovland, C.I. "Computer simulation of thinking", Am. Psychol., 15, 1960, 687-693.
John, I.D. "Stimulus discriminability and anchor effects in judgments of lifted weights", Austral. J. Psych., 23, 1971.
Kochen, M. "Automatic question-answering of English-like questions about simple diagrams", J. ACM, 16, 1969, 26-48.
Kochen, M. "Cognitive learning processes: an explication", in B. Meltzer and N. Findler (eds.), Artificial Intelligence and Heuristic Programming, Edinburgh: Edinburgh University Press, 1971. Revised German translation, 1973.
Kochen, M. and Badre, A.N. "Question-asking and shifts of representation in problem-solving", Am. J. Psych. (forthcoming), 1973.

Kochen, M. and Badre, A.N. "On the precision of adjectives which denote fuzzy sets", submitted to *Journal of Psycholinguistic Research*, 1973.

Kochen, M., Badre, A.N. and Badre, B. "On the formulation of mathematical problems: assessment and improvement", *J. Structural Learning* (forthcoming). Presented at the Sixth Annual Structural Learning Meeting, Philadelphia, April, 1973.

Kohler, W. *The Mentality of Apes*. Humanities Press, 1927; Vintage Books, 1926.

Lakoff, G. "Hedges: a study in meaning criterial and the logic of fuzzy concepts", *Proc. Chicago Linguistics Soc.*, *8*, 1972, 183-228.

Miller, G.A., Galanter, E. and Pribram, K. *Plans and Structure of Behavior*, New York: H. Holt, 1960.

Minsky, M. "Form and content in computer science", *J. ACM*, 17, 1970, 197-215.

Mizumoto, M. Toyoda, J. and Tanaka, K. "Some considerations on fuzzy automata", *J. Com. Syst. Sci.*, *3*, 1969, 409-422.

Newell, A., Simon, H. and Shaw, J.C. "Elements of a th... of human problem solving", *Psych. Rev.*, *65*, 1965, 151-166.

Osgood, C.E., Suci, G. and Tannenbaum, P.H. *The Measurement of Meaning*. Urbana, Ill: University of Illinois Press, 1957.

Polya, G. *Mathematical Discovery*. New York: Wiley, 1962.

Suppes, P. "Applications of mathematical models of learning in education", in Wold, H. O. A. (ed.), Monaco: Union Europeanne d'Edition, 1964, 39-49.

Trabasso, T.R. and Bower, G.H. *Attention in Learning*. New York: Wiley, 1968.

Zadeh, L.A. "Fuzzy sets", *Info. and Cont.*, *8*, 1965, 338-353.

Zadeh, L.A. "Quantitative fuzzy semantics", *Info. Sci.*, *3*. 1971, 159-176.