

## ON RECOGNIZING AND FORMULATING MATHEMATICAL PROBLEMS\*

MANFRED KOCHEN, ALBERT N. BADRE and BARBARA BADRE

*University of Michigan, Ann Arbor  
and Georgia Institute of Technology, Atlanta*

---

### ABSTRACT

When mathematics is used to help people cope with real-life situations, a three-stage intellectual process is involved. First, a person becomes aware of a problem-situation which stimulates him to generate a problem-statement, a verbal story-problem. This may be in writing, expressed orally, or merely thought and evidenced by other behavior. Second, he transforms the verbal problem-statement into a mathematical formulation. Third, he analyzes this mathematically stated problem into subproblems to which the solution is more immediate.

We propose an operational instrument for assessing how a person moves from the first to the second and third stage. This involves coding questions asked, actions taken, hesitancy and latency. Results of using this procedure in experimental situations were used to determine reliability. We also propose and compare methods for effecting and accelerating motion from the first to the third stage. These involve operant conditioning, need arousal cues, controlled verbal instructions. Experience with one such technique to teach problem-formulation in actual teaching practice is reported.

The theoretical basis for these experiments is developed by showing how to program a computer to go through these stages for a reasonably large and interesting class of problem-situations to which mathematics is applicable.

---

### 1. Introduction

Formulating a mathematical “story-problem” is more difficult than solving one already formulated mathematically. It is difficult because it involves recognizing a problem that needs formulation and which arises naturally in a person’s daily activities as one that resembles a mathematical story-problem. 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).

This same problem could arise as a real world activity in the following

---

\* This research was partially supported by an Office of Education Grant OEG5–72–0050(509).

situation: A person in Houston learns at 2 p.m. that a certain event important to him (the Mardi Gras) will start in New Orleans at 8 p.m. that evening. He could drive, if he thinks he can make it. He must now formulate the problem for himself. He may learn, by asking questions, that the distance is 385 miles, that he can average 65 mph for the first 195 miles, and 55 mph for the rest.

Research on problem-solving has focussed primarily on processes by which a person (or computer-program) can and does pass from a preformulated problem to a solution. In this paper, we propose to study the process of initially recognizing and attaining a verbal formulation of a problem. We conceptualize this process with the help of what we call representation theory. We report the results of a pilot experiment using a novel instrument for assessing problem-recognition and formulation performance and a technique for improving it.

#### IMPORTANCE OF PROBLEM-FORMULATION

From kindergarten to graduate school we train people primarily to solve well-defined problems that were posed and presented to them. Scientists and engineers are taught in school to solve such preformulated problems. Only from their professional experience do they learn to recognize and formulate real problems on their own.

We believe that because of the immensity of problems that we face but have not recognized or properly formulated, it is imperative that formal education does not wait until the "scientist" becomes a "professional" in problem-solving before he gets his training in the recognition and formulation of problems. To accomplish such a shift in training priorities, we need a better understanding of how one can and does recognize and formulate a real problem.


A few leaders (Wilcox, 1973) are beginning to recognize that the separation of research and education from real problems of contemporary relevance has become too great. Mathematics, for example, has been likened to chess; and "the world hardly has time to play chess anymore."

#### BACKGROUND

Gestalt and organization theories (Tulving and Donaldson, 1972; Köhler, 1926) are concerned with how, for example, a chimpanzee<sup>1</sup> acquires the "insight" to join 2 poles for reaching a banana that is beyond the reach

---

<sup>1</sup> In a recent TV documentary, a rather impressive and similar experiment with macaques was performed in which one of the monkeys discovered that by pulling a string to which an orange was suspended, the fruit could be obtained.

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 2 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. Psycholinguistics 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. Only the work on "representation theory"<sup>2</sup> (Kochen, 1960a, b, 1964, 1971; Kochen and Badre, 1974; Badre, 1973, 1974; Uhr and Kochen, 1965, 1969) has addressed itself to problems like the one faced by the traveler from Houston to New Orleans, that requires recognition and verbal formulation.

In brief, representation theory is concerned with how a learner represents his environment so that he can recognize and cope with an increasing variety of opportunities and traps that present themselves. This differs greatly from the concern of researchers in artificial intelligence (Amarel, 1971; Simon and Siklossy, 1972) with how to represent well-defined problems to a computer so it can solve them.

#### SPECIFIC PURPOSE

In this paper we pose the problem-formulation question in terms of problems requiring mathematics. We specialize representation theory to environments that generate tasks requiring mathematical formulations in order to develop the theory in the direction of making it more testable. 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, 1974). 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; it resembles the game "Twenty Questions".

---

<sup>2</sup> There is more on representation problems in the computer science literature, notably by Amarel, and Newell, as well as in the management science literature. The closest work in psychology are models developed by Restle, Bower, and Trabasso based on ideas of Bruner.

The central point is an experimental verification of the hypothesis that a large population of children can be taught to improve in recognizing and describing 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.

## 2. 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*<sup>3</sup> which motivated the 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<sup>4</sup> 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,  $s_1, s_2, \dots$ , each of which ranges over some dimension of state-space, and varies with time  $t$ , measured in hours. For simplicity of exposition, suppose that  $s_i(t)$  is the name of a town on the route from Houston to New Orleans (or  $\phi$  to denote no town) where  $L$  might be  $t$  hours after  $L$  left Houston. That is,  $s_1$  ranges over all the town-names along the route. Initially,  $s_1(0) = \text{Houston}$ . The state  $s_1(t) = \text{New Orleans}$  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

<sup>3</sup> The distinction between problems *created* by the real world and problem-statements *presented* by an experimenter, computer-user, teacher or textbook authors seems obvious to us. Those who view a verbal statement of a real-world problem as merely a surrogate or another version of the problem should consider the difference between problems faced by the doctor in the clinic, the business manager in the field or the lawyer in court, in contrast to their facing textbook versions of these problems.

<sup>4</sup> We do not mean to insist on a model which assumes a central role for a passive kind of memory in which rules, hypotheses or other items are stored until retrieved for use. We do assume a store of programs that can generate hypotheses as needed as well as selected hypotheses, plus programs that generate other programs. (Farnham-Diggory, 1972).

stimulus or datum, a question to be initiated. 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,  $a_1, a_2, \dots$ . 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:

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

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  $V_N$ , two 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  $s_1(t) = \text{Houston}$  and  $a_1(0) = 40$  mph, then  $s_1(t+1) = \text{Austin}$ ; weight .8, saliency 1", and  $H2 =$  "If  $s_1(t) = \text{New Orleans}$  and  $t = \text{time the Mardi Gras in New Orleans starts}$ , 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, both 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<sup>5</sup>, classifying it if it is not.<sup>6</sup> The initial input in the above example might be:  $s_1(0) = \text{Houston}$   $s_3(0) =$  "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

<sup>5</sup> This may be the answer to a question, a question or instruction by an experimenter in an experimental situation.

<sup>6</sup> Physical stimuli such as audio visual signals signifying to  $L$  that "he" is in Houston.

parse an input sentence or if the sentence has a word not in  $V_T$ ,  $L$  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<sup>7</sup> 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  $s_1(0) = \text{Houston}$  and  $a_1(0) = 80$ , then  $s(t) = \text{New Orleans}$  for some  $t \leq 6$ ". If that is present, the output is: the decision,  $a_2(0) = \text{"drive to New Orleans"}$ ;  $a_1(0) = 80$  mph; and  $a_3(0) = \text{no further questions}$ . 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  $a_2 = \text{"don't drive"}$  has increased while the weight of an hypothesis containing  $a_2 = \text{"drive to New Orleans"}$  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  $v \cdot T$* ."<sup>8</sup> Once  $L$  has formed this hypothesis – particularly the italicised phrase – he has *recognized and formulated the mathematical* problem which must be posed and solved for  $L$  to make a rational decision. 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  $y$ ?" etc. 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,  $v_1 \cdot T_1 + v_2 \cdot T_2 + \dots$ )? 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

<sup>7</sup> A program for this has been implemented in SNOBOL4.

<sup>8</sup> In words, "If I drive at  $v$  mph from town  $x$  to town  $y$  in  $T$  hours, then I will travel  $v \cdot T$  miles from  $x$  to  $y$ ; this statement holds for any  $v$ ,  $T$ ,  $x$ ,  $y$  and any starting time.

nour of time is associated with  $\nu$  miles of distance, and  $n$  hours are chosen, then the  $n$  miles are associated with  $n \cdot \nu$  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 algorithms in representation theory.

*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.*

If  $L$  is in an environment which stimulates question-asking, such as the above,  $L$  should enrich its memory with general hypotheses. More specifically, if the environment includes conversation, graphics and instrumental tasks involving geometry, arithmetic and algebra, we should obtain increased speed in performance on test tasks in these topics, if such hypotheses are useful. If representations that include such hypotheses are not formed or relevant, no improvement on the test tasks would be expected. It follows that if there is improvement, then the hypotheses have been formed.

### 3. Experiment

#### HYPOTHESIS

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 assumption.

## 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 (Ann Arbor, Michigan). From a group of thirty 4th and 5th graders, two equal groups of 10 each were randomly selected.

## IMPROVEMENT METHOD

The experimental "treatment" group underwent 6 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 20 questions, using various topics, e.g. cryptograms, hidden objects, guessing numbers, etc. 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 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 with the experimental group of 10 children were started. Each child was tested individually for 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, though not in the same room. The testing procedure is detailed next.

### 4. 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<sup>9</sup> ability for recognizing a situation as one requiring certain mathematical operations, such as: (1) “multiplication of a number of items with property 1 by a number of items of property 2 that are associated with one item of property 1 and adding”; (2) “multiplying the dimensions of a rectangle or box to obtain its area or volume”; (3) “determination of a named but as yet non-explicit number, where that name occurs in two statements that together determine its value”.

A description of the four-tasks test follows.

## SET-UP

The subject entered the test room to find 4 tables, D1, D2, D3, D4, and 8 chairs. Each table had associated with it two chairs facing each other on either side of it. The experimenter, E, sat in one chair facing the table. The subject, S,<sup>10</sup> sat in the other chair, facing E. 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” × 23”; (b) three cardboard

<sup>9</sup> As distinct from capacity, which would require a different measuring instrument and a different tabulation.

<sup>10</sup> This corresponds to what we called *L* (for learner) in Section 2.

houses on (a) labeled MacDonald's, 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) three signs placed at the three different roads: Sign 1 read: "Speed Limit is 2 seconds per inch, distance to MacDonald's is 18 inches"; Sign 2 read: "Speed Limit is 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) three spools of orange, white and black wire; (b) price tags — "White wire, 11c per inch", "Orange wire, 13c per inch", "Green wire, 7c per inch".

D4 had on it: (a) a 36" straight steel wire,  $\frac{1}{16}$  inch in diameter; (b) 80 1 inch square cardboard tiles.

#### TASKS

The items on D1 were associated with Task 1, T1; D2 with T2, and D3 with T3, and D4 with T4. 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.

Task 4 — to make a picture frame. The picture should be made up of all these tiles inside the frame. The frame should be made with this wire and use up all the wire with nothing left over."

After *S* was seated, *E* told *S* this was a game and that he had in mind four tasks involving the objects on the four 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 to find out." This continued until either half

an hour was up or *S* had asked questions indicating that he had figured out the 4 tasks in a way that was *equivalent* to the following 4 statements:

1. The *time* (in seconds) for the car to go from the start to the Bank is the *speed* allowed on the Start-MacDonald's 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 (cube 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 per inch, *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.

4. The picture frame we should bend this wire into is a *rectangle* with an *area*, in square inches, equal to the *number of tiles* if they are 1 inch squares and a perimeter equal to the length of the wire; the *length* and *width* of the rectangle must be such that their *product* is the number of tiles, and *twice their sums* is the length of the wire.

#### 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", etc. 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* has recognized and formulated a problem equivalent to statements 1 – 4.

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. A sample of these is given in the appendix.

#### 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.

## 5. Results and Discussion

TABLE I

Summary of Analysis of Variance

Source of Variation	Sum of squares	Degrees of freedom	Mean square	$F$
Between methods	41042	1	41042	11.95*
Experimental error	61844	18	3435.78	
Total	102886	19		

\*  $F_{.99}(1, 18) = 11.95$ , we obtain a significant  $F$  at the .01 level.

Table 1 shows the results of one-way analysis of variance of this data. It proves that the null hypothesis,  $EX_C = EX_T$ , has to be rejected at a .01 level of significance. 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 those 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  $\times$  distance) given that he did well on Task 3 (arithmetic on price  $\times$  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  $X_{3T}$  give 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, though variance should also be taken into account.

That none of the 20 subjects were able to recognize and formulate Task 4 (algebra) within the time limit may be due to the nature of the task. It does not necessarily indicate that 4th and 5th graders cannot formulate algebra problems. The idea of creating sentences about area and perimeter simply did not occur to them. It is also curious that while most of the children asked if the wire frame should form a "square", none asked if it should be a "rectangle". Perhaps children must solve such algebra problems before they can recognize and formulate them.

During the training sessions, some of the children were able to form such algebra problems as "The product of two numbers is 35 and they differ by 2. What are they?". They "saw" the answer rather than applied algebraic operations which, of course, they had never encountered. But they could invent analogous problems, especially in concrete situations rather than such highly abstract arithmetic problems.

Children on their own (e.g. during recess) can go to a playground (that they may not have seen before) with children not previously known to them and create (rather than just follow) rules in inventing original games. A similar creative competence seems to be at work when children are encouraged to inquire freely and to invent problem-statements and rules in tasks requiring mathematical thinking. Our observations of the children confirm that this kind of creative approach to mathematical thinking is more

reinforcing than the more conventional aim of making them proficient at manipulating mathematical symbols (competing with a computer).

One of our motivations for studying problem-formulation in children was our prior experience with attempting to teach this to graduate students. We had<sup>11</sup> to help students of urban/regional planning (Kochen, 1972) learn mathematical thinking to pass a Ph.D. qualifying examination in analytics. Some of these students had exposure to 4th and 5th grade mathematics (but little more) decades ago, and had either forgotten or never learned it (perhaps because they hated it). A few seem never to have been exposed even to that. We observed that not only these, but even some of the best Ph.D. candidates in mathematics, had the greatest difficulty in picking and formulating problems by themselves. Even greater was the difficulty of encountering or sensing a real problem, which has a natural mathematical formulation, and recognizing it as denoted by the kind of mathematical story-problems that has been seen in textbooks, courses and even in seminars. Formulating a thesis problem is therefore a great hurdle for many graduate students. It is very hard to get them to learn this so late in their educational development. That is why we explored how easily children could learn to pick and formulate their own problems.

If our view is sound that problem-formulation is the result of forming and using hypotheses by anything like the processes we suggest, then a further analysis plus experimental corroboration of this "representation theory" could lead to more systematically designed improvement (Churchman, 1972) and testing procedures. Variants of the experiments reported here should be repeated with larger samples from different populations and other tasks. More crucial experiments to lead to strong inferences about the psychological relevance of representation theory and its application to education are planned.

## 6. Summary of Conclusions

We conceptualized the process of recognizing and formulating real problems as mathematical story-problems 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 4th and 5th 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.

---

<sup>11</sup> We did prepare a sequence of graded exercises to help them translate in a step-by-step procedure, preformulated mathematical story problems into mathematical formulations. Some of these problem-statements contained fewer givens than necessary to solve the problem; some contained more. In some the givens were contradictory.

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 it 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.

## References

- Amarel, S. (1971). "Representations and modeling of problems of program formulation," in *Machine Intelligence*, (B. Meltzer and D. Michie, eds.), Edinburgh: Edinburgh University Press.
- Badre, Albert N. (1973). "On hypotheses and representational shifting in ill-defined problem-situations," Ph.D. thesis, University of Michigan, Ann Arbor.
- Badre, Albert N. (1974). "The effects of hypothesis shifting on the solving of insight problems," in *Information Utilities*, Proceedings of the 37th ASIS Annual Meeting, Atlanta, Georgia, Vol. II, pp. 194–196.
- Carroll, J. B. and Freedle, R. O. (eds.) (1972). *Language Comprehension and the Acquisition of Knowledge*, New York: Wiley.
- Churchman, C. West (1972). *The Design of Inquiring Systems*, New York: Basic Books.
- Eichholz, R. E. and D'Affer, P. G. (1964); *Elementary School Mathematics* (Books 4 and 5), Palo Alto, California: Addison-Wesley Publishing Company.
- Farnham-Diggory, S. (ed.) (1972). *Information Processing in Children*, New York: Academic Press, (esp. Chapters by Newell, p. 128, 132 and by Simon, p. 22, 76–78).
- Kochen, M. (1960a). *Cognitive Mechanisms*, New York: IBM, Yorktown Heights.
- Kochen, M. (1960b). "Experimental study of 'hypothesis-formation' by computer," in *Information Theory* (C. Cherry, ed.), London: Butterworth Company.
- Kochen, M. (1969); "Automatic question-answering of English-like questions about simple diagrams," *Journal of the Association for Computing Machinery*, 16: (January) 26–48.
- Kochen, M. (1971). "Cognitive learning processes: an explication," in *Artificial Intelligence and Heuristic Programming* (N. Findler and B. Meltzer, eds) Edinburgh University Press, Edinburgh, 261–317; Revised version in German translation, 1973.
- Kochen, M. and Badre, A. N. (1974). "Questions and shifts of representation in problem-solving," *American Journal of Psychology* 87, 3: 369–383.
- Kochen, M. and Dreyfuss, G. (1972). "An experiment in teaching college mathematics," *International Journal of Mathematical Education in Science and Technology*, 3: 315–328.
- Kochen, M. and Uhr, L. (1965). "A model for the process of learning to comprehend," pp. 94–104 in *Some Problems in Information Science*, New York: Scarecrow Press.
- Kochen, M., MacKay, D. M., Maron, M. E., Scriven, M., and Uhr, L. (1964). "Computers and comprehension," *RAND Report*; also in *The Growth of Knowledge* (Kochen, ed.), Wiley, New York, 1965; see also "Adaptive mechanisms in digital 'concept' processing," same book.





T: Good. Yes. Now what does this tell us about the possibility? What number in the 500s, 600s, 700, 800, and 900s, *can* be the same?

— — — is 555, 666, 777, 888, 999

S: Are all the numbers even?

T: What do you mean by even?

S: like 4, 6, 8.

T: No, so — — — is an odd number; 555 or 777 or 999.

S: Is it 777?

T: No. — — — is either 555 or 999.

S: 999?

T: Yes.

### Examples of “story-problems” children formulated during treatment

The weight of 2 shoes is equal to the weight of 1 ball + 20 grams. The weight of 1 shoe is equal to \_\_\_\_\_ .

A man has 8 cars and 3 engines. How many ways can he combine the cars and engines without doing it twice?

If the weight of 2 shoes is equal to 10 pounds, how much is the weight of one shoe if both shoes weigh the same?

Joe weighs 114 pounds and this is 3 times what Jerry weighs. What is the weight of Jerry?

Nine innings in a baseball game. 24 people ate 3 hot dogs each in each inning. Each hot dog cost 10c. How much money was spent?

If  $\frac{2}{4}$  of the apples are rotten and  $\frac{1}{2}$  are green, and there are 10 apples, how many are rotten?

A brick on one scale weighs 10 pounds and you have to balance it on the scale. What bricks would you use if you have 5, 4, 8, 2, and 1 pound bricks?

If you had 4 sheep and you called a tail a leg, how many legs would you have?