At this moment, there are many promising research efforts on the relation of language structure to the comprehension of prose. Considerable attention has been paid to the comprehension of simple, relatively stereotyped units such as folk tales (Colby, 1973), problem instructions (Hayes & Simon, 1974), and simple narratives (Kintsch, this volume; Kintsch & van Dijk, 1975; Rumelhart, 1975). In this paper, we will present data on the comprehension of those popular 20th century fables called algebra word problems.

Algebra word problems are interesting objects of study for a number of reasons. First, they incorporate rigorous standards for accurate comprehension, which is not always true for other types of materials, such as proverbs. For example, there is wide disagreement about the meaning of the adage “A rolling stone gathers no moss,” even though it is very often used. (The disagreement is about whether it is good or bad to be a rolling stone.) Misinterpreting an algebra word problem, however, typically leads to an error in the solution that can readily be detected. Second, there is a standard and widely known representation used for solving algebra word problems – algebraic equations. This representation is useful for tracing the process of comprehension in individual subjects and for detecting differences in the comprehension process between subjects. Finally, the study of comprehension of algebra word problems can have important practical applications in education.
The processing strategies used for algebra word problems can be related to more general issues of prose comprehension. While there are important similarities among the various theoretical discussions of prose comprehension (Bartlett, 1932; Kintsch, this volume; Kintsch & van Dijk, 1975; Minsky, 1975; Rumelhart, 1975; Thorndyke, 1975; van Dijk, this volume), we find two kinds of differences in their approaches. One is a difference in emphasis on content vs. form. Kintsch and van Dijk, and Rumelhart as well, emphasize the importance of formal structuring of the narrative, as in a sequence of episodes followed by a moral. Others, like Minsky, emphasize the importance of the reader’s knowledge of specific semantic content in comprehending a passage. For example, comprehending a passage about a birthday party presumably requires that the reader invoke his specific knowledge about such events.

The second difference concerns the distribution of decision processes involved in text comprehension. According to Minsky, the comprehension process involves major decisions made early, followed by a number of relatively minor decisions to fill in details. The reader presumably evokes the appropriate knowledge fairly early. For Kintsch and van Dijk and for Rumelhart, major decisions or branch points are distributed throughout the text comprehension process as readers identify various structural aspects of the text. These investigators do not place emphasis on major early decisions.

We have chosen to designate an approach as a “text grammar” approach if it emphasizes the importance of formal structure and postulates distributed decision processes, and as a “schema” approach if it emphasizes the importance of semantic knowledge and major decisions occurring early in the comprehension process. We recognize that these designations are, to a great extent, arbitrary. Our purpose in proposing them is not to describe existing positions but rather to sharpen theoretical distinctions to aid in identifying researchable questions.

In this paper, we will be concerned with only one of the differences between the two approaches, the difference in the distribution of decision processes. The empirical question is whether the data lead us to emphasize the subject’s choice of alternative processing strategies made early in reading, to emphasize a single processing strategy in which the idea of choice early in reading is not important.

Both of these strategies were incorporated in a program called STUDENT (Bobrow, 1968) that was designed to solve algebra word problems. STUDENT attacks most problems by using a process called direct translation, in which it translates successive sentences of the problem text into equations and then tries to solve the equations. The direct translation process involves no choice at the time of reading; it is applied in essentially the same way to all problems. Thus, this aspect of the theory is compatible with the idea of a single processing strategy.

The direct translation process is composed of two sub-processes: mandatory substitution and function tagging. Mandatory substitution replaces strings such
as "twice" by "two times," using a dictionary of synonyms. Function tagging labels some of the words with tags that fall into three main categories:

(1) Object names — the quantities that will be variables in the algebraic equations.
(2) Substitutions — quantities that may be replaced by another string. Some substitutions are mandatory, while others are only tried if the problem cannot be solved without them.
(3) Operators — words that signal that some type of operation should be performed, e.g., plus, times, equals, etc.

The program uses these categories and a simple syntactical parsing scheme to extract the algebraic equations from the text. As an example of how STUDENT translates a text, consider the following simple problem:

The number of applicants is twice the number of jobs.
The number of jobs is 5.
What is the number of applicants?

To solve this problem, STUDENT would translate these sentences into the equivalent of the following equations and solve them.

(\text{The number of applicants}) = 2 \times (\text{The number of jobs})
(\text{The number of jobs}) = 5
X = (\text{The number of applicants})

In the simplest cases, STUDENT makes very little use of semantics. A variable, such as "The number of applicants," is simply an arbitrary noun phrase to STUDENT. "The number of job applicants," since it is a different noun phrase, would be treated as a different variable.

If the direct translation process fails to yield a solution, STUDENT will search its memory for relevant global knowledge. For example, if the problem contains a keyword such as "distance" or "miles," STUDENT will retrieve the equation, "distance = rate \times time," add it to the list of equations produced by direct translation, and try again. The direct translation process, with or without retrieval of global knowledge, involves no choice at the time of reading. The direct translation process can serve as a good first approximation in describing some human behavior in solving algebra word problems (Paige & Simon, 1966). However, direct translation, as it is implemented in STUDENT, cannot account for human solution processes which rely on semantic knowledge. For example, STUDENT cannot recognize that a problem such as the following (from Paige & Simon) is contradictory:

The number of quarters a man has is seven times the number of dimes he has.
The value of the dimes exceeds the value of the quarters by two dollars and fifty cents. How many has he of each coin?
The contradictory nature of this problem cannot be recognized without recourse to semantic knowledge. To be a reasonable prospect as a model for human behavior, direct translation must at least be augmented by semantic knowledge. Thus, we view the direct translation process as similar to a text grammar, because it operates line by line and involves no choice of schema at the time of reading. It is unlike a text grammar in that it requires neither story structure nor semantic consistency.

In addition to the direct translation process, STUDENT includes a process that does involve choice of schema at the time of reading. If the problem is an "age" problem, as indicated by phrases such as "as old as," "years old," and "age," then the processing of the problem is altered by the addition of special heuristics. These special heuristics change the objects the program attends to and the way names are assigned to variables. They take into account the fact that objects age equal amounts in equal times and that "number of years ago" is a variable; and they accomplish several other processing changes appropriate to age problems. The triggering at reading time of this special set of heuristics for processing age problems is, in our view, equivalent to the introduction of an age-problem schema.

If schemas are an important aspect of human processes for solving algebra word problems, then one should be able to verify the following four assertions:

1. **People can categorize problems into types.** If each problem were unique, there would be no relevant information about problem types to retrieve from long term memory.
2. **People can categorize problems without completely formulating them for solution.** If the category is to be used to cue a schema for formulating the problem, the schema must be retrieved before formulation is complete.
3. **People have a body of information about each problem type which is potentially useful in formulating problems of that type for solution.** This information can aid the problem solver in such ways as directing attention to important problem elements, making relevance judgments, retrieving information concerning relevant equations, etc. We would expect such a body of information to act as a schema in Piaget's sense, and to assimilate information in the text to itself.
4. **People use category identifications to formulate problems in the course of actually solving them.** Thus, it must be shown not just that such categories exist and are associated with potentially useful information, but that information associated with categories is in fact used in formulating the problem when the task is to solve problems, rather than merely to categorize them.

It will be our task in this paper to try to verify these four assertions.
EXPRESSMENT 1

To determine whether or not people could categorize algebra word problems into characteristic types, we asked subjects to sort problems selected from a high school algebra text. We selected 76 algebra word problems from the text (Dolciani, Berman, & Wooton, 1973), with approximately 5 selected from each of the 16 chapters, and presented the problems in a random order to each of 14 subjects, high school and college students selected for knowledge of algebra. The subjects were asked to sort the cards into piles by problem type, where “Problem type” was not further specified. When they had finished the sorting task, which typically took an hour, they were asked to describe the properties of the problems in each of the categories they had identified.

Results. On the average, the subjects identified 13.5 problem categories containing more than one problem each. Approximately 5% of problems were placed either in categories with a single member or in a “wastebasket” category. The problems clustered into between 16 and 18 distinct categories with considerable agreement among subjects in categorizing the problems. Table 4.1

<table>
<thead>
<tr>
<th>Table 4.1: Representative Problem from Each of the 18 Clusters</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Triangle</td>
</tr>
<tr>
<td>2. DRT</td>
</tr>
<tr>
<td>3. Averages</td>
</tr>
<tr>
<td>4. Scale conversion</td>
</tr>
<tr>
<td>5. Ratio</td>
</tr>
<tr>
<td>6. Interest</td>
</tr>
</tbody>
</table>
### TABLE 4.1 (continued)

<table>
<thead>
<tr>
<th>7. Area</th>
<th>A box containing 180 cubic inches is constructed by cutting from each corner of a cardboard square a small square with side 5 inches, and then turning up the sides. Find the area of the original piece of cardboard.</th>
</tr>
</thead>
<tbody>
<tr>
<td>8. Max-min</td>
<td>A real-estate operator estimates that the monthly profit $p$ in dollars from a building $s$ stories high is given by $p = -2s^2 + 88s$. What height building would he consider most profitable?</td>
</tr>
<tr>
<td>9. Mixture</td>
<td>One vegetable oil contains 6% saturated fats and a second contains 26% saturated fats. In making a salad dressing how many ounces of the second may be added to 10 ounces of the first if the percent of saturated fats is not to exceed 16%?</td>
</tr>
<tr>
<td>10. River current</td>
<td>A river steamer travels 36 miles downstream in the same time that it travels 24 miles upstream. The steamer's engines drive in still water at a rate which is 12 miles an hour more than the rate of the current. Find the rate of the current.</td>
</tr>
<tr>
<td>11. Probability</td>
<td>In an extra-sensory-perception experiment, a blindfolded subject has two rows of blocks before him. Each row has blocks numbered 1 to 10 arranged in random order. The subject is to place one hand on a block in the first row and then try to place his other hand on the block having the same numeral in the second row. If the subject has no ESP, what is the probability of his making a match on the first try?</td>
</tr>
<tr>
<td>12. Number</td>
<td>The units digit is 1 more than 3 times the tens digit. The number represented when the digits are interchanged is 8 times the sum of the digits.</td>
</tr>
<tr>
<td>13. Work</td>
<td>Mr. Russo takes 3 minutes less than Mr. Lloyd to pack a case when each works alone. One day, after Mr. Russo spent 6 minutes in packing a case, the boss called him away, and Mr. Lloyd finished packing in 4 more minutes. How many minutes would it take Mr. Russo alone to pack a case?</td>
</tr>
<tr>
<td>14. Navigation</td>
<td>A pilot leaves an aircraft carrier and flies south at 360 m.p.h., while the carrier proceeds N30W at 30 m.p.h. If the pilot has enough fuel to fly 4 hours, how far south can he fly before returning to his ship?</td>
</tr>
<tr>
<td>15. Progressions</td>
<td>From two towns 363 miles apart, Jack and Jill set out to meet each other. If Jill travels 1 mile the first day, 3 the second, 5 the third, and so on, and Jack travels 2 miles the first day, 6 the second, 10 the third, and so on, when will they meet?</td>
</tr>
<tr>
<td>16. Progressions-2</td>
<td>Find the sum of the first 25 odd positive integers.</td>
</tr>
<tr>
<td>17. Physics</td>
<td>The speed of a body falling freely from rest is directly proportional to the length of time that it falls. If a body was falling at 144 ft. per second 4 1/2 seconds after beginning its fall, how fast was it falling 3 3/4 seconds later?</td>
</tr>
<tr>
<td>18. Exponentials</td>
<td>The diameter of each successive layer of a wedding cake is 2/3 the previous layer. If the diameter of the first layer of a 5-layer cake is 15 inches, find the sum of the circumferences of all the layers.</td>
</tr>
</tbody>
</table>
shows a representative problem in each cluster. The number of subjects who placed the two members of a pair in the same category was tabulated for each of the approximately 3000 possible problem pairs. We will call this number, which ranges in value from 0 to 14, the similarity index. A problem cluster is defined as a set of problems such that the average similarity index for all pairs of problems in the cluster is equal to or greater than some criterion number.

With the criterion set at 10, 16 clusters, which included a total of 50 of the 76 problems, were identified. When the criterion was reduced to 7, two new clusters, physics and exponentials, were identified, and problems were added to some of the old clusters. However, no two of the old clusters were combined into one cluster by this change in criterion. The resulting 18 clusters included 64 of the 76 problems.

It is clear that people can categorize algebra word problems and that when they do so, they show considerable agreement about what the categories are. The fact that such problem categories exist is not evidence that they are used in formulating problems for solution. It might be that a subject had to formulate the problem for solution in order to categorize it. Certainly there was nothing in Experiment 1 to prevent the subjects from doing this. Experiment 2 was performed to determine how early in the course of reading the subjects could categorize the problem.

EXPERIMENT 2

If we could show that subjects can categorize a problem very early in the course of reading the problem text, we could be reasonably certain that to categorize the problem the subject did not need first to formulate it for solution. The following experiment was run to investigate this question.

Eight problems were chosen from Table 4.1 representing the problem clusters probability, interest, river current, triangles, max-min, scale conversion, progressions, and work, plus an additional interest problem and an additional probability problem. The six subjects, all familiar with algebra, were tested in individual sessions.

Problems were read to the subjects one part at a time. In some cases a part was as short as a noun phrase, e.g., “A river steamer . . . ,” or a dependent clause, e.g., “At the end of each month, . . . .” In other cases, the parts were complete sentences. At the subject’s request the experimenter would review the entire text read up to the time of the request. The subject was asked after each text segment to attempt to categorize the problem and to state what information he expected to receive and what question he expected to be asked. After the entire set of problems had been presented, the subject was asked to define the problem types he had mentioned in discussing the problems. For example, if he had
described a problem as a distance problem, he was asked, "What is a distance problem?"

One problem, the max-min problem, was categorized correctly by only one of the subjects even with complete information. This problem was therefore dropped from the analysis. For three of the problems (one each out of the categories physics, DRT, and triangle), 81% of the categorizations were correct when only the initial noun phrase had been presented.

Results. Averaged over the nine problems, half of the subjects categorized the problems correctly after hearing less than one-fifth of the text. Table 4.2 shows the number of words which had been presented at the time that more than half of the subjects correctly categorized the problem. The category thresholds are given in number of words read (column 2) and in percent of the total problem read (column 4).

For some problems correct categorization may occur in 2 stages. This happens because some problem categories are specializations of other categories. For example, a river current problem is a particular kind of distance-rate-time problem. The subject may recognize that a problem is of the more general type before he recognizes it as a case of the more particular type. Thus, he may categorize a river current problem first as a distance problem and later as a distance problem of the river current type. Four of the problems may be viewed as specializations of more general problem types. The triangle and river-current problems are special kinds of distance problems and the two interest problems are special kinds of money problems. All of the subjects who did not initially categorize these problems into their more special type categorized them initially into their more general type.

<table>
<thead>
<tr>
<th>Problem type</th>
<th>Words read at threshold</th>
<th>Total words</th>
<th>Percent of total read at threshold</th>
</tr>
</thead>
<tbody>
<tr>
<td>Triangle</td>
<td>9</td>
<td>54</td>
<td>17</td>
</tr>
<tr>
<td>Scale conversion</td>
<td>3</td>
<td>58</td>
<td>5</td>
</tr>
<tr>
<td>Interest</td>
<td>4</td>
<td>22</td>
<td>18</td>
</tr>
<tr>
<td>Max-min</td>
<td>–</td>
<td>36</td>
<td>–</td>
</tr>
<tr>
<td>River current</td>
<td>3</td>
<td>46</td>
<td>7</td>
</tr>
<tr>
<td>Probability</td>
<td>6</td>
<td>80</td>
<td>8</td>
</tr>
<tr>
<td>Work Area</td>
<td>17</td>
<td>56</td>
<td>30</td>
</tr>
<tr>
<td>Progressions Area</td>
<td>15</td>
<td>53</td>
<td>28</td>
</tr>
<tr>
<td>Interest Area</td>
<td>8</td>
<td>26</td>
<td>31</td>
</tr>
<tr>
<td>Probability</td>
<td>10</td>
<td>66</td>
<td>15</td>
</tr>
<tr>
<td><strong>Av.</strong></td>
<td><strong>44.7</strong></td>
<td></td>
<td><strong>17.7</strong></td>
</tr>
</tbody>
</table>
These results verify the assumption that subjects can categorize problems without completely formulating them for solution. The next question we must answer is what information subjects have stored about a category at the time of categorization. Our answer to this question must be incomplete because of the nature of the data. We can only interpret our subjects' comments as specifying a very conservative lower bound on what they knew. Some subjects gave no information other than the category name at the time of categorization. Others described the problem type, gave an instance of the type, and specified what sorts of information they expected to receive and what question they would be asked. For example, after hearing the three words, "A river steamer..." from a river current problem, one subject said, "It's going to be one of those river things with upstream, downstream, and still water. You are going to compare times upstream and downstream — or if the time is constant, it will be the distance." Another subject said, "It is going to be a linear algebra problem of the current type — like it takes four hours to go upstream and two hours to go downstream. What is the current — or else it's a trig problem — the boat may go across the current and get swept downstream." After hearing five words of a triangle problem, one subject said, "This may be something about 'How far is he from his goal' using the Pythagorean theorem."

It is clear that at the time of categorization, at least some of the subjects have knowledge about problem types that can be useful in formulating the problem for solution. However, we have yet to show that subjects use this knowledge in solving algebra word problems. Experiment 3 was designed to provide evidence concerning this issue.

EXPERIMENT 3

Two expectations guided the design of this experiment. The first was that if knowledge about problems influenced the solution process, those influences would be apparent in problem-solving protocols. The second expectation was that problem categorization would be influenced more by the problem's superficial structure — its cover story — than by its underlying algebraic structure. Thus, if two problems have the same algebraic structure but different cover stories suggesting different problem categorizations, they will show evidence in problem-solving protocols of characteristic differences in the way they are solved. To test this expectation, protocols were collected from two graduate students solving a set of nine problems.

Nine problems were generated as variations on three standard problem types — the river current type, the age type, and the work type. By a "standard" problem type, we mean a problem which has a characteristic matching of cover story to underlying problem structure. Thus, a standard river current problem is
one with a cover story involving vehicles (usually boats) moving in a current and
a structure involving the equations:

\[
\text{Distance} = \text{rate} \times \text{time},
\]
\[
\text{Rate downstream} = (\text{rate in still water}) + (\text{rate of current}),
\]
\[
\text{rate upstream} = (\text{rate in still water}) - (\text{rate of current}).
\]

A typical age problem is one with a cover story involving ages and a structure in
which increases in one variable, e.g., time, cause uniform increases for all
relevant objects, e.g., people, in another variable, e.g., age. A typical work
problem is one with a cover story involving production and work, and a
structure involving the relation:

\[
\text{Amount produced} = \text{rate of work} \times \text{time}.
\]

The nine problems consisted of three sets of three isomorphic problems in a 3
\(\times\) 3 matrix. This matrix was constructed so that all the problems in a row were
isomorphic to each other but had different cover stories (age problem, work
problem, or river current problem), while all the problems in a column had the
same type of cover story (e.g., all were 'age' problems) but differed in their
algebraic structure. Thus, the diagonal problems were standard problems and the
off-diagonal problems were non-standard problems which were isomorphs of the
standard problems.

Results. There appear to be two different procedures used by the subjects in
extracting the appropriate equations from the text of the problems. With one
procedure, the subject reads the entire problem before he formulates any
equations or notes any relationships explicitly. The second procedure is a
line-by-line approach which formulates equations and relationships explicitly
while reading the text. This procedure is illustrated in the protocol of Subject 1's
solution of an isomorph of a DRT problem with an age cover story:

In a 'Fathers and sons' tennis match, the Greenes are playing doubles against
the Browns. Mr. Greene is four times as old as his son and Mr. Brown is five
times as old as his son. Greene's son is one year older than Brown's son. If Mr.
Greene is the same age as Mr. Brown, how old is Brown's son?

This resulted in the following protocol:

'In a fathers and sons tennis match the Greenes are playing doubles against the
Browns.' Greenes, Browns. Mr. Brown is four times older, so Mr. G. is four
times older than son. So Gf equals four Gb. Uh, and Mr. Brown, Bf, is one, 'is
five times', this equals five, of Bs . . . Ok, 'Greene's son,' Gs, 'is one year older',
is equal Bs plus 1. 'If Mr. Greene is the same age as Mr. Brown,' Bf equals Gf,
'how old is Brown's son?' Ok, so what we want to know is, we want to know
what Bs is. So I guess one thing we could do is we could just substitute the
relation of the Bf's and the Gf's of equal. So we have, five Bs is equal four Gb
and, so since \( Gb \) equals \( Bs \) plus 1, we'll say four \( Bs \) plus 1 is equal five \( Bs \). And that's going to give us, \( Bs \) equals four.

The subjects' use of stored information in formulating a problem may be illustrated by comparing the solutions of the tennis problem with the solutions of its isomorph, a DRT problem shown below:

A candle factory has two workers, Jones and Smith. Jones makes candles at the rate of 60 candles per hour and Smith, at the rate of 75 candles per hour. Jones spends one hour more each day making candles than Smith. If Jones makes the same number of candles each day as Smith, how long does Smith work?

Subject 1's protocol follows:

Sixty per hour and Smith makes 'em at 75 per hour. Jones spends... OK, Jones, Smith, Jones spends one more hour each day making candles than Smith. So \( X + 1 \) versus \( X \). If Jones makes the same number of candles each day as Smith, how long does Smith work. OK, so what we basically want to say is that 60 hours times \( X + 1 \) equals 75 hours times \( X \).

The formula \( 60(X + 1) = 75X \) appears to be put together in a single step in this protocol, while the corresponding formula in the tennis problem, \( 4(Bs + 1) = 5Bs \), takes at least 3 steps to assemble. Essentially the same differences in processing between these two problems were observed for Subject 2.

One may interpret these differences as resulting from the ready availability of the relation “Amount = Rate of work \(*\) time” for this problem when it is presented with a work cover story, but not when it is presented with an age cover story.

In addition to the evidence of line-by-line processing, there is also evidence that the subjects did classify some problems early in the course of solving them and did retrieve useful information about them from memory. Subjects sometimes made explicit statements of the problem type, such as, “Sounds like another \( x + y \).” and there are also more implicit recognitions of the problem types, as in “Oh (expletive deleted), one of these (expletive deleted)! I always get confused.” Not only does this show that the problem type is recognized, but that there is some information stored as to its difficulty or the subject’s experience with this type of problem in the past. Clearly, information of this sort could be helpful in the solution of algebra word problems.

It appears that the line-by-line procedure is utilized mainly in the solution of the non-standard problems: 9 of the 12 non-standard problems were solved using this procedure, but only one of the six standard problems. It also appears that the subject recognizes that this line-by-line process is different from his normal procedure. For example, Subject 1, after reading the first line, stops and formulates two equations. Then he says, “Notice that I haven't read the problem
through to the end though." One interpretation for these results is that the
line-by-line procedure is a default process which is used if the problem is not
successfully matched to one of the subject's schemas.

Experiment 3 has provided evidence that:

(1) Subjects may use a line-by-line procedure, especially in solving the non-
standard problems,
(2) Subjects classified some of the problems while solving them, and
(3) In solving standard problems, problem formulation was facilitated by
information retrieved from memory.

These results might be taken as indicating that schema are important only in the
formulation of standard problems—problems in which the semantics of the cover
story match the problem structure in an expected way. To test this conjecture,
in Experiment 4 we moved as far from standard, familiar problems as we could.

EXPERIMENT 4

If problem categorization depends on the semantics of the cover story, then we
might expect that such categorization would be absent in "nonsense" problems.
Pilot observations, however, suggested that this was not the case, and that in fact
subjects sometimes guess at a categorization and supply an appropriate cover
story in order to "interpret" the nonsense problems. These observations seemed
to provide strong evidence of the existence of problem schemas. This experiment
was performed to confirm the pilot observations.

Protocols were collected from two psychology graduate students solving the
following two "nonsense problems." These problems were constructed by taking
algebra word problems and replacing some of the content words with nonsense
words. For example, substituting nonsense words in a standard mixture problem
yielded the following:

According to ferbent, the optimally fuselt grix of voipe umolts five stens of
viope thrump 95 bines per sten. In order to embler some wuss voipe, each grix
will umolt one sten at 70 bines per sten. If the grix is to be optimally fuselt,
what should the bines per sten of the rest of the voipe be?

The following is a standard work problem in which some words have been
replaced by nonsense words:

Chort and Frey are stimpling 150 fands. Chort stimples at the rate of four
fands per yump and Frey at the rate of six fands per yump. Assuming that
Chort and Frey stimple the same number of yumps, how many fands will
Chort have stimpled when they finish?

Results. Both subjects solved both of the nonsense problems correctly and in
three of the four cases explicitly categorized the problem prior to solution. For
example, Subject 1 said of the ‘Chort and Frey’ problem, “Sounds like a straightforward proportion.”

In the following segment of protocol, Subject 2 seems to treat the problem either as a work problem or as some other variety of rate problem:

You have one time unit and you have two guys going at it at different rates. You have a given total of things that have to be stimpled and you have two different rates going into it.

One of the most interesting phenomena to observe when subjects solve nonsense problems is their paraphrasing of the problem as a familiar problem type. Subject 2 paraphrased the ‘grix of voipe’ problem in the following manner:

There is some sort of machine, let’s call it the grix. Which receives some sort of fuel, called a sten. And it produces, let’s say work, out of that sten, usually whips out 90 bines per unit of sten. Now for some obscure reason one of those input units will be converted at 70, the grix is to be optimally fuselt...

Commenting on what he was doing in solving the nonsense problems, Subject 2 said:

What I was trying to do all the time was to try to fit it (the problem) into some normal schema... As soon as I could pop it into a frame where I could deal with things normally, it was much easier to deal with.

Of the two nonsense problems used, the ‘Chort and Frey’ problem was solved quickly and easily by both subjects. Both indicated that the structure of the problem was familiar and they seemed to retrieve an equation for use in its solution. The simplicity of this problem was probably due to the fact that only three nonsense words were used and this left the question, which is a very important and salient feature of an algebra word problem, relatively intact. The subjects experienced more difficulty, however, with the ‘grix of voipe’ problem. Both subjects immediately noticed the rates of bines per sten, and this seemed to provide them with a key to use in attempting to decipher the problem. About one-third of the way through his solution, Subject 2 says,

Well, the only thing I’ve managed to extract so far is that bines per sten, sten is not a time unit which I thought it was originally, but seems to involve objects.

The results of this study show that subjects may use schemas in solving even very non-standard problems. It also yields evidence that there are problem schemas stored in memory which can be used in problem solution.

EXPERIMENT 5

Although Experiment 3 provided some useful evidence on the influence of problem schemas on the solution process, we felt that more and perhaps
different kinds of information could be obtained by two changes in experimental design. The first change was to use a more complex problem involving some irrelevant material to provide an opportunity for relevance judgements to emerge in the protocols. The second change was to build a second theme into the problem which would suggest an alternate but incorrect problem categorization. We hoped to provide ourselves with the opportunity to compare the behavior of subjects who solved the same problem but who categorized it differently.

Protocols were collected from six subjects solving the “Smalltown” problem shown in Table 4.3. The problem is a standard distance-rate-time problem to which some irrelevant information concerning the triangular relation of three problem elements has been added. This was done to mislead subjects into classifying the problem as a triangle problem.

Results. Three of the subjects, S2, S3, and S5, attended to the irrelevant triangle information and three, S1, S4, and S6, did not. Of the subjects who attended to the triangle information, all three specifically mentioned the triangle, and two noticed that it was a right triangle. All three drew triangles and all three identified the five-mile distance as the hypotenuse of the triangle. One subject, S2, retrieved the Pythagorean theorem and another tried to do so. Subject S2 misread “four minutes” as “four miles,” interpreted this distance as the length of one of the legs of the triangle and applied the Pythagorean theorem to compute the length of the other leg. None of the subjects explicitly categorized the problem as a triangle problem. However, S5 said, “Oh, so the . . . five miles apart . . . is going to be the diagonal. This is turning out to be a different type of problem than I thought.” S3 said, “I suppose the critical thing to know—I think—is how far these are [the legs of the triangle] if this [the hypotenuse] is five miles.”

Of the three subjects who did not attend to the triangle information, two explicitly categorized the problem very early as a variety of distance-rate-time

<table>
<thead>
<tr>
<th>TABLE 4.3</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Smalltown Problem</td>
</tr>
<tr>
<td>Because of their quiet ways, the inhabitants of Smalltown were especially upset by the terrible New Year's Eve auto accident which claimed the life of one Smalltown resident. The facts were these. Both Smith and Jones were New Year's Eve babies and each had planned a surprise visit to the other on their mutual birthday. Jones had started out for Smith's house traveling due east on Route 210 just two minutes after Smith had left for Jones's house. Smith was traveling directly south on Route 140. Jones was traveling 30 miles per hour faster than Smith even though their houses were only five miles apart as the crow flies. Their cars crashed at the right angle intersection of the two highways. Officer Franklin, who observed the crash, determined that Jones was traveling half again as fast as Smith at the time of the crash. Smith had been driving for just four minutes at the time of the crash. The crash occurred nearer to the house of the dead man than to the house of the survivor. What was the name of the dead man.</td>
</tr>
</tbody>
</table>
problem. S1 said, "It looks like a distance problem. So Jones is going east two minutes after Smith is going west. So it might be an 'overtake' problem." S4 said, "It sounds to me like this is a problem involving the speed of where they are going and when they meet in the middle ... So I'm beginning to frame up all those, uh, those kind of speed and distance problem things." None of the three subjects mentioned a triangle, none drew a triangle, and, of course, none mentioned the five-mile distance as the hypotenuse of the triangle. Concerning the five-mile distance, S4 never mentioned it, S1 said, "I don't think that tells you anything," and S6 said, "It's not really important." All three subjects initially formulated the problem as involving one driver going east and the other going west. S4 and S6 eventually corrected this error but S1 solved the problem without ever realizing that the second driver was heading south rather than west.

We interpret the results of Experiment 5 as indicating that there are two schemas which the subjects may use as templates in formulating the Smalltown problem — a "triangle" schema and a "distance-rate-time" schema. In the problem, these schemas seemed to act as alternatives to each other, three subjects using one of the schemas and three using the other. We should note that these schemas may not always be treated as alternatives and that navigation problems in particular require subjects to use both. The contrast between the two schemas is evident in Table 4.4. The schemas directed what the subject attended to in the problem, what information he expected, what information he regarded as relevant, and even what errors he made in reading the text.

Under the influence of the triangle schema, subjects spoke about and drew triangles, and in general, attended to those parts of the text relevant to the triangular relation among Smith's house, Jones's house, and the intersection. In particular, they attended to the 5-mile distance constituting the hypotenuse of the triangle, and to the fact that the triangle was a right triangle. Further, they hypothesized that the triangle was significantly involved in the solution of the problem.

<table>
<thead>
<tr>
<th>Behavior Using Triangle vs. DRT Frame</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject mentions triangle</td>
</tr>
<tr>
<td>S draws a triangle</td>
</tr>
<tr>
<td>5 miles taken as hypotenuse</td>
</tr>
<tr>
<td>S attempts to retrieve Pythagorean theorem</td>
</tr>
<tr>
<td>S categorizes problem as a distance-rate-time problem</td>
</tr>
<tr>
<td>S says that the 5-mile distance is not important</td>
</tr>
<tr>
<td>S forms an initial east-west representation</td>
</tr>
</tbody>
</table>
problem and attempted to retrieve equations from long-term memory relevant to triangles.

Under the influence of the "Distance-rate-time" schema, subjects did not give special attention to triangle information, and in fact, judged such information irrelevant to the problem elements. Thus, they expected that the easterly motion of one vehicle would be matched by a westerly motion of the other vehicle. This strong expectation clearly led Subject 1 to misread the text.

GENERAL DISCUSSION

In initiating our research, we asked the question whether the data lead us to emphasize the subject's choice of alternative processing strategies made early in reading or do they lead us to emphasize a single processing strategy in which the idea of choice early in reading is not important. Our answer for algebra word problems is clearly "Both of the above."

We have established that people do employ specialized processing strategies chosen at the time of reading by showing:

1. Subjects do recognize problem categories. There are at least 16 problem categories which our subjects agree about. Presumably, more could be identified if we were to search more thoroughly.

2. Subjects can in many cases recognize a problem's category early in reading the problem. Sometimes reading as little as the initial noun phrase is sufficient.

3. Subjects have information about the problem categories which is useful for formulating problems for solution. This information includes knowledge about useful equations and diagrams and appropriate procedures for making relevance judgments.

4. Subjects can and often do use this information in solving algebra word problems when their instructions are simply to solve the problems and do not in any way call special attention to problem classification.

While we have shown that subjects do employ specialized processing strategies for many problems, we have also observed that subjects solve some problems in a manner similar to the direct translation process used in STUDENT. Such solutions are observed most frequently in non-standard problems—that is, in problems for which the cover story does not match the structure of a known problem category.

We conclude that humans, like STUDENT, have two ways of approaching algebra word problems. First, if they recognize early in formulation that a problem is one of a known category, they will employ special heuristics specifically useful in formulating and solving problems of that category. Second, if they do not recognize a category for the problem, they will employ a general
solution procedure as a fallback strategy. This procedure resembles the direct translation procedure employed in STUDENT in that it operates line by line. In the light of the observations of Paige and Simon (1966), however, we must assume that, unlike STUDENT, this process makes use of semantic information.

All of the approaches to prose comprehension that we have discussed have made use of a single approach to comprehension. Our results have forced us to conclude that people use more than a single approach in comprehending algebra word problems. Perhaps they also use more than one approach in comprehending prose in general. Some have rejected the schema approach to prose comprehension because they believe that schemas are not adequate to account for the comprehension of all texts (Feldman, 1975). However, if humans do have more than one approach to prose comprehension generally, then such criticism loses much of its force.

ACKNOWLEDGMENTS

The research reported herein was supported at the Department of Psychology, Carnegie-Mellon University, by Public Health Services Grant MH-7722 from the National Institute of Mental Health and by Social Sciences Grant SOC73-09230 AO1 from the National Science Foundation.

We gratefully acknowledge Susan Robinson and Dorothea Simon for their help in collecting the data and in providing comment.

REFERENCES
