DOCUMENT RESUME

ED 220 826 CS 006 822

AUTHOR Keenan, Janice M.
TITLE Development of Comprehension Skills in the Middle Grades. Final Report, September 1, 1979-February 27, 1982.
INSTITUTION Denver Univ., Colo.
SPONS AGENCY National Inst. of Education (ED), Washington, DC.
PUB DATE 82
GRANT NIE-G-78-0173
NOTE 114p.

EDRS PRICE MF01/PC05 Plus Postage.
DESCRIPTORS Academic Aptitude; Beginning Reading; Child Development; Cognitive Development; Coherence; Elementary Education; Intermediate Grades; *Learning Theories; *Reading Comprehension; *Reading Processes; Reading Rate; *Reading Research; *Remedial Reading
IDENTIFIERS Kintsch (Walter); van Dijk (T A)

ABSTRACT

Four studies used the theoretically based understanding of comprehension developed by W. Kintsch and T. A. van Dijk to look for processes that might undergo development as a child becomes more practiced and skilled in reading so that an understanding of these processes might become the basis for reading remediation programs. The first three studies produced the following four conclusions: (1) third grade students used the same units of meaning to construct a mental representation of the meaning of a text as fifth grade students and adults; (2) beginning readers construct these units just as quickly as more skilled readers; (3) beginning readers connect these units into the same organizational structures as skilled readers; and (4) large developmental differences in working memory capacity and speed of lexical access do exist between fourth and sixth grade students. Assuming that the paucity of developmental differences found in these studies is the result more of theoretical inadequacy than the nonexistence of such differences, the final study examined the role of perspective in encoding and retrieval and the basis for coherence in texts. (JL)

FINAL REPORT

DEVELOPMENT OF COMPREHENSION SKILLS IN THE MIDDLE GRADES

NIE Project No. 8-0123

Grant No. NIE-G-78-0173

Period: September 1, 1979 - February 27, 1982

Name of Institution: University of Denver

Principal Investigator: Janice M. Keenan

Non-technical Abstract

The purpose of the research carried out under this grant was to take a theoretically-based understanding of the processes involved in comprehension developed by Kintsch (1974) and Kintsch and van Dijk (1978) and look for processes that might undergo development as a child becomes more practiced and skilled in reading. The general motivation behind this research was the belief that once these processes were identified they might serve as the basis for developing reading remediation programs. We first examined the unit of meaning used in constructing a mental representation of the meaning of a text and found that beginning third-grade readers use the same units as fifth graders and adults; futhermore, they construct these units with the same speed as more skilled readers. Thus, the units used during comprehension do not seem to undergo development with increased reading skill. We then examined whether beginning readers connect these units into the same organizational structures as those used by skilled readers. This is done by looking at what types of information they recall best from a text. We found no developmental differences here either. Finally, we examined the effect of the number of different concepts in a text on reading speed and memory for the text. This variable reflects the speed with which the reader can activate concepts in the mental lexicon and the capacity to handle information in working memory. Here we found large developmental differences between fourth- and sixth-grade readers. Thus, the only processes which seem to develop during the middle grades are working memory capacity and speed of lexical access. In the belief that the small number of developmental differences uncovered by this research may have been due to inadequacies in the theory rather than a true lack of developmental changes, the remainder of the research carried out under this grant attempted to advance our theoretical understanding of text comprehension by examining the role of perspective in encoding and retrieval and the basis for coherence in texts.

TABLE OF CONTENTS

CHAPTER 1

Development of Reading Comprehension Skills: Children's Reading Rate and Retention As a Function of the Number of Propositions in a Text.

Janice M. Keenan   Polly Brown

Abstract

This study examined whether there are differences between beginning and skilled readers in the units used to represent the meaning of a text. For skilled readers, the time to read a sentence is a function of the number of propositions contained in the sentence. In order to assess whether beginning readers also use propositional representations, the reading times of beginning third graders were compared to those of fifth graders on sentences that all had the same number of words but varied in the number of underlying propositions. Although third graders read more slowly than fifth graders, they showed the same increase in reading time with increasing numbers of propositions, suggesting that they use the same propositional units and process them at the same rate as fifth graders. This held for both long sentences, which place an increased load on memory and attentional resources, as well as short sentences, and for average as well as high ability readers. Immediate recall of the sentences for both grade levels showed that the higher the proposition in the propositional hierarchy, the more likely it was recalled. Although average readers recalled higher level propositions as well as good readers, they recalled significantly fewer of the lowest level propositions.

CHAPTER 1

Development of Reading Comprehension Skills: Children's Reading-Rate and Retention As a Function of the Number of Propositions in a Text

Of the two basic sets of skills involved in reading -- decoding and comprehension, most research concerned with learning to read has focused on decoding (Chall, 1976). One of the most interesting findings to emerge from this research is that the difference in decoding between beginning and more mature readers is not just a quantitative difference, such that beginning readers are able to decode less than skilled readers, but also a qualitative difference. Qualitative changes which accompany the development of decoding skills are evidenced by a changing pattern in error types, from semantically-based to visual/phonemically-based to semantic plus visual/phonemically-based, as the child becomes more proficient in decoding (Biemiller, 1970). They are also evidenced by the fact that the decoding skills of beginning readers require attentional resources, whereas those of more mature readers are automatic (LaBerge & Samuels, 1974; Guttentag & Haith, 1978).

Much less is known about developmental changes in comprehension. Numerous studies find quantitative differences in comprehension between beginning and more mature readers; for example, more mature readers score higher than beginning readers on measures such as percent of comprehension questions answered correctly and percent recalled. However, it is not known if these quantitative differences are due to qualitative differences in the comprehension process.

Because children are engaged in comprehending discourse for several years prior to encountering the reading situation, their comprehension skills might be said to be fairly well developed by the time they begin to read. It is therefore possible that the process of learning to read does not induce any qualitative changes in comprehension but only quantitative changes due to the accumulation of more world knowledge and a larger vocabulary. On the other hand, it is equally possible that the reading situation provides new challenges to the child which do induce qualitative changes in comprehension.

One way in which the reading situation poses challenges not present in the listening situation is that it requires the child to decode visual symbols. The demands of this decoding task might induce changes in comprehension and hence cause the comprehension processes of beginning readers to differ from those of mature readers. As mentioned previously, the decoding skills of beginning readers, unlike those of skilled readers, require attentional resources; in fact, even after a child has acquired decoding operations, practice is required before these skills are automatized to the point that they require few attentional resources (LaBerge, 1976; LaBerge & Samuels, 1974). Thus, beginning readers have fewer resources available to allocate to comprehension than mature readers. It is possible that the processes of comprehension are qualitatively different under conditions of limited resources than when more resources are available.

Compounding this problem of fewer resources available to allocate to comprehension processes is the fact that the reading situation provides far less contextual support for the interpretation of meaning than the typical listening situation in which the child is accustomed to comprehending discourse. For example, the referents of words are often physically present in conversational settings for children, whereas in reading, with the exception of pictures, there are no simultaneous, external events onto which the child can map the words he or she reads. Thus, in reading, children must learn to construct their own referents for these words and accomodate to the increased memory burden that this represents.

There are many ways in which these decreases in attentional resources and increases in memory load could conceivably affect the comprehension processes of beginning readers so as to produce qualitative differences in comprehension processes between beginning and skilled readers. This paper examines whether they affect the units used to represent the meaning of text. It is known that the comprehension processes of skilled readers involve accessing the conceptual representations of lexical units and grouping these concepts into basic units of meaning called propositions. The research reported here seeks to determine whether beginning readers also use propositions to represent the meaning of sentences, or whether the memory and decoding demands they face limit the amount of availabe processing resources to the point that not enough are available to integrate word concepts into higher-order, propositional units.

Propositional Representations of Meaning

Our text analyses are based on the propositional representation system of Kintsch (1974). This system, like most others, assumes that the smallest units of meaning, concepts, are grouped into relational structures called propositions. A proposition is defined as consisting of a single relational term and the concepts or arguments which it relates. It is assumed that the number of arguments contained in a proposition is not a constant, but rather varies with the particular relation; in short, relations are assumed to be _n_-place terms.

The elements of a proposition are concepts, not words, though they are expressd in the surface structure as words or phrases. A relation is expressed in the surface structure by either a verb, adjective, adverb, or conjunction; arguments can be expressed by words of any grammatical class and can also be other propositions. A proposition does not necessarily correspond to a sentence in a text, because sentences may, and often do, incorporate multiple propositions.

To illustrate the nature of the propositional representation consider the sentence, The little girl wanted to bake a birthday cake. This sentence contains four relational terms and thus consists of four propositions which are represented as:

1. (LITTLE, GIRL)
2. (WANTED, GIRL, 3)
3. (BAKE, GIRL, CAKE)
4. (BIRTHDAY, CAKE)

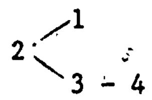
The concepts of the propositions are written in upper-case to highlight the fact that they are not words but rather their conceptual counterparts. The propositions are numbered for convenience so that if one proposition is embedded in another, as in proposition 2, we can simply refer to it by number rather than write it out again. Further details concerning the analysis of a text into its propositions can be obtained by consulting Kintsch (1974) and Turner and Greene (1978).

The first study to demonstrate the psychological validity of the proposition as a basic unit of processing in text comprehension and memory was performed by Kintsch and Keenan (1973). They used a reading time paradigm in which college-age subjects were timed as they read each of a large set of sentences. Subjects were told to read for meaning, and after reading each sentence, they were asked to immediately recall it. The sentences were constructed so that they all had the same number of words (16 words) but varied in the number of propositions they contained from four to nine. Kintsch and Keenan reasoned that if reading comprehension involves grouping word concepts into propositions, then the more propositions that need to be constructed, the longer it should take to comprehend the sentence. In other words, reading times should systematically increase with the number of propositions contained in the sentence. On the other hand, if comprehension simply involves activating the conceptual representations of the words in the sentence and does not involve propositional units, then there should be no systematic relation between number of propositions and reading time; rather, reading times should be roughly equivalent across sentences since all contained the same number of words.

Kintsch and Keenan found that reading times did in fact increase with the number of propositions contained in the sentences; subjects took about one additional second of reading time per proposition. Subjects did not, however, always recall a sentence perfectly. Because there was very little, if any, chance for forgetting in this immediate recall paradigm, it was assumed that these partial recalls of sentences reflected the fact that subjects did not always process all the propositions that were presented in the sentence. On this assumption, Kintsch and Keenan argued that a better indicator of the number of propositions the subject actually processed was not the number of propositions presented but rather the number of propositions recalled. Since reading times were supposed to be a function of the number of propositions processed, they therefore also examined reading time as a function of the number of propositions recalled. Reading times were found to systematically increase with number of propositions recalled, with subjects taking about 1.5 additional seconds for each proposition recalled. Thus, regardless of whether reading times are examined in terms of the number of propositions presented in the sentence or the number of propositions recalled by the subject, it appears that the proposition is a psychologically valid unit of meaning since the more propositions, the longer the reading time. More recent studies using recall (e.g., Kintsch & Glass, 1974) and priming paradigms (Ratcliff & McKoon, 1978) have provided further corroborating support for the psychological validity of propositions.

The representation of the meaning of a text does not consist solely of a listing of its propositions; rather, it must also include connections between the propositions so as to render the meaning representation coherent. One proposal for establishing connections between propositions is on the basis of shared referents or arguments; this is called the argument repetition rule (Kintsch, 1974; Kintsch & van Dijk, 1978). According to this rule, propositions are said to be coherent if they share a common argument.

It is assumed that the coherence graph of propositions which represents the meaning of a text is hierarchical. This assumtion of a hierarchy is intended to capture the fact that some ideas are perceived as more central to the meaning of a text than others. The hierarchy is constructed by first selecting the proposition which expresses the main idea of the text to head the hierarchy. For single-sentence texts, this is the proposition expressing the main action of the sentence; for longer texts, it is usually the title of the text or the main action of the first sentence. The hierarchy is then constructed by using the argument repetition rule to connect propositions to this most superordinate proposition either directly or indirectly through shared arguments with more subordinate propositions. Propositions which share arguments in common with the most superordinate proposition are directly subordinated to it; propositions which do not share arguments with the most superordinate proposition, but which share arguments with the next level of propositions are subordinated to them, and so on. To illustrate with the sample sentence given above, proposition 2 would be selected to head the hierarchy (Level 1) since it expresses the main idea of the sentence. Propositions 1 and 3 would be connected to proposition 2 on the next level of the hierarchy (Level 2) because they both share a common argument with proposition 2, namely the argument GIRL. Because proposition 4 does not share any arguments with proposition 2, it cannot be connected to it; rather, it would be connected to proposition 3 on the next level of the hierarchy (Level 3) because of the common argument, CAKE. The resulting hierarchy of propositions can be depicted as:



As mentioned above, Kintsch and Keenan (1973) found that recall of the sentences in their study was not always perfect. They therefore examined recall of the propositions as a function of their level in the hierarchy. They found that percent recall steadily increased across levels such that the higher the proposition in the hierarchy, the more likely it was to be recalled. This result has been replicated in many studies (e.g., Kintsch, Kozminsky, Streby, McKoon, & Keenan, 1975; Meyer, 1975; Thorndyke, 1976). It has also been shown that verification times are faster for superordinate than subordinate propositions (McKoon, 1976). Together these studies provide strong support for the notion that comprehension in the mature reader involves connecting propositions together in a hierarchical structure in memory.

In the present study we set out to determine whether beginning readers also group concepts into propositions and connect these propositions into a hierarchical representation. We used the same paradigm with beginning third and fifth graders as that used by Kintsch and Keenan (1973) with adults. Reading times were measured for sentences that all had the same number of words but varied in their number of underlying propositions; immediate recall was obtained for each sentence. In order to make the task meaningful and enjoyable for the children it was cast within the following scenario. The children were told to pretend that they were controlling a computerized center for outer space exploration. Explorer space ships were sending messages to their computer screen one at a time. They were to read the message and press a button to signal the space ship when they understood it; this allowed reading times to be collected on each sentence. Because the button press caused the message to be erased from the screen, and because the children were told it was necessary to preserve the messages, they were instructed to "save" each message by reciting it from memory into a microphone connected to a tape recorder; this saving-the-messages task thus served as the immediate free recall test.

If the increased cognitive demands placed on the beginning reader (i.e., the attentional demands of decoding and the memory load resulting from the relatively impoverished semantic environment of the reading situation compared to the child's accustomed listening situation) have an effect on the unit of processing such that not enough resources are available to group concepts into propositions, then one might expect the reading times of third graders to be less systematically related to number of propositions than the reading times of the adults observed by Kintsch and Keenan (1973) and their recall to be less related to the level of the proposition in the hierarchy than that observed by Kintsch and Keenan with adults. Because fifth graders have had more practice in reading, however, their reading time and recall data are expected to be more similar to that found for adults.

It is possible that the ability of the beginning reader to integrate concepts into propositions may depend on the complexity of the text. Specifically, when the number of word concepts that need to be grouped into propositions is small, beginning readers may be able to construct propositional representations; however, when the number of word concepts is large, the memory and attentional load is greater and the ability to group concepts into propositions may break down. We examined this possibility by employing two sets of sentences. One set consisted of sentences that were all seven words long but varied in the number of underlying propositions from two to four. The other set consisted of sentences that were all 16 words long but varied in number of propositions from four to nine.

It is also possible that the ability to construct propositional representations may be related to the child's overall level of reading skill as measured by standardized reading tests. Beginning readers who are reading above grade level might therefore show the same increase in reading time with number of propositions as more mature readers, and the lack of relation between number of propositions and reading time might only be evident in beginning readers who are reading at or below grade level. Support for this hypothesis stems from Cromer's finding that a subset of low reading ability junior college students could be made to comprehend as well as good readers by spatially grouping the words of sentences into meaningful phrases (cf. Mason & Kendall, 1978). Because these phrases often corresponded to propositional units, the results suggest that one of the problems underlying these subjects' low level of reading ability was their failure to group concepts into propositions. Although standardized reading scores were not available for all subjects used in this study, we did examine this possibility for those subjects for whom scores were available.

Method

Subjects

Forty third graders and forty fifth graders, with 20 males and 20 females in each grade, participated in the study. Half of the subjects at each grade level were tested in the summer preceding their entry into the third or fifth grade; the other half were tested during the school year, either in late fall or early winter. Subjects were recruited by letter from a large number of elementary schools in the greater Denver metropolitan area and were paid for their participation. Six additional third graders attempted to participate in the study but were unable to complete it because their decoding skills were so poor.

Thirty-two of the 40 third graders and 18 of the 40 fifth graders were included in the analyses examining the effects of reading ability. These were the only subjects for whom comparable standardized reading scores were available. All of these subjects had been tested at the beginning of the third grade on either the Metropolitan Achievement Test or the Stanford Achievement Test. Grade equivalent scores from the reading comprehension portion of these tests were used in the analyses. For the purposes of performing ANOVAs, it was necessary to classify the children at each grade level into two groups of reading ability; this was done using a median split on the grade equivalent scores. Somewhat surprisingly, the median split resulted in approximately equal numbers of males and females in each group: 5 males and 4 females in the higher reading ability group and 4 males and 5 females in the lower reading ability group for the fifth grade; 8 males and 8 females in the higher and 9 males and 7 females in the lower reading ability groups for the third grade. The mean grade equivalent scores for both the lower third- and lower fifth-grade groups were at or slightly above grade level so these groups are most appropriately labeled average rather than poor readers. The higher reading ability fifth graders averaged 2 grade equivalents higher than the group of average fifth graders (5.97 vs. 3.97 on a test taken in third grade), and the good third-grade readers averaged 2.8 grade equivalents higher than the average third graders (7.0 vs. 4.2).

Materials

The materials consisted of two sets of sentences. One set consisted of nine sentences that were all seven words long; three of these contained two propositions, three contained three propositions, and three contained four propositions. The other set consisted of 18 sentences that were all 16 words long; the number of propositions in these sentences ranged from four to nine, with three sentences at each level of number of propositions.

The sentences were constructed by the authors in collaboration with a second grader and a fourth grader to ensure their meaningfulness to children. They were designed to be about possible things one could observe in outer space and on other planets. The vocabulary items used in the sentences were normed on six pilot third graders to ensure that they were all words in the children's reading vocabulary. This was done by typing each word on an index card and having the child go through the deck of cards reading each word out loud and either explaining its meaning or using it in a sentence. Four words were found that were not familiar to all the children; these were replaced with substitute words suggested by the children.

Tables 1 and 2 present sample sentences from the 7-word and the 16-word sentence sets respectively. Included in these tables are the propositions for each sentence as well as their hierarchical arrangement. The analysis into propositions follows the propositional representation system of Kintsch (1974). A full description of the propositional analysis as well as the notational system can be found in Turner and Greene (1978). The only notation shown in these tables which may not be readily understood is the abbreviation LOC which refers to location.

Insert Tables 1 and 2 about here

Procedure

The presentation of sentences and recording of reading times was controlled by a PDP 11/03 microprocessor connected to an ADDS 580 CRT. The keyboard of the CRT was covered with a wooden frame which contained one response button for the subject to use. This wooden frame served to prevent the subject from pressing keys that could have aborted the experimental program. A microphone, connected to a tape recorder, was situated to the left of the CRT and was used to record the subject's recall of the sentences.

Subjects were tested individually in a small room set up to simulate a mission control center for outer space exploration. When the child arrived for the experiment, he/she and the accompanying parent were given a demonstration of the equipment in the room and the adjoining computer room in an attempt to make them feel comfortable and excited about working on the computer. The parent was then asked to wait in a waiting room while the child participated in the experiment.

The children were asked to pretend that they were working in the mission control center. They were told that their job was to receive messages from space ships that were exploring outer space and sending their observations back to earth through the subject's computer. They were told to read each message that occurred on the screen and press the button as soon as they understood it. The button press was to let the space ship that sent the message know that the message had been received and understood.

Subjects were told that it was important to save the messages. However, they could not be saved on the screen since it was necessary to clear the screen to allow for new messages to come in. In fact, when the child pressed the button signalling comprehension of the message, it went off the screen. The children were therefore told that the only way to save a message was for them to recall it from memory and record it on the tape recorder right after pressing the button signalling comprehension. Subjects were told that if they could not recall the message word for word, that was all right. They were told that what was important was to get as much of the meaning recorded as possible and to feel free to use their own words to do so.

The experimental session began with three practice sentences, two short and one long. They were: "Welcome to the space center," "Get ready for messages from outer space," and "We hope you like this space game and we hope you do very well." The sentences were presented one at a time and the subjects were told to read them silently. As soon as the subject understood the sentence, he/she pressed the response button. This button erased the sentence from the screen and replaced it with a message to recall the sentence. The button press also caused the reading time for the sentence to be recorded. The subject then attempted to orally recall the sentence into the tape recorder. When the subjects finished recalling as much as they could, they pressed the response button. This caused the screen to display the message, "Are you ready for the next message?" When the child was ready, he/she pressed the button and the next sentence was displayed.

Following the practice sentences, the nine 7-word sentences were presented in a different random order for each subject. The 7-word sentences were presented first because they were easier and allowed the children to gain confidence in their ability to do the task. Following the 7-word sentences, the 18 16-word sentences were presented in random order.

Design and Analyses

The 7-word sentences were analyzed separately from the 16-word sentences. For each sentence set, an analysis of reading times was performed as a function of the number of propositions presented in the sentence. Two additional analyses were performed on the 16-word sentence set. One dealt with the reading time data. Because recall of the 16-word sentences, unlike that of the 7-word sentences, was generally less than perfect, it is likely that a better indicator of the number of propositions actually processed by a subject is the number of propositions that the subject recalled rather than the number of propostions presented in the sentence. Thus, the second analysis of the reading times for the 16-word sentences examined reading times as a function of the number of propositions the subject recalled. This analysis was not performed on the 7-word sentences because the number of propositions recalled was almost always the same as the number of propositions presented. The other analysis performed on the 16-word sentence set that was not performed on the 7-word sentence set was an analysis of the recall data as a function of the level of the proposition in the propositional hierarchy. Again, this analysis was not performed on the 7-word sentences because the near perfect recall of these sentences simply yields ceiling effects in the analysis of level effects on recall. The following describes the design for each of these analyses.

Reading Time As a Function of Number of Propositions Presented. For the 7-word sentences. the number of propositions presented has three levels: 2, 3. or 4 propositions. For the 16-word sentences, the number of propositions presented has six levels: 4, 5, 6. 7, 8, or 9 propositions. The reading times for the three sentences nested within each level of number of propositions were averaged so as to produce a mean reading time for each subject for each level of number of propositions presented. For both the 7-word sentences analysis and the 16-word sentences analysis, subjects were nested in the crossing of grade (third or fifth) with sex with time of experiment (summer or school year).

Reading Time As a Function of Number of Propositions Recalled. This analysis was performed only on the 16-word sentences. The number of propositions that can be recalled from a 4-proposition sentence ranges from 0 to 4; the number that can be recalled from a 5-proposition sentence ranges from 0 to 5, and so on. Reading times for each of a subject's 18 16-word sentences were classified according to the number of propositions the subject recalled, irrespective of the number of propositions the sentence actually contained. Mean reading times were then computed for each subject for each level of number of propositions, ranging from 0 to 9. Because there were relatively few instances of recalling less than 3 propositions or more than 7, the range was truncated so that the analysis examined only five levels of number of propositions recalled: 3 or less, 4, 5, 6, 7 or more. Subjects did not always have entries in all cells; in fact, there were a total of 39 out of the 400 cells that were empty. We therefore used Winer's (1962) method of estimation to obtain values for these cells and adjusted the degrees of freedom in the ANOVA accordingly. In sum, the design of this analysis consisted of number of propositions recalled crossed with subjects which were nested in the crossing of grade with sex with test time.

Recall As a Function of Level in the Hierarchy. The tapes containing the children's recall protocols were transcribed and then scored independently by two judges. The interrater reliability was .94, with the disagreements turning out to be errors by one or the other judge which were resolved in conference. For each sentence it was determined which of the propositions were recalled. Paraphrases of the original wording were accepted as correct, as long as the meaning was accurately expressed. If a subject made an error in a superordinate proposition which then reappeared in a subordinate proposition, the subordinate proposition was accepted as correctly recalled, while the superordinate proposition was scored as incorrect. For example, suppose a subject recalled the four-proposition sentence in Table 1 as Big fish float down the muddy rivers. Although bird is contained in propositions 1, 2, and 3, only proposition 2 would be scored as incorrect because of the substitution of fish for birds. Propositions 1 and 3 would be scored as correct because the subject correctly recalled that whatever the superordinate proposition was concerned with, it was big and it floated down the rivers.

Using the hierarchical representations of meaning like those shown in Tables 1 and 2, it was determined that, across the 18 16-word sentences, there were a total of 20 propositions at Level 1 (the most superordinate level), 56 propositions at Level 2, 31 at Level 3, 8 at Level 4, and 2 at Level 5. Because of the small number of propositions at Levels 4 and 5, they were combined. For each subject we computed the percent of propositions recalled at Levels 1, 2, 3, 4/5. The design of the analysis of variance consisted of level in the hierarchy crossed with subjects which were nested in the crossing of grade with sex with time of testing.

Reading Ability Analyses. A number of analyses were performed to examine the effects of reading ability on reading time and recall. First, correlations were computed relating reading ability and overall reading time for both the 7-word sentences and the 16-word sentences. Second, in order to determine whether the relationship of increasing reading times with increasing numbers of propositions held for average as well as good readers, ANOVAs were performed to determine whether reading ability interacted with number of propositions. Three ANOVAs were performed: one on the 7-word sentences with number of propositions presented as the main independent variable, one on the 16-word sentences with number of propositions presented as the main independent variable, and one on the 16-word sentences with number of propositions recalled as the main independent variable. Finally, an ANOVA was performed examining the effects of reading ability on recall as a function of the level of the proposition in the hierarchy. In all these analyses, subjects were nested in the crossing of grade with reading ability (average or good, as determined by the median split); subjects were crossed then with either number of propositions or level in the hierarchy, depending on the analysis. The variables of sex and time of testing were dropped from these designs because the division of subjects into reading ability groups resulted in slightly unequal Ns for sex and time of test.

Results and Discussion

Reading Time As a Function of Number of Propositions Presented

7-Word Sentences. Figure 1 presents the average reading times of each grade level for the 7-word sentences as a function of the number of propositions presented in each sentence. Although third graders read significantly more slowly than fifth graders ($F(1,72) = 13.02$, $p < .001$, $MSe = 4.0251$), there was a highly significant effect of number of propositions ($F(2,144) = 11.98$, $p < .001$, $MSe = .6008$) which held for both the third graders ($F(2,72) = 7.63$, $p < .001$, $MSe = .93248$) and the fifth graders ($F(2,72) = 4.72$, $p < .02$, $MSe = .26914$). As can be seen in Figure 1, reading times for both groups monotonically increase as the number of propositions presented in the sentence increases. It thus appears that, at least for these short sentences, third graders use the same units of meaning -- namely propositions -- as fifth graders.

Although the interaction of grade level and number of propositions was not significant ($F(2,144) = 1.97$), it can be seen in Figure 1, that the slope of the function relating reading time to number of propositions tends to be steeper for the third graders than the fifth graders. Using linear regression techniques to compute these functions, with RT standing for reading time and PP for number of propositions presented, we found that the function for third graders is $RT = 3.37 + .42PP$, and that for fifth graders is $RT = 3.17 + .18PP$. Kintsch and

Keenan (1973) argued that the slopes of these regression lines represent the time required to construct and process each additional proposition and that the intercepts reflect the time for all the other processes involved in reading, such as decoding and performing syntactic analyses. Applying this interpretation to the present functions, it would appear that the difference in reading speed between third and fifth graders on 7-word sentences is attributable to both a small difference in the speed of decoding and syntactic operations and a somewhat larger difference in the speed of constructing propositions. However, since the interaction of grade level and number of propositions was not significant, the difference in slopes cannot be regarded as significant. Consequently, we averaged over grade level to determine the amount of time required to process each additional proposition. The resulting regression line is, RT = 3.27 + .30PP. Hence, for 7-word sentences, an additional .3 seconds of reading time is required for each additional proposition presented.

Insert Figure 1 about here

Sex had no significant effect on reading time ($F(1,72) < 1$), nor did it interact with any other variables. There was also no significant difference between the reading times of subjects tested in the summer preceding the school year and those tested during the year ($F(1,72) = 1.82$), nor did this factor interact with any of the other variables.

Correlations between average reading time for the 7-word sentences and the subject's grade equivalent score from the standardized reading tests showed, as one might expect, that higher grade equivalent scores were associated with lower overall reading times. Specifically, we found a highly significant negative correlation for the fifth graders ($r = -.75$, $p < .001$) and a marginally significant negative correlation for the third graders ($r = -.32$, $p < .10$). Of most interest are the results from the ANOVA involving reading ability. Again, there was a highly significant effect of number of propositions ($F(2,64) = 6.72$, $p < .005$). And although average readers read more slowly than good readers ($F(1,32) = 4.11$, $p = .05$, MSe = 5.2610), there was no interaction of reading ability and number of propositions. As can be seen in Figure 2, the average readers, as well as the good readers, show an increase in reading times with increasing numbers of propositions.

Insert Figure 2 about here

16-Word Sentences. Figure 3 presents the average reading times of each grade level for the 16-word sentences as a function of the number of propositions presented in each sentence. Again, third graders read significantly more slowly than fifth graders ($F(1,72) = 14.44$, $p < .001$. MSe = 105.5291). This time, however, there was a significant interaction between grade and sex

F(1,72) = 5.60, $p < .05$). For third graders, males read more slowly (16.35 sec) than females (13.71), while for fifth graders, females read more slowly (12.33) than males (10.41). There was also a significant interaction between time of testing the subjects and grade level. Third graders who were tested in the summer preceding their entry into third grade read significantly more slowly than third graders tested during the year ($F(1,36) = 4.31$, $p < .05$, MSe = 119.3233). There was no significant difference, however, between the reading times of fifth graders tested during the summer and during the year ($F(1,36) = 2.05$, $p > .15$). A couple of additional months of reading experience thus makes a large difference in the overall reading speed of third graders but not fifth graders.

Most important is the finding shown in Figure 3 that, for both age groups, reading time increases as the number of propositions in the sentence increases. The effect of number of propositions presented was highly significant ($F(5,360) = 20.08$, $p < .001$, MSe = 7.9392), and held for both the third graders ($F(5,180) = 10.99$, $p < .001$, MSe = 9.4009) and the fifth graders ($F(5,180) = 10.06$, $p < .001$, MSe = 6.4774). The interaction of grade level with number of propositions presented was not significant ($F(5,360) = 1.15$). Thus, third graders, like fifth graders, appear to use propositional units not only in comprehending short sentences but also in comprehending these longer, 16-word sentences.

Insert Figure 3 about here

Figure 3 also shows the regression lines for each age group. Remarkably, the slopes of these functions are identical for the two age groups; only the intercepts differ. The function for third graders is, RT = 10.6 + .66PP, and that for fifth graders is, RT = 7.06 + .66PP. Thus, just as was found with the 7-word sentences, the difference in reading speed between third and fifth graders lies not in the time required to process the propositions but in the time required to decode and do syntactic analyses.

Kintsch and Keenan (1973) obtained a function of RT = 6.37 + .94PP for college-aged subjects on 16-word sentences. Although it is always difficult to make comparisons across studies, a few comments are in order. The fact that the intercept that they obtained for college students is lower than that for our third and fifth graders is reasonable if we assume that the speed of decoding and syntactic analyses is a function of practice and that college students are more practiced in these skills than grade school children. The fact that the slope is higher for college students than for our subjects at first seems unreasonable. One would expect college students to require no more time per proposition than grade school children. There are, however, at least two reasons why one might expect the slopes for the college students to be higher.

The most likely reason is that there are large differences between the two studies in the familiarity of the subjects with the content of the sentences. Kintsch and Keenan used sentences whose content was purposely designed to be unfamiliar to the subjects. Their sentences expressed little known facts from ancient history. In an attempt to make the task meaningful and enjoyable to the children, we used sentences that were about objects that were familiar to the children. It is therefore likely that the contents of our sentences were more familiar to our subjects than the facts used by Kintsch and Keenan were to their subjects. It is known that familiarity of content has a strong effect on reading time (Kintsch, et al., 1975); hence the steeper slopes obtained with the college students probably reflect this difference in familiarity. Another possible reason for the difference in slopes is that Kintsch and Keenan's college students recalled more propositions on the average (80%) than our subjects (72%). Perhaps if our subjects had recalled as much as theirs they would have required more time per proposition. The analyses in the next section which examine reading rate as a function of the number of propositions recalled control for this factor. As will be shown, however, the difference in slopes between the two studies remains even when reading times are examined in terms of number of propositions recalled; hence, it would appear that it is due to a difference in familiarity of the content.

Correlations between average reading time for the 16-word sentences and the subject's grade equivalent score from the standardized reading tests showed a highly significant negative correlation for the fifth graders ($r = -.67$, $p < .01$), but only a nonsignificant negative correlation for the third graders ($r = -.12$). Thus grade equivalent scores are predictive of overall reading speed on 16-word sentences for fifth graders but not third graders.

The ANOVA examining reading ability and number of propositions presented again showed a highly significant effect of number of propositions ($F(5,160) = 10.78$, $p < .001$, $MSe = 12.5888$). But neither the interaction of reading ability with number of propositions ($F(5,160) < 1$), nor the three-way interaction of grade, reading blty, and number of propositions ($F(5,160) < 1$) was significant. Thus, for average readers, as well as good readers, the greater the number of propositions in the sentence, the longer the reading time.

<u>Reading Time As a Function of the Number of Propositions Recalled</u>

As Kintsch and Keenan (1973) argued, partial recalls in this immediate memory task are due not so much to forgetting, but rather to a failure to sufficiently process all of the propositions presented. Thus the number of propositions actually processed by the subject is best indicated not by the number of propositions presented but rather the number of propositions recalled. When reading times are examined as a function of the number of propositions recalled, we again find a highly significant effect of number of propositions ($F(4.248) = 29.06$, p

< .001, MSe = 4.5759), which holds for the third graders (F(4,127) = 9.19, p < .001, MSe = 5.7211) as well as the fifth graders (F(4,122) = 13.66, p < .001, MSe = 8.7793); the degrees of freedom in the denominators of the Fs have been adjusted for missing cells. As can be seen in Figure 4, the more propositions the subject recalled, the longer the reading time.

Insert Figure 4 about here

Also shown in Figure 4 are the regression lines for each grade level. As in the analyses of reading times as a function of number of propositions presented, we again find a significant difference between grade levels in the intercepts (F(1,72) = 11.72, p < .001, MSe = 88.4303). For third graders the intercept is 11.8 seconds, while for fifth graders it is only 7.3 seconds. Presumably, this difference reflects the fact that third graders spend more time than fifth graders on decoding and/or syntactic operations. Interestingly, we again find, as in the analysis of reading times as a function of number of propositions presented, that there is no significant difference between the grade levels in the slopes, i.e. in the time required to process each additional proposition; the interaction of grade level and number of propositions recalled was not significant (F(4,248) = 1.97). On the average, an additional .76 seconds of reading time was required for each additional proposition recalled.

In their analyses of reading times as a function of number of propositions recalled, Kintsch and Keenan (1973) found that their college students required an additional 1.5 seconds for each additional proposition recalled. The fact that college students required twice as much time for each additional proposition as our grade school children must be due to a difference in the relative difficulty of the content of the sentences; as pointed out earlier, their sentences concerned unfamiliar historical facts, as opposed to the more familiar content of our sentences. An important conclusion to be drawn from these comparisons, therefore, is that the time to construct a proposition is not a fixed parameter of the comprehension system; rather, it varies with the familiarity of the to-be-related concepts and the length of the sentence, since 7-word sentences result in a much smaller slope than 16-word sentences.

The analyses involving reading ability showed neither a significant interaction of reading ability with number of propositions recalled (F(4,108) < 1) nor a significant three-way interaction of ~~read~~ reading ability, grade, and number of propositions recalled (F(4,108) = 1.53). As can be seen in Figure 5, however, the good readers at both grade levels show a steady increase in reading time with each additional proposition recalled, while the average readers show much less systematicity. This pattern suggests that perhaps there may be some difference across reading levels in the function relating reading time to number of propositions recalled for long sentences which we were unable to observe in the statistical tests, either because we did not have enough poor, as opposed to average, readers in the sample or because of the relatively small number of subjects in these analyses. This point will be discussed further in the conclusion of the paper.

Insert Figure 5 about here

Recall As a Function of Level in the Propositional Hierarchy

Although the fifth graders recalled slightly more propositions (73%) from the 16-word sentences than the third graders (71%), the difference was not significant ($F(1,72) < 1$). There was also no significant effect of sex ($F(1,72) < 1$) or time of testing ($F(1,72) < 1$) on amount recalled. The only factor which had an effect on recall was the level of the proposition in the hierarchical representation; this had a highly significant effect ($F(3,216) = 56.15$, $p < .001$, $MSe = .01333$). As can be seen in Figure 6, the higher the level of the proposition in the hierarchy, the more likely it was recalled. This finding replicates that of Kintsch and Keenan (1973) and Kintsch et al. (1975).

The function relating recall to levels is not linear. Rather, there is only a small decrease in recall between Levels 1 and 2 and between Levels 3 and 4/5, while there is a dramatic decrease from Level 2 to Level 3. Because propositions at the higher levels contain the main actions and relations of the sentence, while lower level propositions tend to contain only modifying details, the large difference in recall between Levels 2 and 3 probably reflects the difference in memorability between main actions and modifying details (cf. Brown & Smiley, 1977).

Insert Figure 6 about here

The fact that the levels effect on recall is the same for third graders as for fifth graders supports the conclusion drawn from the reading time analyses; namely, third graders, like fifth graders and adults, appear to use propositional representations to represent the meanings of sentences. Furthermore, the recall results show that not only are they constructing the underlying propositions of the text during comprehension, but they are also connecting them in the same hierarchical fashion as that used by more mature readers.

Although other studies of children's comprehension (Christie & Schumacher, 1975; Brown & Smiley, 1977; Waters, 1978) have shown a levels effect in their recall, even for children as young as kindergarten age, these studies were all concerned with oral comprehension, not reading comprehension. Furthermore, with the exception of Waters' (1978) study which used the same propositional analysis used in the present study, the basis for classifying ideas into levels of importance in these studies was essentially atheoretical; they used ratings of relevance or importance derived from skilled readers. The present results thus represent an important extension of these earlier studies by showing that there is a theoretical basis for distinguishing levels of importance wich has the same effects on children's recall as adults' and by showing that the processes involved in reading do not disrupt the levels effect found in children's oral comprehension.

The analyses involving reading ability showed no effect of reading ability on amount recalled ($F(1,32) = 1.48$). However, there was an inteeting significant interaction between reading ability and level in the propositional hierarchy ($F(3,96) = 2.77$, $p < .05$, $MSe = .01129$). As can be seen in Figure 7, there is no significant difference in recall between average and good readers for propositions at Levels 1, 2, or 3; however, there is a large difference between the two groups in their recall for propositions at the lowest level, Level 4/5. Whereas good readers recall these low level details as well as Level 3 propositions, average readers recall them significantly worse. Dunn, Mathews, and Bieger (1979) have also found interactions between reading ability and level of information, such that high ability readers tend to recall lower level information better than average or below average readers; however, their results are somewhat more difficult to interpret since they failed to find a levels effect in their high ability readers.

Insert Figure 7 about here

Conclusions

This study represented an initial attempt to determine whether there are qualitative differences in the comprehension processes of beginning and more skilled readers by examining the units used to represent the meaning of a text. A considerable amount of research has shown that skilled readers integrate the conceptual representations of lexical units into basic units of meaning called propositions. It was hypothesized that beginning readers, however, may not integrate concepts into propositions because of insufficient attentional resources due to, among other things, the relatively greater attentional demands of decoding.

The results of this study suggest that if there are qualitative differences in comprehension between beginning third-grade readers and the more skilled fifth-grade readers, it is unlikely that they lie in the units used to represent the meaning of sentences. Third graders showed the same effect of number of propositions on reading time as that of fifth graders and that of the adults observed by Kintsch and Keenan (1973). Although all of the sentences in each set contained the same number of words, reading times for third graders, as well as fifth graders, increased as the number of propositions contained in the sentence increased. This result obtained for the long sentences, which presumably place an increased demand on attentional resources, as well as the short sentences. When reading times were examined in terms of the number of propositions actually constructed, as reflected by the number recalled, rather than the number contained in the sentence, the same result obtained; the greater the number of propositions processed, the longer the reading time. The fact that third graders showed the same decrease as fifth graders in recall of propositions as a function of their level in the hierarchy further supports the contention that they were utilizing the same propositional representations to represent the meanings of the sentences.

One of the most interesting results to emerge from the reading time analyses was the finding that the significant difference in reading speed between third and fifth graders is due not to the time required to process each additional proposition but rather to the time required to perform the other processes involved in reading, such as decoding and syntactic analyses. In all of the analyses, the slope of the function relating reading time to number of propositions was not significantly different, and in one case was identical, for third and fifth graders. Thus, not only do third graders use the same units as fifth graders, but they construct them at the same speed.

Although it was hypothesized that the ability to construct propositional representations may be related to reading skill, the results did not provide much support for this hypothesis. For the 7-word sentences, it was clear that the average readers showed the same increase in reading time with number of propositions as the good readers. For the 16-word sentences, there was also no interaction of reading ability with number of propositions, but the pattern of reading times for the average readers showed less of a systematic relation to number of propositions than that of the good readers. Given the absence of a statistically significant effect, not much can be made of this result. It does, however, seem to warrant further investigation using a group of poor readers rather than average readers. One of the problems with using poor readers in this task, however, is that they must not be so poor as to be unable to decode the words. Otherwise, the reading times are meaningless. When we attempted to include more poor third-grade readers in our sample, we found that they could not do the task because they could not decode.

One of the most interesting findings to emerge from the reading ability analyses was the finding that the only difference between good and average readers in their memory for the sentences was in their recall of the lowest level propositions; good readers recalled significantly more of these propositions than the average readers. It would be interesting to pursue this finding in future research using cueing procedures to determine whether it is due to retrieval or encoding differences.

The present study found no difference in the units used to represent the meanings of sentences between third and fifth graders. Although this result could be taken as evidence that there is no difference in comprehension units between beginning and skilled readers, it is also possible that the decoding skills of these third graders were sufficiently automated that they do not represent a fair test of the hypothesis that the decrease in attentional resources due to nonautomatic decoding skills in beginning readers results in qualitative differences in their comprehension processes. Perhaps if one examined children at an earlier stage in the reading process, for example beginning second graders, the hypothesized differences between beginning and skilled readers might be found. The major difficulty in doing this research, however, is that which we mentioned previously in the discussion of poor readers; namely finding beginning readers with decoding skills that are sufficient to read the sentences but at the same time not yet automated.

References

Biemiller, A. J. The development of the use of graphic and contextual information as children learn to read. Reading Research Quarterly, 1970, 6, 75-96.

Brown, A. L., & Smiley, S. S. Rating the importance of structural units of prose passages: A problem of metacognitive development. Child Development, 1977, 48, 1-8.

Chall, J. S. Reading and development. Newark, Delaware: International Reading Association, 1976.

Christie, D. J., & Schumacher, G. M. Developmental trends in the abstraction and recall of relevant versus irrelevant thematic information from connected verbal materials. Child Development, 1975, 46, 598-602.

Cromer, W. The difference model: A new explantation for some reading difficulties. Journal of Educational Psychology, 1970, 61, 471-483.

Dunn, B. R., Mathews, S. R., & Bieger, G. Individual differences in the recall of lower-level textual information. Center for the Study of Reading Technical Report No. 150, University of Illinois, 1979.

Guttentag, R. E., & Haith, M. M. Automatic processing as a function of age and reading ability. Child Development, 1978, 49, 707-716.

Kintsch, W. The representation of meaning in memory. Hillsdale, N.J.: Erlbaum, 1974.

Kintsch, W., & Glass, G. Effects of propositional structure upon sentence recall. In W. Kintsch, The representation of meaning in memory. Hillsdale, N.J.: Erlbaum, 1974.

Kintsch, W., & Keenan, J. M. Reading rate and retention as a function of the number of propositions in the base structure of sentences. Cognitive Psychology, 1973, 5, 257-274.

Kintsch, W., Kozminsky, E., Streby, W., McKoon, G., & Keenan, J. M. Comprehension and recall of text as a function of content variables. Journal of Verbal Learning and Verbal Behavior, 1975, 14, 196-214.

Kintsch, W., & van Dijk, T. A. Toward a model of text comprehension and production. Psychological Review, 1978, 85, 363-394.

LaBerge, D. Perceptual learning and memory. In W. K. Estes (Ed.), Handbook of learning and cognitive processing. Vol. 4. Attention and memory. Hillsdale, N.J.: Erlbaum, 1976.

LaBerge, D., & Samuels, S. J. Toward a theory of automatic information processing in reading. Cognitive Psychology, 1974, 6, 293-323.

Mason, J. M., & Kendall, J. R. Facilitating reading comprehension through text structure manipulation. Center for the Study of Reading Technical Report No. 92, University of Illinois, 1978.

McKoon, G. Organization of information in text memory. Journal of Verbal Learning and Verbal Behavior, 1977, 16, 247-260.

Meyer, B. J. F. The organization of prose and its effects on memory. Amsterdam: North-Holland Publishing Company, 1975.

Ratcliff, R., & McKoon, G. Priming in item recognition: Evidence for the propositional structure of sentences. Journal of Verbal Learning and Verbal Behavior, 1978, 17, 403-417.

Thorndyke, P. W. Cognitive structures in comprehension and memory of narrative discourse. Cognitive Psychology, 1977, 9, 77-110.

Turner, A., & Greene, E. The construction of a propositional text base. JSAS Catalog of Selected Documents in Psychology, 1978.

Waters, H. S. Superordinate-subordinate structure in semantic memory: The role of comprehension and retrieval processes. Journal of Verbal Learning and Verbal Behavior, 1978, 17, 587-598.

Winer, B. J. Statistical principles in experimental design. New York: McGraw-Hill, 1962.

TABLE 1

Samples of the 7-Word Sentences Used in the Experiment Together With Their Propositional Representations and Propositional Hierarchies

2 PROPOSITIONS:

The rocket will spin around the sun.

1. (SPIN, ROCKET)
2. (LOC: AROUND, 1, SUN)

1 — 2

3 PROPOSITIONS:

The earth is a billion miles away.

1. (LOC: AWAY, EARTH)
2. (MILE, AWAY)
3. (BILLION, MILE)

1 — 2 — 3

4 PROPOSITIONS:

Big birds float down the muddy rivers.

1. (BIG, BIRD)
2. (FLOAT, BIRD)
3. (LOC: DOWN, 2, RIVER)
4. (MUDDY, RIVER)

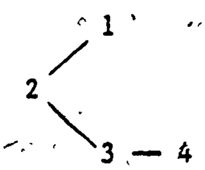


TABLE 2

Samples of the 16-Word Sentences Used in the Experiment Together With Their Propositional Representations and Propositional Hierarchies

4 PROPOSITIONS:
The earth is sending messages to the planet over the radio but nobody can understand them.

1. (SEND, EARTH, MESSAGE, PLANET, RADIO)
2. (CONCESSION: BUT, 1, 3)
3. (CAN, NOBODY, 4)
4. (UNDERSTAND, NOBODY, MESSAGE)

1 — 2 — 3 — 4

9 PROPOSITIONS:
The animals' ears are better here than on earth so they can hear sounds far away.

1. (POSSESS, ANIMAL, EAR_1)
2. (BETTER, EAR_1, EAR_2)
3. (LOC: HERE, EAR_1)
4. (LOC: ON, EAR_2, EARTH)
5. (SO, 2, 6)
6. (ABLE TO, ANIMAL, 7)
7. (HEAR, ANIMAL, SOUND)
8. (LOC: AWAY, SOUND)
9. (FAR, AWAY)

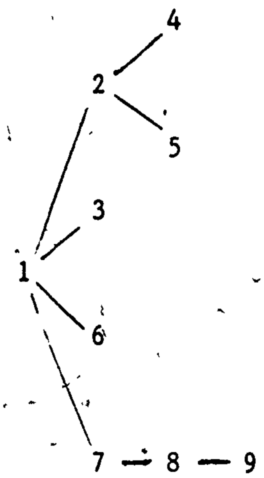


Figure Captions

Figure 1. Reading times for the 7-word sentences as a function of the number of propositions contained in the sentences.

Figure 2. Reading times for the 7-word sentences as a function of the number of propositions contained in the sentences, broken down by grade and reading ability.

Figure 3. Reading times for the 16-word sentences as a function of the number of propositions contained in the sentences, together with their best fitting regression lines.

Figure 4. Reading times for the 16-word sentences as a function of the number of propositions recalled.

Figure 5. Reading times for the 16-word sentences as a function of the number of propositions recalled, broken down by grade and reading ability.

Figure 6. Proportion of propositions recalled at each level of the propositional hierarchy for third and fifth graders.

Figure 7. Proportion of propositions recalled at each level of the propositional hierarchy for average and good readers.

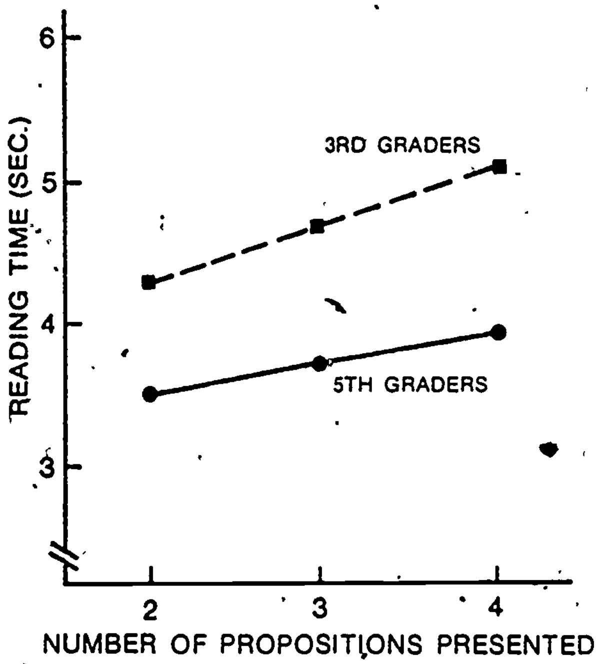

6
5
4
3
READING TIME (SEC.)
3RD GRADERS
5TH GRADERS
2
3
4
NUMBER OF PROPOSITIONS PRESENTED

Figure 1

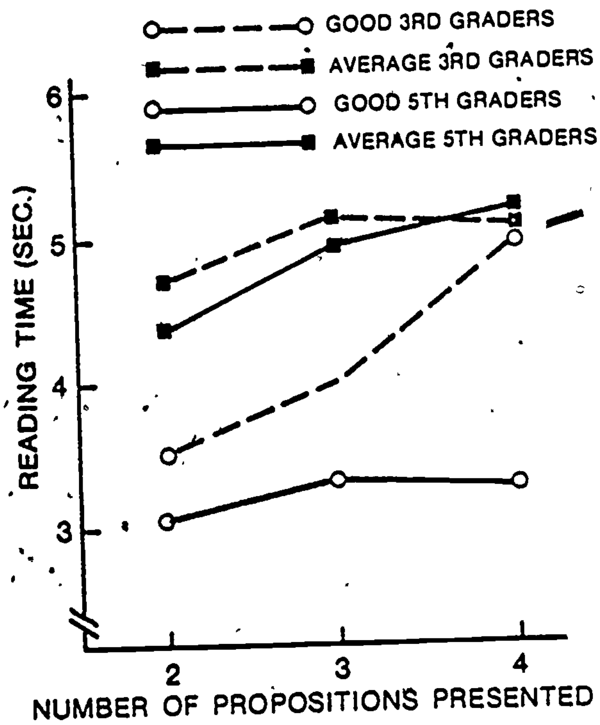
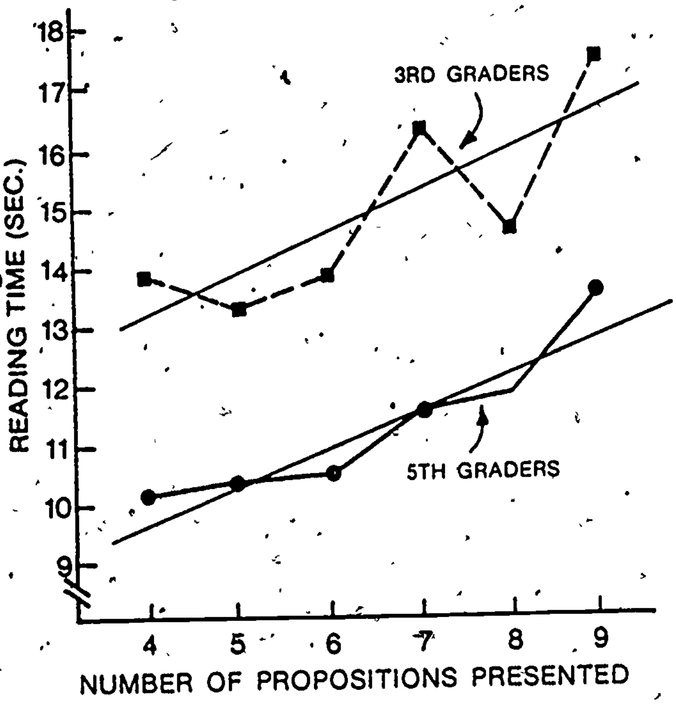

GOOD 3RD GRADERS
AVERAGE 3RD GRADERS
GOOD 5TH GRADERS
AVERAGE 5TH GRADERS
READING TIME (SEC.)
6
5
4
3
2
3
4
NUMBER OF PROPOSITIONS PRESENTED

Figure 2



3RD GRADERS
5TH GRADERS
READING TIME (SEC.)
18
17
16
15
14
13
12
11
10
9
4
5
6
7
8
9
NUMBER OF PROPOSITIONS PRESENTED

Figure 3

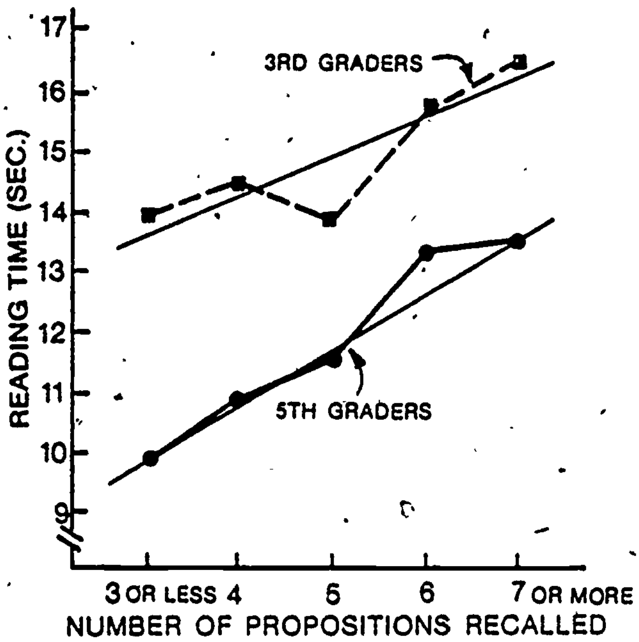

3RD GRADERS
5TH GRADERS
READING TIME (SEC.)
17
16
15
14
13
12
11
10
9
3 OR LESS
4
5
6
7 OR MORE
NUMBER OF PROPOSITIONS RECALLED

Figure 4

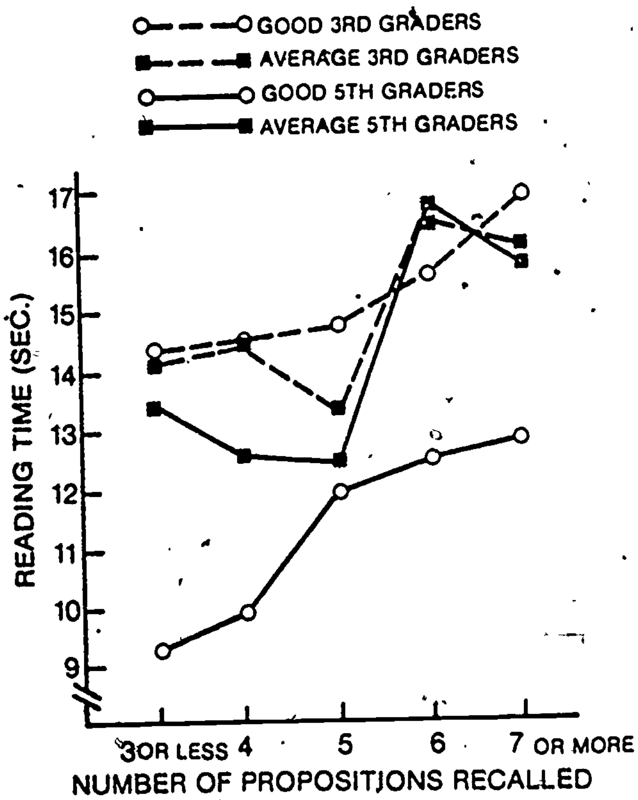

GOOD 3RD GRADERS
AVERAGE 3RD GRADERS
GOOD 5TH GRADERS
AVERAGE 5TH GRADERS
READING TIME (SEC.)
17
16
15
14
13
12
11
10
9
3 OR LESS
4
5
6
7 OR MORE
NUMBER OF PROPOSITIONS RECALLED

Figure 5

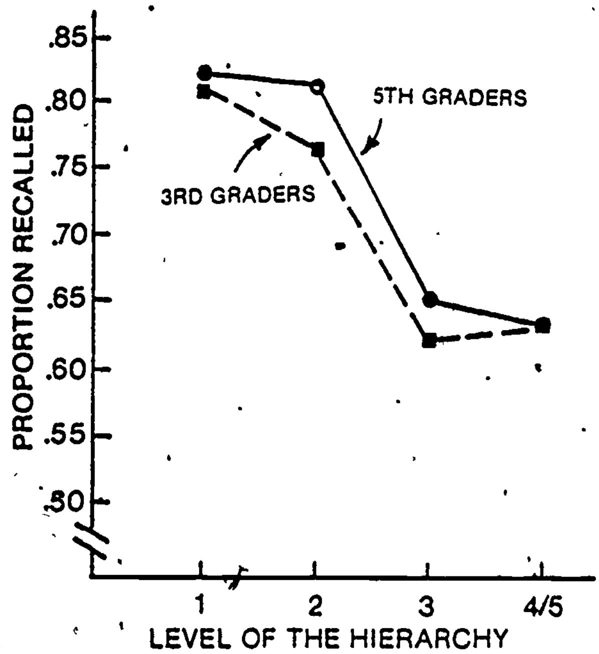
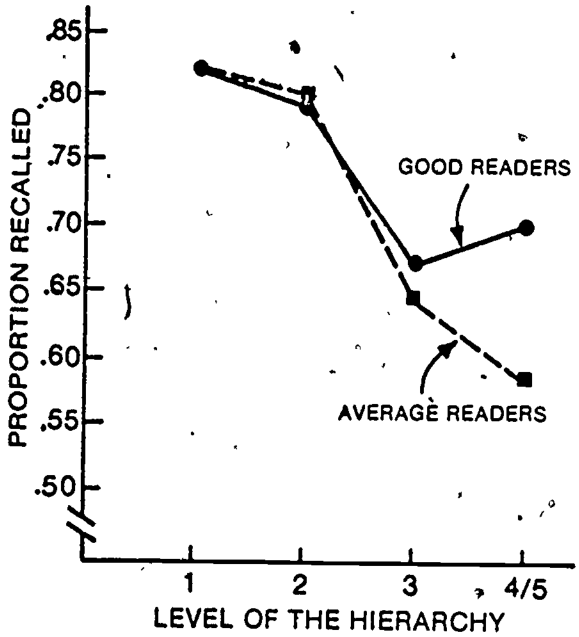

PROPORTION RECALLED
.85
.80
.75
.70
.65
.60
.55
.50
5TH GRADERS
3RD GRADERS
1
2
3
4/5
LEVEL OF THE HIERARCHY

Figure 6



PROPORTION RECALLED
.85
.80
.75
.70
.65
.60
.55
.50
GOOD READERS
AVERAGE READERS
1
2
3
4/5
LEVEL OF THE HIERARCHY

Figure 7

CHAPTER 2

Children's Reading Rate and Recall As a Function of the Number of Different Arguments in the Text Base

Janice M. Keenan

Abstract

It is known that mature readers require more processing time per proposition recalled for texts containing many different arguments than for texts containing few different arguments. This is because the greater the number of different arguments, the more the lexicon needs to be accessed and the greater the memory load. On the assumption that memory capacity and speed of lexical access increase with increasing familiarity with concepts, it was hypothesized that an interaction should obtain between grade or reading skill and number of different arguments, such that younger or less skilled readers would show a greater difference in comprehension between many and few different arguments than older, more skilled readers. To test this, fourth and sixth graders of high and average reading ability read and immediately recalled 24 paragraphs, half containing few and half many different arguments, from three content areas: history, science, and familiar experiences such as birthday parties. The results clearly supported the hypothesized interaction between number of arguments and grade level, making number of arguments one of the few variables in reading comprehension research to yield an age interaction. It was also found that familiar experience passages were read more quickly and recalled better than history or science passages, even though they were the same length. Fourth graders recalled less than sixth graders on history and science passages, but they recalled familiar experience passages just as well. Nonetheless, they continued to show a greater difference between many and few different argument passages than sixth graders even on these familiar experience passages.

CHAPTER 2

Children's Reading Rate and Recall As a Function of the Number of Different Arguments in the Text Base

In a 1976 speech to the International Reading Association, Chall noted that although there had been a significant improvement over the past decade in national reading test scores for children in the primary grades where the tests primarily concern decoding skills, there had been no concommitant improvement in scores for children in the middle grades where the tests primarily concern comprehension abilities. She noted that what was needed to rectify this state of affairs was a deeper, theoretically-based understanding of the processes involved in comprehension that could be used to look for processes which might undergo development as the child becomes more practiced and skilled in reading and thus serve as the basis for remediation programs.

Partly in response to this need, there has been a considerable increase in recent years in the amount of research regarding reading comprehension. This research has resulted in models which show a fair amount of agreement about the component processes of reading comprehension. These processes can be divided into those involved in building a representation of the textual input (microstructure processes) and those which operate on this representation as a whole (macrostructure processes) for purposes such as summarization, inferencing, and integration with existing knowledge (Kintsch & van Dijk, 1978). In this paper we confine our attention to the more basic microstructure processes in an attempt to determine whether they undergo development as a child learns to read.

According to Kintsch (1974) and Kintsch and van Dijk (1978), there are three basic steps involved in building a representation of the meaning of a text: (1) activating the conceptual representations of words, (2) grouping these word concepts into propositions. and (3) connecting the propositions into a coherent hierarchical structure. These processes operate on a series of inputs, each consisting of a phrase or a short sentence. Given such an input, the first thing that happens is that the words in the input activate their conceptual representations in memory. Then, using knowledge about the semantic relations that can obtain between concepts in conjunction with syntactic knowledge, these concepts are grouped into idea units called propositions.

Propositions are n-tuples of word concepts in which one, and only one, serves as the predicator; the others serve as arguments which each fulfill a unique semantic role. The concepts which serve as predicators are relational terms which typically appear in the surface structure as either verbs, adjectives, adverbs. or serve as predicators are relational terms which typically appear in the surface structure as either verbs, adjectives, adverbs, or conjunctions. For example, the sentence, John sent Mary flowers, consists of a single proposition (because it contains only one relational term) which has three arguments. The proposition is expressed as (SEND, Agent: JOHN, Receiver: MARY, Object: FLOWERS). The concepts are written in upper-case to highlight the fact that the elements of a proposition are concepts, not words. The terms Agent, Receiver, and Object in this example identify the semantic roles of each of the arguments.

As propositions are constructed, it is necessary to connect them together in order to form a coherent representation of the meaning of the text. Kintsch (1974) and Kintsch and van Dijk (1978) assume that the propositions are connected on the basis of shared referents into a hierarchical structure called the text base. The proposition which heads the hierarchy is that which expresses the title of the text, or in the absence of a title, the proposition expressing the main action of the first sentence. The hierarchy is then constructed by subordinating each proposition to the highest level proposition with which it has an argument in common. (If a proposition does not share arguments with any other propositions, memory is searched for information that will allow one to inferentially connect the propositions together.) The examples of text bases which will be presented below should suffice to show the main features of the representational system. Further details can be obtained by consulting Kintsch (1974) and Turner and Greene (1978).

Initial empirical support for these three processes came from studies of mature readers. Research relevant to the first — activation of concepts — stems from work by Kintsch, Kozminsky, Streby, McKoon. & Keenan (1975) and Manelis and Yekovitch (1976). They compared reading times and recall for texts that all had the same length, both in terms of number of words and number of propositions, but that differed in the number of different arguments they contained. They found that texts which had few different arguments were read more quickly and. in some instances, recalled better than texts which had many different arguments, demonstrating that the number of different lexical activations required is an important determinant of reading time.

The psychological validity of the proposition as a unit of meaning which is constructed during comprehension was initially established with a study by Kintsch and Keenan (1973). They measured reading times for sentences that all had the same number of words but which varied in the number of propositions they contained. They found that reading time was determined, not by the number of words, but by the number of propositions that needed to be constructed to comprehend the sentence; the greater the number of propositions, the longer the reading time. Subsequent studies using recall (e.g., Kintsch & Glass, 1974) and priming techniques (Ratcliff & McKoon, 1978) have provided further corroborating support for the psychological validity of propositions.

Evidence for the hierarchical nature of the textual representation stems from numerous studies which have examined recall of propositions as a function of their level in the hierarchy (e.g., Kintsch & Keenan, 1973; Kintsch, et al., 1975; Meyer, 1975; Thorndyke, 1976). In all of these studies, it has been found that the higher the proposition in the hierarchy, the more likely it is to be recalled. McKoon (1976) provided further support for the hierarchical nature of the representation by showing that superordinate propositions are also verified faster than subordinate propositions.

Of these three basic processes, the only one to have received much attention in investigations of children's comprehension is the construction of a hierarchical representation. Although there are developmental differences in the ability to rate ideas in terms of their importance to a passage (Brown & Smiley, 1977), there seem to be no developmental differences in the construction of the hierarchical representation. Even children as young as kindergarten age show the same levels effect in their recall as adults. The levels effect obtains in children's recall regardless of whether levels are defined on an atheoretical basis, using ratings of relevance or importance derived from mature readers (Christie & Schumacher, 1975; Brown & Smiley, 1977), or whether they are defined using Kintsch's (1974) propositional hierarchy system described above (Waters, 1978; Keenan & Brown, Chapter 1). Furthermore, it holds for children's reading comprehension (Keenan & Brown, Chapter 1) as well as for their oral comprehension (Christie & Schumacher, 1975; Brown & Smiley, 1977; Waters, 1978).

There also seem to be no developmental differences in the use of propositions to represent the meaning of texts (Keenan & Brown, Chapter 1). On the assumption that beginning readers need to allocate more attentional resources than skilled readers to decoding operations, Keenan and Brown hypothesized that beginning readers might not have sufficient resources available to group concepts into propositions. They tested this hypothesis by examining the reading times of third and fifth graders on sentences that all had the same number of words but that varied in the number of underlying propositions. They found that although third graders read more slowly than fifth graders, they were just as sensitive to the propositional structure as the fifth graders, with their reading times linearly increasing as the number of propositions in the sentence increased. Furthermore, the time required to process each additional proposition was the same for both grade levels.

To date, there have been no investigations of the effect of number of different lexical activations on children's comprehension. The experiment presented below therefore addresses this issue. Fourth and sixth graders' reading times and recall are examined for passages that have the same length.

both in terms of number of words and number of propositions, but that vary in the number of different arguments they contain. It is known that mature readers take longer to read passages with many different arguments than passages with few different arguments (Kintsch, et al., 1975; Manelis & Yekovitch, 1976). The question of interest here is whether there is an interaction between age and number of different arguments, such that fourth graders show a greater difference between many and few different argument passages than sixth graders.

The fact that there have been no observed developmental differences in either the ability to use propositions or the ability to connect them into hierarchies might lead one to think it unreasonable to expect developmental differences in this process either. However, there are at least two reasons why one might expect developmental differences here, despite their absence in the other two processes. First, because speed of lexical activation is partly a function of familiarity, and because familiarity or knowledge about concepts develops with age and reading experience, it is possible that the speed of lexical access for fourth graders is slower than that for sixth graders. If this is so, and because passages with many different arguments require more lexical activations than passages with few different arguments, then the difference in reading times between many and few different argument passages should be greater for fourth than for sixth graders. Second, a passage that has many different arguments places a greater load on working memory than a passage with few different arguments; this is because there are more concepts to maintain in memory and also because there are more arguments that need to be searched in attempting to connect an incoming propositon to earlier propositions. On the assumption that memory capacity develops with age, we would therefore expect that the greater memory load of many different argument passages would more adversely affect the reading times and recall of fourth graders than those of sixth graders.

We have characterized the hypothesized interaction between age and number of different arguments as stemming from age differences on two dimensions -- familiarity of concepts and memory capacity. Studies by Chi (197?) and Lindberg (1980), however, have shown that age-related differences in memory capacity are often due to age differences in knowledge about the to-be-remembered items. Therefore, it is probably best to characterize the hypothesized interaction as stemming from age differences only in knowledge about the concepts which can affect either the speed of lexical activation, the capacity to maintain and utilize the concepts in working memory, or both. Given this characterization, it is reasonable to expect that the interaction might not obtain on passages whose content is as familiar to fourth graders as to sixth graders. The present study tests this hypothesis by employing passages concerning familiar experiences, such as going to MacDonalds and birthday parties, as well as passages whose content is expected to be less familiar to fourth graders than to sixth graders, namely history and science passages. We expect the difference in reading times and number of propositions recalled between many and few different argument passages to be similar for fourth and sixth graders on the passages concerning familiar experiences and to be greater for fourth than for sixth graders on the history and science passages.

To the extent that familiarity and knowledge about concepts derives from reading experience and not simply from age, we might expect reading ability to have an effect on the size of the difference between many and few different argument passages on reading time and recall. The present study examines this possibility by comparing high ability readers with average ability readers at both grade levels.

Method

Subjects

Eighteen fourth graders and eighteen sixth graders, recruited from the Van Dellen elementary school in Denver participated in the study for payment. A median split on grade equivalent scores from the reading subtest of the California Test of Basic Skills was used to divide the students at each grade level into two reading ability groups. The average grade equivalent score for the sixth graders above the median was 9.76 and for those below the median was 6.96. The average grade equivalent score for the fourth graders above the median was 7.85 and for those below was 4.8. Thus, the two groups at each grade level can be characterized as high ability and average ability readers.

Materials

The materials consisted of 24 paragraphs: 8 history, 8 science, and 8 familiar experiences. Half of the paragraphs of each content type contained few different arguments and half contained many different arguments. The number of different arguments was exactly 4 for each of the few different argument passages and exactly 9 for the many different argument passages. The length of the paragraphs was strictly controlled by having each contain exactly 14 propositions and by restricting the number of words to between 34 and 37.

Tables 1 through 3 present examples of each paragraph type. Included in these tables are a listing of the different arguments, a listing of the propositions contained in the paragraph, and the arrangement of the propositions into their hierarchy.

The history and science paragraphs were adapted from the following children's books: Childcraft's _How Things Work_, Childcraft's _World and Space_, Educational Research Council's _Industry: Man and the Machine_, Foster's _George Washington's World_, and Hillyer's _A Child's History of the World_. The paragraphs concerning familiar experiences were constructed by the author. Several pilot subjects were run in order to adjust the vocabulary of the texts to be familiar to all of the subjects.

Insert Tables 1, 2, and 3 about here

Procedure

The presentation of paragraphs and recording of reading times was controlled by a PDP 1110 minicomputer. The subject was seated in a room adjoining the computer room in a sound-shielded subject station which contained a video monitor for paragraph presentation, a subject response button box, and a microphone connected to a tape recorder that was used to record the subject's recall.

The experimental session began with a practice paragraph that concerned the formation of clouds. Following the practice, the experimental paragraphs were presented in two blocks of 12 paragraphs each, with a 10-minute break given between the two blocks to prevent fatigue. Each block contained two paragraphs from each of the six cells generated by crossing the three levels of content with the two levels of munber of different arguments. The ordering of the two blocks was counterbalanced across subjects.

Within each block of 12 paragraphs, the four paragraphs (2 few and 2 many different argument passages) of each content type were blocked, but the order of presentation of the blocks was randomized for each subject. Thus, a given subject might read 4 history paragraphs, then 4 science paragraphs, and then 4 familiar experience paragraphs. The ordering of the four paragraphs within each content type was counterbalanced across subjects.

For each paragraph trial, the subject pressed a button when he or she was ready to receive a paragraph. The pragraph was then displayed until the subject prssed a button signalling comprehension. Subjects were told to read the paragraph carefully so that they would be able to remember it. They were also told to read it only once and to not try to memorize it. The subject's button press, signalling the end of reading, simultaneously caused the reading time to be recorded and the screen to display a message telling the subject to recall the paragraph. Subjects were told to speak into the microphone and to recall as much as they could; they were told that they need not recall the passage word-for-word, but to feel free to use their own words. When the subject finished recalling, he or she pressed the response button; the screen then displayed a message asking the subject to press the button when he or she was ready for the next paragraph.

Subjects were tested individually. Although the computer controlled the experiment, an experimenter was in the room with the child throughout the session to offer assistance and encouragement.

Design

There were two main independent variables manipulated within subjects. They were number of different arguments (few or many) and content type (history, science, or familiar experience). These two factors were factorially combined, with four passages nested within each of the resulting six cells for a total of 24 passages. All subjects read all 24 passages and the data from the four passages nested within each cell were averaged prior to the analyses. Subjects were nested in grade (4th or 6th) crossed with reading ability (high or average), with 9 subjects per group.

Scoring

The tapes containing the children's recall protocols were transcribed and then scored independently by two judges. The interrater reliability was .96, with the disagreements turning out to be errors by one or the other judge which were resolved in conference. For each sentence it was determined which of the propositions were recalled. Paraphrases of the original wording were accepted as correct, as long as the meaning was accurately expressed. If a subject made an error in a superordinate proposition which then reappeared in a subordinate proposition, the subordinate proposition was accepted as correctly recalled, while the superordinate proposition was scored as incorrect. For example, suppose a subject recalled the first sentence of the familiar experience few argument passage shown in Table 3 as *Billy's new baby brother just came home from the hospital*. Because of the substitution of *brother* for *sister*, the superordinate Proposition 4 would be scored as incorrect. However, the subordinate propositions 1, 2, and 3, which also contain the argument *sister*, would be scored as correct because the subject correctly recalled that whatever the superordinate proposition was concerned with, it was *Billy's*, it was *new*, and it was a *baby*.

Results

Reading Times

Mean reading times are presented in Table 4 as a function of content type and number of different arguments. As can be seen from this table, in every case, subjects took longer to read passages with many different arguments than those with few different arguments ($F(1,32) = 38.40$, $MSe = 7.1193$, $p < .00001$), even though the two passages were exactly the same length both in terms of number of words and number of propositions. We expected that the difference in reading times between many and few different argument passages would be greater for the fourth than for the sixth graders; however, this result did not obtain, as evidenced by the lack of an interaction between number of arguments and grade level, $F(1,32) < 1$.

Insert Table 4 about here

Reading ability had a significant effect on reading times ($F(1,32) = 5.63$, $MSe = 218.08796$, $p < .03$), with high ability readers averaging 16.84 seconds per paragraph as compared to 21.61 seconds for average ability readers. Also, the content of the passages had a highly significant effect on reading speed ($F(2,64) = 18.71$, $Mse = 8.72261$, $p < .0001$), with familiar experience passages being read significantly faster (17.49 sec) than either the history (20.20 sec) or the science (19.99 sec) passages, which were not significantly different from each other. However, neither reading ability nor content interacted with number of different arguments (for both, $F < 1$), nor were any of the higher level interactions significant (for all, $F < 1$).

Recall

The average number of propositions recalled by each subject group are presented in Table 5 as a function of content type and number of different arguments. Whereas the reading time data showed a significant effect of number of different arguments and no significant interaction between number of arguments and grade, the recall data show the opposite. Specifically, the main effect of number of different arguments is not significant ($F(1,32) = 1.29$, $MSe = 1.8122$), but the interaction with grade is ($F(1,32) = 3.40$, $p = .07$). Furthermore, the interaction is in exact accord with the predictions in that while sixth graders recalled just as many propositions from many different argument passages (8.16) as from few different argument passages (8.03), $F(1,16) < 1$, fourth graders recalled significantly less from the many different argument passages (7.47) than from the few different argument passages (8.02), $F(1,16) = 4.93$, $MSe = 1.6351$, $p < .05$.

Insert Table 5 about here

The different effects of number of different arguments on the reading and recall data highlight a trade off that is always possible between reading time and amount recalled. Thus, in order to assess the true nature of the effect of number of different arguments on comprehension, we combined the reading time and recall data into a single measure; this measure was the average reading time per proposition recalled. We calculated such values for each subject and submitted them to an analysis of variance.

The mean reading times per proposition recalled are presented in Table 6 for each grade level and for each type of passage. As the table shows, the processing time per proposition is significantly less for few argument passages (2.45 sec) than for many argument passages (2.94), $F(1,32) = 8.05$, $MSe = 1.62889$, $p < .01$. Most importantly, the difference in processing times per proposition between few and many different argument passages is significantly greater for fourth graders (2.50 versus 3.39) than for sixth graders (2.40 versus 2.49), $F(1,32) = 5.26$, $p = .028$. Thus, as we predicted, the ability to handle many different arguments during comprehension increases with development.

Insert Table 6 about here

Like the reading times, the recall data show a highly significant effect of content type, $F(2,64) = 43.43$, $MSe = 1.1620$, $p < .00001$. Passages concerning familiar experiences were recalled significantly better (8.89 propositions) than either history (7.50) or science (7.39) passages which were not significantly different from each other. When the reading time data are combined with the recall data, we find that familiar experience passages require only 2.13 seconds of processing time per proposition recalled, compared to a processing rate per proposition recalled of 2.93 for history passages and 3.01 for science passages.

Of great interest in the recall data was the finding of a significant interaction between content and grade, $F(2,64) = 3.24$, $p < .05$. On familiar experience passages, the content of which might be expected to be just as familiar to fourth graders as to sixth graders, there is no significant difference between fourth and sixth graders in amount recalled (8.75 versus 8.92); however, on history and science passages, fourth graders recalled significantly less than sixth graders (for history, 7.08 versus 7.92; for science, 7.22 versus 7.53). We speculated at the beginning of the paper that since the interaction between grade and number of different arguments is likely due to age differences in familiarity with concepts, it might hold only for the history and science passages and not for the familiar experiences. However, this result did not obtain. The interaction of grade, number of different arguments, and content was not significant, $F(2,64) < 1$.

High ability readers tended to recall more (8.44 propositions) than average readers (7.4), although the difference was not significant. However, when the reading time and recall data are combined in the reading time per proposition recalled measure, then there is a highly significant effect of reading ability, $F(1,32) = 12.93$, $MSe = 7.80175$, $p = .001$. Also, there is a significant interaction between reading ability and content ($F(2,64) = 4.35$, $MSe = 0.36007$, $p = .015$), which is similar to the interaction between grade and content; namely, the difference between high and average ability readers is less for familiar experience passages (1.03 seconds per proposition) than for either history (1.56) or science (1.51) passages.

In a separate analysis of variance, recall was examined as a function of the level of the proposition in the hierarchy. There is considerable variability in the number of propositions occurring at each level; consequently, the analysis used proportion of propositions recalled at each level. For each passage, the propositional hierarchy as shown in Tables 1 - 3, was used to assign propositions to levels, with the most superordinate proposition being assigned to Level 1. Because there were so few propositions occurring at Levels 4 and above, these data were pooled with Level 3; the resulting category is referred to as Level 3+.

Level in the hierarchy had a highly significant effect on recall, $F(2,64) = 169.21$, $MSe = 0.03057$, $p < .00001$. As is typically the case, Level 1 propositions were recalled the best (82%); in addition, Level 2 propositions (64%) were recalled significantly better than Level 3+ propositions (51%). This held for ~~both~~ grade levels, for both reading ability groups, and across all paragraph types. There was, however, an interesting interaction between levels and content which can be seen in Table 7, $F(4,128) = 11.70$, $p < .001$. Recall from the earlier analyses that both grades recalled significantly more from familiar experience passages than from history and science passages. What Table 7 shows is that the superior recall of the familiar experience passages does not hold across all levels. Rather, at Level 1 history and science propositions are just as memorable as familiar experience propositions; only at the other levels are they less memorable.

Insert Table 7 about here

Discussion

Like the adults studied by Kintsch, et al. (1975) and Manelis and Yekovitch (1976), the fourth and sixth graders in this study found passages that have many different arguments to be more difficult to comprehend than passages with few different arguments. However, the fourth graders found them relatively more difficult than the sixth graders in that the difference in processing times per proposition recalled between many and few different argument passages was significantly greater for them than for sixth graders. This is an important finding because it is the first evidence for developmental differences in the microstructure processes involved in constructing the representation of a text during reading comprehension.

Other studies of children's reading comprehension have shown that younger, less skilled readers construct essentially the same representation as mature readers. They use the same propositional units (Keenan & Brown, Chapter 1) as mature readers, and they connect the propositions into the same hierarchical representations (the levels effect in the present study and in Chapter 1; Waters, 1978; Brown & Smiley, 1977). The interaction of grade and number of arguments observed in the present study suggests, however, that even though they may construct the same representation as mature readers, the processes they use to do so do not operate with the same efficiency, so that many different arguments present more of a problem for them than for more skilled readers.

In the introduction of this chapter we outlined the processes that can be affected by an increase in the number of different arguments. Specifically, each new argument requires lexical access to activate its conceptual representation and its associated information which is relevant to linking it to other concepts. Furthermore, a new argument substantially increases the load placed on working memory relative to the repetition of a previously occurring argument. This is because not only is it another item that must be kept track of, but also, because it increases the number of items in working memory, it can affect the ease with which other processes can be accomplished, such as the search of memory for arguments with which incoming propositions can be connected.

The fact that fourth graders are more adversely affected by an increase in number of arguments than sixth graders suggests that either their processes for accessing the lexicon or their working memory capacities or both operate less efficiently than those of sixth graders. It seems reasonable to expect that at least one cause for these developmental differences, be they in lexical activation, memory, or both, is a difference in knowledge about the concepts; since fourth graders have less experience with and hence knowledge about the concepts, they cannot access them as quickly or maintain them in memory as easily. The fact that the interaction between number of different arguments and grade level continued to obtain even for familiar experience passages, however, suggests that knowledge about the concepts is not the only factor responsible for the age interaction. It appears that general familiarity with the reading task itself may also be a factor.

References

Brown, A. L., & Smiley, S. S. Rating the importance of structural units of prose passages: A problem of metacognitive development. Child Development, 1977, 48, 1-8.

Chall, J. S. Reading and development. Newark, Delaware: International Reading Association, 1976.

Chi, M. T. H. Knowledge structures and memory development. In R. Siegler (Ed.), Children's thinking: What develops? Hillsdale, N. J.: Erlbaum, 1978.

Childcraft. How things work. Chicago: Field Enterprises, 1976.

Childcraft. World and space. Chicago: Field Enterprises, 1976.

Christie, D. J., & Schumacher, G. M. Developmental trends in the abstraction and recall of relevant versus irrelevant thematic information from connected verbal materials. Child Development, 1975, 46, 598-602.

Educational Research Council: Social Sciences Program. Industry: Man and the Machine. Boston: Allyn & Bacon, 1971.

Foster, G. George Washington's world. New York: Charles Scribner's Sons, 1941.

Hillyer, V. M. A child's history of the world. New York: Appleton Century Crofts, Inc., 1951.

Keenan, J. M., & Brown, P. Development of reading comprehension skills: Children's reading rate and retention as a function of the number of propositions in a text. Chapter 1 of NIE Final Report.

Kintsch, W. The representation of meaning in memory. Hillsdale, N.J.: Erlbaum, 1974.

Kintsch, W., & Glass, G. Effects of propositional structure upon sentence recall. In W. Kintsch, The representation of meaning in memory. Hillsdale, N.J.: Erlbaum, 1974.

Kintsch, W., & Keenan, J. M. Reading rate and retention as a function of the number of propositions in the base structure of sentences. Cognitive Psychology, 1973, 5, 257-274.

Kintsch, W., Kozminsky, E., Streby, W., McKoon, G., & Keenan, J. M. Comprehension and recall of text as a function of content variables. Journal of Verbal Learning and Verbal Behavior, 1975, 14, 196-214.

Kintsch, W., & van Dijk, T. A. Toward a model of text comprehension and production. Psychological Review, 1978, 85, 363-394.

Lindberg, M. A. Is knowledge base development a necessary and sufficient condition for memory development? Journal of Experimental Child Psychology, 1980, 30, 401-410.

Manelis, L., & Yekovich, F. R. Repetitions of propositional arguments in sentences. Journal of Verbal Learning and Verbal Behavior, 1976, 15, 301-312.

McKoon, G. Organization of information in text memory. Journal of Verbal Learning and Verbal Behavior, 1977, 16, 247-260.

Meyer, B. J. F. The organization of prose and its effects on memory. Amsterdam: North-Holland Publishing Company, 1975.

Ratcliff, R., & McKoon, G. Priming in item recognition: Evidence for the propositional structure of sentences. Journal of Verbal Learning and Verbal Behavior, 1978, 17, 403-417.

Thorndyke, P. W. Cognitive structures in comprehension and memory of narrative discourse. Cognitive Psychology, 1977, 9, 77-110.

Turner, A., & Greene, E. The construction of a propositional text base. JSAS Catalog of Selected Documents in Psychology, 1978.

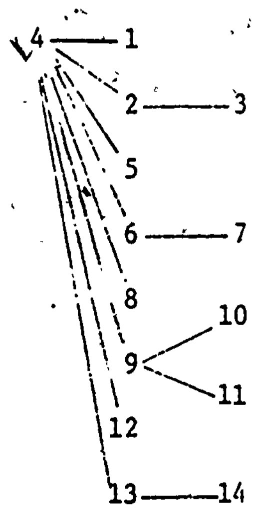
Waters, H. S. Superordinate-subordinate structure in semantic memory: The role of comprehension and retrieval processes. Journal of Verbal Learning and Verbal Behavior, 1978, 17, 587-598.

TABLE 1

Sample of a History Passage with Few Different Arguments Together With Its Propositions and Propositional Hierarchy

Long ago, soldiers who were fighting in the battlefields did not eat very well. They had to live mostly on food that was stale or spoiled. More soldiers died from bad food than from gunshots. (35 words)

1. (TIME: WHEN, 4, LONG AGO)
2. (FIGHT, SOLDIERS)
3. (LOC: IN, 2, BATTLEFIELD)
4. (EAT, SOLDIERS)
5. (NOT, 4)
6. (WELL, EAT)
7. (VERY, WELL)
8. (MUST, SOLDIERS, 9)
9. (LIVE-ON, SOLDIERS, FOOD)
10. (STALE, FOOD)
11. (SPOILED, FOOD)
12. (DIE, $SOLDIERS_1$, FOOD)
13. (DIE, $SOLDIERS_2$, GUNSHOTS)
14. (CONTRAST: MORE THAN, $SOLDIERS_1$, $SOLDIERS_2$)



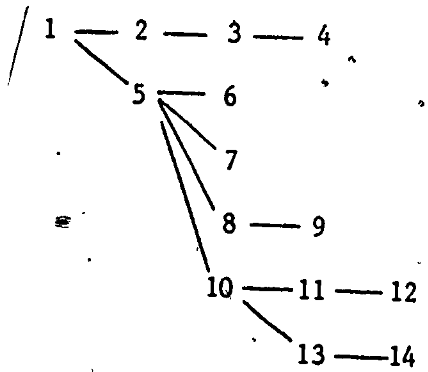
<u>ARGUMENTS</u>: SOLDIERS, BATTLEFIELD, FOOD, GUNSHOTS (4)

TABLE 1 (cont'd)

Sample of a History Passage with Many Different Arguments

Winston Churchill ruled England during the second World War. He made many speeches over the radio. His speeches were important because they gave the army and the people of England courage to fight the Germans. (35 words)

1. (RULE, WINSTON CHURCHILL, ENGLAND)
2. (TIME: DURING, 1, WAR)
3. (WORLD, WAR)
4. (SECOND, 3)
5. (MAKE, WINSTON CHURCHILL, SPEECHES)
6. (MANY, SPEECHES)
7. (LOC: OVER, 5, RADIO)
8. (IMPORTANT, 5)
9. (BECAUSE, 8, 10)
10. (GIVE, SPEECHES, 11, COURAGE)
11. (AND, ARMY, PEOPLE)
12. (LOC: IN, 11, ENGLAND)
13. (PURPOSE, COURAGE, 14)
14. (FIGHT, PEOPLE, GERMANS)



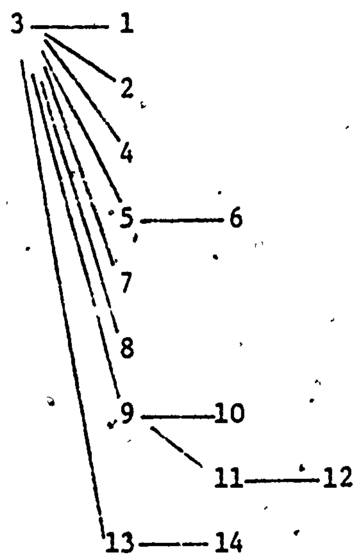
ARGUMENTS: WINSTON CHURCHILL, ENGLAND, WAR, SPEECHES, RADIO, ARMY, PEOPLE, COURAGE, GERMANS (9)

TABLE 2

Sample of a Science Passage with Few Different Arguments
Together with Its Propositions and Propositional Hierarchy

Coal is black. But it comes from green plants that lived long ago. The plants died and they rotted. They were then buried deep into the earth. The earth's weight slowly turned them into coal. (35 words)

1. (BLACK, COAL)
2. (CONCESSION: BUT, 1, 3)
3. (COMES FROM, COAL, PLANTS)
4. (GREEN, PLANTS)
5. (LIVE, PLANTS)
6. (TIME: WHEN, 5, LONG AGO)
7. (DIE, PLANTS)
8. (ROT, PLANTS)
9. (BURY, $, PLANTS)
10. (DEEP, BURY)
11. (LOC: INTO, 9, EARTH)
12. (QUALITY OF, WEIGHT, EARTH)
13. (TURN, WEIGHT, PLANTS, COAL)
14. (SLOWLY, TURN)



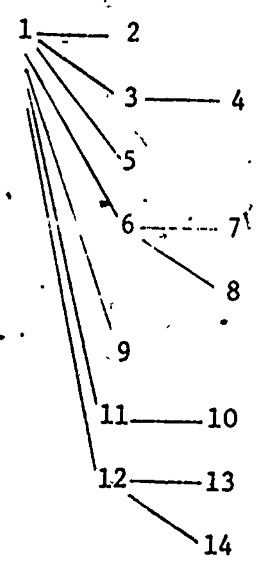
ARGUMENTS: COAL, PLANTS, EARTH, WEIGHT (4)

TABLE 2 (cont'd)

Sample of a Science Passage with Many Different Arguments

Divers carry metal tanks on their backs to help them breathe under water for hours. The tanks are full of air. Two hoses connect the tank's mouthpiece which the diver holds between his teeth. (36 words)

1. (CARRY, DIVERS, TANKS)
2. (METAL, TANKS)
3. (LOC: ON, 1, BACKS)
4. (OF DIVERS, BACKS)
5. (HELP, TANKS, 6)
6. (BREATHE, DIVERS)
7. (LOC: UNDER, 6, WATER)
8. (TIME: DURATION, 6, HOURS)
9. (FULL OF, TANKS, AIR)
10. (TWO, HOSES)
11. (CONNECT, HOSES, TANKS, MOUTHPIECE)
12. (HOLD, DIVER, MOUTHPIECE)
13. (LOC: BETWEEN, 12, TEETH)
14. (OF DIVER, TEETH)



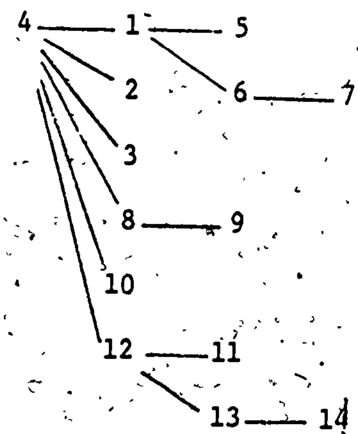
ARGUMENTS: DIVERS, TANKS, BACKS, WATER, HOURS, AIR, HOSES, MOUTHPIECE, TEETH (9)

TABLE 3

Sample of a Familiar Experience Passage with Few Different Arguments Together With Its Propositions and Propositional Hierarchy

Billy's new baby sister has just come home from the hospital. He can hardly wait for her to get big so that he can play with her. Right now she just sleeps and eats. (34 words)

1. (POSSESS, BILLY, SISTER)
2. (NEW, SISTER)
3. (BABY, SISTER)
4. (COME, SISTER, HOME, HOSPITAL)
5. (CAN, BILLY, 6)
6. (WAIT-FOR, BILLY, 8)
7. (HARDLY, 6)
8. (BECOME, SISTER, BIG)
9. (ENABLE, 8, 10)
10. (PLAY, BILLY, SISTER)
11. (TIME: NOW, 12)
12. (13, SISTER)
13. (AND, SLEEP, EAT)
14. (ONLY, 13)

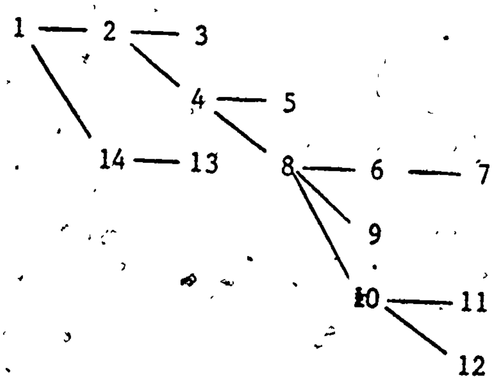


ARGUMENTS: BILLY, SISTER, HOME, HOSPITAL (4)

TABLE 3 (cont'd)
Sample of a Familiar Experience Passage with
Many Different Arguments

Jim went to the zoo with a friend. He saw many animals in cages. But in the bird house, the birds were not in cages; they flew freely among the visitors. One even landed on Jim's shoulder! (37 words)

1. (GO, JIM, ZOO, FRIEND)
2. (SEE, JIM, ANIMALS)
3. (MANY, ANIMALS)
4. (LOC: IN, ANIMALS, CAGES)
5. (CONCESSION: BUT, 4, 9)
6. (LOC: IN, 8, HOUSE)
7. (BIRD, HOUSE)
8. (LOC: IN, BIRDS, CAGES)
9. (NOT, 8)
10. (FLY, BIRDS)
11. (FREELY, FLY)
12. (LOC: AMONG, 10, VISITORS)
13. (LAND, BIRD, SHOULDER)
14. (PART OF, SHOULDER, JIM)



ARGUMENTS: JIM, ZOO, FRIEND, ANIMALS, CAGES, HOUSE, BIRDS, VISITORS, SHOULDER (9)

TABLE 4

Mean reading times (sec) as a function of contact and number of different arguments for good and average readers in fourth and sixth grade

		HISTORY		SCIENCE		FAMILIAR EXPERIENCES	
		FEW	MANY	FEW	MANY	FEW	MANY
4th	Good	17.21	20.40	17.26	20.23	16.17	17.52
	Average	22.09	23.65	22.59	23.10	17.86	22.24
	X̄	19.65	22.03	19.93	21.67	17.02	19.88
6th	Good	15.92	17.71	15.38	17.40	12.59	14.28
	Average	20.95	23.67	20.62	23.29	18.57	20.69
	X̄	18.43	20.69	18.00	20.35	15.58	17.49

TABLE 5

Average number of propositions recalled (out of 14) as a function of content and number of different arguments for fourth and sixth-grade good and average readers.

		HISTORY		SCIENCE		FAMILIAR EXPERIENCES	
		FEW	MANY	FEW	MANY	FEW	MANY
4th grade	Good	8.47	8.00	8.33	7.89	10.03	9.44
	Average	6.06	5.81	7.03	5.69	8.22	7.72
	$\bar{X}$	7.27	6.91	7.68	6.79	9.13	8.58
6th grade	Good	8.42	8.64	7.92	7.94	9.58	9.11
	Average	7.28	7.39	7.31	7.00	7.81	8.89
	$\bar{X}$	7.85	8.02	7.62	7.47	8.70	9:00

TABLE 6

Mean reading time per proposition recalled.

	HISTORY		SCIENCE		FAMILIAR EXPERIENCES	
	FEW	MANY	FEW	MANY	FEW	MANY
4th Grade	2.94	3.65	2.62	3.89	1.92	2.63
6th Grade	2.54	2.60	2.65	2.89	2.00	1.99

TABLE 7

Proportion of propositions recalled as a function of level in the hierarchy and content of passage.

	LEVEL 1	LEVEL 2	LEVEL 3
HISTORY	.80	.60	.45
SCIENCE	.82	.57	.51
FAMILIAR EXPERIENCES	.83	.74	.57

CHAPTER 3

The Effects of Causal Cohesion on Comprehension and Memory

Janice M. Keenan   Susan D. Baillet   Polly Brown

Abstract

It is argued that coreference is neither necessary nor sufficient for integrating sentences in comprehension; rather, the basis for integration is establishing a knowledge-based relation between propositions, one type of which is causal relatedness. Two experiments are reported in which sentence-by-sentence reading times were collected on two-sentence paragraphs, where the first sentence specified a cause for the event in the second sentence. Each paragraph had four versions. All versions had the same second sentence and were referentially coherent; they differed, however, in the causal relatedness of the two sentences. Despite referential coherence, reading times for second sentences were shown to steadily increase as causal relatedness decreased. Recognition and recall memory for the causes was poorest for the most and least related causes and best for causes of intermediate levels of relatedness.

CHAPTER 3

The Effects of Causal Cohesion on Comprehension and Memory

One of the main goals in the study of discourse comprehension is to specify the factors which affect the coherence of a text and hence the ease with which it can be comprehended. To date, psychological efforts in this regard have focused on coreference. Two units of text are said to cohere if they share a common referent, what is also known as the argument repetition rule.

This emphasis on coreference as the basis for textual cohesion in psychological models is due in part to a belief in its relative importance. This belief is most clearly stated by Kintsch and van Dijk (1978). They state, "Referential cohesion is probably the most important single criterion for the coherence of text bases" (p. 367). It is also due to the ease with which coreference can be incorporated into models of comprehension. Unlike other criteria for cohesion, coreference is easily specified, readily agreed upon by all comprehenders, and capable of being defined without reference to particular text units. Thus, even though Kintsch and van Dijk acknowledge that coreference may be "neither a necessary nor a sufficient criterion linguistically", they use it as the sole basis for coherence in their model of comprehension because "the fact that in many texts other factors tend to be correlated with it makes it a useful indicator of coherence that can be checked easily, quickly, and reliably" (p. 367).

Recent work by textlinguists suggests that this reliance on coreference as the primary indicator of coherence may be misleading. Van Dijk (1977) himself has claimed that from the point of view of text grammars coreference is actually quite secondary. Because thaere is nothing to prevent referentially coherent texts from being contradictory, van Dijk argues that "referential identity, although often paid attention to in discourse studies, is neither necessary nor sufficient to determine the meaningfulness of composite expressions" (p. 10). He concludes that the most important criterion for coherence is not coreference but rather a relation, typically conditional, between propositions as they denote related facts in some possible world. In other words, the basis for text cohesion is in terms of shared knowledge structures and not simply repetition of textual elements.

In formulating his position on the importance of knowledge-based cohesion over coreference, van Dijk (1977) was working within the framework of developing a text grammar using only linguistic criteria. He was not developing a process model of comprehension using psychological data and at the time, there was not a sufficient psychological data base to know whether his position applied equally well to process models. It was probably this lack of psychological data that was responsible for Kintsch and van Dijk's (1978) statement that "coreference is neither a necessary nor a sufficient condition linguistically" (p. 367, emphasis added) and for the hedging in van Dijk's (1977) statement that "Ultimately, the connection between propositions is determined by the relatedness of the facts denoted by them, it seems" (p. 47). One of the goals of this paper, therefore, is to review existing evidence as well as present some new data to show that in terms of the factors which affect comprehension time, coreference is neither necessary nor sufficient.

If the coherence of a text is determined by the relatedness of the facts it expresses, rather than coreference, it is necessary to begin examining which types of relations have an effect on comprehension time. Thus, a second goal of this paper is to extend our knowledge concerning the types of relations affecting comprehension times by examining the effects of causal relatedness. We examine reading times for two-sentence texts in which the event in the first sentence can be viewed as a cause for the event in the second sentence. We use the term "cause" to include both types of causality discussed by van Dijk (1977). That is, it includes strict causal relations, such that A is a cause of B if A is a sufficient condition for B; for example, lack of rain causes arid soil. It also includes cases where although A cannot properly be said to cause B, it is a sufficient reason for the action in B; for example, lack of rain causes us to irrigate our fields. We manipulate the degree of causal relatedness between the sentences by changing the event in the first sentence so that it ranges from a highly probable to an improbable, but plausible cause for the event in the second sentence. By examining reading times for the second sentences of these texts, which are all referentially coherent but which differ in their degree of causal relatedness, we simultaneously examine whether coreference is a sufficient condition for text comprehension and whether degree of causal relatedness affects comprehension time.

Is Coreference Necessary or Sufficient for Text Comprehension?

A principal method for determining whether a given type of information is necessary for comprehension is to compare comprehension times when the information is present with those when the information is absent. If comprehension times are longer when the information is absent than when it is present. it is concluded that this information is necessary for comprehension and that the extra time involved is the result of having to search memory and construct an inference in order to provide the needed missing information. If there is no difference in comprehension times between the case when the information is present and when it is absent, then it is concluded that the information is not necessary for comprehension.

Several studies using this method have shown that lack of coreference results in longer comprehension times, thus supporting the view that coreference is necessary for text comprehension (Haviland & Clark, 1974; Garrod & Sanford, 1977; Carpenter & Just, 1977; Singer, 1979; Sanford & Garrod, 1980). For example, Haviland and Clark (1974) constructed pairs of two-sentence paragraphs like the following, in which the only difference between the pairs lay in whether or not the first sentence explicitly mentioned the referent of the second sentence.

(1) John left the beer in the car.
The beer was too warm to drink.

(2) John left the picnic supplies in the car.
The beer was too warm to drink.

In (1) the two sentences are referentially coherent; in (2) they are not. If referential coherence is a necessary condition for comprehension, then in order to comprehend (2), a time-consuming search of memory must occur to provide the information needed to referentially bridge picnic supplies and beer. Because the time to comprehend the second sentence was found to be approximately 200 milliseconds longer in (2) than in (1), it appears that coreferential cohesion is necessary for comprehension.

A recent study reported by Garrod and Sanford (1982), however, suggests that coreferential cohesion is not always necessary for comprehension. Using the same paradigm as that described above, they constructed paragraphs so that the first sentence either explicitly mentioned the referent of the second sentence or merely implied it. The following is an example of one of their paragraph pairs.

(3) Mary put the baby's clothes on.
The clothes were made of pink wool.

(4) Mary dressed the baby.
The clothes were made of pink wool.

Although (3) is referentially coherent and (4) is not, it took no longer to read and comprehend (4) than (3). Across a wide range of materials, only a nonsignificant 7 msec difference between the two types of paragraphs was observed. Thus, coreferential cohesion does not seem to be as necessary for comprehension as the earlier studies led us to believe.

The critical difference between Haviland and Clark's (1974) materials and Garrod and Sanford's (1982) materials seems to lie in the degree to which the first sentence implies the antecedent of the second sentence. The difference is not the result of the structuring of the two sentences; rather, it is the result of one's knowledge about the objects and events mentioned in the sentences. To illustrate, the concept, to dress, in Garrod and Sanford's example, necessarily entails the presence of clothes; one cannot dress without clothes. However, the concept, picnic supplies, in Haviland and Clark's example, does not necessarily entail beer and, in fact, for the more temperate among us, beer would not even be suggested by picnic supplies.

Combining this analysis of the difference in the semantics of the two paragraphs with their different results in the comprehension-time paradigm yields the following conclusion. What is necessary for comprehension is not the presence of explicit coreference, but rather the presence of knowledge-based cohesion. Because the first sentence establishes a scenario that has as one of its parts a concept or event mentioned in the second sentence, the second sentence coheres by virtue of being involved in the same knowledge structure.

It might be argued that Garrod and Sanford's (1982) results merely show that some lexical decomposition is required in the representation of verbs and that coreference -- either explicit or by virtue of decomposition -- is still necessary for the comprehension of texts. However, there are other nonreferentially coherent texts to which such an argument could not apply. Consider the following example.

(5) Things were getting very tense.
Suddenly John punched George and knocked him out.
Mary started screaming.
I ran to the phone and called the police.
Kathy ran for the doctor.

Although this text is generally considered to be coherent, none of the five sentences are referentially coherent. Furthermore, no simple decomposition of any of the verbs would render them referentially coherent. Thus, it appears that referential coherence -- either explicit or by decomposition -- is not a necessary condition for cohesion in text comprehension (cf. Sanford & Garrod, 1981).

The next question we need to consider is whether coreference is a sufficient condition for text comprehension. Coreference can be said to be a sufficient condition for text comprehension if it can be shown that when two text units are referentially coherent, no inferential processes are required to comprehend them.

We are aware of only one study in the literature which addresses this issue. This is a study by Haberlandt and Bingham (1978). They used the same comprehension time paradigm as that described above, with the exceptions that the paragraphs were three sentences in length and both members of the paragraph pair were referentially coherent. The only difference between the members of the pair was in the verb of the second sentence. In one member of the pair, the verb was chosen to be causally coherent with the preceding and following sentences; in the other member, the verb was chosen to be unrelated to the action of the preceding and following sentences. The following is an example of one of their paragraph pairs.

(6) 1. Brian punched George.
2. George called the doctor.

3. The doctor arrived.

(7) 1'. Brian punched George.
2'. George liked the doctor.
3'. The doctor arrived.

Of interest was the reading time for the third sentence. If referential coherence is a sufficient basis for integrating texts in comprehension, there should be no difference in reading times between (3) and (3'). If referential cohesion is not sufficient and knowledge-based coherence is also necessary, then (3') should take longer to read than (3) since it lacks causal coherence. In fact, Haberlandt and Bingham found that (3') took much longer than (3) to read and comprehend, suggesting that referential coherence is not a sufficient condition for text comprehension.

A potential problem with the interpretation of Haberlandt and Bingham's (1978) results stems from the fact that comprehension times differed not only for the third sentences of each pair, but also for the second sentences. As McKoon and Ratcliff (1980) point out, there are sequential effects in reading times such that a slow reading on one sentence can produce another slow reading on the subsequent sentence by virtue of a simple "spillover" of reading time. This raises the possibility that increased reading times for the third sentences in Haberlandt and Bingham's causally unrelated triples were due to "spillover" effects from the second sentence, rather than lack of causal cohesion.

One of the goals of the experiments reported in this paper is to examine whether coreferential cohesion is a sufficient condition for text comprehension under conditions that exclude an interpretation in terms of "spillover" effects. A second goal of this research is to provide a firmer basis than that provided by Haberlandt and Bingham's study for the claim that the speed with which propositions can be integrated is determined by the relatedness of the facts denoted by them. Thus, instead of contrasting the effects of presence or absence of causal relatedness on comprehensiuon time, we manipulate causal relatedness over four levels of relatedness, ranging from highly probable to improbable, but plausible. We assume that the time to establish a causal connection is a function of the a <u>priori</u> relatedness of the events in memory such that the more highly related the two events are in memory, the faster one can construct a causal connection between them and hence comprehend them. If it can be shown that comprehension time decreases as causal relatedness increases, this should provide very compelling evidence that text comprehension is affected by the degree of causal cohesion between events.

The experiments use the comprehension paradigm described above with two-sentence paragraphs such as that shown in Table 1. There are four versions of each paragraph. All versions have the same second sentence, and in all versions, the second sentence is coreferential with the first, as indicated by the use of a pronoun in the second sentence to refer to the proper noun of the first sentence. In each version the first sentence describes an event that can be construed as the cause of the event described in the second sentence. The differences between the four versions lie in the ranked relatedness of the event specified by the first sentence to the event in the second sentence, with Level 1 referring to the most related and Level 4 referring to the least related. Levels of relatedness reflect differences in the probability of the first event causing the second event.

If coreferential cohesion is a sufficient condition for text comprehension, then there should be no difference in comprehension time for the second sentence across the four versions because in each case it is the same sentence and is coreferential with the first sentence. On the other hand, if coreferential cohesion is not sufficient for comprehension and causal cohesion must also be established, then comprehension time for the second sentence should increase as its relatedness to the first sentence decreases.

Insert Table 1 about here

Experiment 1

Method

Materials. Construction of the materials required two phases. The first phase involved finding a set of events which had multiple causes and whose causes spanned a broad range of relatedness to the events. A large set of events, together with their possible causes, was compiled and subjected to a norming study in order to obtain likelihood ratings for each of the causes. Forty subjects, who did not participate in the present experiments, were given booklets, where each page of the booklet listed an event at the top (e.g., a house fire), followed by a list of causes (e.g., a lightning strike, falling asleep with a burning cigarette, faulty wiring, child playing with matches, unattended grease fire in kitchen, spontaneous combustion in garage). Subjects were asked to rate each cause on a five-point scale in terms of how likely it was as a cause of the event. It turned out that for many of the events, the average difference in the likelihood of the causes was either too small or too variable to be suitable for use in a reading time paradigm. Such was the case, for instance, with the house-fire example given above, where across subjects, the various causes turned out to be approximately equally likely to cause the event. There were, however, eight events whose causes consistently spanned a broad range of relatedness; and it was these eight that were used to construct the paragraphs for the experiments. (The fact that there were only eight reflects more on the poverty of our imaginations than on the uniqueness of these events.) The eight events were: a badly bruised child, taking out a bank loan, going to see a doctor, a politician deciding not to run for re-election, joining the Peace Corps, becoming unconscious, getting undressed, having red bumps on the skin.

The second phase consisted of constructing four versions of each paragraph, with each version involving a different cause of the target event, and then norming the relatedness of each paragraph version. This second norming study was deemed necessary in order to ensure that by expressing the events as referentially coherent sentences in which the cause preceded the event we did not alter the rank ordering of the causes obtained when they were judged as simple event phrases with the event preceding the cause.

For each of the eight events obtained from the first phase, we selected the four causes that represented the broadest range of rated likelihood. Each of the four causes was expressed in a sentence which became the first sentence of a paragraph version. The event itself was expressed as the second and target sentence of the paragraph. To illustrate, the example in Table 1 was constructed on the basis that repeated punching was rated as the most likely cause of a child being bruised. Falling off a bike was rated as next most likely, followed by parental abuse; going out to play was considered to be a fairly unlikely cause of a child being bruised.

All second sentences were constructed to be referentially coherent with the first sentence by using a pronoun in the second sentence to refer to the proper noun of the first sentence. Furthermore, all second sentences were constructed to be exactly 12 syllables in length, so as to minimize variability in reading times across the different paragraph sets; the number of words in these sentences ranged from seven to twelve. Equating sentences, in terms of number of syllables rather than number of words is the method recommended by Haviland and Clark (1974). It was impossible to similarly constrain the length of the first sentences because the different causes often varied considerably in the number of words needed to express them. We therefore simply ensured that there was no systematic increase in number of syllables or words in the first sentence with decreased relatedness to the second sentence. The average number of syllables in the first sentences at each level from most highly related to least related were 13, 22, 23, and 11. Because the first sentences of least related causes are the shortest in length, this should eliminate the possibility that any "spillover" effects (McKoon & Ratcliff, 1980) could account for any observed increases in reading times for the second sentences of these pairs. Care was also taken to ensure that frequency and imagability of the vocabulary used to express the causes was similar across relatedness levels.

After the eight paragraph sets were constructed, they were given to 48 new subjects, who did not participate in any of the other experiments, to rank the versions in terms of the relatedness of the two sentences. The subjects were given booklets, where each page listed the four versions of a given paragraph in random order. Four different random orderings were used. The results of this ranking task showed that expressing the events in referentially coherent sentences did not alter the ordering of the relatedness among the versions.

Procedure. The experiment was implemented on a PDP 11/10 computer which is equipped with software to control four independent, sound-shielded, subject stations simultaneously. Each subject station consists of a CRT and button box for collecting responses.

The paragraphs were presented on the CRT, one sentence at a time. Each paragraph was preceded by a screen stating "New Paragraph." The subject then pressed a button to receive the first sentence of the paragraph. Subjects were told to read the sentence both quickly and carefully and to press a button when they understood it. As soon as the subject pressed the button, the second sentence was presented. When the subject pressed the button signaling comprehension of the second sentence, a screen was displayed which requested the subject to write a sentence which continued the thought of the first two sentences. This was done to ensure that the subjects were actually reading the sentences for comprehension and not just pushing buttons. Continuation sentences were written on sheets of paper provided at the subject stations. When the subjects finished writing the continuation sentence, they pressed a button which caused the "New Paragraph" screen to be displayed and the whole sequence of events to be repeated with the next paragraph.

Subjects received one practice paragraph and one version of each of the eight experimental paragraphs. Only one practice paragraph was needed because this experiment followed another experiment which served to familiarize the subjects with the procedure of reading text on the CRT and pressing buttons. The eight experimental paragraphs were presented in a different random order for each subject.

After all the paragraphs had been presented, the subjects were asked to rate the relatedness of each version of each paragraph using a five-point scale (1 = Highly Related and 5 = Not Very Related). This was done to ensure that the difference in relatedness between the versions held for these subjects and to provide a more sensitive measure of relatedness than the ranks obtained in the earlier norming study. The ratings were collected from booklets, where each page listed the four versions of a given paragraph in random order. Four different random orderings were used.

Design. The main independent variable, degree of causal relatedness, had four levels and was manipulated within subjects. Each subject read only one version of each of the eight experimental paragraphs, with two paragraphs at each of the four levels of relatedness. Assignment of paragraphs to level of relatedness was counterbalanced across subjects by using a Latin-square design with four groups of subjects and eight subjects per group.

Subjects. Thirty-two adults from the University of Denver community were paid for their participation in the experiment.

Results

Relatedness Ratings. Because every subject rated every version of every paragraph, it was possible to analyze the relatedness ratings treating both paragraphs and subjects as random factors; level of relatedness from the earlier ranking study was the only fixed factor. The results of this analysis showed a highly significant effect of level of relatedness. $F'(3, 31) = 82.57$, $p < .001$, $MSe = 4.075$. The average ratings for each level are presented in Table 2. As Table 2 shows, rated relatedness steadily decreases across the four levels. Newman-Keuls tests showed that Level 1 versions were rated as significantly more related than Level 2 versions ($q(2, 31) = 4.37$, $p < .01$); Level 2 versions were rated significantly more related than Level 3 versions ($q(2, 31) = 4.18$, $p < .01$); and Level 3 versions were significantly more related than Level 4 versions ($q(2, 31) = 12.54$, $p < .01$). There was also a significant effect of paragraphs ($F(7, 217) = 9.85$, $p < .01$, $MSe = 0.8463$), and a significant interaction of paragraphs with level of relatedness ($F(21, 651) = 5.56$, $p < .01$, $MSe = 0.5904$). Although every paragraph showed a steady decrease in rated relatedness across the four levels, with the exception of one paragraph that had a tie for levels 2 and 3, the interaction was produced by the fact that the slope of this decreasing function varied across paragraphs. The fact that the rank orderings of the versions derived from the relatedness ratings exactly corresponds to the earlier rankings validates the perceived differences in the causal relatedness of the four versions. The question of interest is whether these perceived differences affect reading times for the second sentence or whether reading times are only a function of coreferential coherence.

Insert Table 2 about here

Reading Times. The average reading times for second sentences as a function of causal relatedness to the first sentences are also presented in Table 2. As this table shows, the time required to read and comprehend a sentence is significantly affected by the relatedness of the sentence which precedes it, $F(3, 84) = 9.61$, $p < .001$, $MSe = 0.2484$. Sentences preceded by Level 1 sentences were read significantly faster than those preceded by Level 2 sentences. $q(2, 84) = 3.49$, $p < .01$. The difference between Level 2 and Level 3 is not significant but is in the right direction. Sentences preceded by Level 3 sentences were read significantly faster than those preceded by Level 4 sentences, $q(2, 84) = 3.73$, $p < .01$. The linear trend is highly significant, $F(1, 84) = 26.61$, $p < .001$; and there is no significant quadratic or cubic trend.

There was also a significant difference in first sentence reading times as a function of level of relatedness, $F(3, 84) = 22.50$, $p < .001$, $MSe = 1.0342$. However, the pattern of reading times here is quite different from that observed for second sentences. Most importantly, whereas Level 4 second sentences were read the slowest, Level 4 first sentences were read the fastest. This result therefore eliminates the possibility that spillover effects from the first sentence caused the increased reading time on second sentences of Level 4 versions. The reading times for first sentences directly reflect the differences in the number of syllables for these sentences. Level 4 first sentences had the fewest syllables (11) and were read the fastest (3012 milliseconds); followed by Level 1 sentences (13 syllables, 3263 milliseconds), then Level 2 sentences (22 syllables, 4564 milliseconds), and Level 3 sentences (23 syllables, 4641 milliseconds).

Discussion

The purpose of this experiment was to determine whether or not coreferential cohesion is a sufficient condition for text comprehension. The evidence strongly suggests that it is not. Although second sentences were always coreferential with the first sentence. reading times for these sentences varied with the content of the first sentence. Specifically, the higher the *a priori* relatedness of the event in the second sentence to the event in the first sentence, the faster the second sentence could be read. This result suggests that the presence of coreferential cohesion is not sufficient for comprehension. It shows that the integration of sentences in comprehension must involve establishing some relation between them, since the time to comprehend is systematically related to the strength of this relation.

One might object to this conclusion, however, on the grounds that reading times for the second sentences reflected not the time to comprehend them, as we have been suggesting, but rather the time to think of a continuation sentence. Recall that after reading each paragraph, subjects were required to construct a sentence which continued the thought of the paragraph. Although they were told not to construct the continuation sentence until they had pressed the button signalling comprehension of the second sentence, it is possible that subjects did not always follow these instructions.

It is reasonable to assume that the time required to construct a continuation sentence would be a function of the relatedness of the events in the first two sentences. Thus, if subjects were constructing continuation sentences before pressing the button to terminate the second sentence, the obtained relation between reading time and causal relatedness may simply reflect the effects of causal relatedness on the ease of constructing continuation sentences. In other words, coreferential cohesion may in fact be a sufficient condition for text comprehension; and causal relatedness may only be relevant to the construction of continuation sentences. If this is true, however, then if one eliminates the requirement of constructing continuation sentences, there should be no difference in reading times for second sentences as a function of the causal relatedness of the first sentence. On the other hand, if causal cohesion is necessary to comprehension per se, apart from any involvement in construction of continuation sentences, then even when the continuation task is eliminated, reading times for second sentences should vary as a function of the relatedness of the first sentence. Experiment 2 is designed to resolve this issue by repeating Experiment 1 without the continuation-sentence task.

Experiment 2

This experiment differs from Experiment 1 in two respects. First, a set of "yes/no" comprehension questions were used in place of the sentence continuation task as a means of ensuring that subjects were comprehending the paragraphs and not just pressing buttons. The questions were designed so that subjects could not anticipate them ahead of time. In this way, reading times for the sentences reflect only the time to comprehend the sentence and are not open to the possibility, as they were in Experiment 1, of including the time required to execute the comprehension test.

The second way in which this experiment differs from Experiment 1 is that it includes both a recall and a recognition test for the causes (specified in the first sentences) of the target events. It has often been claimed that causally coherent texts are more memorable than causally unrelated texts (e.g., Schank, 1975; Mandler & Johnson, 1977), and recent evidence by Black and Bern (1981), using free recall and cued recall procedures, supports this contention. However, there has been no investigation of memory as a function of degree of causal relatedness. Does the strength of the causal connection affect recall memory, such that the more related the cause is to the event, the more likely it will be recalled? Or is it the case that the only thing that determines memory is the presence of some relationship, so that related texts are recalled better than unrelated texts with no effect of degree of relatedness?

Another question we were interested in was whether the relationship between the relatedness of the cause and memory for the cause would differ for recall and recognition tests. There is some evidence in the literature to support this hypothesis. For example, Bower, Black, and Turner (1979) and Graesser, Gordon and Sawyer (1979) have shown that recognition is better for actions that are unrelated to a text script than for typical script actions, while Black and Bern (1981) have shown that recall is poorer for unrelated texts than related texts. Thus, we might expect to find highly related causes better recalled, but more poorly recognized, than lowly related causes. In other words, subjects may be more likely to recall how Joey's body got covered with bruises when the cause is something fairly common, such as a fight, than something somewhat unusual, such as child abuse; however, they may be more likely to recognize something unusual, such as child abuse, than something more common, such as a fight.

Method

Materials. The materials consisted of the same eight paragraph sets as used in Experiment 1. In addition, there were eight "yes/no" comprehension questions, one for each paragraph. Four of the questions required a "yes" response and four a "no" response. The questions were designed so that the answer was the same regardless of which version of the paragraph was read. This was accomplished by having the question refer to some aspect of the event in the second sentence, which was the same in all versions. For example, the comprehension question for the example given in Table 1 was, "Was Joey likely to be feeling pain?"

Nine filler paragraphs were also included, with one serving as the practice paragraph. Like the experimental paragraphs, these paragraphs were two sentences long and were followed by a comprehension question; but unlike the experimental paragraphs, the first sentences of these paragraphs did not specify a cause for the event in the second sentences. Instead, a variety of other relations, such as instantiation and scenario inclusion, conjoined the two sentences. The fillers were included for two reasons. One was to make sure that the results of Experiment 1 were not due to subjects adopting some special strategy as a result of discerning the cause-event form of all the paragraphs. Including filler paragraphs whose second sentences could take a variety of relations to the first makes it less likely that subjects will adopt an expectation or strategy for cause-event relatedness. The other reason for including the fillers was to lengthen the experimental session and thus add time to the retention interval between reading and the memory tests.

The memory tests concerned only the eight experimental paragraphs and examined memory for the causes specified in the first sentences. The recall test was a cued recall test, with the events in the second sentences serving as the cues. The eight cues were in the form of questions; for example, "How did Joey get his bruises?" The recognition test consisted of eight, four-alternative, forced-choice items. It used the same question cues as in the recall test, e.g., "How did Joey get his bruises." followed by alternatives which corresponded to the four possible causes from the four versions. e.g., "(a) his brother beat him up, (b) he fell off his bike, (c) his mother beat him, (d) he got into trouble at a neighbor's house."

Procedure. The procedure was similar to that of Experiment 1, with the paragraphs being presented on a CRT screen. one sentence at a time. The difference was that as soon as the subject pressed the button to signal comprehension of the second sentence, the comprehension question for that paragraph appeared on the screen, rather than instructions to write a continuation sentence. Subjects responded "yes" or "no" to the comprehension question by pressing either of two labeled buttons on the response panel. The response panel was equipped with lights above each of the buttons so that feedback on the accuracy of responses to the questions could be given by briefly lighting the light above the correct answer as soon as the subject responded.

The presentation order of the paragraphs was randomized for each subject with the constraint that the first three paragraphs following the practice paragraph be fillers. This was done to eliminate any warm-up effects on the apparatus because, unlike the subjects in Experiment 1, the only experience these subjects had with reading text off the CRT and pressing buttons was in the instruction phase of the experiment.

When the subjects completed reading the eight experimental and eight filler paragraphs and answering their questions, they participated in two other text comprehension tasks which took approximately 35 minutes to complete. They were then given the recall test followed by the recognition test.

Design. The main independent variable, degree of causal relatedness, had four levels and was manipulated within subjects. Each subject read only one version of each of the eight experimental paragraphs, with two paragraphs at each of the four levels of relatedness. Assignment of paragraphs to levels of relatedness was counterbalanced across subjects by using a Latin-square design with four groups of subjects and nine subjects per group.

Subjects. Thirty-six University of Denver undergraduates participated in the experiment in return for either payment or class credit. Nine additional subjects were run in the experiment but excluded from the analyses because they made more than one error on the comprehension questions.

Results

Reading Times. The average reading times for the second sentences are presented in Table 3 according to level of causal relatedness. The fact that these times are considerably shorter ($\bar{X}$ = 2407 milliseconds) than those in Experiment 1 ($\bar{X}$ = 2938) lends credence to the notion that reading times in Experiment 1 may have included some of the time required to construct continuation sentences. However, the fact that reading times in this experiment again show an increase with decreased levels of causal relatedness demonstrates that causal cohesion has its effects not simply on the construction of continuation sentences, but also on the comprehension process per se. The main effect of causal relatedness on second sentence reading times was highly significant, $F(3, 96) = 5.76$, $p < .005$, $MSe = 0.122$. The linear trend was also highly significant, $F(1, 96) = 15.99$, $p < .001$; the quadratic and cubic trends were not significant. Newman-Keuls tests showed that target sentences preceded by Level 1 causes were read significantly faster than targets preceded by Level 3 causes ($q(3, 96) = 3.46$, $p < .05$) and Level 4 causes

($q(4, 96) = 4.92$, $p < .01$), but not Level 2 causes. Targets preceded by Level 2 causes were read significantly faster than those preceded by Level 3 causes ($q(2, 96) = 3.12$, $p < .05$) and Level 4 causes ($q(3, 96) = 4.57$, $p < .01$). The difference between Level 3 and Level 4 was not significant.

Comprehension Questions. Causal relatedness had no effect on the subjects' ability to correctly answer the comprehension questions. Mean percent correct for the four levels from most related to least related was 93, 93, 95, and 96.

Insert Table 3 about here

Recognition. Table 3 also presents the recognition data according to level of causal relatedness. Level of relatedness had a highly significant effect on subjects' ability to recognize the causes (specified in the first sentences) of the target events, $F(3, 96) = 5.03$, $p < .005$, $MSe = 0.27025$. However, in contrast to the effect on reading times, the effect on recognition performance was not linear; the linear trend analysis produced a nonsignificant $F(1, 96) = 1.48$. Rather, there was a highly significant quadratic trend, $F(1, 96) = 13.59$, $p < .001$. Newman-Keuls tests showed that Level 2 and Level 3 causes were recognized approximately equally well and significantly better ($q(3, 96) = 4.17$, $p < .01$ and $q(3, 96) = 3.21$, $p < .10$) than Level 1 and Level 4 causes, which were not significantly different. Thus, recognition is best for causes of intermediate levels of relatedness and poorest for highly related and unrelated causes.

Recall. The recall data were scored for gist by two independent judges; discrepancies were resolved by a third judge. The results are shown in Table 3. Level of relatedness again had a highly significant effect. $F(3, 96) = 10.62$, $p < .001$, $MSe = 0.27199$. As in the recognition data, causes of intermediate levels of relatedness were remembered the best. With the recall data, however, both the quadratic and the linear trends were significant; for the quadratic trend, $F(1, 96) = 20.02$, $p < .001$, and for the linear trend, $F(1, 96) = 11.77$, $p < .01$. Newman-Keuls tests showed that Level 4 causes were recalled significantly more poorly than all the others ($q(4, 96) = 7.67$, $p < .01$; $q(3, 96) = 5.75$, $p < .01$; $q(2, 96) = 4.47$, $p < .01$). Also, the difference between Level 1 and Level 2 was marginally significant, $q(3, 96) = 3.20$, $p < .10$.

Discussion

At the end of Experiment 1, an objection was raised that reading times for second sentences may have been contaminated by the time required to think of continuation sentences. If so, this would permit the possible interpretation that comprehension times are not necessarily influenced by causal cohesion, but only coreferential cohesion, and that effects of causal cohesion are confined to the time required to construct continuation sentences. The results of this experiment, however, eliminate this interpretation. Because subjects could not anticipate ahead of time the nature of the comprehension questions, it is reasonable to assume that the only thing they were doing while the second sentence was exposed was comprehending it. And because comprehension times were shown to steadily increase as the causal relatedness of the events decreased, it can be concluded that coreferential cohesion is not a sufficient condition for text comprehension; rather, one must also try to establish causal cohesion.

The results of the memory tests provided some unexpected findings. First, both recall and recognition provided somewhat similar patterns of results; this was in contrast to our expectation that lowly related causes would be recalled more poorly than highly related causes but recognized as well, if not better, than highly related causes. Least related causes were least well recalled, supporting our expectations and the results of Black and Bern (1981); however, they were also the least well recognized. It is possible that the discrepancy between our expectations for the recognition data and the actual results is due to the fact that the rcognition process was contaminated by having performed the recall test first. As Mandler (1981) has argued, recognition can be accomplished on the basis of either reinstating the context, as in recall, or familiarity. Having just taken a recall test on the same information may have biased the subjects to recognize the items in terms of the same strategy they used in recall -- reinstating the context -- thus yielding the same pattern of results on the recognition test as on the recall test. Future research will need to determine whether this pattern continues to maintain when recognition is not preceded by recall.

It should be noted that there was some difference between the patterns of recall and recognition data. Specifically, the difference between least related and most related causes on the recognition test was small and not significant; whereas on the recall test, it was large and significant. In this respect, then, there is a difference between recognition and recall performance, and the difference coincides with our expectations, in that lowly related statements fare worse on recall relative to the other statements than on recognition.

Perhaps the most surprising result to emerge from the memory tests was the finding that highly related causes were not recalled very well. Although they were recalled significantly better than the least related causes, they were nonetheless recalled significantly worse than the causes of intermediate levels of relatedness. There may be several ways to explain this finding, but perhaps the most reasonable explanation lies in the amount of processing required to establish the causal connection. Jacoby and his colleagues (Jacoby, 1978; Jacoby, Craik, & Begg, 1979) have shown that the memorability of an event is often a function of the amount of processing it undergoes in encoding: the more processing, the more memorable the event. It could be argued that highly related causes undergo relatively little processing in encoding and for this reason are less recallable than causes of intermediate levels of relatedness.

Why might Level 1 causes undergo less processing? Level 1 causes often so strongly suggest the target event in the second sentence, as for example, repeated punching by someone bigger suggests bruises, that the target event can be expected or activated before the second sentence is even presented. Thus, when the second sentence is presented, it requires little processing, as reflected by the reading times. Because no search process is required to find the causal connection, the events undergo relatively little processing and are thus not very memorable.

In contrast, Level 2 and Level 3 causes never strongly suggest the target event in the second sentence; for example, one can fall off a bike or have mother get angry without ever incurring bruises. Thus, comprehension of these causes does not activate an expectation of the target event to facilitate comprehension of the second sentence. Instead, comprehension requires searching memory for the inferential connection; consequently, reading times are longer and the events are more memorable because they undergo greater processing.

Note, however, that there is a definite constraint on the extent to which more processing leads to better memory. Level 4 causes must involve the greatest amount of processing, and yet they are the least memorable. Thus, it appears that a second factor must also be involved in determining the retrievability of the causes; namely, the strength of the inferential connection. If increased processing only results in a tenuous causal connnection, then it cannot facilitate memory.

Further research will certainly be necessary to establish the exact nature of the relationship between memory and causal relatedness. One of the factors which makes our results tentative is that the number of words used to express the causes was not tightly controlled. Because the average number of words in intermediate level causes was greater than that for highly related and least related causes, it will be necessary in future research to exclude the possibility that memorability is determined by number of words.

Conclusions

Two conclusions can be drawn from this paper. One is that with respect to the processes involved in integrating texts in comprehension, coreference is neither sufficient nor necessary. The second is that the causal relatedness of events affects comprehension time such that the stronger the causal link between a cause and an event, the more quickly the event can be comprehended.

Further research is needed to explicate the types of information that are relevant to determining the strength of causal relations. In this study, causal relatedness was specified in terms of ratings of the likelihood of the causes being sufficient conditions or sufficient reasons for the events. But, other factors undoubtedly affected these likelihood ratings, such as the length of the inferential chain connecting the two events and the salience of the events due to personal experience with them. In future research we hope to assess the independent contributions of these various factors.

Further research is also needed to explicate the rules governing the pragmatics of discourse which determine such things as the range of causal relatedness values acceptable in discourse. In this study we purposely avoided using either texts in which the event was a necessary consequent of the cause, as in She was in Denver consequently she was in Colorado, or texts in which there was no possible causal connection, because such texts seem to violate the pragmatics of discourse. Thus, our Level 1 causes were always highly probable but not necessarily entailed; similarly, our Level 4 causes were improbable but not implausible. What is needed is a theory of the types of expectations comprehenders have about the level of description in discourse which, together with their knowledge about possible relations between events, governs the ease with which events are comprehended.

References

Black. J. B., & Bern, H. Causal coherence and memory for events in narratives. Journal of Verbal Learning and Verbal Behavior, 1981, 20, 267-275.

Bower, G. H., Black, J. B., & Turner, T. J. Scripts in memory for text. Cognitive Psychology, 1979. 11. 177-220.

Carpenter, P. A., & Just, M. A. Reading comprehension as eyes see it. In M. Just & P. Carpenter (Eds.), Cognitive processes in comprehension. Hillsdale, N.J.: Erlbaum, 1977.

Garrod, S. C., & Sanford, A. J. Interpreting anaphoric relations: The integration of semantic information while reading. Journal of Verbal Learning and Verbal Behavior, 1977, 16, 77-90.

Garrod. S. C., & Sanford, A. J. Bridging inferences and the extended domain of reference. In J. Long & A. Baddeley (Eds.), Attention and performance IX, in press (a).

Graesser, A. C., Gordon, S. E., & Sawyer, J. D. Recognition memory for typical and atypical actions in scripted activities: Tests of a script pointer + tag hypothesis. Journal of Verbal Learning and Verbal Behavior, 1979, 18, 319-332.

Haberlandt. K., & Bingham, G. Verbs contribute to the coherence of brief narratives: Reading related and unrelated sentence triples. Journal of Verbal Learning and Verbal Behavior, 1978. 17. 419-425.

Haviland. S. E., & Clark, H. H. What's new? Acquiring new information as a process in comprehension. Journal of Verbal Learning and Verbal Behavior, 1974. 13. 512-521.

Jacoby. L. L. On interpreting the effects of repetition: Solving a problem versus remembering a solution. Journal of Verbal Learning and Verbal Behavior, 1978, 17, 649-667.

Jacoby. L. L., Craik, F. I. M., & Begg, I. Effects of decision difficulty on recognition and recall. Journal of Verbal Learning and Verbal Behavior. 1979. 18, 585-600.

Kintsch. W., & van Dijk, T. A. Toward a model of text comprehension and production. Psychological Review, 1978, 85, 363-394.

Mandler, G. The recognition of previous encounters. American Scientist, 1981, 69. 211-218

Mandler, J. M., & Johnson, N. S. Rembrance of things parsed: Story structure and recall. Cognitive Psychology. 1977, 9, 111-151.

McKoon. G., & Ratcliff, R. The comprehension processes and memory structures involved in anaphoric reference. Journal of Verbal Learning and Verbal Behavior. 1980, 19, 668-682.

Sanford, A. J., & Garrod, S. C. Memory and attention in text comprehension: The problem of reference. In R. S. Nickerson (Ed.), Attention and performance VIII. Hillsdale, N.J.: Erlbaum, 1980.

Sanford, A. J., & Garrod, S. C Understanding written language. Chichester: John Wiley & Sons. 1981.

Singer, M. Processes of inference during sentence encoding. Memory and Cognition. 1979, 7, 192-200.

van Dijk, T. A. Text and context: Explorations in the semantics and pragmatics of discourse. London: Longman, 1977.

TABLE 1

Example of One of the Paragraph Sets Used in Experiments 1 and 2.
(Level refers to the rank relatedness of the version,
with 1 = Most Related)

Level	
1	Joey's big brother punched him again and again. The next day his body was covered with bruises. (1.19)*
2	Racing down the hill, Joey fell off his bike. The next day his body was covered with bruises. (1.59)
3	Joey's crazy mother became furiously angry with him. The next day his body was covered with bruises. (2.72)
4	Joey went to a neighbor's house to play. The next day his body was covered with bruises. (3.94)

*Relatedness ratings obtained in the present study.

TABLE 2

Average Relatedness Ratings and Second Sentence Reading Times (msec)

As a Function of Relatedness Level (1 = Most Related)

Level	Relatedness Rating	Reading Time
1	1.21	2611
2	1.76	2916
3	2.29	2947
4	3.87	3277

TABLE 3

Average Second Sentence Reading Times (msec) and Average Percent of Causes Recognized and Recalled As a Function of Relatedness Level (1 = Most Related)

Level	Reading Time	% Recognized	% Recalled
1	2280	78	69
2	2300	91	83
3	2481	89	75
4	2566	71	50

CHAPTER 4

The Role of Encoding and Retrieval Processes in the Recall of Text

Susan D. Baillet   Janice M. Keenan

Abstract

This study is a modified replication of a study by Anderson and Pichert (1978). Subjects read three stories taking a particular perspective for each, recalled each story from that perspective, and either immediately or after one week, recalled the stories again from a new perspective. It is concluded that what is recalled from a text is a function of selection processes operating at <u>both</u> the time of encoding and the time of retrieval. Specifically, the encoding framework induces selective elaboration of textual information which results in differential accessibility of this information in memory. Although the retrieval framework can also operate selectively in making certain information more accessible for output, it is ultimately constrained by the accessibility of information as determined by the encoding framework.

CHAPTER 4

The Role of Encoding and Retrieval Processes in the Recall of Text

Much current research on memory for text has been based on the assumption that encoding processes are the primary determinant of what is remembered from a text (e.g., Kintsch, 1974; Anderson & Biddle, 1975; Craik & Tulving, 1975; Schallert, 1976; Kozminsky, 1977; and Thorndyke, 1977). Although some research on text has investigated possible effects of a retrieval context on memory, the results of that research have supported the belief in the importance of encoding processes (Bransford & Johnson, 1972; Dooling & Mullet, 1973; Thorndyke, 1977). Thus, current research on text memory has accepted this view and has concerned itself with examining the factors which control the types of processing occurring during encoding. However, this prevailing belief in the dominance of encoding factors has been challenged recently by an experiment by Anderson & Pichert (1978).

In Anderson & Pichert's (1978) experiment, subjects read a 373-word story, taking one of two possible points of view, or perspectives, as they did so. The story contained two sets of target information such that one set was relevant to each perspective. When the subjects recalled the story, they recalled more information that was relevant to the perspective they had taken while reading the story than to the alternate perspective. However, when the subjects were given the alternate perspective as a retrieval cue for a second recall trial, they recalled more information relevant to the new cue than to the original perspective. Previously unrecallable information was recalled in response to a new perspective, indicating that information <u>unrelated</u> to the original perspective was stored in memory as well as information that <u>was</u> relevant to that perspective.

Anderson & Pichert (1978) interpreted their results in light of a retrieval schema set up by the new retrieval cue, which allowed access to previously unrecallable, but encoded, information. While they state that it is likely that both encoding and retrieval effects play a role in memory, they add that in their study, "readers must have developed a richer representation for the story material than could be accounted for solely in terms of the dominant schema brought into play by the perspective instructions. Otherwise there would have been no information in the recesses of the mind which could be recovered when the perspective shifted." (p. 10)

Even more recently, Britton, Meyer, Hodge, & Glynn (1980) also found that retrieval processes are responsible for selective recall, rather than encoding processes. Britton et al. (1980) compared free recall to cued recall of paragraphs that were high or low in the text hierarchy (see Meyer, 1975, for a description of how the test hierarchy is formed). A paragraph low in the text hierarchy contains information which is important to the text as a whole, and which is likely to be recalled. A paragraph that is high in the text hierarchy contains less important information, which is less likely to be recalled.

Britton et al. (1980) found that in the free recall condition, more information was recalled from the paragraph that was high in the text structure than from the paragraph that was low in the text structure. In the cued recall conditions, however, memory for the low paragraph was as good as that for the high paragraph, suggesting that information from the low paragraph was in memory but was not accessible without the appropriate retrieval cue. Thus, the results of the story by Britton et al. seem to support the conclusions presented by Anderson & Pichert (1978) concerning the role of retrieval processes in text memory.

Since Anderson & Pichert's (1978) results and those of Britton et al. (1980) are contrary to earlier research on both word lists and texts (Thomson & Tulving, 1970; Tulving & Thomson, 1973; Thorndyke, 1977), and to the prevailing view that encoding processes are the source of selective recall, it is important to scrutinize their procedures and results very carefully. In doing so, we find a number of methodological factors which may be responsible for their findings.

The principle factor which may be responsible for giving the impression that so much information is available in memory regardless of the encoding perspective is the length of the delay interval. In Anderson & Pichert's study, the first recall trial occurred twelve minutes after the story had been read, and the second recall trial was within the half hour. Britton et al. also used immediate recall. It may be that the encoding perspective did have a significant influence on the type of processing that information in the text underwent, but that with such a short delay, items with even a minimal amount of processing were available and accessible in memory. Keenan (1975) has shown that memory for even the surface form of a text can be retained for up to twenty minutes after reading. The accessibility of these less-processed items could then have been determined by the retrieval perspective. However, if a longer delay had been used, the availability of all information would be reduced, thus rendering differences in availability more apparent, and a difference in recall as a function of the encoding perspective might have been demonstrated.

The second methodological factor that may have contributed to the results obtained by Anderson & Pichert (1978) and by Britton et al. (1980) is the use of intentional memory instructions. Given explicit instructions that a memory test will follow the story, a subject is likely to try to remember as much of the story as possible, regardless of the encoding framework. If an incidental task were used, this problem would be avoided.

The third point which may have contributed to the results is the length of the story used. The texts used by these researchers were so short that it is likely that much or all of them could be encoded fairly easily, especially when the instructions specifically requested this. If a longer test were used, the reader may be forced to encode only a part of the information, and an effect of an encoding perspective could be more clearly determined. Thus, in these two studies, the techniques used to measure the effect of the retrieval process leave doubts as to the generality of the conclusions.

The arguments presented above might lead us to reject the results of the studies by Anderson & Pichert (1978) and by Britton et al. (1980). However, a study by Hirst (1980), which does not suffer from the methodological flaws present in the two studies discussed above, provides evidence supporting the notion that encoding processes may not be the source of selective recall.

Hirst's (1980) experiment is a study of memory for mathematical proofs. Hirst reasoned that the ability to work out a proof is facilitated when a problem can be identified as belonging to a class of problems for which the general form of the proof is known. Once the general rule is known, a new problem of the same form can be easily solved. Hirst sought to teach his subjects rules for various types of problems by presenting instructions about the rule either before a to-be-studied proof, after the proof, or not at all. He viewed the instructions as a form of context analogous to Bransford & Johnson's (1972) picture context, because knowledge of the general rule provides a framework with which to comprehend the proof; given after the proof, the rule's effect would have to be on retrieval.

Measuring recall of the first step of the study proofs, Hirst found no difference between the instructions-before and the instructions-after conditions; though subjects performed better with instructions than they did on problems where they received no instructions. Similarly, subjects could generalize the rule to new problems equally well whether the instructions had come before or after the study proof, and they performed better than they did on new problems for which no rule had been learned.

Hirst interprets these results as evidence for construction occurring at the time of retrieval, since a rule given after the study proof facilitated recall of the first step of that proof as well as did instructions given before the study proof. However, it should be noted that when the rules were presented after the proofs, the delay was the length of a five-minute rest break, while the test was the following day. Furthermore, the subjects had been given 15 minutes to study and memorize the study proofs, unlike Bransford & Johnson's (1972) subjects who heard a passage once. After this much study, it is possible that the proofs would be encoded strongly enough for the subjects to be able to apply the rules presented after the proofs as well as the rules presented before the proofs. Because of the very short delay used, and the amount of learning before the rules were given, it is not clear that Hirst's results unequivocally demonstrate a retrieval phenomen of the kind that Anderson and Pichert (1978) postulate. Nevertheless, Hirst's results are suggestive that retrieval processes may have more of a role in memory than is currently believed by most researchers.

The experiment described here is designed to determine the validity of the claims that selective recall is a function of processes operating at the time of retrieval rather than during encoding. The experimental procedure is somewhat similar to that of Anderson & Pichert (1978). The materials involved stories that were written such that they can be read from one of two perspectives. Subjects read the stories from one perspective, and were then asked to recall the stories. Following the first recall trial, they were given the alternate perspective for each story (this manipulation is referred to as a perspective shift) and were asked to recall each story again. Of major interest are the data from this second recall trial. In particular, we are interested in determining whether our subjects, like Anderson & Pichert's, can recall more information relevant to the second perspective, following the perspective shift, than they could recall on the first recall trial.

However, this procedure differs from that of Anderson & Pichert's in several ways. First, the delay interval between reading the stories and the perspective shift for the second recall trial was manipulated. Half of the subjects were given the alternative perspective and were asked to recall the stories again approximately twenty minutes after their first recall trial, while the other half waited a week between the first and second recall trials. By introducing a week delay between reading and recall, we are greatly reducing the overall level of accessibility of information and are thus permitting any differences in accessibility, which may be due to the encoding perspective, to be demonstrated.

A second way in which this study differs from that of Anderson & Pichert is that instead of instructing subjects to read the stories in preparation for a memory test, our subjects were instructed to read the stories in order to make a judgment concerning a question that relates to each story. For example, one of the stories is a murder mystery story, and the subjects were told to read the story in order to determine the evidence against a particular character, as if they were lawyers desiring to convict that suspect. Because of the demand characteristics of a memory experiment, a subject told to read a story in preparation for a memory test is likely to try to remember as much as possible. regardless of any encoding perspective. By asking the subject to read the story in order to answer a question concerning the encoding perspective, we are maximizing the probability that an effect of the encoding perspective on memory will be demonstrated in the data.

Third, this study uses more stories. more complex stories, and longer stories than did Anderson & Pichert, who used a short, fairly simple story, most of which could be encoded after one reading. The effect of an encoding perspective would not be evident if the entire story had been encoded. Because the stories used here are quite a bit longer (908, 1286, and 21 words) and are more complex, it would be impossible to remember all of the information from all three stories after one reading. Therefore, a possible effect of the encoding schema is more likely to be shown here than in an experiment with a short, simple story.

Finally, in this study both the title of the story and the goal for which the subject is reading are manipulated during encoding. Each story was written such that both perspectives are directly related to the content of the story. Either perspective is a plausible title, since each perspective is directly related to the plot. Similarly, each perspective provides a reasonable question to be presented to the subject as a goal for reading the story. Therefore, it is possible to independently manipulate the title and goal given to the subject, and still maintain a reasonable "real-life" task.

When the title and goal are mainipulated independently, they can either refer to the same perspective or they can each refer to a different perspective. In this experiment, both of these conditions are examined. Let us first take the case where the title and goal refer to the same encoding perspective. This condition is referred to as the Consistent condition, since the title and goal are consistent with each other. When subjects in this group initially read the story, the title of the story and the question for which they are asked to read the story refer to the same subset of information in the story. For example, they may read that the title of the story is "A Nephew Kills his Uncle to Gain Inheritance," and they may be asked to see what evidence could be used to convict Walters, the nephew, of the crime. Later, however, before the subjects are asked to recall the story a second time, they are given the alternative perspective in the form of a new goal, which they are asked to think about prior to recalling the story again. In the example above, the new goal would be to think about how to convict Boardman, the secretary (the other suspect). So, before the second recall trial, subjects read the original title plus the new goal, which refer to different sets of information.

The case when the encoding title and goal differ is referred to as the Inconsistent condition. Before reading the story, subjects in this condition read a title and goal that each refer to a different perspective. For example, the title might be "A

Nephew Kills his Uncle to Gain Inheritance", and the subject might be asked to determine how a lawyer could convict Boardman. the secretary, of the crime. Then, before the second recall trial, the new goal relating to convicting Walters, the nephew, would be presented, along with the original title. So, in the Inconsistent condition. the title and goal differ at encoding, but refer to the same perspective at the second recall trial.

A third condition studied in this experiment is the case where neither perspective is referred to, and the title and goal are neutral. This condition is called the Control condition. The control title for the story referred to above is "The Murder of Ellington Breese," and the goal is to determine which suspect is the criminal. Control subjects do not participate in the perspective shift manipulation.

The reason that it is important to separate the effect of the title from the effect of the reader's goal is as follows. It has frequently been demonstrated that the title of a story, or its text structure, can influence what is recalled, presumably by manipulating the importance of certain elements of the story to the text as a whole (Kozminsky, 1977; Meyer, 1975). However, Meyer and Freedle (1979) accidentally found that a reader's interests can override the structurally determined importance of elements in a text when their subjects, public school teachers, read a text which referred to the dismissal of public school coaches. The information recalled by these teachers did not fit the predicted pattern. because their particular interests. rather than the text structure, determined what information was important. Therefore, since the interests or goals of the reader can affect what is recalled independently of the topic of the text as indicated by the title or text structure, we seek to determine which aspect of the reader's perspective is more important in affecting what is recalled.

The major question we are interested in is whether recall is determined by selective processing which occurs during encoding or by selective retrieval mechanisms. To answer this question we will look at memory for information relevant to the two perspectives during the first and second recall trials, and will compare the performance of subjects in the immediate condition to that of subjects in the delay condition. If retrieval processes determine recall, then when subjects are given a new perspective for the second recall trial, we would expect them to recall more information relevant to that perspective than they did on the first recall trial. Furthermore. we would expect the same result for subjects in both the immediate and delay conditions.

If, however, only encoding processes determine recall, then information related to the encoding perspective should be recalled better than other information during both recall trials and at both delays. If this occurs, then it would suggest that the results of those researchers who have found evidence for retrieval processes may be due to the particular materials and methodology.

A third possibility is that both encoding and retrieval processes have a role in determining what is recalled. Since in the immediate condition, the second recall trial occurs so soon after encoding, it is possible that some information related to the new perspective would be accessible in memory and would be recalled if the appropriate retrieval cue is presented, even if this information were less strongly encoded than information related to the encoding perspective. If this were the case, subjects in the immediate condition would replicate the results of Anderson & Pichert (1978). However, if the information related to the new perspective is only weakly encoded, it may not be accessible after a week, in which case those subjects in the delay condition would demonstrate only an effect of the encoding perspective. Thus, a difference between the immediate and delay conditions would suggest that both encoding and retrieval processes are involved in determining recall.

If the data establish that encoding factors help to determine recall, then we may ask whether the selection of the information to be stored is controlled by the title of the story or by the reader's goal. For this question we must compare the performance of the Consistent and Inconsistent groups. relative to the Control groups, who receive neutral titles and goals. If the title is more important in determining the perspective, then there will be little difference between the Consistent and Incosistent groups, since the title cue does not change between the two recall trials for either group. Recall for both groups would be primarily determined by the original title received before the story was read. If the goal is more important, again there will be little difference between the Consistent and Inconsistent groups, since for both conditions the goal does change between the first and second recall trials, and recall will be primarily determined by the goal.

However, if both the title and the goal contribute to the perspective taken by the reader, then the selective effect of the encoding schema on the first recall trial will be attenuated for the Inconsistent group, since both perspectives are cued when the title and goal differ; and information about both topics would be encoded. Similarly, the possible short-term effect of the retrieval schema on the second recall trial for subjects in the immediate condition will be attenuated for the Consistent group, when both perspectives are cued by the presentation of the origianl title and the new goal. Since previous research (Anderson & Biddle. 1975; Kozminsky, 1977; Greene. 1977) has demonstrated that when tested individually, the title and the reader's purpose can both affect recall. it is expected that in this experiment, both the title and the goal will affect recall.

Method

Subjects

Ninety people from the University of Denver and the University of Portland areas served as subjects. They ranged in age from 15 to approximately 55. Twenty-one additional subjects were run but their data were discarded; ten of these subjects chose not to complete the experiment, and eleven answered the question posed in the goal rather than recalling the stories (after these early subjects were detected, more emphatic instructions were introduced). Subjects were paid and some received course credit as well.

Design

Subjects were randomly assigned to one of six groups representing the combination of Group (Consistent, Inconsistent, and Control) and Delay (twenty minutes and one week). The data were collapsed across the three stories. For the two experimental groups (Consistent and Inconsistent), there were two main factors under investigation: Orientation and Information Set. Orientation refers to the combination of the title and the goal presented refers with each story. For the Consistent group, this was either Title 1 and Goal 1 or Title 2 and Goal 2, while the Inconsistent group read either Title 1 and Goal 2 or Title 2 and Goal 1. For each story, there were two Information Sets, each referring to specific statements from the story. Also included in the analysis was the factor of Order, with three levels, representing a counterbalanced order of story presentation. Order, Orientation, Delay, and Information Set were crossed; subjects were tested in Order, Orientation, and Delay. Thus, Information Set was a within-subject factor, while Order, Orientation, and Delay were between-subject factors.

For the Control group, the design was the same as for the experimental groups, with the exception of the Orientation factor. The Control subjects all received the same, nonbiasing titles and goals, so there was no Orientation factor. (See Table 1 for an outline of the design for the Consistent and Inconsistent groups.)

Insert Table 1 About Here

Thirty-six subjects served in each of the two experimental groups: eighteen per delay, nine per orientation, and three per order. Eighteen subjects served in the Control group: nine per delay and three per order.

Materials

The materials consisted of three stories, written or adapted from other sources in such a way that each could be about one of two possible topics. The Housewalk story, written by one of the authors, is 1286 words long, and concerns a young couple who went on a housewalk, or tour of three houses. Within the story, the identities of the two young people are unspecified, allowing the title to manipulate who they were and their purpose in attending the tour. One experimental title identified the couple as interior designers, studying the decor of the houses ("Interior Design Students Study Classic Homes"), while the other experimental title named them burglars looking at possible houses to rob ("Enterprising Burglars Check Out Future Jobs"). The control title was neutral toward the two possible topics ("Young Reporters Investigate 'How the Other Half Lives'").

The Murder story, adapted from "The Murder of Ellington Breese," from the Baffle Books (Wren & McKay, 1929). is 908 words long, and contains information implicating two men in the murder of a wealthy businessman, though there is insufficient evidence to convict either man. Each of this story's experimental titles points at one of the two suspects as the killer: "A Nephew Kills his Uncle to Gain Inheritance", and "A Secretary Kills his Employer to Gain Power". The Control title was "The Murder of Ellington Breese".

The Boardinghouse story, adapted from James Joyce's "The Boarding House", portrays a series of events primarily involving two characters, Mrs. Mooney (the proprietor of the boardinghouse) and Mr. Doran (a boarder). This story was the longest of the three, at 2171 words. Each of the two experimental titles highlights the actions of one of the two characters: the titles were "Mrs. Mooney Marries Off her Daughter", and "Mr. Doran's conscience Forces him to Marry". The Control title was "The Boardinghouse".

Subjects were also given a goal for which to read each story, which could complement or differ from the story's title, for the experimental groups, or which was neutral, for the Control group. For the Housewalk story, the Control group was asked to just read the story in anticipation of being asked to make an unspecified judgment. The Consistent and Inconsistent groups received either Goal 1, which was to decide which house was decorated most attractively and which would be the nicest to live in, or Goal 2, which was to decide which house would be the easiest to rob, with easiest access and the best loot.

For the Murder story, the Control group's goal was to decide which suspect was the culprit. Goal 1 for the experimental groups was to decide how a lawyer would convict Walters, the nephew, of murdering his uncle. Goal 2 was to decide how a lawyer would convict Boardman, the secretary, of killing his employer.

The Control group's goal for the Boardinghouse story was to think about the motives and behavior of the people in the story for a question which would be asked after the story was read. Goal 1 for the experimental groups was to think about the motives for the behavior of Mrs. Mooney, while Goal 2 was to think about the motives for the behavior of Mr. Doran.

Each story contains information relevant to the point of view determined by each of the titles and goals. In the Housewalk story, the decor of the houses (interesting to interior designers) and valuable, portable objects (interesting to burglars) are described. In the Murder story, evidence and motives relevant to both the nephew and the secretary are included. And in the Boardinghouse story, the behavior and motives of both of the primary characters are portrayed. The contents of the two information sets for each story were determined by means of a rating study.

In the rating study, a panel of eight judges read and evaluated the stories in terms of the two perspectives for each story. First, the Murder story and the Boardinghouse story were edited somewhat from the original versions, to reduce their length and, in the case of the Murder story, to delete a clue which could be used to identify the true killer. Next, the sentences in each story were broken down into their constituent clauses, which were marked by placing a slash between each pair of clauses. Then the eight judges read each story with the purpose of rating the relevancy of each clause to each perspective.

To rate the stories, each judge read each story a minimum of three times. The first time through, the judge read the story to familiarize him or herself with it. The judge read the story a second time in order to rate each clause in terms of its importance to one point of view, and a third reading allowed the judge to rate each clause in terms of the alternate perspective. Thus, each clause was rated twice, once in terms of each perspective.

Each clause was rated on a scale of 1 to 5, with a score of 1 indicating that the clause easily be eliminated, and a score of 5 indicating that the clause was very important to the point of view. The scores for each rating were summed over the eight scorers. If a clause scored at least 30 points on one perspective <u>and</u> no more than 19 on the other, then it was chosen to be part of the target information set for the perspective on which it received the high scores. Therefore, the target clauses were rated as relevant to only one of the two perspectives.

By this rating system, none of the stories had the same number of clauses per information set, so proportions were used in the analysis of the results.
The Murder story had 10 clauses related to one suspect (Walters) and 11 clauses related to the other suspect (Boardman). The Housewalk story had 17 clauses related to the Interior Designers and 11 related to the Burglars. The Boardinghouse story had 22 items relevant to Mrs. Mooney, the proprietor of the boardinghouse, and 47 related to Mr. Doran, the boarder.

Finally, a questionnaire concerning the subjects' own perceptions of their behavior during the experiment was completed by each subject. The purpose of this questionnaire was to determine whether the subjects had followed the instructions to maintain a perspective while reading and recalling, and to determine whether they felt they had, indeed, written everything they could recall.

<u>Procedure</u>

Subjects were tested in groups of one to ten. Each subject worked independently, following written instructions which informed the subjects that they were to read the stories in order to answer particular questions. The subjects were then allowed to read the stories at their own pace. After the subjects finished reading the stories, they were told that there would be a short task before the story-related questions were to be answered. A "Memory Precision Task" was distributed and the subjects were asked to spend approximately ten minutes writing from memory two pieces of "literature" such as the National Anthem, the Pledge of Allegiance, a prayer, or a poem; this filler task is similar to that used by Anderson & Pichert (1978). The purpose of this task was to introduce a short delay in order to minimize verbatim recall of the stories.

After completion of the memory Precision Task, each subject was given the first recall task. The written instructions emphasized the necessity of writing down everything everything the subject could remember, and the experimenter reiterated this point orally. As recall cues, the subjects were provided with both the title of the stories and the goals for which they had read the stories in the same order as the stories had been read. The subjects wrote the stories from memory at their own pace.

After recalling the three stories, the subjects in the Delay condition left, while those in the Immediate condition were given a short break if they desired. The next task was a filler task designed to introduce a delay of approximately 20 minutes. This task involved ordering scrambled statements or pictures into a story, determining the importance of each item to the story, and writing a summary of the story. The story used in this task was not one of the three the subjects had read earlier.

Following the filler task, the subjects were given the second recall test; this time, as recall cues, they were provided with the title they had read at encoding and the goal they had not read earlier. For the Consistent group subjects, this meant that the title and the goal now concerned different information sets, while for the Inconsistent group subjects, the title and goal now referred to the same information set. The Control groups, having read a neutral goal at encoding, received only the stories' titles as retrieval cues on both recall trials. Again, the instructions emphasized that the subjects were to write down as much as they could remember, and explained that the new goal was presented only to remind them of the story, not as a question to be answered. Then the subjects were free to recall the stories at their own pace.

Finally, the subjects filled out the debriefing questionnaire. The entire experiment took between two and three hours for the Immediate group subjects.

When the Delay group subjects returned after a week, they were given the second recall test and the debriefing questionnaire. Thus, the procedure for the two groups was identical except for the amount of delay between the first and second recall trials, and the omission of the story-ordering filler task for the Delay condition subjects.

Results

Scoring

The two protocols produced by each subject were examined for presence of the target clauses for each information set that had been identified in the rating procedure. Statements in the protocols that did not correspond to the two information sets were not scored. Thus, in talking about amount of recall, we will always be referring only to the amount of target information recalled. Protocols were scored for the gist of a target clause rather than for verbatim recall. Recall of part of a target clause was counted as a correct recall of the target. The protocols were coded to prevent the scorer from knowing the predicted results for each subject. After approximately one-third of the protocols had been scored, the scorer checked those already scored to ensure consistency of scoring. After all of the protocols were scored, an independent judge, also ignorant of the predictions, read and checked twelve protocols for accuracy and consistency of scoring. There was a question about the scoring of only one clause, which was resolved by discussion.

In order to analyze the data, the number of target clauses recalled for each information set for each story was converted into a proportion by dividing the number recalled by the total number of target clauses in that information set, in order to equate for unequal target set size. Analyses were performed on proportions, but the data are presented here in percentages for ease of exposition.

First Recall Trial

Table 2 presents the data from the first recall trial. Of major interest in examining the data is the question of whether the titles and goals had any influence on the amount recalled from each of the two information sets. In order to answer this question, however, we first need to examine the data from the Control group to see if the two information sets are equally memorable in the absence of any biasing titles or goals. As it turned out, the two sets were not equally memorable, $F(1,12) = 9.71$, $MSe = .002$, $p < .01$. Information Set 1, which included items related to the Interior Designers, Mrs. Mooney, and Walters, was more memorable (34.8%) than Information Set 2, which included items related to the Burglars, Mr. Doran, and Boardman (30.2%).

Despite the presence of this initial bias in memorability, nonetheless, for the experimental groups, the orientation created by the title and goal proved to be the controlling factor in determining the amount recalled from each of the two information sets. This is most evident in the data for the Consistent groups, where there was a highly significant interaction between Orientation and Information Set ($F(1,24) = 22.48$, $MSe = .0029$, $p < .001$). Title Set 1 - Goal Set 1 subjects recalled 40.2% of Information Set 1 and only 36.6% of Information Set 2. While Title set 2 - Goal Set 2 subjects recalled 41.6% of Information Set 2 and only 33% of Information Set I.

The influence of titles and goals on the selectivity of recall is also apparent in the data for the Inconsistent groups, although the effect here is somewhat less straightforward. Recall that in the Inconsistent groups the title highlights one set of information while the goal highlights the other set. The effect of this manipulation was to make both sets of information equally memorable as evidenced by the lack of a main effect of Information Set and the lack of an interaction between Orientation and InformationSet. The fact that Information Set 2 was less memorable than Information Set 1 for the Control group but was equally memorable for the Inconsistent groups shows that the titles and goals had an effect on recall in that they overpowered the initial bias toward Information Set 1.

Insert Table 2 About Here

In terms of the overall amount recalled for each of the three groups, there was a tendency for subjects in the experimental groups to recall more of the target information than subjects in the Control group. Overall amount recalled was 37.9% for the Consistent groups, 36.7% for the Inconsistent groups, and only 32.6% for the Control group. Although this result suggests that reading for a specific goal increases overall recall compared to the general goals given to the Control group, this difference in recall did not prove to be statistically significant, $F(2,87) = 1.48$, $MS = .01$.

Second Recall Trial

The data for the second recall trial, that is, the recall trial following a perspective shift, are presented in Table 3. The questions of major interest in examining these data are whether the perspective shift results in a greater amount of recall from the information set referred to by the retrieval perspective than the information set referred to by the encoding perspective, as was found by Anderson and Pichert (1978), and if so, whether the effect obtains in the delay condition as well as in the immediate condition.

Insert Table 3 About Here

Let us first examine the data from the Immediate condition. Looking first at the Consistent groups, we find that when subjects who encoded the stories with Title Set 1 - Goal Set 1 were given Goal Set 2 as the retrieval perspective, they recalled more from Information Set 2 (36.9%) than from Information Set 1 (31.5%). While this difference is not statistically significant, recall that on the first recall trial, when these subjects were given Goal set 1 as the retrieval perspective, they recalled more from Information Set 1 (40%) than from Information Set 2 (38%). Furthermore, a comparison of the amount recalled from each Information Set in the two recall trials reveals a significant interaction between Recall Trial and Information Set ($F(1,6) = 7.4$, $MSe = .002$, $p < .05$). Thus, this group demonstrates the same shift in recall pattern as a function of the retrieval perspective (here determined by the goal Set) as that reported by Anderson and Pichert (1978). However, it is important to note that the percent of items in the information set cued by the new retrieval perspective recalled in the second recall trial does not increase overall relative to recall of that information set in the first recall trial; rather, the results suggest an omission of unrelated information. This contradicts Anderson and Pichert's (1978) proposal that new, previously unavailable information is accessed by the retrieval cue.

A similar pattern is found in the results of subjects who encoded the stories with Title Set 2 - Goal Set 2, and were given Goal Set 1 as a retrieval cue for the second recall trial. On that trial, they recalled more information from Information Set 1 (35.1%) than from Information Set 2 (34.7%). Although this difference is not significant, it represents a substantial shift from the pattern of recall obtained from these subjects on the first recall trial when Goals Set 2 was the retrieval perspective; on that trial, they recalled more from Information Set 2 41.4%) than from Information Set 1 (33.6%). When the results from the first and second recall trials are compared for this group, we find a significant interaction between Recall Trial and Information Set ($F(1,6) = 22.04$, $MSe = .0007$, $p < .01$). This group also demonstrates a shift in recall as a function of retrieval perspective. While recall of Information Set 1 does increase from recall trial 1 to recall trial 2, the increase in items recalled is smaller (33.6% to 35.1%) than is the decrease in recall of items from Information Set 2 (41.4% to 34.7%), suggesting once again that the perspective shift may be due to omission of irrelevant information rather than an increase in accessibility of relevant information.

Examining the data from the Inconsistent groups in the Immediate condition, we find that when subjects who encoded the stories with Title Set 1 - Goal Set 2 are given Goal Set 1 as the retrieval cue, they recall more from Information Set 1 (31.1%) than from Information Set 2 (25.9%); however, this difference was not significant. Neither was there a significant interaction between Information Set and Recall Trial, although the pattern of results indicates the same kind of perspective shift shown by the Consistent groups. Recall of Information Set 1 decreased from 34.1% to 31.1% from recall trial 1 to recall trial 2, while recall of Information Set 2 decreased much more sharply, from 36.1% in recall trial 1 to 25.9% in recall trial 2. Thus, there is a suggestion of an effect of the retrieval perspective (Goal Set 1) at the second recall trial.

The results from the Inconsistent subjects in the Immediate condition who encoded the stories with Title Set 2 - Goal Set 1 and who were given Goal Set 2 as a retrieval perspective also indicate a shift in recall as a function of perspective. These subjects recalled significantly more from Information Set 2 (30.5%) than from Information Set 1 (20.7%), $F(1,6) = 8.59$, MSe = .005, $p < .05$. When these data are compared to the data from the first recall trial, we find a significant interaction between Information Set and Recall Trial ($F(1,6) = 30.4$, MSe = .001, $p < .01$). Again, retrieval perspective seems not to increase the recall of the related information, but to lessen the decrease of recall between trials compared to the unrelated items.

The length of the interval between first and second recall trials in the immediate condition of this experiment corresponds to the interval used by Anderson and Pichert (1978), and, as we have seen, our data nicely replicate their result of a shift in recall as a function of perspective shift. We have noted, however, that while there is a shift in the pattern of recall, there is little evidence that the new retrieval perspective leads to an increase in the overall percent of related information recalled compared to the first recall trial. The question of major interest now is whether the results of our delay condition will also show that what is recalled is a function of the retrieval perspective.

When subjects in the Consistent group who encoded the stories with Title Set 1 - Goal Set 1 were given Goal Set 2 as a retrieval cue after a delay of one week, recall of Information Set 2 (32.8%) was not superior to recall of Information Set 1 (33.6%), as would be predicted if a perspective shift influenced recall. The difference between recall of Information Set 1 and Information Set 2 was not statistically significant, $F(1,6) < 1$, MSe = .0047. When performance on the second recall trial is compared to the first recall trial, no interaction is found between Recall Trial and Information Set (as can be seen from Figure 3; $F(1,6) = 2.05$, MSe = .002), unlike the results from subjects in the Immediate condition.

Similarly, Delay subjects who encoded the stories with Title Set 2 and Goal Set 2 and who were given Goal Set 1 as a retrieval perspective recalled more of Information Set 2 (29.1%) than of Information Set 1 (27%), a nonsignificant difference ($F(1,6)$ 1, MSe = .0034). When these data are compared to the results from the first recall trial, there is no significant interaction between Recall Trial and Information Set ($F(1,6) = 3.1$, MSe = .0037), unlike the results from the Immediate condition. Thus, from the Consistent groups, there is no evidence of an effect of the retrieval perspective on recall in the Delay condition.

Turning to the Inconsistent groups, we see a pattern of results similar to those from the Consistent groups. When subjects who encoded the stories with Title Set 1 and Goal Set 2 were given Goal Set 1 as a retrieval perspective, more was recalled from Information Set 2 (34.1%) than from Information Set 1 (31.5%); this difference was not significant, $F(1,6)$ 1, MSe .0033. Combining these data with the data from the first recall trial does not lead to a significant interaction between

Information Set and Recall Trial ($F(1,6) = 3.07$, $MSe = .0009$), as is illustrated in Figure 4.

Similarly, Delay subjects who encoded the stories with Title Set 2 and Goal Set 1 and who were given Goal Set 2 as a retrieval perspective showed no significant difference in recall of the two Information Sets, $F(1,6) = 2.6$, $MSe = .0025$; the subjects recalled 31.1% from Information Set 1 and 34.9% from Information Set 2. When these data are combined with the data from the first recall trial, there is no significant interaction between Information Set and Recall Trial, $F(1,6)$ 1, $MSe = .0015$. Thus, from the Inconsistent groups as well as from the Consistent groups, we have no evidence for the effect of the retrieval perspective on recall at a delay of one week.

When the results from the Immediate and Delay conditions for both Consistent and Inconsistent groups are compared directly (using the goal given at retrieval as a basis for combining groups), we find a significant triple interaction between Delay, Orientation, and Information Set ($F(1,48) = 6.26$, $MSe = .0045$, $p < .05$). Inspection of Table 4 reveals that the interaction between Orientation and Information Set occurs only for the Immediate condition ($F(1,24) = 8.68$, $MSe = .0056$, $p < .01$), not for the Delay condition ($F(1,24) = .095$, $MSe = .0035$). The interaction for the Immediate groups demonstrates the effect of the retrieval perspective on recall, while the lack of interaction for the Delay groups demonstrates the lack of such an effect after a delay of one week.

Insert Table 4 About Here

Changes in Particular Clauses Recalled

The analyses described so far have dealt with the total percentages of clauses recalled, without examining which particular clauses were recalled or whether the items recalled changed from the first recall trial to the second. It is possible that the encoding and/or retrieval cues would affect the degree to which new clauses, not previously recalled, would be included in the second recall trial, and the degree to which clauses recalled on the first recall trial would be omitted during the second trial, without changing the overall pattern of recall enough to lead to significant effects in the total amount recalled. Therefore, two additional measures were taken: the proportion of "New" clauses that were recalled on the second trial but not on the first, and the proportion of "Omitted" clauses that were recalled on the first trial but not on the second.

To perform the analyses of the New and Omitted measures, the Consistent and Inconsistent groups were combined, using the goal given at retrieval as the basis for combining groups. For clarity of exposition of this analysis, the groups will be referred to by Goal Set only, i.e., Goal Set 1 and Goal Set 2. Analyses of New and Omitted clauses were also performed on the

Control group's data to serve as a baseline for change in recall when the goal is not changed. The data for all of the groups are presented in Table 5.

Insert Table 5 About Here

New Clauses

Of major interest here are the questions of whether the retrieval perspective affects the proportion of clauses that were recalled on the second recall trial but not on the first, and whether this effect, if any, occurs at both the Immediate and Delay conditions. Relevant to these questions are the significant interactions found between Goal Set and Information Set ($F(1,48) = 8.01$, $MSe = .0009$, $p < .01$), and between Delay, Goal Set, and Information Set ($F(1, 48) = 5.65$, $MSe = .0009$, $p < .01$). In the Immediate condition, subjects with Goal Set 1 at retrieval recalled 6.46% New clauses from Information Set 1, but only 4.13% from Information Set 2, while Goal Set 2 subjects recalled 6.7% from Information Set 2 but only 3.85% from Information set 1, leading to a significant interaction ($F(1,24) = 10.4$, $MSe = .0012$, $p < .01$). However, in the Delay condition, there was not interaction between Goal Set and Information Set ($F(1,24)$ 1, $MSe = .0006$. Subjects with Goal Set 1 at retrieval recalled 4.21% New items from Information Set 1 and 3.04% from Information Set 2, and subjects with Goal Set 2 gave the same pattern or results, recalling 3.72% New items from Information Set 1 and only 3.0% from Information Set 2. Subjects in the Immediate condition recalled more New clauses (5.29%) than did subjects in the Delay condition (3.49%), $F(1,48) = 8.1$, $MSe = .0014$, $p < .01$. Thus, it appears that the presentation of a new retrieval perspective or cue can influence the retrieval of new information from memory after a short delay, but not after a long one.

Analysis of the data from the Control group showed that the proportions of New clauses recalled from the two Information Sets were not significantly different, $F(1,12) = 1.33$, $MSe = .0003$. Subjects recalled 4.69% New clauses from Information Set 1, and 4.0% from Information Set 2.

Omitted Clauses

As in the case of the New clause analysis, an interaction involving Goal Set and Information Set would be relevant to the question of the effect of the retrieval perspective on recall. For the analysis of Omitted clauses, there was a significant interaction between Goal Set and Information Set, $F(1,48) = 27.86$, $MSe = .0033$, $p < .001$. Subjects with Goal Set 1 at retrieval omitted 8.08% from Information Set 1 and 12.78% from Information Set 2, while subjects with Goal Set 2 omitted 12.84% from Information Set 1 but only 7.38% from Information Set 2. These data are presented in Table 5. Even after a week, the retrieval perspective affects what is omitted from the recall protocol, since there was no significant interaction between Goal

Set, Information Set, and Delay. If a set of information is not cued by the retrieval perspective, it is more likely to be omitted than a set of information that is related to the retrieval perspective, even after a delay of one week.

Subjects in the Control group omitted equal amounts from Information Set 1 (6.71%) and Information Set 2 (5.63%), $F(1,12) = 1.10$, $MSe = .001$.

Debriefing

The debriefing questionnaire was intended to provide some information about the subjects' perceptions of their own performance. While we are reluctant to accept self-reports with the same assurance as other, more objective data (see Nesbitt and Wilson, 1977), it will be interesting to see whether the subjects' perceptions are consistent with the evidence from the recall data.

The subjects were asked whether they had kept the titles and goals in mind while reading and recalling the stories. Since this question is informative primarily about the experimental groups, the data below refer to the 72 subjects in the experimental groups only. Although they were presented with both a title and a goal before reading and recalling the stories, they seemed to be better able to keep the goals in mind. While reading the stories, 67 (93%) kept the goal in mind, compared to 32 (45%) for the title. During the first recall trial, 46 (66%) kept the goal in mind, compared to 26 (36%) for the title (the title and goal were the same as those presented during encoding). During the second recall trial, 56 (78%) kept the goal in mind, compared to 27 (39%) for the title. Therefore, the goal seems to have been a more powerful determinant of the perspective taken by these subjects than the title.

Important to the question of whether the retrieval cue determines what is recalled is whether the subjects actually wrote down everything they could remember. Again, the data below refer only to the experimental groups. For the first recall trial, 22 (61%) of the 36 subjects in the Immediate condition said they did, and 25 (69%) of the 36 subjects in the Delay condition said they did. On the second recall trial, 17 (47%) of the subjects in the Immediate condition said they did, while 25 (69%) of the subjects in the Delay condition said they did. A lower score for the subjects in the Immediate condition would be consistent with the notion that part of the effect of the perspective shift found in the Immediate condition is due to response suppression.

Those 43 subjects who felt that they did not write down everything they could remember in one or both of the recall trials were asked to specify why they had not done so, and what they had omitted. Their responses fall into four categories. First, 14, (33%) of the subjects responding to this question reported that they forgot or couldn't remember some things and so didn't write them down. Second, 12 (28%) responded that they omitted parts because they were too tired, bored, or it would take too long. Third, 13 (30%) stated that they had omitted details or things that seemed unimportant. The remaining 4 (9%) in the fourth group cited miscellaneous reasons such as "I had other things on my mind," or "I got the houses confused," which did not seem to indicate deliberate omissions of selected material.

If the second and third groups described above are taken together as representing subjects who knowingly edited their responses, they total 58% of the subjects who felt that they had not written down everything in memory. Because response editing is a possible explanation for the effect of the perspective shift, an important question is whether subjects in the Immediate and Delay groups adopted different editing strategies. Of the 25 subjects in the second and third groups described above, 13 are in the Immediate condition, with the remaining 12 in the Delay condition, suggesting that the subjects in the two groups did not adopt different editing strategies. However, without knowing specifically what the subjects in the two conditions omitted, it is impossible to determine more exactly the extent of response editing as a factor in the effect of the perspective shift. We can only state that there was possibly some editing by subjects in both conditions.

Discussion

Of major importance is the question of whether information is selectively encoded, or whether much or most information is encoded and recall depends on selective retrieval. The results of this study, which demonstrate a perspective shift for those subjects in the Immediate condition but not for those in the delay condition, do not support an hypothesis which postulates only selective retrieval. Instead, they suggest that the encoding perspective has a strong influence on what information is stored in the long-term store, and that any effect of a retrieval cue at a short delay is due to a combination of retrieval from a short-term store and an omission of information unrelated to the retrieval cue.

When we compared the recall of Information Sets 1 and 2 for the first and second recall trials for the subjects in each Title Set - Goal Set combination, we found evidence for an effect of the perspective shift for subjects in the Immediate condition but not for subjects in the Delay condition. The information recalled was indeed a function of the retrieval cue for subjects in the Immediate group. However, the major effect of the retrieval perspective was to lesson the <u>decrease</u> of recall of perspective-related information over time, compared to unrelated information; not to increase recall of perspective-related information. The hypothesis that all information is stored and needs only the appropriate cue to be accessed would predict an increase in the recall of perspective-related information. Instead, we found that the effect of the perspective shift after a short delay is primarily due to an omission of previously recalled information that is unrelated to the new retrieval perspective.

Furthermore, a retrieval hypothesis such such as that of Anderson & Pichert (1978) would predict an effect of the perspective shift after a long delay as well as a short one. We found no evidence of a perspective shift after a delay of one week, providing further strength to the argument of selective encoding.

While the perspective shift in the Immediate condition seems to stem primarily from the ommission of information, there is also evidence that a small amount of previously unrecalled information is actually accessed by the new retrieval cue. The analysis of the particular clauses recalled demonstrates that in the Immediate condition, the new retrieval cue provided before the second recall trial led to the recall of a greater number of New clauses related to that cue than New clauses not related to that cue, and to the omission of fewer clauses related to the cue than of those not related. But in the Delay condition, the presentation of a new retrieval cue did not lead to the recall of more New clauses related to that cue. It did, however, reduce the percent of relevant clauses that were omitted, compared to irrelevant clauses, just as we found in the Immediate condition. Therefore, it seems that while presentation of the cue can reduce the amount of omissions at either delay interval, it will only increase the recall of not-previously-recalled items if the delay is short, and then only by a small amount.

Taken together, the evidence suggests that even with long, complex materials and an incidental task, recall of previously unrecallable information is possible after a fairly short delay, if an appropriate retrieval cue is present, but such recall does not occur after a longer delay. Instead, when a subject is required to think about a certain set of information in order to make a judgment during encoding, that information receives greater elaboration and, consequently, becomes more accessible to retrieval. This finding is consistent with the research of Keenan & Baillet (1980), who found that the richness of the conceptual structure into which an event is encoded affects memory for that event.

The conceptual structure referred to by Keenan & Baillet (1980) existed already in the subject's mind prior to the experiment in the form of previous knowledge about particular people, but such a structure could also be created during encoding by the elaborative processes which take place as one analyses information for its relevance to a question, using stored knowledge as well as the stimulus materials to perform the task. Keenan & Baillet propose that the elaboration of a memory trace depends on three aspects of memory operation: the encoding task, the meaningfulness of the event, and the richness of the semantic structure. Based on the results of this study, we propose that the congruency or match between the stimulus event and, the goal of the task is a fourth aspect of memory operation that affects the elaboration of the event in memory. This view is consistent with the results of Schulman (1974) and of Craik & Tulving (1975), who proposed that questions leading to congruous answers led to better memory for those answers than for incongruous answers due to the greater elaboration given to the congruous answers during processing. Similarly, Jacoby (1978) and Jacoby, Craik, & Begg (1979) found that the greater the difficulty of a semantic judgment, or the more operations required in making the judgment, the greater the memorability of the information, because of the more distinct memory trace that is established.

In the present study, the task of making a judgment was constant across subjects; the meaningfulness of the stimulus events, i.e., the target information sets, was controlled by balancing; and the richness of the cognitive structures for the target information, which was determined by the content of the stories, was the same for all subjects; and still, memory performance varied with the relationship between the information set and the goal or judgment. That information which was congruent with the goal, and therefore related to the judgment, went through greater elaboration during its evaluation in terms of the task, and consequently, the probability of encoding that information increased.

We propose that the mechanism for the difference in memory as a function of relevance to the judgment is as follows. In the process of evaluating a judgment in terms of new information, some information already in memory is referred to, and the new information that is relevant to the judgment is encoded with many links or associations to that information in memory which has been called into play. New information that is not relevant to the judgment does not establish many associations to the information in memory. New information that has many links to stored knowledge could be well recalled for two reasons. First, the probability of long-term storage is increased. And second, the probablility of being accessed by a retrieval cue is increased by the large number of links to other information in memory. In either case, recall of such information is a direct function of the operations that occurred during encoding.

Why, then, was any material that was not directly relevant to the judgment task stored in memory? This result can be explained by assuming that the task was complex enough to require that all of the information in the text be processed semantically, and that all of it be compared, at least initially, to the goal or judgment in order to compute its relevance to the judgment. The fact that incidental learning can lead to a fair amount of recall even after a week suggests that the processes required by evan a minimal comparison can lead to some elaboration in memory.

In answer to the question of whether the title or the goal determines the reader's perspective, it appears that although both contribute to the perspective, the reader's goal has the stronger effect. The data from the first and second recall trials (Tables 2 and 3) show that for the Consistent groups, recall was higher for the information set cued by the title and goal than for the uncued set on the first recall trial, and that a perspective shift occurred when the goal was changed for the second recall trial, in the Immediate condition. For the Inconsistent groups, where the title and goal are pitted against each other during encoding, we can see that both the title and the goal influence encoding, since memory for the two information sets didn't differ at the first recall trial even though the Control group's data indicate that the two sets are not equally memorable. For the Inconsistent groups, like the Consistent groups, a perspective shift follows a shift in the goal. As further support for the stronger effect of the goal, more subjects found that they could keep the goal in mind during reading and recalling then the title.

References

Anderson, R.C. & Biddle, W.B. On asking people questions about what they are reading. In G.H. Bower (Ed.), The Psychology of Learning and Motivation, Vol. 9. New York: Academic Press, 1975.

Anderson, R.C. & Pichert, J.W. Recall of previously unrecallable information following a shift in perspective. Journal of Verbal Learning and Verbal Behavior, 1978, 17, 1-12.

Bransford, J.D. & Johnson, M.K. Contextual prerequisites for understanding: some investigations of comprehension and recall. Journal of Verbal Learning and Verbal Behavior, 1972, 11, 717-726.

Britton, B.K., Meyer, B.J., Hodge, M.H., & Glynn, S.M. Effects of the organization of text on memory: Tests of retrieval and response criterion hypotheses. Journal of Experimental Psychology: Human Learning and Memory, 1980, 6, 620-629.

Craik, F.I.M. & Tulving, E. Depth of processing and the retention of words in episodic memory. Journal of Experimental Psychology: General, 1975, 104, 268-294.

Dooling, D.J. & Mullet, R.L. Locus of thematic effects in retention of prose. Journal of Experimental Psychology, 1973, 97, 404-406.

Greene, E. Memory for evidential information from text. Unpublished master's thesis, University of Colorado, 1977.

Hirst, W. The locus of constructive activity in memory for mathematical proofs. Journal of Experimental Psychology: Human Learning and Memory, 1980, 6, 119-126.

Jacoby, L.L. On interpreting the effects of repetition in solving a problem versus remembering a solution. Journal of Verbal Learning and Verbal Behavior, 1978, 17, 649-667.

Jacoby, L.L., Craik, F.I.M., & Begg, I. Effects of decision difficulty on recognition and recall. Journal of Verbal Learning and Verbal Behavior, 1979, 18, 585-600.

Joyce, J. The Boarding House. In J. Joyce, Dubliners (R. Scholes & A.W. Litz, Eds.). New York: The Viking Press, 1969. (Originally published by B.W. Huesbsch, 1916.)

Keenan, J.M. The role of episodic information in the assessment of semantic memory representations for sentences. Unpublished doctoral dissertation, University of Colorado, 1975.

Keenan, J.M. & Baillet, S.D. Memory for personally and socially significant events. In R.S. Nickerson (Ed.), Attention and Performance VIII. Hillside, N.J.: Lawrence Erlbaum Associates, 1980.

Kintsch, W. The Representation of Meaning in Memory. Hillside, N.J.: Lawrence Erlbaum Associates, 1974.

Kozminsky, E. Altering comprehension: The effects of biasing titles on text comprehension. Memory and Cognition, 1977, 5, 482-490.

Meyer, B.J.F. The Organization of Prose and its Effects on Memory. Amsterdam: North-Holland Publishing Co., 1975.

Meyer, B.J.F. & Freedle, R.O. Effects of discourse type on recall. Research Report No. 6, Department of Education, Arizona State University, 1979.

Nisbett, R.E. & Wilson, T.D. Telling more than we can know: Verbal reports on mental processes. Psychological Review, 1977, 84, 231-259.

Schallert, D.L. Improving memory for prose: The relationship between depth of processing and context. Journal of Verbal Learning and Verbal Behavior, 1976, 15, 621-632.

Schulman, A.I. Memory for words recently classified. Memory and Cognition, 1974, 2, 47-52.

Thomson, D.M. & Tulving, E. Associative encoding and retrieval: Weak and strong cues. Journal of Experimental Psychology, 1970, 86, 255-262.

Thorndyke, P.W. Cognitive structures in comprehension and memory of narrative discourse. Cognitive Psychology, 1977, 9, 77-110.

Tulving, E. & Thomson, D.M. Encoding specificity and retrieval processes in episodic memory. Psychological Review, 1973, 80, 352-373.

Table 1

Example of the Experimental Design for the Consistent Group

	Story 1		Story 2		Story 3	
	Information Set 1	Information Set 2	Information Set 1	Information Set 2	Information Set 1	Information Set 2
IMMEDIATE						
Title Set 1 - Goal Set 1						
Order 1: Subjects 1-3						
Order 2: Subjects 4-6						
Order 3: Subjects 7-9						
Title Set 2 - Goal Set 2						
Order 1: Subjects 10-12						
Order 2: Subjects 13-15						
Order 3: Subjects 16-18						
DELAY						
Title Set 1 - Goal Set 1						
Order 1: Subjects 19-21						
Order 2: Subjects 22-24						
Order 3: Subjects 25-27						
Title Set 2 - Goal Set 2						
Order 1: Subjects 28-30						
Order 2: Subjects 31-33						
Order 3: Subjects 34-36						

Table 2

Percent of Target Information Recalled on the First Recall Trial

Information	Information Set 1	Set 2
Control	34.8	30.2
Consistent		
Title Set 1 - Goal Set 1	40.2	36.6
Title Set 2 - Goal Set 2	33.1	41.6
Inconsistent		
Title Set 1- Goal Set 2	34.9	39.0
Title Set 2 - Goal Set 1	36.6	36.4

Table 3

Percent of Target Information Recalled on the Second Recall Trial

	Information Set 1	Information Set 2
IMMEDIATE CONDITION		
Control	33.0	30.4
Consistent		
Title Set 1 - Goal Set 2	31.5	36.9
Title Set 2 - Goal Set 1	35.1	34.7
Inconsistent		
Title Set 1 - Goal Set 1	31.1	25.9
Title Set 2 - Goal Set 2	20.7	30.5
DELAY CONDITION		
Control	32.6	26.8
Consistent		
Title Set 1 - Goal Set 2	33.6	32.8
Title Set 2 - Goal Set 1	27.0	29.1
Inconsistent		
Title Set 1 - Goal Set 1	31.5	34.1
Title Set 2 - Goal Set 2	31.1	34.9

Table 4

Percent of Target Information Recalled on the Second Recall Trial

Data for the Consistent and Inconsistent Groups Combined

	Information Set 1	Information Set 2
IMMEDIATE CONDITION		
Retrieval Goal Set 1	33.1	30.3
Retrieval Goal Set 2	26.1	33.7
DELAY CONDITION		
Retrieval Goal Set 1	29.2	31.6
Retrieval Goal Set 2	32.4	33.9

Table 5

Percent Change in Particular Clauses Recalled from First to Second Recall Trials

	Information Set 1	Information Set 2
NEW CLAUSES		
IMMEDIATE CONDITION		
Control	5.2	3.9
Consistent and Inconsistent Groups		
Retrieval Goal 1	6.5	4.1
Retrieval Goal 2	3.9	6.7
DELAY CONDITION		
Control	4.2	4.1
Consistent and Inconsistent Groups		
Retrieval Goal 1	4.2	3.0
Retrieval Goal 2	3.7	3.0
OMITTED CLAUSES		
IMMEDIATE CONDITION		
Control	7.2	6.4
Consistent and Inconsistent Groups		
Retrieval Goal 1	7.2	12.7
Retreival Goal 2	15.1	7.2
DELAY CONDITION		
Control	6.2	4.9
Consistent and Inconsistent Groups		
Retrieval Goal 1	8.9	12.8
Retrieval Goal 2	10.5	7.6

Publications

In addition to the papers presented in this report, the following publications have been supported in part by this grant.

Keenan, J. M. Beyond Bartlett: Issues in the study of comprehension. In H. W. Dechert & M. Raupach (Eds.), Psycholinguistic models of production. Hillsdale, N. J.: Lawrence Erlbaum Associates, in press.

Keenan, J. M., & MacWhinney, B. Understanding the relationship between comprehension and production. In H. W. Dechert & M. Raupach (Eds.), Psycholinguistic models of production. Hillsdale, N. J.: Lawrence Erlbaum Associates, in press.

MacWhinney, B., Keenan, J. M., & Reinke, P. The role of arousal in memory for conversation. Memory & Cognition, in press.

Zimler, J., & Keenan, J. M. Imagery in the congenitally blind: How visual are visual images? Journal of Experimental Psychology: Learning, Memory, & Cognition, in press.