Technical Report No. 530

ELABORATIVE INFERENCES ON AN EXPOSITORY TEXT

Aydin Y. Durgunoğlu
University of Illinois at Urbana-Champaign

Jihn-Chang J. Jehng
Southwest Research Institute

April 1991

Center for the Study of Reading

TECHNICAL REPORTS

College of Education
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN
174 Children's Research Center
51 Gerty Drive
Champaign, Illinois 61820

CENTER FOR THE STUDY OF READING

Technical Report No. 530

ELABORATIVE INFERENCES ON AN EXPOSITORY TEXT

Aydin Y. Durgunoğlu
University of Illinois at Urbana-Champaign

Jihn-Chang J. Jehng
Southwest Research Institute

April 1991

University of Illinois at Urbana-Champaign
51 Gerty Drive
Champaign, Illinois 61820

The work upon which this publication was based was supported in part by the Office of Educational Research and Improvement under Cooperative Agreement No. G0087-C1001-90 with the Reading Research and Education Center, in part by U.S. Army Research Institute Office of Basic Research Grant No. MDA 903-86-K-0443 to Rand J. Spiro, and in part by a grant from the Mellon Foundation. The publication does not necessarily reflect the views of the agencies supporting the research.

Abstract

Remembering the information in a text is different from learning from a text and applying the acquired knowledge (e.g., by making inferences). This distinction was investigated with a dissociation paradigm. After reading an expository text, subjects performed either a memory (recognition) or an inferencing (verification) test. The effects of the same variables on the performance on the two tasks were compared. Text organization tended to affect recognition but not verification test performance. When verifying nonstudied items by inferencing, the richness of the available text information and the type of processing required to make the inference were important. The educational implications of this dissociation between memory and inference tests are discussed.

ELABORATIVE INFERENCES ON AN EXPOSITORY TEXT

One of the purposes of reading expository texts is to learn from them. Learning from a text does not necessarily mean comprehending and even memorizing the text, but rather, being able to use and apply the information presented in it. Our goal in this report is to demonstrate that comprehending and memorizing a text do not necessarily imply understanding it well enough to use and manipulate the information in it. One indication of how well readers understand the information in a text is to determine if they apply the studied information to make elaborative inferences.

When reading a text, a reader is never given all the information that is necessary to form a coherent representation of that text and all of its possible implications and elaborations. Instead, the reader needs to rely on his or her background knowledge and on the clues from the text to draw inferences and to close the gaps in the presented information as well as to go beyond the information actually presented. Therefore, from an educator's point of view, the ability of a reader to determine the missing information and to elaborate beyond the available information by making inferences is a prerequisite of reading comprehension (Anderson & Pearson, 1984; Bransford, 1979; Pearson, Hansen, & Gordon, 1979).

Recently, investigators have begun to distinguish between several different kinds of inferences a reader can generate (e.g., Graesser, Haberlandt, & Koizumi, 1987; Long, Golding, Graesser, & Clark, 1990; Potts, Keenan, & Golding, 1988). According to these researchers, *bridging inferences* are necessary to form a coherent representation of the text, and they are essential for comprehension. It is likely that such inferences are generated online, that is, while a reader is reading the text. In contrast, *elaborative inferences* are not essential for a coherent integrated representation of the text, but rather they "refine or embellish the text representation" (Long et al., 1990). Such inferences are more likely to be generated at retrieval, for example, when the inference is necessary for a reader to answer a question about the information in the text (McKoon & Ratcliff, 1990; Whitney, 1987). That is, elaborative inferences are not readily available in memory for direct retrieval but need to be computed given the demands of the (testing) situation. This does not mean, however, that elaborative inferences are trivial or unimportant, because these inferences involve going beyond the text and deriving unspecified text aspects such as consequences of events (McKoon & Ratcliff, 1986; Potts et al., 1988) and specification of ongoing states (Seifert, Robertson, & Black, 1985). Of course, as research on transfer of knowledge indicates, a crucial aspect of knowledge transfer is being able to *use* the presented information and understand its implications rather than merely remembering the text (e.g., Kintsch, 1986; Spiro, Vispoel, Samarapungavan, Schmitz, & Boerger, 1987). Hence, elaborative inferences are important for a deeper understanding of the presented material.

The research reported here had two goals: To demonstrate that elaborative inferencing involves processes that are different from those involved in direct retrieval of propositions from memory, and to examine the variables affecting elaborative inferences and discuss their educational implications. The first goal has been intuitively obvious for many years to educators who emphasized that learning from text cannot be equated with remembering the text. A student may be able to remember all the information in a passage and yet be unable to transfer this knowledge and apply it to new situations (Spiro et al., 1987), as indicated by problems in embellishing, extending, and applying information in the text, that is, in drawing elaborative inferences. Therefore, it is important to investigate (a) the kinds of variables that affect the elaborative inferencing process and (b) how these variables affect inferencing tests differently as compared to memory tests.

Reder and Kintsch and their respective colleagues (Kintsch, 1986; 1988; Perrig & Kintsch, 1985; Reder, 1982; 1987; Reder, Wible, & Martin, 1986) have begun to investigate this latter issue. Because the present research design is based on the findings of these researchers, we will first briefly summarize the findings of their studies.

In two experiments by Reder (1982), subjects first read a story followed by the presentation of the test probes (test probes were single sentences). One group of subjects decided whether the test probe had been presented in the story, that is, they made a recognition judgment, whereas another group of subjects decided whether a test probe was plausible given the story, that is, they made a plausibility judgment. In other words, the first group had a memory task and the second group had a task that required some inferencing.

Reder manipulated the delay between the presentation of the stories and the test probes. The recognition judgments were very much affected by the delay between the initial reading and the test. As the memory became poorer with delay, both the accuracy and the speed of the recognition judgments dropped considerably. In contrast, the plausibility judgments were not affected as much by delay. More important, except in the no-delay condition, plausibility judgments were faster and more accurate than were the recognition judgments. Consequently, Reder (1982, 1987; Reder et al., 1986) suggested that searching memory for a proposition is a different process than computing an inference. She also convincingly argued that when verifying the truth of statements, judging plausibility by making inferences is a more efficient strategy than direct retrieval of facts stored in memory.

Even though their theoretical orientation is somewhat different, Perrig and Kintsch (1985; Kintsch, 1986) also suggested that memory retrieval and inferencing tests are not alike. In their experiments, they have presented the description of a town either in geographical terms in what they call the "survey text" (e.g., the church is north of the inn) or in driving instructions format in what they call the "route text" (e.g., coming from the church you make a left turn to reach the inn). Because the overall coherence of the route text was higher, the subjects in that group recalled more information. However, when they were asked to make inferences to verify a spatial orientation that was not explicitly mentioned in the text (e.g., the highway is south of the church), both groups had similar levels of performance. The critical variable affecting the inference task was the congruency between the type of text read and the type of inference item presented for verification. The subjects who had the geographical description were better on inference statements given in geographical terms, whereas the subjects who had the route text performed better on items written in the route description format. (Although the gender of the subjects interacted with this finding, it is not important for our purposes). In short, performance on direct memory and on inference tasks was affected by different variables. The coherence of the text affected performance on the memory test, whereas the congruency between the type of text and the type of test questions affected inferencing performance. However, instead of proposing different processes involved in memory retrieval and inference tasks (as Reder had suggested), Perrig and Kintsch postulated that each type of task relies on a different type of representation. They suggested that a propositional representation of a text is sufficient for free recall or recognition tests. However, to make inferences, a more global situation model (van Dijk & Kintsch, 1983) or "mental model" (Johnson-Laird, 1983) is required.

Our goal is not to compare Reder's process and Kintsch's representational models, because as Anderson (1978) has argued, it is quite difficult to separate the question of representation from the question of processing. It is possible, for example, that the kind of information in the situation model proposed by Kintsch and his colleagues invites more plausibility computations, whereas the propositional representation of the text invites more direct retrieval. Our goal, however, is to explore further the difference between memory and inference tasks using a dissociation paradigm under well-controlled conditions. The dissociation paradigm involves comparing the effects of the same variables across the tasks of interest. To have unambiguous results with such a paradigm, the conditions during study and test phases need to be as equivalent as possible across the two tasks (Neely, 1989).

Reder investigated the effects of identical variables on two different tasks, but only used short stories. Perrig and Kintsch used an expository text, but they did not systematically observe the effects of identical variables across the memory and inference tasks (because it wasn't their main focus). In our study, we have attempted to fill the gaps in the currently available data. That is, we have used a relatively long expository text and manipulated the same variables across both a memory and an inference task. We have kept the study conditions identical across the two tasks and compared the performance on the two tasks on identical materials.

We used a recognition test as the memory test and a sentence verification test as the inference test. The recognition test did not require generating inferences, because subjects were asked to judge whether a given statement was explicitly studied before. To use Reder's model, direct retrieval was encouraged. Also, because the delay between the acquisition and the retrieval phases of the experiment was not very long, it was likely that in this task the subjects would not try to generate inferences. In contrast, the verification test required judging whether a statement was true or false. Instructions for the verification test emphasized that a statement could be true or false irrespective of whether it has been studied before. In sum, the only difference between the two tasks was in the instructions given *after* the subjects read the text: The memory group was asked to judge whether the given test statement was studied, whereas the inference group was asked to judge whether the same statement was true.

We compared the effects of the same two variables described below across these two tasks. If a variable has similar effects on performance in recognition and verification tests, then there will not be evidence for a dissociation between the two tasks. Such a result would indicate that the elaborative inferencing process involved in verifying a statement is similar in nature to processes involved in direct retrieval from memory. In contrast, differing effects of a variable on the two tasks would indicate that the two tasks differ. Finding such a dissociation between the two tasks has educational implications. If memory and inference tasks are affected differently by several variables, then it would be a sound educational strategy to emphasize variables that affect inferencing just as much as those that affect memory.

The text used in our experiment was an expository passage about an ancient civilization (Phoenicians) that was quite unfamiliar to the undergraduate students in our subject pool. In a pilot experiment with different subjects from the same pool, we had asked five open-ended questions, worth 2 points each, about the Phoenicians (e.g., when and where they lived, their economy, their contribution to modern civilization), as well as direct questions about whether the students had ever studied about the Phoenicians. The 32 subjects in the pilot group had an average score of 1.12 out of 10 on the five questions. Only 12 subjects reported ever studying about Phoenicians, but 7 of those individuals qualified their answer by saying that they had studied about Phoenicians very briefly or that Phoenicians were only mentioned in their history class. In short, our pilot data indicated that the topic of our texts was quite unfamiliar to the subjects in our pool. The reason for selecting an unfamiliar topic was to discourage the use of background information and to encourage subjects to rely on the presented information to draw their inferences. As Ceci and McNellis (1987) have discussed, in some experiments, the differences observed in processing may actually be caused by the differences in the knowledge base. Hence, choosing an unfamiliar text reduces (but does not completely eliminate) the problem of differing knowledge bases accounting for differences observed in the inferencing process.

The first variable manipulated across the recognition and verification tasks was the style of presentation of the text. Subjects read the text in one of the three formats: For the "structured" group, the text was hierarchically organized under headings such as "History," "Culture and civilization" and so on. This group read the same structured text twice. For the "unstructured" group, the text was coherent, but it was not organized hierarchically. To do this, the *paragraphs* in the hierarchically organized (structured) text were reordered and the headings were removed. This group read the same unstructured text during the two presentations. Finally, for the "mixed" group, the passage was structured during the first presentation, but it was unstructured during the second presentation.

The presentation style was predicted to affect the memory for the text. Research on text structure has indicated that subjects can recall more information--especially main ideas--from well-organized expository texts as compared to scrambled texts (Taylor & Samuels, 1983; Richgels, McGee, Lomax, & Sheard, 1987). In addition, well-organized texts are more helpful when the topic is unfamiliar (for a review see Roller, 1990). However, when the same scrambled passage is given twice, some improvement in memory occurs (Danner, 1976). In our study, reading twice the text organized the same way (structured *or* unstructured during both presentations) was predicted to lead to better memory for the material as compared to a mixed presentation, which in turn, was predicted to help in the recognition test. However, in the verification test such direct retrieval of information is not required and hence the memory manipulation should have little effect in this test (see also Perrig & Kintsch, 1985).

The second variable compared across the two tasks was the nature of the inference items. As discussed before, readers do not seem to generate a lot of elaborative inferences while reading a text, but do so during retrieval, for example in response to a test question (for a review see Weaver & Kintsch, 1990). This implies that both the richness of the text information available for possible use in inferencing as well as the nature of the test questions that necessitated inferencing are important. Hence, in our study, following the study of texts, we gave our subjects test statements and asked for recognition or verification judgments. The nonstudied items that necessitated inferencing were classified on two dimensions labeled as richness and distance. The "richness" dimension reflected the richness of the text information available to generate inferences. An inference could be drawn based on a limited set of information in the text (i.e., Restricted item), or based on a richer data base from the text (i.e., Broad item). The richness of the sources available to make an inference should affect the verification task. In fact, Anderson and Reder (1987) demonstrated that subjects were faster in making sensibility judgments if they knew many facts about a concept. However, the richness dimension is not predicted to affect the recognition task because this task does not require manipulation of the information in the text to draw inferences.

The "distance" dimension reflected the type of inferencing that was required by the test question. Following the framework of Armbruster and Ostertag (1989), we classified each inference item as Near or Far in terms of the type of inference necessary to verify the question. Armbruster and Ostertag were dissatisfied by classification schemes that identify a question only as textually explicit or implicit because the textually implicit category embodied many different kinds of inferences. Instead, Armbruster and Ostertag used a modification of the taxonomies of Bloom and his colleagues (Bloom, Engelhart, Furst, Hill, & Krathwohl, 1976) and of Barrett (1976) to reflect the differential cognitive processing required of different types of inferences. They distinguished between inferences that require combining information across a text and inferences that require applying information to a novel situation or inferences that involve predictions, hypothesis generation and analogies. We classified the first type of inference as Near and the latter two as Far inferences for our purposes. That is, if the inference only involved putting together the explicit information across the text to verify a statement, then it was classified as a Near item. If an item required applying the studied information to a novel situation or to make predictions about a situation, then it was classified as a Far item. It was predicted that Far items would lead to better recognition performance than Near items because the inferred information could be easily distinguished from actually studied information, whereas the opposite pattern was predicted for the verification test.

To summarize, we predicted that the nature of an inference item--both the "richness" and "distance" dimensions--as well as the style of text presentation would affect recognition and verification tasks differently, thus producing a dissociation.

Method

Subjects

Subjects were 110 undergraduates participating to fulfill a course requirement. All were native English speakers. There were 55 students in each of the two tasks, verification and recognition.

Design and Materials

In the experiment, test format (recognition vs. verification), and the presentation format (structured, unstructured, or mixed) were between-subjects variables and the type of inference item was a within-subjects variable.

The basic text was a passage about the Phoenicians compiled by the experimenters from several ancient history texts. It was approximately 1,300 words long. In the structured text, there were five sections under the following headings: I - Introduction II - People and region III - History IV - Culture and civilization V - Industry, commerce, and exploration. In the unstructured text, identical material was included, but the headings were eliminated. Also, in the unstructured text, the *paragraphs* were re-ordered, but without compromising the coherence of the text. (Copies of the texts are available from the first author.)

The tests used at the end of the experiment were created by the experimenters. There were 44 statements on the verification test. Subjects decided whether a statement was true or false based on the text they had read. Out of these 44 items, 22 were inference (i.e., nonstudied) items because the answers to these questions were not explicitly stated in the text; instead, the answer needed to be determined using different sources of information from the text. For example, the item "Phoenicians were expert shipbuilders" was considered an inference question because the text did not include this information. This statement is true because several sources of information in the text support it. The text contains information that Phoenicians lived by the sea, did extensive sea travel as traders, had many different kinds of ships, had access to forests, and were good woodworkers. Combining these various pieces of information, this inference question can be answered as "true." The answers to the inference questions can be found in the history books. However they were not included in our text, but rather left to our subjects to infer. In short, the "correctness" of a response to an inference question could be judged by comparing a subject's response to the answer found in the books.

The remaining 22 statements on the verification test were called studied (i.e., old) items, because their answers were explicitly available in the text. The subjects did not need to combine various pieces of information across the text to verify these statements. For example, the statement "In the Phoenician alphabet, pictures were used to depict words similar to the Egyptian hieroglyphs" is false because the text explicitly mentions that the Phoenician alphabet consisted of 22 consonants that reflected the sounds of the language. However, the studied items were not verbatim copies of the material in the texts. In short, the studied items were conceptually identical but paraphrased versions of the information in the text. Out of the 22 inference items, 13 were true and 9 were false. Out of the 22 old (studied) items, 14 were true and 8 were false.

The same items were included in the recognition test except following Reder (1982), the false items were excluded. This was done to discourage subjects from first making plausibility judgments in the recognition test and thus using the decision strategy that an item must be nonstudied if it is false. There were 27 true items (14 studied and 13 inference) in the recognition test[1].

All inference (i.e., nonstudied) items were classified a priori on the two dimensions of distance and richness. The distance dimension reflected whether the required inference could be generated by combining information across the text (i.e., Near) or by making predictions and/or by applying the studied information to novel situations (i.e., Far). For example, the statement "Phoenicians were highly influenced by the Egyptian culture" was classified as a Near item, because the reasoning chain only involves combining several pieces of information in the text. The text mentions that the Phoenicians lived under the Egyptian rule at one time, and that they improved the glass-making techniques they learned from the Egyptians. Consequently, the inference reasoning chain can go as follows: Phoenicians lived under the Egyptian rule, Phoenicians learned how to make glass from the Egyptians, therefore they must have been influenced by the Egyptians and this statement must be true.

An example of a Far item is the statement "Finding Phoenician products such as jewelry, trinkets, and silver bowls in an area is good archeological evidence for a Phoenician settlement or city in that area." To verify this statement, the inference chain can go as follows: The text mentioned that Phoenicians were craftsmen and traders. The text also mentioned that they handcrafted many products such as jewelry, trinkets, glass, etc., and that they traded these products all around the Mediterranean. Another civilization could have bought and used the Phoenician products. Therefore, finding such products is not good evidence for a city or a settlement because the products may have been traded goods. One needs better evidence to decide that an ancient Phoenician city existed in an area, and hence this statement is false. In other words, the information about Phoenicians is used to generate hypotheses and/or to predict the kind of archeological evidence that indicates a Phoenician settlement.

The second dimension, richness, reflected the number of sources of information that could be used to answer the inference question. If a test item had many sources of information in the text that could converge on the inference, then we called it a Broad item. In contrast, if the sources of information in the text that were useful for making the inference were very few, then we called it a Restricted item. For example, the statement about Phoenicians being expert shipbuilders was classified as a Broad item because there were many sources of information about the forests, woodworking, trading, living by the sea, etc., that permitted an inference. In contrast, the item "The traditional Phoenician garment was a long, white starched dress" was classified as a Restricted item. The only information in the text that allows subjects to correctly classify this statement as false is that Phoenicians had a very well-developed textile-dyeing industry.

To classify the items a priori on the richness dimension we presented the text and the verification test to our pilot subjects ($n=32$). All subjects decided whether a statement was true or false. In addition, half explained their reasons, using an If . . . Then . . . type of construct. The remaining subjects listed the pieces of information in the text that could be used to answer each question. Both groups of subjects could always look back to the text as they were completing the verification test. We identified the number of propositions from the text the subjects mentioned for each of the inference items. Across all 32 subjects, if an item had five or more distinct propositions used to determine the answer, then it was classified as a Broad item, otherwise it was classified as a Restricted item.

The background knowledge of the subjects was assessed by a short questionnaire. On this questionnaire there were two questions about the subjects' history education and four very general questions about ancient civilizations (e.g., Where did Alexander the Great live? What is the greatest contribution of King Hammurabi of Babylon? What happened to the Roman Empire in 476 A.D.?) We did not include any specific questions about the Phoenicians, because we did not want to affect the initial reading strategies of our subjects. In terms of testing logistics, this also gave us the advantage of being able to give this background test even after our subjects had read the text about Phoenicians. We believe that individuals who cannot answer simple, well-known questions about Greeks and Romans are less likely to know about the Phoenicians. An examination of several history books suggests that compared to Greeks and Romans, the space allotted to the discussion of Phoenicians is very limited. As an example, in the 750-page volume *History of the Ancient World* (Starr, 1965), the discussion of Phoenicians is only seven pages long. For our pilot subjects who took both this general background test and answered specific questions about the Phoenicians, the correlation between the performance on the two tests was significant, $r = 0.56$ ($n=32$, $p < .0001$). Hence, we believe that the general background test is a good predictor of whether our subjects are familiar with ancient civilizations, especially the Phoenicians. The score on this test was used as a covariate in all the analyses reported in the Results section.

Procedure

Subjects first completed a short test asking them to match names of places and civilizations around the Mediterranean to acclimate them to the topic. They then studied the text about Phoenicians for 10 minutes. For the structured and mixed groups, this text was structured, but for the unstructured group, it was not. Then the subjects answered the questions on the background test assessing their knowledge about ancient history. Following the background test, they read the text for the second time for 5 minutes. During this presentation, the structured and unstructured groups read the texts they had studied before, whereas the mixed group read an unstructured text. Following the study of texts, subjects had a filler task of completing a learning style and attitude questionnaire for 5 minutes. Finally, they either took the verification or the recognition test. For the group performing the recognition test, it was emphasized that the statements would not be identical to the material in the text (see note 1), but they were to mark an item as "studied" if that information was explicitly described in the text. In addition, subjects were told that all of the items on the recognition test were true. Therefore they did not need to use the strategy that if an item is false, it must be nonstudied (Reder, 1982). In short, subjects needed only to distinguish between studied and nonstudied information. After making a decision about the study status of an item, the subjects were asked to rate how confident they were of their decision (1=very uncertain, 5=very certain), although the confidence ratings were not analyzed.

For the subjects performing the verification test, it was stressed that an item could be true or false independent of whether it was studied before. The subjects in this group wrote a short statement explaining the rationale for their true/false decision.

For the recognition group, a score of 1 indicated a hit or a correct rejection and a score of 0 indicated a false alarm or a miss. For the verification group, a score of 1 indicated a correct response and a score of 0 indicated an incorrect response[2]. The authors scored the sheets independently (interrater reliability = .98) and any conflicts were resolved by discussion.

Results and Discussion

Separate statistical analyses were performed on the data from each task. All results are reported at the .05 level of significance unless otherwise indicated. The results are organized around the two variables of interest: presentation type and item effects.

Effects of the Presentation Type

The effects of the type of text during study, structured-structured (SS) for the structured group, structured-unstructured (SU) for the mixed group, unstructured-unstructured (UU) for the unstructured group, were examined for both tasks using the score on the background knowledge questionnaire as a covariate.

Verification task. The verification scores were analyzed in an analysis of covariance (ANCOVA), with response type (true vs. false) and study status (studied vs. inferred) as within-subjects variables and presentation type (SS, SU, UU) as the between-subjects variable. The results, as well as the number of subjects in each condition, are presented in Table 1. The effect of study status was significant, $F(1,52) = 182.56$, $MS_e = .02$; indicating that the studied items were verified more accurately than inferred items, .848 and .622, respectively. Neither the main effect of the presentation type nor its interactions with response type and study status were significant, all Fs < 1.

To summarize, the presentation type which can presumably influence the memory for the text did not affect performance on the verification task. Hence, the goodness of the memory for the text was not a significant variable in the verification task.

[Insert Table 1 about here.]

Recognition task. Because only true items were included in the recognition test, response type was not a variable any longer, thus only the study status and the presentation type were included in the analyses.

The results are presented in Table 2. Hit and false alarm rates, indicating correct recognition of studied items and false recognition of inferred items, respectively, were analyzed separately with presentation style (SS, SU and UU) as the between-subjects variable. When the hit rates were considered, the effect of the presentation type was not significant, $F < 1$, indicating that all subjects were equally adept in recognizing studied items.

[Insert Table 2 about here]

However, the false alarm rates showed a different pattern. Here the three groups differed, $F(2,51) = 3.16$, $MS_e = .027$. The groups that studied the same text twice (SS and UU) had low false alarm rates, 0.266 and 0.251, respectively. The group that studied two versions of the text (SU) had a significantly higher false alarm rate (0.367), indicating a poorer memory and more difficulty in rejecting nonstudied items.

Across the recognition and verification tasks, text presentation did not produce very strong effects. However, any effects of text presentation were only observed in the recognition but not in the verification test performance.

Item Effects

The test items that required the generation of an inference rather than direct retrieval of intact information from memory were analyzed further using the dimensions of richness and distance. Because there were only true items in the recognition task, the false items were excluded from the verification task to compare the same inference (nonstudied) items across the two tasks. However, in the last section, the distance and richness effects for both true and false items in the verification task are also discussed.

Recognition task. Separate analyses were performed on the correct rejection rates of inferred (nonstudied) items with presentation type as the between-subjects variable and distance as the within-subjects variable in the first analysis and richness as the within-subjects variable in the second one. The results are presented in Table 3. In the first analysis the presentation type yielded no main effects, $F(2,51) = 2.47$, $MS_e = .04$, and did not interact with the distance variable, $F(2,52) = 1.52$, $MS_e = .03$. However, there was a significant main effect of distance, $F(1,52) = 70.76$, $MS_e = .03$, indicating that performance on the Far items was much better than the performance on the Near items (.903 versus .645, respectively). As expected, it was harder to reject inferred items as nonstudied when they were conceptually closer to the actually studied material.

[Insert Table 3 about here]

The analysis with the richness variable yielded no significant effects of presentation type, $F(2,51) = 2.45$, $MS_e = .07$, or a Presentation Type x Richness interaction, $F(2,52) = 1.30$, $MS_e = .03$. There was a marginally significant effect of the richness variable $F(1,52) = 3.06$, $MS_e = .03$, $p < .09$. Correct rejection rates were .723 for Restricted items and .664 for Broad items. As predicted, the number of facts that needed to be manipulated to compute an inference (i.e., richness) did not have significant effects in the recognition task. In fact, if anything, Restricted items tended to be rejected more accurately.

Verification task. A repeated measures analysis with distance (Near or Far) as the within-subjects variable and presentation type as the between-subjects variable yielded no main effects of presentation type or any interaction of presentation type and distance, both Fs < 1. There was a significant effect of distance, $F(1,52) = 16.80$, $MS_e = .04$, with Near items being verified more accurately than Far items, .649 and .485, respectively; indicating that it is easier to make an inference if the required analysis involves integrating the material in the text rather than using or applying the studied material.

A separate analysis with the richness (Restricted or Broad) variable yielded no main effects or interactions of presentation type either, both *F*s < 1. There was a significant main effect of richness, $F(1,52) = 36.01$, $MS_e = .04$. As expected, the performance on the Broad inference items that can be verified using many sources of information from the text was considerably more accurate than performance on the Restricted inference items with a small number of sources of information, .773 and .539, respectively.

The same analyses were repeated, this time including both true and false items from the verification test. The pattern was identical to that found when only true items were included. Presentation type did not produce a main effect or interact with the distance variable, both *F*s < 1.60. There was a main effect of distance, $F(2,52) = 43.35$, $MS_e = .03$, with Near items being verified more accurately than Far items, .675 and .476, respectively.

Presentation type did not interact with the richness variable either, F <1, but there was a main effect of richness $F(1,52) = 72.55$, $MS_e = .02$ with Broad items verified more accurately than Restricted items, .790 and .542, respectively.

Compared on the same inference items (all true), performance on the recognition and verification tasks were affected differently by the type of inference item indicating a dissociation between the recognition and verification tasks. If the inference involved combining different sources of information from the text (i.e., Near) rather than applying the information from the text to a new situation (i.e., Far) then the verification performance was better. This variable had an opposite effect in the recognition test. The Near items were rejected less accurately than Far items. As Neely (1989) has summarized, the most convincing type of dissociation between two tasks is when changes in a variable (e.g., distance) improves performance in one task but debilitates performance in the other task. A dissociation was also observed with the richness dimension. Even though Broad items were verified more accurately than Restricted items in the verification task, this variable did not produce any significant effects in the recognition test. (If anything, there was a trend in the opposite direction with Restricted items being rejected more accurately than Broad items in the recognition task).

To summarize, if the required inference is closer to the text information in memory (distance variable) or if many sources of information and several routes are available to compute the inference (richness variable), then the verification but not the recognition performance, improves.

General Discussion

The results of the experiment can be summarized simply: The variables that affect performance on a recognition test are not similar to the variables that affect performance on a verification test. This finding conceptually replicates and extends the dissociations reported by Kintsch, Reder and their respective colleagues.

On studied information, performance was quite good for both recognition and verification task groups. This result indicates that for both tasks, the studied information was most likely to be retrieved from memory. Hence the text presentation manipulation--which presumably affected the memory for the text --did not produce any strong dissociations between the two tasks on studied items. However, the two tasks required different processes on nonstudied (inference) items. The recognition judgments could still be made by checking the memory to see if a proposition was studied. In contrast, an inference needed to be computed to make a verification judgment. Our results indicate that both the text presentation as well as the richness and distance dimensions affected the two tasks differently on those nonstudied items. Although not very strong, a mixed text presentation caused more false alarms but did not affect verification judgments. We will discuss the effects of the richness and distance dimensions in more detail below.

Online research has indicated that elaborative inferences are not routinely made while reading the text. Instead such inferences are usually generated as a function of the comprehension questions that are asked. Under these circumstances, both the available information in memory that can be used to generate inferences as well as the nature of the questions that invites inferencing become important. If the information in memory is rich enough (i.e., broad not restricted), then verification performance improves considerably. In contrast, the richness of the available information does not improve recognition performance. In our study, we deliberately chose a topic that was unfamiliar to our subjects so that the inferences could be based on the information in the text rather than on their background knowledge. However, if a reader has enough background information on a topic, even an impoverished text can be represented richly and lead to better inferencing. For example Yekovich, Walker, Ogle, and Thompson (1990) demonstrated that low-ability high school students were much better at drawing inferences and determining the main idea of a passage if they had a lot of background information on a topic. In contrast, recalling factual information from a passage was much less affected by the level of background knowledge. This pattern conceptually replicates our finding that the richness of the available information affects inferencing, but not direct retrieval from memory.

The nature of the question that necessitates inferencing is another dimension of interest. Questions that require integrating material across a text are verified more accurately than questions that require applying the knowledge (near vs. far). This pattern demonstrates a difficulty articulated by educators many times: Students have a lot more difficulty in applying and transferring studied information to new contexts as compared to integrating studied information (for a review see Spiro et al., 1987). However, the distance dimension has the opposite effect on recognition performance. Nonstudied but Near items possibly have more intersections with the information in the text that makes them harder to reject as nonstudied.

The dissociation between memory and inferencing tasks has another educational implication. One of the important issues in education is reading texts in order to learn from them. Especially with expository texts, the goal is not so much to recall a text well, but to learn from the text and apply this knowledge. Hence variables that improve inferencing performance should be emphasized just as much as those variables that only improve memory performance. Along the same lines it is interesting to note that educational researchers have mostly focused on variables that affect comprehension and recall of a text rather than on variables that affect learning from text (Kintsch, 1986).

As the dissociations in the studies by Perrig & Kintsch (1985) and Reder (1982) as well as our study indicate, having a good memory of the studied information does not necessarily mean that the studied information can be used to make inferences. The challenge for an educator is to distinguish between educational practices that help improve memory of a text versus those that help inferencing, because these two outcomes, remembering a text well as compared to learning from a text are not facilitated by identical variables.

References

Anderson, J. R. (1978). Arguments concerning representations for mental imagery. *Psychology Review, 85*, 249-277.

Anderson, J. R., & Reder, L. M. (1987). Effects of number of facts studied on recognition versus sensibility judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*, 355-367.

Anderson, R. C., & Pearson, P. D. (1984). A schema-theoretic view of basic processes in reading comprehension. In P. D. Pearson (Ed.), *Handbook of reading research* (pp. 255-291). New York: Longman.

Armbruster, B. B., & Ostertag, J. (1989). *Questions in elementary science and social studies textbooks* (Tech. Rep. No. 463). Urbana-Champaign: University of Illinois, Center for the Study of Reading.

Barrett, T. (1976). A taxonomy of reading comprehension. In R. Smith & T. C. Barrett (Eds.), *Teaching reading in the middle grades*. Reading, MA: Addison-Wesley.

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, W. H. (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain.* New York: Longman.

Bransford, J. D. (1979). *Human cognition*. Belmont, CA: Wadsworth Publishing Co.

Ceci, S. J., & McNellis, K. L. (1987, April). *Entangling knowledge and process*. Paper presented at the annual meeting of the American Educational Research Association. Washington, D.C.

Danner, F. W. (1976). Children's understanding of intersentence organization in the recall of short discriptive passages. *Journal of Educational Psychology, 68*, 174-183.

Graesser, A. C., Haberlandt, K., & Koizumi, D. (1987). How is reading time influenced by knowledge-based inferences and world knowledge? In B. K. Britton & S. M. Glynn (Eds.), *Executive control processes in reading* (pp. 217-251). Hillsdale, NJ: Erlbaum.

Johnson-Laird, P. N. (1983). *Mental models*. Cambridge, MA: Harvard University Press.

Kintsch, W. (1986). Learning from text. *Cognition and Instruction, 3*, 87-108.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review, 95*, 163-182.

Long, D. L., Golding, J. M., Graesser, A. C., & Clark, L. F. (1990). Goal, event and state inferences: An investigation of inference generation during story comprehension. In A. C. Graesser & G. H. Bower (Eds.), *The psychology of learning and motivation* (pp. 89-102). New York: Academic Press.

McKoon, G., & Ratcliff, R. (1986). Inferences about predictable events. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12*, 82-91.

McKoon, G., & Ratcliff, R. (1990). Dimensions of inference. In A. C. Graesser & G. H. Bower (Eds.), *The psychology of learning and motivation* (pp. 313-328). New York: Academic Press.

Neely, J. H. (1989). Experimental dissociations and the episodic/semantic memory distinction. In H. L. Roediger & F. I. M. Craik (Eds.), *Varieties of memory and consciousness* (pp. 224-270). Hillsdale, NJ: Erlbaum.

Pearson, P. D., Hansen, J., & Gordon, C. (1979). The effect of background knowledge on young children's comprehension of explicit and implicit information. *Journal of Reading Behavior, 11*, 201-209.

Perrig, W., & Kintsch, W. (1985). Propositional and situational representations of text. *Journal of Memory and Language, 24*, 503-518.

Potts, G. R., Keenan, J. M., & Golding, J. M. (1988). Assessing the occurrence of elaborative inferences: Lexical decision versus naming. *Journal of Memory and Language, 27*, 399-415.

Reder, L. M. (1982). Plausibility judgment versus fact retrieval: Alternative strategies for sentence verification. *Psychological Review, 89*, 250-280.

Reder, L. M. (1987). Strategy selection in question answering. *Cognitive Psychology, 19*, 90-138.

Reder, L. M., Wible, C., & Martin, J. (1986). Differential memory changes with age: Exact retrieval versus plausible inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12*, 72-81.

Richgels, D. J., McGee, L. M., Lomax, R. G., & Sheard, C. (1987). Awareness of four text structures: Effects on recall of expository text. *Reading Research Quarterly, 22*, 177-196.

Roller, C.M. (1990). Commentary: The interaction of knowledge and structure variables in the processing of expository prose. *Reading Research Quarterly, 25*, 79-89.

Seifert, C.M., Robertson, S.P., & Black, J.B. (1985). Types of inferences generated during reading. *Journal of Memory and Language, 24*, 405-422.

Spiro, R. J., Vispoel, W. L. Samarapungavan, A., Schmitz, J., & Boerger, A. E. (1987). Knowledge acquisition for application: Cognitive flexibility and transfer in complex content domains. In B. K. Britton & S. M. Glynn (Eds.), *Executive control processes in reading* (pp. 177-199). Hillsdale, NJ: Erlbaum.

Starr, C. G. (1965). *A history of the ancient world*. New York: Oxford University Press.

Taylor, B. M., & Samuels, S. J. (1983). Children's use of text structure in the recall of expository material. *American Educational Research Journal, 20*, 517-528.

van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.

Weaver, C. A., & Kintsch, W. (1990). Expository text. In R. Barr, M. L. Kamil, P. Mosenthal, & P. D. Pearson (Eds.), *Handbook of reading research* (Vol. 2, pp. 230-245). New York: Longman.

Whitney, P. (1987). Psychological theories of elaborative inferences: Implications for schema-theoretic views of comprehension. *Reading Research Quarterly, 22*, 229-310.

Yekovich, F. R., Walker, C. H., Ogle, L. T., & Thompson, M. A. (1990). The influence of domain knowledge on inferencing in low-aptitude individuals. In A. C. Graesser & G. H. Bower (Eds.), *The psychology of learning and motivation* (pp. 259-278). New York: Academic Press.

Author Note

Correspondence concerning this report should be addressed to: Aydin Durgunoğlu, Center for the Study of Reading, University of Illinois, 51 Gerty Drive, Champaign, IL 61820

Footnotes

[1] Recall that the studied items were paraphrased versions of the information in the text. This manipulation actually works against our predictions by bringing recognition test closer to the verification test and reducing the differences between the two tasks, thus making it more difficult for us to find dissociations between the two tasks.

[2] We have also scored the verification test data considering both the accuracy of the true-false decision and the accuracy and completeness of the given rationale. The overall pattern of results did not change.

Table 1

Mean Scores (and Standard Deviations) on the Verification Task as a Function of Presentation Type

	Presentation Type[a]		
	SS *n*=18	SU *n*=19	UU *n*=18
Studied items	.840 (.119)	.839 (.133)	.866 (.120)
Nonstudied items	.606 (.147)	.627 (.205)	.636 (.159)

[a]S=Structured text U=Unstructured text

Table 2

Mean Scores (and Standard Deviations) on the Recognition Test as a Function of Presentation Type

	Presentation Type[a]		
	SS $n=18$	SU $n=19$	UU $n=18$
Studied items[b]	.785 (.148)	.759 (.117)	.728 (.184)
Nonstudied items[c]	.266 (.164)	.367 (.156)	.251 (.180)

[a]S = Structured text U = Unstructured text

[b]Correct acceptance, hit rates

[c]Incorrect acceptance, false alarm rates

Table 3

Mean Scores (and Standard Deviations) on the Nonstudied (Inference) Items as a Function of Test and the Type of Item

	Distance		Richness	
	Near	Far	Restricted	Broad
Verification[a]	.649 (.176)	.485 (.263)	.539 (.182)	.773 (.247)
Recognition[b]	.645 (.206)	.903 (.166)	.723 (.167)	.664 (.277)

[a]= Verification scores

[b]= Correct rejection rates

This page is intentionally blank.