Institute of Cognitive Science

Technical Report

University of Colorado, Boulder

Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text.

Danielle S. McNamara
Eileen Kintsch
Nancy Butler Songer
Walter Kintsch

Institute of Cognitive Science
University of Colorado at Boulder
Boulder, CO 80309

ICS TECHNICAL REPORT #93-04

Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text

in press, *Cognition & Instruction*

Danielle S. McNamara, Eileen Kintsch,
Nancy Butler Songer, and Walter Kintsch

The University of Colorado, Boulder

Send Correspondence to:
Dr. Danielle S. McNamara
Institute of Cognitive Science, CB 344
University of Colorado
Boulder, CO 80309-0344
mcnamara@clipr.colorado.edu

Abstract

Two experiments, theoretically motivated by the construction-integration model of text comprehension (Kintsch, 1988), investigated the role of text coherence in the comprehension of science texts. In Experiment 1, junior-high students' comprehension of one of three versions of a biology text was examined via free recall, written questions, and a keyword sorting task. This study demonstrates advantages for globally coherent text and for more explanatory text. In Experiment 2, interactions between local and global text coherence, readers' background knowledge, and levels of understanding were examined. Using the same methods as in Experiment 1, we examined students' comprehension of one of four versions of a text, orthogonally varying local and global coherence. We found that readers who know little about the domain of the text benefit from a coherent text, whereas high-knowledge readers benefit from a minimally coherent text. We argue that the poorly written text forces the knowledgeable readers to engage in compensatory processing to infer unstated relationships in the text. These findings, however, depended on the level of understanding, textbase or situational, being measured by the three comprehension tasks. Whereas the free-recall measure and text-based questions primarily tapped readers' superficial understanding of the text, the inference questions, problem solving questions, and sorting task relied on a situational understanding of the text. This study provides evidence that the rewards to be gained from active processing are primarily at the level of the situation model, rather than at the superficial level of textbase understanding.

Are good texts always better?
Interactions of text coherence, background knowledge, and levels of understanding in learning from text

People learn a great deal from texts -- story books, textbooks, newspapers, or manuals. The process of learning from texts is complex, however, and as yet not completely understood. In order to optimize learning, should one make the comprehension process as easy as possible? Or should one, as many educators insist, ensure that the learner participates actively and intentionally in the process of constructing the meaning of a text? The traditional view of learning and instructional design has long been governed by the notion that the learner's path should be made as troublefree as possible. However, recent research on the acquisition of simple skills (e.g., Schmidt & Bjork, 1992) indicates that this approach does not always lead to optimal learning. The studies reported here are concerned with whether there are analogous effects in complex cognitive processing, such as text understanding. Specifically, should we facilitate the readers' task by improving the comprehensibility of a text, or can we increase their active involvement by placing obstacles in their way? In the latter case, what sort of obstacles might have beneficial effects on learning and under what conditions?

Two sets of research findings provide an empirical background for the present research. The first set of findings refers to the impact of the coherence of a text on comprehension and the second considers the benefits of active processing for learning.

<u>Coherence</u>. The coherence of a text is a major factor in comprehension (Kintsch & Vipond, 1979; van Dijk & Kintsch, 1983). Recent studies have demonstrated that revisions which increased the structural and explanatory coherence of texts resulted in substantial increases in recall among 5th-grade students (Beck, McKeown, Sinatra, & Loxterman, 1991) and college students (Beyer, 1987, 1990; Britton & Gulgoz, 1991). In Beck et al. the revisions consisted of adding clarifying and elaborative content to supply background information and explain causal relationships. The revisions in Beyer and Britton and Gulgoz were based on a computer simulation of the comprehension of the original text (Kintsch & van Dijk, 1978; Miller & Kintsch, 1980). Whenever the simulation broke down due to coherence gaps from one proposition to the next, text was added to make the relationship explicit. These text additions ranged from low-level information, such as identifying anaphoric referents, synonymous terms, or connective ties, to supplying background information left unstated in the text. Filling in this kind of information, which an author would normally expect readers to provide from their own background knowledge, proved to be very effective, particularly since the college students in their study lacked the domain knowledge to do so on their own.

<u>Active processing</u>. As described above, one effect of increasing the coherence of text is that the reader comprehends it better. However, another effect that this text manipulation might have is to reduce the amount of active processing during reading. That is, easing the reader's burden of figuring out the meaning of the text could result in less effective learning. According to E. Kintsch (1990), this was the case for more knowledgeable and more skilled readers who wrote better summaries of a poorly organized text than of a well organized one. In general, researchers have found that people remember information that they have actively generated better than presented information and they are better able to put such knowledge to use in novel situations. Advantages of active processing for memory and learning are well established in the list-learning literature (e.g., Slamecka & Graf, 1978), as well as for text processing and strategy use (e.g., Bereiter & Scardamalia, 1989; Chi & Bassok, 1989). Active inferencing during reading, if it is successful, can benefit learning because it results in the formation of more links between

the incoming information and information in the personal knowledge base (Mannes & Kintsch, 1987). Due to multiple ties to other concepts, well-integrated knowledge is more quickly accessed and available for use when needed.

In summary, an easy-to-read, coherent text is generally easier to recall. On the other hand, active processing, such as that necessary for difficult text, facilitates learning. These contradictory findings regarding easy versus difficult text may be better understood in the context of Kintsch's construction-integration (CI) model (1988, 1992) of text comprehension. One important feature of this model is that it distinguishes several different levels in the mental representation of a text that readers construct. The two levels of understanding that are relevant here are the *textbase* and the *situation model.*

The *textbase* contains the information that is directly expressed in the text, organized and structured in the way that the author had organized the material. It has a local structure (the *microstructure)*, as well as a global structure (the *macrostructure*). The construction of the textbase involves the extraction of semantic information from a text. In the CI model, this takes the form of an interrelated network of propositions. The process of transformation from words to meaning units involves a certain amount of inferential activity. For example, the referents of pronouns must be identified, synonymous terms must be matched, glaring gaps in the coherence of the network must be filled by bridging inferences, and so on. In general, this is the minimum amount of processing that a cooperative and motivated reader would perform. The result is a locally and globally, well-structured memory representation of the text -- an episodic text memory in the form of a coherent propositional network. This representation is capable of supporting several activities. On the basis of the textbase readers can verify statements they have read, they can answer questions about the text, they can recall the text, or they can summarize it.

However, knowing the text at the level of the textbase does not necessarily ensure that the reader understands it at a deeper level. Especially in the case of a demanding scientific text, more is required for understanding than just the ability to reproduce the text itself. It is usually the case that the reader must contribute information that was not stated explicitly in the text from his or her own store of knowledge about the domain in question. Furthermore, considerable active inferencing might be required to link the text with the reader's prior knowledge (an example will be discussed below). The result of such inferencing is the *situation model.* The situation model integrates the information provided by the text with prior knowledge, often reorganizing and restructuring it in terms of the reader's understanding of the knowledge domain as a whole rather than the particular text just read. The resulting mental representation allows for a deeper understanding of the text, which is no longer a separate, episodic memory unit but is linked to the reader's long-term memory and knowledge. The textual information thus becomes retrievable not just via an episodic memory trace resulting from reading the text, but via the shared elements by which it is linked in the reader's long-term memory. The new knowledge is therefore usable in novel environments and for unanticipated problem solving tasks.

For the purpose of reproducing a text in one way or another, a good textbase suffices; learning from text implies more than that and requires a good situation model linked with the reader's long-term knowledge (Kintsch, 1994). One should not think of textbase and situation model as two distinct mental structures, however. The reader constructs a single mental structure, but for the purpose of analysis it is convenient to distinguish aspects of that structure that are directly derived from the text, and aspects that are more knowledge and inference based. The same elements usually will participate in both types of relationships, but in somewhat different ways. Indeed, a number of studies have succeeded in demonstrating differential effects of these two levels of representation on various tasks (e.g., navigating a route, Perrig & Kintsch, 1985; a computer programming task, Schmalhofer & Glavanov, 1986; inferential reasoning, Mannes & Kintsch, 1987). More informally, consider the difficulty one may experience in following directions -- for example, through unfamiliar terrain, or in order to assemble an unfamiliar device -- despite an easily comprehended description. The problem does not lie at the textbase level of understanding, but rather in the inability to construct from that information a representation of the situation depicted. This theoretical distinction has important implications for our understanding of the processes by which readers acquire usable information from texts.

With respect to the design of instructional text, the CI theory implies that textbooks should induce students to form appropriate situation models, not just textbase representations good only for reproduction. The formation of such situation models requires two things as a basis: active inferencing, and adequate prior knowledge. Without adequate prior knowledge, students are limited in their constructive processes. Nevertheless, a well written, fully coherent, explicit text at least will allow them to form a good textbase. In contrast, students who have the requisite background should be encouraged to use it. Accordingly, one way to promote deeper understanding -- the construction of an adequate situation model -- might be to disrupt the coherence of the text at both the local, intersentential level and at the level of its global organization, or macrostructure. If there is a gap in comprehension that the reader cannot bridge with an inference, this forms an impasse. It is likely to be noticed, and a repair process can be initiated. This repair process necessarily forces the reader to construct a situation model because prior knowledge is necessary to do so. Hence, the repairs will be particularly beneficial because the reader might otherwise remain satisfied with a superficial textbase understanding.

An example from the text used in Experiment 2 might clarify this argument. Consider the following sentence pair from a fully coherent text:

> "The blood cannot get rid of enough carbon dioxide
> through the lungs. Therefore, the blood becomes purplish."

A reader can easily comprehend these sentences without ever realizing just why excess carbon dioxide would make the blood purplish. A superficial understanding may result because the reader lacks the necessary background knowledge, or just doesn't bother to figure it out. Readers all too frequently take the path of least resistance. To generate the low-coherence text, the "therefore" was deleted. Without this connective, the two sentences are simply juxtaposed. It is, of course, possible to accept these two sentences as statements of facts belonging to the same topic, but which are otherwise unrelated. However, it is more likely that the reader will notice the gap and attempt to fill it. This is not possible on the basis of the textual content alone: Given merely these two statements, there is a large set of potential sentence connectives, such as *therefore, and, because, when, but,* and so on, and no linguistic cues that would allow a reader to choose the right one. Hence, superficial solutions to the impasse are not available. The only way to bridge the gap is by knowing certain facts, for example, that the blood that flows from the heart to the lungs is low in oxygen, high in carbon dioxide, and looks blue, whereas the blood that flows out of the lungs is high in oxygen, low in carbon dioxide, and looks bright red. Knowing these facts enables the reader to infer *therefore* as the right connective between these two sentences.

In summary, the two sets of empirical findings regarding easy and difficult text understood in the context of the CI model suggests a possible way to improve comprehension and learning from a textbook. Although a coherent text is generally easier to recall, more difficult text may induce active processing and thus facilitate learning. Kintsch's (e.g., 1988) model of text comprehension suggests that one way to avoid a superficial understanding of a text may be to make the text more difficult by disrupting its coherence. Coherence is of crucial importance for understanding, but deeper understanding results when readers make their own bridging inferences and derive their own macrostructure. Although filling in such linguistic gaps requires a certain degree of background knowledge, doing so may force more knowledgeable readers out of a passive processing mode induced by easy text. Consequently, we hypothesize that the same text cannot be optimal for everyone: Low-knowledge readers will profit from a fully coherent text, whereas high-knowledge readers will learn better with a text that stimulates more active processing.

Experiment 1

The second experiment reported here tested our hypothesis suggested above that text coherence would differentially affect readers with low and high background knowledge. However, testing this hypothesis required first building a more solid empirical platform. Hence, Experiment 1 was designed to address several questions fundamental to investigating the complex interactions predicted by our hypothesis. First, we wanted to replicate the results of Britton and Gulgoz (1991) with a younger population of students and with a different kind of text. Britton and Gulgoz found advantages for more coherent text with college students who read a fairly difficult, and unfamiliar, historical text about the Vietnam War. One goal in Experiment 1 was to test whether improving the coherence of text would generalize to younger students who are just embarking on the formal study of a new domain. Thus, our target population consisted of middle school students. Moreover, because we were interested in devising ways to improve textbooks designed for instruction, it was important to determine whether previous advantages for coherent text could be replicated with instructional texts. Thus, our target text was an instructional biology text written for the appropriate grade level.

A second goal of Experiment 1 was to compare the individual contributions of local, global, and explanatory coherence. A text is locally coherent if each sentence is explicitly related to the next, for example, via argument overlap. A text is globally coherent if the paragraphs and larger sections of text are clearly related to each other and to the overall topic. Explanatory coherence refers to content that supplies background knowledge needed to understand the text which the reader may not have. Thus, three versions of a chapter on the traits of mammals, taken from a biology text, were used to examine the effects of these three types of coherence: the original school text which was locally coherent to begin with, but which lacked global coherence; a revised text which added macrosignals alerting the reader to the underlying structure; and an expanded text which added explanatory coherence to the text.

Our third purpose in conducting Experiment 1 was methodologically motivated. In previous research the effects of text coherence have primarily been assessed by recall which is a measure of readers' textbase knowledge: It tests how well they remember the actual content of the original text and the way it is organized. However, we are also interested in comprehension measures that tap into the reader's understanding of the situation described by the text, that is, the mental model which the reader constructs from the textual content. Experiment 1 uses a sorting task as a measure of situational understanding. This task, which involves organizing a set of key concepts into categories, poses problem solving demands that require more than a superficial, textbase understanding of the text.

Method

Students' pre-knowledge about the traits of mammals was assessed by means of the sorting task and a set of questions about the text. They then read one of three text versions, recalled it immediately afterwards, performed the sorting task again, and answered another set of questions similar to the previous one. The different measures used are intended to reflect different levels of understanding. Recall and a subset of the questions based entirely on the text content (i.e., textbased questions) were used to assess subjects' ability to reproduce this content. They provide a relatively superficial index of understanding, based primarily on the textbase that has been constructed. In contrast, the sorting task and a subset of the questions based on both text content and prior knowledge (i.e., inference questions) were used to assess subjects' deeper understanding: how their text memory (textbase as well as situation model) modified their perception of the relations among the concepts to be sorted, and how well the situation model could be used for inferring novel relations.

Subjects

Thirty-six subjects living in Boulder, Colorado were each paid $10 to participate in this study. They were between the ages of 11 and 15 years (M = 13 years, 2 months) and were entering grades 7 through 9. Twelve subjects were assigned to each of three text conditions (i.e., original, revised, expanded texts) based on the order of testing. The mean age of the subjects in the original, revised, expanded text conditions were 13.1, 13.0, and 13.5, respectively. The mean grade levels for the three conditions were 7.8, 7.6, and 8.0. There were no significant differences between the three groups of subjects in terms of their age or grade level.

Materials

Texts. Three versions of a biology text about traits of mammals were used: an *original* version, a *revised* version, and an *expanded* version. The original biology text (Silver Burdett & Ginn, 1987) discussed traits of mammals, a subject area reasonably familiar to students in grades 6-8. The text was 1.5 pages long (590 words). Following the method of Britton and Gulgoz (1991), comprehension of the text was simulated by means of a computer program (Kintsch & van Dijk, 1978; Miller & Kintsch, 1980). A breakdown in the simulation due to lack of overlapping arguments signalled a gap in the coherence structure of the text. However, since our text was instructional text written at grade level, there were, in fact no such gaps at the local level. Unlike the text used by Britton and Gulgoz, not a single bridging inference was required to run the simulation. On the other hand, the text had a list-like macrostructure which was less coherent. That is, the various subtopics were not always clearly marked in their relationship to the overall topic and to each other. Furthermore, the paragraph structure of the text did not correspond entirely to its macrostructure.

The revision consisted primarily of adding material that explicitly identified the major subtopics as traits of mammals. We use the term *macrosignal* to refer to these linguistic pointers to the macrostructure. This should be distinguished from the term *macroproposition* which refers to the major subtopics discussed in all three text versions (i.e., the 7 mammal traits in this case), whether they are signalled or not. The result of our revision was a 2-page text of 821 words in which each paragraph corresponded to one of the traits discussed. The original and revised texts therefore have approximately the same content. The only difference is that the interconnections between the various content units are made explicit in the revised version, especially at the global level, but to some extent locally as well. For instance, the original text merely says that "mammals are the only animals that have hair or fur", whereas the revision relates this fact to the more global topic concerning mammals' ability to live in many environments.

The expanded version of the text was written because of some obvious shortcomings of both the original and revised versions, not in their structure, but in their content. Identifying various traits that mammals have does not really help a student to understand biology. It is more useful to know how these traits contribute to the success of this group of animals and their ability to survive. Hence, content to this effect was supplied in the expanded version, while at the same time some irrelevant detail was omitted (e.g., the names of the four chambers of the heart). The expanded version was therefore considerably longer, 4 pages or 1167 words. In effect, a few concrete details were exchanged for some rather lengthy discussions about the functions of the traits in terms of the evolutionary success of mammals.

At the local level, all three texts were coherent. At the global level, only the revised and expanded texts were coherent. In terms of content, the situation models implied by the original and the revised versions were essentially the same. That is, the additional content in the revised version was implicit in the original version for a knowledgeable reader; the equivalence would be less complete for low-knowledge readers. In contrast, the expanded version introduced a new, higher-level discussion about the role that the traits of mammals play in a larger context.

An excerpt from each of the three texts is presented in Appendix A. Each of the excerpts is from the section of the text describing mammalian teeth. The coherence additions to the revised text are macrostructure statements which identify specialized teeth as a mammalian trait and relate this trait to the more global topic concerning mammals' ability to live in many environments. The additions to the expanded text include both the macrostructure statements, as in the revised text, and explanatory additions referring to how the characteristics of mammalian teeth have contributed to the survival of the species. The complete texts are reprinted in Kintsch, McNamara, Songer, and Kintsch (1992).

Pretest and posttest questions. There were 32 questions comprised of multiple-choice, true/false, fill-in-the-blank, and short-answer questions concerning animal traits and behaviors. The questions were classified into three different types: (a) text-based questions for which the necessary information was stated in the original text (n=19); (b) inference questions which required some type of inferencing or analytic reasoning (n=9); and (c) non-text questions which dealt with information that did not occur in any of the three texts, but that was related to the general topic (n=4).

The three different types of questions were chosen for several reasons. Text-based questions were designed to assess both prior knowledge of the topic before reading as well as memory for the text information after reading. The answers to text-based questions were explicitly stated in the text and generally required only one sentence from the text to answer. An example of a text-based question pertaining to the excerpt provided in Appendix A is: "The type of teeth a mammal has is related to ?" (Answer: the kind of food the mammal eats). Inference questions were intended to assess differential abilities to apply information in the text to a novel situation, depending on which text was read. Inference questions required linking information from separate sentences and/or paragraphs and applying the information to the stated problem. An example of an inference question is: "Mother coyotes usually give birth to a litter of pups in early spring. However, one year, soon after most of the young pups were born, a series of late storms brought snow and very cold temperatures. What would help the young coyotes survive the cold?" (Possible answers: their fur, mother's warmth, being warm-blooded). If a reader supplied the answer that the coyotes' fur would help them through the winter, this particular answer would require integrating and applying information from one sentence that mammals have fur and from another that the fur provides insulation. Similarly, if a reader supplied the answer that being warm-blooded would help them, this requires linking and applying the separate information that mammals are warm-blooded and that being warm-blooded helps them to maintain a constant temperature. Of course, all of the correct answers to the question also required the prior knowledge that coyotes are indeed mammals. Thus, this type of question required linking together various pieces of information in the text along with an understanding of the information presented in the question itself. Non-text questions were included to ensure that there were no differences between pretest and posttest, nor between groups for general background knowledge. An example of a non-text question is: "Carnivores eat ?" (Answer: meat).

The 32 questions were divided pseudorandomly into two sets of 16 questions, with the constraints that each set include approximately half of each of the three types of questions, and that the questions in each set be matched as nearly as possible for difficulty and content matter. Half of the subjects in each group received one set of questions as the pretest and the other as the posttest, and the remaining subjects were given the tests in the reverse order. For each set, there were two random orders of questions which were also counterbalanced across pretest and posttest.

Sorting task cards. Subjects sorted two sets of cards (3.5" x 1.5"). The first set of cards, consisting of 21 animal names, was a practice task, for which the data were not analyzed. The second set of cards consisted of 16 animal characteristics, each printed on a separate card. This list included 7 characteristics classified a priori as mammalian traits (i.e., *has hair or fur, has several types of teeth, has mammary gland, instinctive and learned behavior, tolerates extreme temperatures, provides a lot of protection for their young, has complex heart*), 4 as traits that could be characteristic of either mammals or non-mammals (i.e., *hibernates, migrates, has lungs, has central nervous system*), and 5 as non-mammalian traits (i.e., *produces many babies at one time, is cold blooded, lays or releases eggs, mainly instinctive behavior, has scales or hard shell*). The keyword sorting task was employed in order to assess the changes in the organization of knowledge that occurred as a result of reading the experimental texts. These kinds of changes would be indicative of the situation model the reader has constructed.

Procedure

Subjects were tested individually, either in their homes or in the university laboratory. The session lasted approximately an hour. The subjects first answered questions concerning personal information (e.g., age, grades, personal interests). The order of the experimental tasks was as follows: (a) sorting tasks, (b) pretest questions, (c) text reading, (d) text recall, (e) sorting tasks, and (f) posttest questions. The pretest questions and sorting tasks may have biased reading and recall by calling attention to particular points discussed in the text. However, these effects were equal across the three text conditions.

Sorting tasks. The subjects sorted the cards containing the keywords before reading the text and again after reading the text. To familiarize themselves with the concepts printed on the cards, they read each card aloud. They were then told to put the cards into piles according to how they thought the concepts should go together. They were told that they could make as few or as many piles or categories as they wished, there needn't be the same number of cards in each pile, they could change their mind and reorganize at any time, and that there was no correct or incorrect way to organize the cards. After stating that they had finished sorting the cards, the subjects were asked why they had put each pile of cards together in this manner and how they would label each set of cards.

Pretest and posttest questions. After completing the sorting task, the subjects were given one of two sets of questions which asked questions concerning mammals and their characteristics. The subjects were encouraged to do their best, but to guess if they were not sure of an answer to a question.

Text reading. The subjects were asked to read silently one of the three texts. After finishing, they were asked to read it a second time. By having them read the text twice we hoped to avoid cursory reading by the children and to ensure as complete an understanding as possible.

Text recall. After reading the text, subjects were asked to recall orally as much of the text as they could. Their recall protocols were recorded on a mini-cassette recorder.

Under the conditions of the present experiment, recall can be considered a measure of the textbase formed by the subject during reading. In general, text recall has both a reproductive and a reconstructive component because there is an admixture of reconstructions derived from the situation model that readers have constructed, especially when they are very familiar with the domain of the text or when the textbase itself can no longer be successfully retrieved. However, for low-knowledge readers or on short-term tests, as in the present experiment, the reproductive component of recall based on the textbase usually dominates.

Results

All reported analyses are reliable with a probability less than .05 unless otherwise stated.

Text recall

Text recall was quantified in terms of the proportion of propositions remembered from the texts. The three texts were propositionalized according to the principles specified in van Dijk and Kintsch (1983). Complex propositions as defined by van Dijk and Kintsch are schematic representations of the meaning of sentences. This schema, shown in Figure 1, comprises a proposition slot and a circumstance slot, specifying the time and place of the action or state. The proposition consists of a predicate and arguments characterizing the internal semantic structure of a sentence. Each of these components may be modified, and modifiers may be further modified themselves, as indicated by the syntactic structure of the sentence to be represented (e.g., embedded clauses, relative clauses, adjectivization). Propositions in a discourse may either be unrelated, in which case the discourse is incoherent; or they may be indirectly related because they share a circumstance (e.g., time, place) or argument; or they may be directly related, either temporally, conditionally, or causally by an explicit sentence connective. The propositional analysis of our experimental texts yielded 55 complex propositions for the original, 57 for the revised, and 73 for the expanded text.

Figure 1

A lenient scoring system was used to score for presence or absence of a complex text proposition or any part thereof containing information related to the topic. Thus, only the reproductive content of the recall protocols was scored, not the inferential or relational content.

Since the three texts differed in length (there were 48, 50, and 66 micropropositions in the original, revised, and expanded texts, plus 7 macropropositions common to all texts), recall could be analyzed in two ways. In one analysis only those propositions common to all three texts were considered. This type of analysis, however, ignores much of the content in the expanded text version. In the second analysis, all propositions were included, but the length of the texts was controlled for by conducting the analyses using proportional recall. The two analyses yielded virtually identical results, and only the results of the second analysis are reported here.

An analysis of variance (ANOVA) including the between-subjects factor of text (original, revised, expanded), and the within-subjects factor of proposition type (microproposition, macroproposition), was conducted on proportional recall. There was a main effect of text condition, $F(2,33) = 3.7$, $MSe = .038$, reflecting greater proportional recall for subjects who read the revised text ($\underline{M} = .53$) than for subjects who read the expanded ($\underline{M} = .43$) and original texts ($\underline{M} = .38$), $F(1,33) = 6.8$. The recall difference between the expanded and original texts was not reliable, $F(1,33) < 1$.

Macropropositions ($\underline{M} = .59$) were recalled significantly better than micropropositions ($\underline{M} = .30$), $F(1,33) = 91.8$, $MSe = .016$. This difference, however, depended on text condition, $F(2,33) = 4.7$. According to a planned comparison using contrast codes, this interaction between text condition and proposition type was due to the superior recall of macropropositions by subjects who read the revised ($\underline{M} = .71$) and expanded ($\underline{M} = .60$) texts compared to subjects who read the original ($\underline{M} = .46$) text, $F(1,33) = 9.2$. Macroproposition recall was not significantly greater for the revised than for the expanded text condition, $F(1,33) < 1$. Differences in the recall of micropropositions were not significant statistically among the three texts (original $\underline{M} = .31$; revised $\underline{M} = .35$; expanded $\underline{M} = .26$).

The results thus show that the revision resulted in better overall recall, but that this improvement was primarily a consequence of the better recall of macropropositions. Greater improvement in the recall of macropropositions than of micropropositions was expected for the revised version, since most of the revisions were made at the level of the macrostructure. The original version was locally coherent to begin with, so not much could be gained in recall of micropropositions. Indeed, recall of micropropositions for the revised version improved by only 13 percent over the original version, whereas a 54 percent improvement was observed for macropropositions.

<u>Question answering</u>

The pretests and posttests were composed of three types of questions: text-based questions, inference questions, and non-text, general knowledge questions. Points were assigned to each question according to the amount of material targeted in the response. Thus, short-answer questions were each worth a total of 2 points, and true/false, fill-in-the-blank, and multiple-choice questions received 1 point each. Partial credit for the questions was assigned wherever appropriate. Each set of questions was scored for percent correct. A total of 18 possible points could be obtained for either test.

Separate ANOVAs for each question type (text-based, inference, non-text) were performed including the between-subjects factor of text and the within-subjects factor of test (pretest, posttest). Each question type was analyzed separately because the inference and non-text questions did not adequately assess changes in knowledge as a result of reading the texts. Inference questions did not improve from pretest to posttest, and there were no differences between the three text conditions. Likewise, non-text questions were answered equally well by all three groups and did not change from pretest to posttest (all Fs < 1).

For the text-based questions, the increase from pretest to posttest in average percent correct was 18 percent, 33 percent and 16 percent for the original, revised, and expanded texts, respectively. However, while the overall improvement was statistically significant, $F(1,33) = 30.6$, $MSe = .025$, the interaction with text condition failed to reach statistical significance, $F(2,33) < 2$, and the superior improvement of the subjects who read the revised text compared to the original and expanded versions also was only marginally reliable, $F(1,33) = 3.1$, $p = .085$.

In summary, the question results proved disappointing. They were consistent with the prediction that subjects who read the revised text should be able to form a good textbase and hence do well on text-based questions. However, the attempt to measure the deeper understanding of the problem domain that these texts communicated to the readers by means of inference questions was unsuccessful, perhaps because there were too few of them (only 4 or 5 inference questions in each set), or perhaps we simply asked the wrong questions. Question answering is the traditional method for assessing learning from text, in school as well as in the laboratory. In the present case, it proved to be inadequate. The sorting task, however, turned out to be a more sensitive measure of situational understanding that yielded more robust data.

Sorting

In order to test the generalizability of our sorting task results, analyses were performed both by items and by subjects. Item analyses were not possible for the recall and question data because the items were not identical for all subjects. The items differed between subjects for the questions due to the counterbalancing of pretest and posttest questions, and for recall because the propositions of the three texts were somewhat different.

Item analysis. For the sorting task item analysis, a similarity matrix indicating how often each item was sorted into the same category with every other item was constructed for each text condition at pretest and posttest. Since we are particularly interested in how the sorting of the keywords changed as a function of reading the three versions of the experimental text, we need a measure of how mammal-like each keyword was perceived initially, and how this perception changed after reading. Such an index can be obtained by counting for each keyword the frequency with which it was sorted together with other keywords that are characteristics of mammals. The subjects were not instructed to sort these keywords into mammal and non-mammal characteristics, but to the extent that this distinction played a role in their knowledge structure, one would expect it to be a (partial) determinant of their sorting behavior.

Of the 16 keywords, 7 were mammalian characteristics, 4 were characteristics shared by mammals and other species, and 5 were characteristics of non-mammalian species. This would constitute the optimal classsification of the items as suggested by the text. Pre- and post-sort scores can be calculated for each item in these three classes. From these scores, change scores can be computed. These change scores indicate how much effect the text (as well as other intervening factors) had on the use of the mammal concept as a sorting principle. Positive change scores indicate that an item was perceived as more mammal-like after reading the text, negative scores indicate that it was perceived less so. Thus, ideally, mammal traits would be classified together more often after reading, whereas non-mammal and shared traits would be less likely to occur together with the mammal traits. Figure 2 shows the average values of these change scores for the three item types for the three text groups.

Figure 2

The item analysis of the sorting task scores included the within-items factors of test and text, and the between-items factor of item type (mammal, shared, and non-mammalian characteristics). The main effect of test was reliable, $F(1,13) = 65.7$, $MSe = 42.5$, reflecting the greater use of mammal as a sorting principle at posttest (M = 40.5) than at pretest (M = 28.2). The main effect of item type was reliable, $F(2,13) = 45.8$, MSe = 201.4, reflecting higher scores for the mammal (M = 42.9) and shared (M = 45.1) characteristics than for the non-mammalian characteristics (M = 13.8), $F(1,13) = 90.7$.

Overall, there was greater change in sorting from pretest to posttest for the revised (M = 14.8) and expanded (M = 14.7) texts than for the original text, $F(1,13) = 5.7$. Additionally, the interaction between test and item type, $F(2,13) = 41.3$, reflects the finding of greater change scores for mammal (M = 23.0) and shared (M = 15.2) items, than for non-mammalian items (M = -5.0), $F(1,13) = 68.8$, and greater change for mammal than for shared items, $F(1,13) = 5.6$. Thus, there was a larger increase in the tendency to sort items together with mammals for those items that are in fact characteristic of mammals, somewhat less so for shared traits, while traits which are not characteristic of mammals were sorted less frequently with mammal items after reading the texts. Most importantly, the expected three-way interaction of test, text, and item type (shown in Figure 2 in terms of change scores) was reliable, $F(4,26) = 4.5$, MSe = 29.36. This result shows that the strengthening of the mammal classification after reading the text was stronger for the revised and expanded groups than for the subjects who read the original text, $F(1,13) = 11.2$.

Subject analysis. For the analysis by subjects, each subject was given a score that indicates how closely his or her sorting approximates the ideal sorting suggested by the text (i.e., sorting all mammal characteristics in one category, shared characteristics into another, non-mammal characteristics into a third category, and associating the first two categories). A measure related to the harmony statistic introduced by Smolensky (1986) was used to indicate the degree of correspondence between each subject's sort and the ideal sort. The measure of harmony that we used is the sum of the inner product between a subject's sorting matrix and a weight matrix indicative of an ideal sorting. This sum was divided by the sum of the positive values in the weight matrix (in this case, 102) so that the sort score varied between +1.0 and -1.0.[1]

The average improvement in sorting scores from pretest to posttest was 0.155 for the original text (pretest M = 0.306, posttest M = 0.461), 0.355 for the revised text (pretest M = 0.199, posttest M = 0.554), and 0.319 for the expanded text (pretest M = 0.220, posttest M = 0.539). However, while the overall change from pretest (M = 0.241) to posttest (M = 0.518) was significant, $F(1,33) = 26.7$, MSe = .052, the amount of change for the three text versions was not significant statistically, $F(2,33) = 1.3$. Thus, the by-subject analysis is in good qualitative agreement with the by-item analysis, in that the tendency to sort the keywords into mammal and non-mammal groups increased from pretest to posttest, and more so for the revised and expanded versions than for the original version. However, the lack of statistical reliability in the by-subject analysis indicates that this effect was not robust. Apparently, not all subjects reacted in the same way to the three versions of the text. In general, the revised and expanded versions of the text had stronger effects on the sorting data, but not for all subjects. In Experiment 2 this difference between individual subjects will be examined further and related to the knowledge background of the subjects.

Discussion

The goals of our first experiment were to investigate conceptual changes in readers that occur as a function of reading a text and to replicate earlier findings that making texts more coherent improves their recall. The study yielded informative results in both respects.

We were particularly interested in the sorting task as an indicator of the students' knowledge structure before and after reading a biology text. We were able to show that reading a text changes the way students organize concepts in predictable ways. Before reading they sort on the basis of their general knowledge; after reading they sort in a somewhat different way -- based on their knowledge plus the episodic text memory.

The present study does not provide evidence concerning long-term changes in the sorting patterns, nor the extent to which the episodic text memory was integrated with the knowledge structure. The sorting task cannot yield a complete or direct picture of the subjects' knowledge structure. How the keywords are sorted is only in part determined by the sorter's knowledge structure. Nevertheless, sorting a set of key phrases related to the text proved to be a sensitive method for assessing the effects of reading. In the present case, items that were discussed in the text as typical of mammals were assigned more strongly to the mammal category after reading than before. Items not discussed in the text or not characteristic of mammals showed the least tendency to be categorized with mammal items on the posttest and were in some cases even more clearly differentiated than before, in spite of the heavy focus on mammals introduced by the text.

The second main question investigated here concerned the effects of manipulating textual coherence. The original and revised versions of the text differed in coherence, primarily at the macro-level. As expected, subjects who read the revised, more coherent version, recalled more of the text, especially more of the macropropositions within the text, than subjects who read the original version. In addition, sorting changes were larger for the revised than for the original text. Subjects who read the revised version also showed more differentiated sorting changes than the readers of the original version: items for which the text emphasized their mammalian character moved more strongly into the mammal category, while uncharacteristic items moved slightly away from it.

The results of Experiment 1 complement those reported by Britton and Gulgoz (1991). These researchers reported a large increase in text recall after revising their text to make it locally coherent. However, the text used in the present study was locally coherent to begin with, though not globally coherent. Hence our changes, and their effects on recall, were primarily at the macro-level. Thus, making texts more coherent at either the micro- or macro-level can have beneficial results on recall.

So far, we have focused on the coherence differences between the original and revised texts. This has the advantage that it allows unambiguous statements about the effects of that single variable. But there is, of course, much more to good textbook writing than issues of coherence. Beck and McKeown (1991), in their review of textbook analysis methods, identify lack of coherence at the macro-level as one of the major problems with instructional text, but they also raise a number of broader issues. They make the important point that for evaluating a textbook, and presumably for writing one, one must have a clear idea of what kind of information is to be conveyed to the student (the underlying situation model). The problem is not only structural, but also one of content.

It was this concern with content that led us to include the expanded version in our study. In addition to the structural improvement made in the revised version, the expanded version attempts to provide a more adequate situation model. The situation model on which the original text (and hence also the revised text) was based appeared to be inadequate because it simply listed traits of mammals, without explaining their importance. In the expanded version, we supplied explicit arguments that stated why each of the traits was being considered. This made the text longer and possibly more difficult (arguments, not concrete facts were added). As a result, recall suffered somewhat in terms of the proportion of the total text that was reproduced. But at the same time the sorting changes induced by the expanded text were larger and more differentiated than those induced by the revised and original texts. This result confirms the conclusions of Mannes and Kintsch (1987), who observed that richer constructive processing could be related to the generation of a more elaborated situation model and to superior performance on a problem solving task.

While the text revisions affected sorting behavior of junior-high students, this was not true for all students: The sorting analysis was significant only in an analysis by item, but not in the analysis by subjects. Some students did better with the less coherent text, possibly because they had good background knowledge and were stimulated into more active processing by the less coherent text, as has been hypothesized above. The design of Experiment 1 did not permit us to test this hypothesis. In Experiment 2, a different text was employed so that coherence could be manipulated both at the local and global level. Further, we developed a separate pretest and posttest. Unlike Experiment 1, the new pretest did not directly query information presented in the text, but rather background knowledge which would be helpful for understanding the text. This pretest then enabled us to classify subjects according to their knowledge background in order to examine more precisely how knowledge and text coherence interact in comprehension and learning from text.

Experiment 2

Experiment 2 was designed to test the prediction that rather than having a fully coherent text to read, some students may learn better when they have to provide the coherence themselves, both at the local and at the global level. Specifically, we predicted that incoherent texts, compared to fully coherent texts, would be more beneficial to readers with the requisite knowledge background. We expect such readers to use their knowledge to generate the information that is missing from the text and, in so doing, to construct a more complete model of the situation it describes. This active processing advantage would manifest itself most clearly on tasks that depend upon the construction of a situation model and less so on tasks that assess memory for the text itself. Therefore, we used tasks that are differentially sensitive to textbase and situation model constructions.

As in the case of Experiment 1, measures that reflect primarily the textbase are free recall and text-based questions. Free recall is by its very nature a measure of memory for the text itself, a textbase measure, particularly when the task presented to the reader is simply to recall the text, in contrast, for example, to explaining the text. Text-based questions require only a single sentence from the text to be answered, thus understanding the relation between two sentences or the text as whole is not necessary. Situation model measures include problem solving questions, bridging inference questions, elaborative inference questions, as well as the sorting task. Problem solving questions require applying information from the text to a novel situation and hence depend on situational understanding. Bridging inference questions require linking information from two or more sentences in the text to answer the question. Inferring the unstated relationship between sentences is also a process that relies on the situation model. Elaborative inference questions require linking textual and outside knowledge information, which requires some but not necessarily a very deep situational understanding. Finally, the sorting task measures the reader's understanding of the relationship between concepts presented in the text which reflect at least in part the situation model.

We propose that the textbase measures tap into a superficial understanding of the text, whereas the situation model measures tap into learning processes. Of course, this distinction is not absolute, but a matter of degree. Free recall can be dependent on a situation model. For example, it is possible to reconstruct a text on the basis of a good situation model, rather than reproducing it from a textbase. Similarly, textbase effects even in inference questions and the sorting task cannot be ruled out. However, it is safe to say that on the whole, one set of measures is more textbase dependent and another set more situation model dependent.

A junior-high level encyclopedia article on heart disease served as our starting point. We chose the topic of heart disease because we believed it would be fairly unfamiliar to most of the students and yet reveal a wide variation in their background knowledge. In Experiment 1, our revised version of the mammal text primarily added coherence at the macro-level, since local-level coherence was already adequate in the original text. In contrast, Experiment 2 examined the effects of coherence at both the local and global macrostructure level. By increasing and decreasing the coherence of the original text we obtained four texts which all had the same biology content but differed in coherence. The text **CM** was maximally coherent at both the local (C) and macro-level (M); the text **Cm** was identical at the local level but lacked some inferable statements which made the macrostructure explicitly coherent; for the text **cM**, these macro-statements were included, but certain sentence connectives and other material important for local coherence were deleted; finally, the text **cm** had coherence gaps at both the local and macro-level. The experimental predictions concerned primarily the differences between texts **CM** and **cm**; the other two text versions were included to explore the relative weights of local and macro coherence in this situation.

The students' background knowledge was assessed by means of a question-answering task about the functioning of the heart. They were classified as having low or high background knowledge on the basis of a median split of the scores on this test. The theory outlined above predicts an interaction for measures that depend strongly on the readers' situation model: for low-knowledge subjects, the fully coherent text **CM** should be best; in contrast, for high-knowledge subjects, the text with coherence gaps, **cm**, should be best. On the other hand, for measures that reflect primarily the readers' textbase, **CM** should always be better than **cm**.

Method

Subjects

Fifty-six subjects were paid $10 for participating in this study. They were between 10 and 15 years old ($\underline{M}$ = 13 years, 11 months) and were entering grades 7 through 10. Fourteen subjects were assigned to each of the four text conditions based on the order of testing. There were no significant differences between the four groups of subjects in terms of their age or grade level.

Materials

Texts. The experimental texts were based on an entry in a science encyclopedia for school-age students (Raintree Illustrated Science Encyclopedia, vol. 8, 1984). Text coherence was orthogonally manipulated at the local and global levels by adding or deleting linguistic coherence signals. This process resulted in four text versions: a maximally coherent text, at both the local and macro-level (**CM**); a text maximally coherent at the local level and minimally coherent at the macro-level (**Cm**); a text minimally coherent at the local level and maximally coherent at the macro-level (**cM**); and a minimally coherent text, at both the local and macro-level (**cm**). The number of words in the **CM**, **Cm**, **cM**, and **cm** texts are 1053, 913, 823, and 683, respectively. Thus, 230 words were added to increase local coherence, and 140 words to increase coherence at the macro-level.

The maximally and minimally coherent texts are presented in Appendix B. The following four types of text revisions were used to maximize local coherence:

(a) replacing pronouns with noun phrases whenever the referent was potentially ambiguous (e.g., replacing *it* with *the heart*);

(b) adding descriptive elaborations which link unfamiliar concepts with familiar ones (e.g., This disease usually follows a sore throat caused by bacteria known as streptococci. *This is often called "strep throat."*);

(c) adding sentence connectives (e.g., *however, therefore, because, so that*) to specify the relationship between sentences or ideas; and

(d) replacing words to increase argument overlap (e.g., replacing *person* and *cases* with *baby(s)*).

Since the actual text was more or less the same in all versions, the same macropropositions could be inferred by an attentive and knowledgeable reader, irrespective of the version read. However, in the high macro-coherence versions of the text, texts **cM** and **CM**, these macropropositions were signalled explicitly by various linguistic means (i.e., macrosignals):

(a) adding topic headers (e.g., *Congenital Heart Disease, Acquired Heart Disease*); and

(b) adding macropropositions serving to link each paragraph to the rest of the text and overall topic (e.g., *There are many kinds of heart disease, some of which are present at birth and some of which are acquired later.*).

The above local and global text manipulations were reversed (e.g., noun phrases were replaced with pronouns, sentence connectives were deleted) to decrease text coherence. The additions specifying the macrostructure are considered macrosignals and were only present in texts **cM** and **CM.** The macropropositions common to all versions of the text correspond to individual heart disorders and treatments.

Prior knowledge questionnaire. Subjects were given a knowledge assessment test to determine individual differences in domain knowledge. We assumed that most readers would have minimal knowledge of the information presented in the text, but might vary in terms of their background knowledge about the human heart. Therefore, the test assessed basic level understanding of the structure and function of the heart, rather than knowledge about the text topic. The test consisted of three sections that were completed in order and removed before the next was presented. The first section required the subject to list as many parts of the human heart as possible and subsequently to describe the function of each part listed. The second section was a recognition test that required the subject to match parts of the heart to a diagram showing a cross section of the heart and its major blood vessels. There were eight terms listed which were actual parts of the heart and two distractor terms. The third section consisted of six fill-in-the-blank and six short-answer questions about the structure and function of the heart.

Posttest questions. Subjects answered two sets of short-answer questions about the content of the text (heart diseases and their treatments). Altogether there were 41 questions which were classified into four different types:

(a) text-based questions (n = 9) for which the necessary information to answer the question was stated within a single sentence of the minimally coherent (**cm**) text (e.g., *What is the main function of the coronary arteries?*);

(b) elaborative inference questions (n = 13) which required linking text information and information from outside knowledge to answer the question (e.g., *Explain why arrhythmia would be a threat to life.*);

(c) bridging inference questions (n = 9) for which the information occurred in the text, but required linking two or more sentences to answer the question (e.g., *Which disease causes scarring of the heart valves?*); and

(d) problem solving questions (n = 10) which required linking information from separate sentences within the text and applying this information to a novel situation (e.g., *Although he is only 28 years old, KT is experiencing heart failure from a disease that is gradually destroying the heart muscle. What kind of treatment would help this patient? Explain your answer.*).

Questions of type (a) can be answered on the basis of the textbase; type (b) requires, in addition, some outside knowledge but not a very specific situation model; answers to types (c) and (d), in contrast, depend on a well-formed situation model.

The 41 questions were divided into two sets with the constraints that the questions in each set be matched as well as possible for difficulty, content matter, and number of possible points. Half of the subjects in each group received one set of questions as the posttest after the first text reading and the other as the posttest after the second reading; the remaining subjects were given the tests in the reverse order. For each set, there were two random orders of questions.

Sorting task. Subjects received one set of cards to sort. The set of cards consisted of 18 concepts each printed on a separate card. There were 8 non-text items and 10 concepts which were keywords from the text. The 10 text items consisted of 3 concepts related to congenital heart disease (i.e., *blue baby, septal defect, carbon dioxide*) and 7 concepts related to acquired heart disease (i.e., *rheumatic fever, streptococci, blood clot, coronary thrombosis, by-pass surgery, arrhythmia, pacemaker*). Of the 8 non-text concepts, 3 referred to parts of the human heart (i.e., *ventricle, mitral valve, pulmonary vein),* 2 were other body organs (i.e., *thyroid, kidney)* and 3 were related to diseases not specific to the heart (i.e., *cancer, multiple sclerosis, malignant*). Our motivation for selecting these items was to provide a group of concepts for which there were several rational sorting principles, but also clearly discernible, text-driven sorting principles.

Procedure

The general procedure and the instructions to the subjects for each of the tasks were identical to those in Experiment 1. The order of the experimental tasks was as follows: (a) sorting task, (b) prior knowledge questionnaire, (c) text reading, (d) text recall, (e) posttest questions, (f) second text reading, (g) second text recall, (h) second set of posttest questions, and (i) sorting task. Subjects read the text and completed the comprehension measures twice in order to ensure as complete an understanding of the text as possible. Reading and task completion times were recorded with a stopwatch. The session lasted approximately an hour and a half.

Results

Prior knowledge questionnaire

The purpose of this measure was to classify subjects as high- or low-knowledge on the basis of a median split procedure on these scores. There was a total of 41 possible points on the knowledge test, 1 point for each correct piece of information supplied. The responses were scored for percent correct.

There were no significant differences among the subjects assigned to the four different text conditions, neither in terms of proportion correct (M = .34; $F(3,52) < 1$) nor in terms of the amount of time to complete the test (M = 12.84 min; $F(3,52) < 3$). The scores for high-knowledge (M = .48) and low-knowledge (M = .20) subjects differed statistically, $F(1,48) = 152.9$, and there was no interaction between the high- and low-knowledge subjects and the four text conditions, $F(3,48) < 3$. Hence, high- and low-knowledge subjects were fairly equally distributed across conditions.

Although the average ages of the high-knowledge (M = 14.4 years) and low-knowledge (M = 13.5 years) subjects differed statistically, $F(1,48) = 6.1$, there was no difference in age between text conditions and there was no interaction between text condition and knowledge group for age, both $F(3,48) < 1$. Hence, the older and younger subjects were equally distributed across conditions. There was no difference in grade between high- and low-knowledge subjects, the average grade level being 7.5 for high- and 7.6 for low-knowledge subjects. Furthermore, grade levels did not differ among texts, nor was there a grade by text interaction (all Fs < 1). Thus, the large difference in knowledge between the high- and low-knowledge subjects was not due to length of schooling and could only in part be due to the relatively small (1 year) difference in age.

Analyses

In order to generalize over both subjects and items, we conducted all analyses both by subjects and by item. However, our primary concern is the reliability of the by-subject analyses, and subsequently, the substantiation of these analyses with the by-item analyses. Therefore, only those F values reliable at $p < .05$ according to the by-subject analyses are reported. Planned comparisons and mean squared errors (MSe) are reported only for the by-subject analyses. Unless otherwise stated, all reported Fs are reliable at $p < .05$.

ANOVAs including the three between-subjects (within-items) factors of local coherence (high, low), global coherence (present, absent), and knowledge (high, low, by median split on the pretest scores), were conducted on the three knowledge assessment measures (recall, questions, sorting). Our local and global text manipulations were orthogonal, allowing us to conduct our analyses including these two factors. For the factor local coherence, high local coherence includes the texts **CM** and **Cm**, and low local coherence refers to the texts **cm** and **cM**. Likewise, the factor global coherence refers to the presence (texts **CM** and **cM**) or absence (texts **Cm** and **cm**) of macrosignals. Other factors were included in the analyses when appropriate.

Reading rates

Reading time for the entire text was measured with a stopwatch. The average reading rates for the **CM**, **cM**, **Cm**, and **cm** texts were 92, 87, 81 and 80 words per minute, respectively. Although the absence of an explicit macrostructure in a text apparently slowed subjects down, the factor global coherence was not significant statistically, $F(1,48) < 2$. In addition, high- and low-knowledge subjects read the texts at about the same rate, F(1,48) < 1. However, there was a significant interaction between global coherence and knowledge, F(1,48) = 8.2, reflecting the finding that low-knowledge subjects read the text more quickly when macrosignals were present (M = 94 wpm) than when absent (M = 70 wpm), but the high-knowledge subjects read the text more slowly when macrosignals were present (M = 85 wpm) than when absent (M = 92 wpm).

Text recall

The four text versions were propositionalized using the same method adopted in Experiment 1. In order to compare recall for the different text versions, the analysis includes only those propositions containing information common to all four texts; that is, those comprising the **cm** text. This scoring method allows by-item analyses to be performed because the propositions that are scored remain the same for all subjects regardless of text. There were 63 micropropositions and 7 macropropositions common to all texts.

Subjects recalled the text twice, once after the first reading and again after reading the text a second time. The two protocols for each subject were pooled and scored collectively. Thus, a composite recall was formed of the propositions provided in the first recall together with any additional (non-repeated) propositions that occurred in the second recall.

ANOVAs by subjects and by items were performed on proportional recall including the factors local coherence, global coherence, knowledge, and proposition type (macroproposition, microproposition). Subjects reproduced all texts quite well. Overall, more than a third of the text propositions were recalled, which, for texts of this length, is quite high. The least coherent text, **cm**, was recalled worst (M = .35), while the other three texts were recalled about equally well (M = .43). Neither these differences nor the overall differences in the recall of high- and low-knowledge subjects were statistically reliable (all Fs < 2). However, the triple interaction shown in Figure 3 was marginally reliable by subjects, F(1,48) = 3.0, MSe = .038, p < .09, and significant by items F(1,68) = 6.9. For high-knowledge subjects, coherence, either at the local or at the global level, made very little difference (F(1,24) < 1); their recall was above 40% for all versions of the text. Apparently, they were able to construct a good textbase with or without the help of explicit linguistic signals. In contrast to the high-knowledge subjects, low-knowledge subjects seemed to form a good textbase with a text that had coherence added at one level or another (**Cm** and **cM**) or at both levels (**CM**). However, as shown by the significantly lower recall of the **cm** text for low-knowledge subjects (F(1,24) = 4.5), they could not do so when they had to deal with coherence gaps at both levels.

Figure 3

The main effect of proposition type indicates that macropropositions were recalled better (M = .43) than micropropositions (M = .38) (F(1,48) = 4.7, MSe = .013, by subjects; F(1,68) < 1, by-items). This difference in recall, however, depended on the global coherence of the text (F(1,48) = 7.3, by subjects; F(1,68) = 8.2, by items). For texts containing macrosignals (**CM** and **cM**), macropropositions (M = .48) were recalled better than micropropositions (M = .38). However, for texts without macrosignals (**Cm** and **cm**), there was no difference in the recall of macropropositions (M = .39) and micropropositions (M = .40). The triple interaction indicates that this effect also depended upon local coherence (F(1,48) = 4.5, by subjects; F(1,68) = 3.9, by items). As is apparent in Table 1, there was very little difference in the recall of micropropositions across the four texts, whereas the three texts with either local or global coherence or both yielded higher macroproposition recall than the minimally coherent text.

Table 1

Posttest questions

Subjects answered a set of questions after each of the two readings of the text. Because there were no differences between the times to complete the first or second posttest questionnaire, these times were combined.

There were no significant differences between the four text conditions in terms of the total amount of time spent answering questions (M = 25 min). However, high-knowledge readers took longer to answer the questions (M = 26 min) than low-knowledge readers (M = 23 min), F(1,48) = 4.7, probably because they simply had more to say.

Table 2

To score the posttest questions, 1 point was awarded for each piece of targeted information. As some questions targeted more than one piece of information, individual questions varied from 1 to 2 points. There was a total of 27 possible points possible for each set of questions. The questions were scored for percent correct. There was no difference between the scores for the first and second postreading tests; these scores were therefore combined for each subject. ANOVAs, both by subjects and items, were performed on the postreading question percentage correct scores, with the factors of local coherence, global coherence, knowledge, and question type (text-based, bridging, elaborative, problem solving). These results are presented in Table 2.

The main effects of local and global coherence were unreliable according to both the by-subjects and by-items analyses. There was, however, a main effect of knowledge (F(1,48) = 11.9, MSe = .086, by subjects; F(1,37) = 61.8, by items); high-knowledge subjects scored better on the posttest questions (M = .49) than low-knowledge subjects (M = .35).

Question type yielded a significant main effect by subjects, F(3,144) = 9.3, MSe = .020, reflecting the finding that more of the elaborative questions were answered correctly (M = .48) than bridging (M = .44), text-based (M = .43), and problem solving questions (M = .34), F(1,48) = 25.0, and that accuracy was lowest for the problem solving questions, F(1,48) = 17.2. These differences were not, however, significant by items, F(3,37) < 1. The interaction between question type and global coherence was reliable according to both the by-subjects analysis, F(3,144) = 3.2, and the by-items analysis, F(3,37) = 3.1. Subjects who had read texts with an explicit macrostructure answered text-based questions more accurately (M = .49) than subjects who read low global coherence texts (M = .37), whereas global coherence did not affect performance on the three other types of questions (low global coherence, M = .42; high global coherence, M = .41), F(1,48) = 6.4. No other factors reliably interacted with question type.

The predicted triple interaction between knowledge, local coherence, and global coherence was reliable by subjects, F(1,48) = 6.4, MSe = .086, and by items, F(1,37) = 42.7. (The three two-way interactions were not reliable by subjects.) The proportion correct averaged across the four question types presented in Table 3 shows the triple interaction. This interaction reflects the finding that high-knowledge subjects performed better on the questions with the **cm** and **CM** texts (M = .58) compared to the **Cm** and **cM** texts (M = .40), F(1,24) = 9.9, and better on the **cm** text overall, F(1,24) = 6.0. On the other hand, although low-knowledge subjects appeared to perform worse on the questions with the **cm** text, this difference (F(1,24) = 2.9) was not reliable for these subjects.

Of primary interest was the interaction between knowledge and text coherence between the maximally (**CM**) and minimally (**cm**) coherent texts for each of the four question types. First, we predicted that low-knowledge subjects would consistently perform better with the coherent text. This prediction is consistent with previous findings and our first experiment. Second, we predicted that the high-knowledge subjects would perform better on the inference and problem solving questions with the **cm** text, and better on the text-based questions with the **CM** text. This prediction was based on the assumption that the text with coherence gaps would force the high-knowledge readers to engage in active processing leading to a better situation model of the text information, and furthermore, that answering the bridging inference and problem solving questions relied on this situation model. To test this prediction, we conducted separate ANOVAs including the two between-subject factors of text (**CM, cm**) and knowledge on each of the four question types. These results are shown in Figure 4.

Figure 4

As predicted, the interaction between text and knowledge was reliable for both the problem solving questions, $F(1,24) = 5.5$, $MSe = .025$, and the bridging inference questions, $F(1,24) = 4.8$, $MSe = .030$. High-knowledge readers performed better on these questions with the low-coherence text, while the opposite was the case for low-knowledge readers. This interaction was not, however, reliable for the elaborative inference questions, $F(1,24) < 2$. Also as predicted, the pattern of results for the text-based questions mirrors that of the recall data: both high and low-knowledge subjects did somewhat better with the coherent text ($F(1,24) = 2.8$, $p = .11$).

Sorting

The sorting data are used to determine how strong an effect reading the text had on the reader's conceptual structure concerning heart diseases. We are interested not in how well or reasonably subjects sort the items, but in the degree to which the information presented in the text influences their sorting. Therefore, as in Experiment 1, the subject's sorting score is a measure of harmony between the subject's sorting matrix and a weight matrix indicative of an ideal sorting. Specifically, the sorting score is the sum of the inner product between the subject's sorting matrix and an ideal sorting matrix. This sum is divided by the total of the positive values in the matrix, in this case 25 (i.e., 12.5 * 2) so that the sort score varied between +1.0 and -1.0. For the ideal sorting matrix we assumed that the most credit should be given when subjects sort into the same category those items that were closely related in the text. Specifically, the following five sets of items were assigned a weight of 1: (a) blue baby, septal defect, and carbon dioxide; (b) rheumatic fever and streptococci; (c) blood clot, coronary thrombosis, and by-pass surgery; (d) arrhythmia and pacemaker; and (e) by-pass surgery and pacemaker). Less credit (a weight of .5) was given for non-text items that were correctly categorized: (e) ventricle, mitral valve, and pulmonary vein; (f) thyroid, and kidney; and (g) cancer, multiple sclerosis, and malignant. As in Experiment 1, all other sorts were assigned a negative value in the ideal sorting matrix such that the sum of the weight matrix was zero.

An ANOVA was performed on the subjects' sorting scores with the between-subject factors of local coherence, global coherence, and knowledge, and the within-subject factors of test and item type (text, non-text). The results of this analysis are shown in Table 3.

Table 3

The main effects of, and interactions between, the three between-subject factors (i.e., local coherence, global coherence, knowledge) were unreliable according to the by-subjects analysis. A main effect of item type (F(1,48) = 30.5, MSe = .012, p < .001, by subjects; F(1,16) = 2.1, NS, by items), reflecting higher scores for text items (M = .164) than for non-text items (M = .085), is basically uninteresting because it reflects at least partially our scoring system which weighted text items twice as much as non-text items. More importantly, item type reliably interacted with test (F(1,48 = 18.3, MSe = .007, by subjects; F(1,16) = 33.2, by items; both p < .001), reflecting greater change for text items (M = .105) than for non-text items (M = .013). Item type did not interact with any of the between-subjects factors; therefore, this difference was relatively constant across text and knowledge conditions.

There was a main effect of test (F(1,48) = 38.1, MSe = .005, by subjects; F(1,16) = 54.2, by items; both p < .001), reflecting the overall change in sorting patterns from pretest (M = .095) to posttest (M = .154) in the direction of the text structure. Most crucially, the predicted four-way interaction of test, local coherence, global coherence, and knowledge was reliable (F(1,48 = 8.1, by subjects; F(1,16) = 8.9, by items; both p < .01). This four-way interaction with test is qualified by reliable two-way interactions with local coherence (F(1,48 = 7.7, by subjects; F(1,16) = 14.4, by items; both p < .01) and global coherence (F(1,48 = 5.7, by subjects; F(1,16) = 8.0, by items; both p < .02), and a three-way interaction with local coherence and knowledge (F(1,48 = 8.7, by subjects; F(1,16) = 20.7, by items; both p < .01). As shown by the overall change scores in Table 4, the **cm** text was most effective in changing the sorting patterns of the high-knowledge subjects, whereas the two locally coherent texts (**CM** and **Cm**) were more effective for the low-knowledge subjects.

Our primary concern is with the difference between the maximally and minimally coherent texts. Thus, we conducted a separate analysis of the **CM** and **cm** texts including the between-subject factors of text (**CM**, **cm**) and knowledge, and the within-subject factors of test and item type. There was a main effect of knowledge, F(1,24) = 6.3, MSe = .038, reflecting higher overall sorting scores for the high-knowledge subjects (M = .336) than for the low-knowledge subjects (M = .205). The overall difference between high and low-knowledge subjects depended on text condition, F(1,24) = 4.8, whereby only a 0.02 difference in sorting scores occurred between knowledge groups for the high coherence text, and a 0.24 difference for the low coherence text. There was a main effect of test, F(1,24) = 27.1, MSe = .008, reflecting the overall change in scores from pretest (M = .208) to posttest (M = .333), and a main effect of item type, F(1,24) = 14.2, MSe = .006, reflecting the higher scores for the text items (M = .17) than for the non-text items (M = .10).

Figure 5

We predicted that low-knowledge subjects would change their sorting patterns more toward the text structure after reading the **CM** text, whereas the high-knowledge subjects would change more with the **cm** text. As shown in Figure 5 in terms of change scores, this prediction was confirmed by a reliable three-way interaction of text, knowledge, and test, F(1,24) = 8.7, MSe = .004.

Discussion

The results of the sorting analysis and the inference question data support the hypothesis that a text which requires gap-filling inferences is beneficial for learning, provided the learner has an appropriate knowledge background. Otherwise, the learner needs a fully coherent, fully explicit text. Figure 4 shows that high-knowledge subjects who read a minimally coherent text did best on questions requiring bridging inferences or problem solving. Having to infer the local coherence of the text and having to construct a macrostructure for the text on their own yielded greater conceptual change for these subjects than when all the necessary cues for the coherence of the text and its macrostructure were explicitly provided by the text. Similarly, with regard to the sorting task, Figure 5 shows that the **cm** text, which required the most active processing, produced the largest change in the high-knowledge subjects' sorting behavior towards congruence with the text structure.

For subjects with a poor knowledge background, this pattern of results is reversed. The fully explicit text is most effective for both textbase and situation model measures. This finding agrees with the prediction that subjects who are unable to do more inferential processing on their own, because they lack the necessary knowledge, require a text that does this processing for them. These data replicate those of Experiment 1, as well as the earlier results obtained by Beyer (1987, 1990) and Britton and Gulgoz (1991), and suggest that the subjects in those experiments were primarily low-knowledge subjects.

We do not claim, of course, that no active processing takes place while reading the coherent text. All readers are necessarily processing the information in the text at some level or another. However, our results do indicate differences in the nature of processing by high- and low-knowledge readers as a function of text type. The benefits of active processing that we are referring to have to do with generating information from the knowledge base and adding it to the information in the text. High-knowledge readers undoubtedly are capable of active processing even when a text is easy to understand, and may very well do so in some situations. Nevertheless, the present data indicate that they are less likely to do so with material that does not pose a challenge.

The absence of both local and global coherence of a text seem to be important in stimulating active processing by the high-knowledge subjects. For both the question data and the sorting data, the active processing advantage for these subjects was obtained only with the low-coherence, implicit-macrostructure text (**cm**). For the low-knowledge subjects, explicitness at either the local or global level aided their performance on the questions, while local coherence was clearly the deciding factor for the sorting task. The low-coherence, explicit-macrostructure text (**cM**) was least effective in the sorting task for both groups of subjects, and for problem solving questions for the high-knowledge subjects. Hence, a text that is overly supportive at the macro level, while at the same time difficult at the local level, may induce a false sense of understanding which masks deficiencies in the construction of an adequate situation model. However, this speculation is tempered by the consideration that the **cM** text was not particularly bad for the other types of questions. The question whether readers become more aware of their own comprehension processes when reading difficult text remains an intriguing question for future research.

The predicted interaction between level of background knowledge and performance was observed in the bridging inference questions, problem solving questions, and the sorting task. These tasks depend primarily upon the adequacy of the situation model constructed by the reader. Low-knowledge readers could not form adequate situation models when the text did not support them, whereas high-knowledge readers actually constructed better situation models when they had to actively process a sub-optimal text.

An interaction with knowledge was also observed in the free-recall data, one of our measures of text memory. However, no advantage for the difficult text was observed among high-knowledge readers. Recall was good among all readers as long as either some sort of coherence (local or global) was present in the text, or subjects had the necessary background knowledge (see Figure 3). Our second measure of text memory, the text-based questions, did not show a reliable interaction with knowledge. Here, maximum coherence clearly provided an advantage for all readers.

In terms of the van Dijk and Kintsch (1983) theory, free recall (or at least its reproductive component), as well as questions specifically directed at text information, depend primarily on the textbase readers construct. Except for the low-knowledge readers who were given the most difficult text, all readers were apparently able to form adequate textbases and recalled the texts quite well. The textbase by its very nature reflects the hierarchical structure of the text. It can, therefore, serve as an efficient retrieval structure, enabling two propositions to be reproduced successfully, regardless of whether the reader understood or remembered the relationship between the two. Thus, reproductive recall was about equal for all versions of the text. On the other hand, to answer an inference question correctly, or to form the correct sorting categories, it is not sufficient to retrieve the relevant propositions. One also needs to know the relation between them, that is, one needs an elaborate situation model that specifies these relationships.

Conclusions

The coherence of a text from which a student is supposed to acquire knowledge is clearly important. Increasing the coherence of a text has been found to improve readers' comprehension by previous investigators (Beck et al., 1991; Beyer, 1987, 1990; Britton & Gulgoz, 1991) as well as in the present study for low-knowledge readers. However, we also found that by giving learners with sufficient background knowledge an incoherent text which forced them to infer unstated relationships, we were able to engage them in compensatory processing at the level of the situation model. This enabled them to understand the text more deeply than if we had given them a more coherent text.

Our finding that it is sometimes better to let readers construct the coherence of a text is not limited to text-based learning. The advantage of active processing has also been observed with other kinds of tasks involving memory, learning and procedural skills. For example, the benefits of generating versus simply reading or hearing the answers are well established in memory tasks (e.g., Slamecka & Graf, 1978) and have recently been demonstrated for learning tasks (McNamara & Healy, in press; but cf. Carroll & Nelson, 1993). Similarly, Battig (1979) observed that increased inter-item interference during learning resulted in better performance in paired-associate learning. Battig's findings led some researchers in the area of skill training to question the long assumed benefits of optimizing performance during training (see, e.g., Healy et al., 1993; Healy & Sinclair, in press; Schmidt & Bjork, 1992, for reviews of this literature). These researchers have shown that facilitating the acquisition of skills or knowledge can impede their long-term retention and transfer. Moreover, long-term retention can be enhanced with training methods that slow down acquisition by making the learning process more difficult and thereby increase the learner's active engagement. In sum, what appears to be an optimal training strategy in the short-run is often detrimental in the long-run. Likewise, in the present study we have found that what appears to be an optimal learning tool at one level can be detrimental at another. We have shown that facilitating comprehension by increasing the coherence of a text can result in a superficial level of comprehension that supersedes a deeper understanding of the text.

With respect to reading, several studies have demonstrated advantages for text-induced active processing (e.g., E. Kintsch, 1990; Mannes & Kintsch, 1987). Even degraded text (e.g., with letters deleted or sentences scrambled) has been found to improve text comprehension under particular conditions (e.g., Einstein, McDaniel, Owen, & Coté, 1990; McDaniel, Einstein, Dunay, & Cobb, 1986; O'Brien & Myers, 1985) and has also been found to increase certain types of inferential processes (Keefe & McDaniel, in press). In an educational application, Rauenbusch and Bereiter (1991) used text with letters deleted from each word as a context for learning and practicing meaning-based reading strategies. When combined with explicit instruction, the students were able to transfer these strategies to reading normal texts. Our findings suggest, however, that the success of this type of approach in further educational applications will require a clearer understanding of how the knowledge level of the reader interacts with the difficulty of the task and the type of comprehension processing engaged in (see also, McKeown, Beck, Sinatra, & Loxterman, 1992; Roller, 1990).

According to the results of our second experiment, constructing coherence without explicit textual support is only possible when students possess adequate background knowledge. The boundary conditions for this effect clearly merit further investigation. The better performance of the high-knowledge subjects with the low-coherence text may be limited to materials of just the right degree of difficulty. If the text is too hard, the construction of coherence may fail; on the other hand, if the text is too easy, the gains through more active processing may be too small (see, e.g., Einstein et al., 1990; O'Brien & Myers, 1985; Keenan, Baillet, & Brown, 1984; Meyers, Shinjo, & Duffy, 1987, for similar arguments). In addition, findings regarding the generation effect in skill and knowledge acquisition likewise indicate that a reader's familiarity with a particular domain is an important consideration: In a learning paradigm, McNamara and Healy (in press) found no advantages for generating when the process was relatively automatic. With respect to the present study, we assume that our readers are acquiring new information from the texts we present to them. That is, at least some of the information is unfamiliar to them, as would be the case for most students reading an instructional text. However, it is possible that many subjects (especially older ones) would be quite familiar with the information about heart disease presented in our text. For such a subject our low-coherence text might appear fully coherent; that is, it would be processed as easily as a fully coherent text because the gap-filling inferences would be automatically generated. In this situation the advantages we found for constructing coherence would presumably disappear.

In short, there is no easy solution to the problem of promoting learning from text, not only due to the complexity of these interactions, but also because comprehending text is not a unitary process. Comprehension may not succeed equally well at all levels, such as the textbase or the situation model. Disproportionate comprehension at different levels of understanding may complicate the student's self-monitoring of comprehension. If comprehension proceeds well at the level of the textbase (e.g., for a high-knowledge reader with a coherent text), the student may register satisfactory progress and thus fail to work harder for a deeper understanding. Our findings suggest that the imperfectly written text, on the other hand, may prevent such self-deception, and hence give rise to constructive processes which result in better comprehension at the level of the situation model.

The implications of findings such as ours, showing benefits from difficult reading, are of educational as well as general theoretical interest. With respect to educational applications, our findings suggest an approach in which the coherence level of the text is adjusted to the student's level of knowledge, such that reading becomes challenging enough to stimulate active processing, but not too difficult so as to break down comprehension. Of course, this would mean constructing several versions of a text in order to accommodate varying levels of knowledge among readers. The idea of "customizing" a textbook is not as impractical as it might seem. For example, textbook publishers now provide instructional texts which are a composite of particular subject areas requested by individual teachers. Moreover, the kind of educational application of customized text that we have in mind is easily within the capability of present-day hypertext computer systems. Text could be presented on a computer screen with interspersed questions or tasks designed to assess a student's comprehension on-line. Instructional text could then be presented at the level of coherence that is appropriate to the student's current level of understanding -- such that it encourages inferencing but also ensures that the reader is indeed able to do so. In this way students are forced to use their knowledge as they read, allowing effective learning from a textbook to be achieved by a much wider range of students than is possible with a single textbook, targeted at a supposed average reader.

Additionally, the notions explored here concerning text coherence, background knowledge, and levels of understanding in text comprehension may be useful in directing attempts to formulate guidelines for textbook writers, editors and lecturers. Tuning the level of detail and coherence to the audience is implicitly expected of writers. Often, however, writers underestimate their audience and end up with an overly explicit text. What we have demonstrated here is that overly explicit writing is not only annoying, it can actually prevent the audience from actively processing what we have to say.

Finally, it is important to see our results within the broader context of other findings on learning and skill acquisition, such as those reported by Schmidt and Bjork (1992). Taken together, these studies imply that for both simple skills and higher-order cognitive processes there is no simple relation between ease of learning and final performance. Focusing on ease and speed of learning, as training programs have been doing in practice, with the approval of learning theorists, is not necessarily optimal. Instead, the optimal training program will have to be considerably more complex, introducing difficulties and stumbling blocks into the learning process, but introducing them in carefully considered and controlled ways. In the present paper we not only provide an empirical basis for this claim, but also a theoretical framework within which this phenomenon can be understood and explored further, at least for the case of learning from text.

References

Battig, W. F. (1975) The flexibility of human memory. In L. S. Cermak & F. I. M. Craik (Eds.) *Levels of processing in human memory* (pp 23-44). Hillsdale, NJ: Erlbaum.

Beck, I. L., McKeown, M. G., Sinatra, G. M., & Loxterman, J. A. (1991). Revising social studies text from a text-processing perspective: Evidence of improved comprehensibility. *Reading Research Quarterly, 27*, 251-276.

Beyer, R. (1987) Psychologische Untersuchungen zur Textverarbeitung unter besonderer Berücksichtigung des Modells von Kintsch und v. Dijk (1978). *Zeitschrift für Psychologie*, Supplement 8.

Beyer, R. (1990) Psychologische Untersuchungen zur Gestaltung von Instruktionstexten. *Wissenschaftliche Zeitschrift der Humboldt-Universität zu Berlin*, Mathematisch-Naturwissenschaftliche Reihe 39, 1, 69-75.

Bereiter, C., & Scardamalia, M. (1989). Intentional learning as a goal of instruction. In L. B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser* (pp. 361-392). Hillsdale, NJ: Erlbaum.

Britton, B. K., & Gulgoz, S. (1991). Using Kintsch's model to improve instructional text: Effects of inference calls on recall and cognitive structures. *Journal of Educational Psychology, 83*, 329-345.

Carroll, M. & Nelson, T.O. (1993). Failure to obtain a generation effect during naturalistic learning. *Memory & Cognition, 21*, 361-366.

Chi, M. T. H., & Bassok, M. (1989). Learning from examples via self-explanations. In L. B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser* (pp. 251-282). Hillsdale, NJ: Erlbaum.

Einstein, G. O., McDaniel, M. A., Owen, P. D., & Coté, N. C. (1990). Encoding and recall of texts: The importance of material appropriate processing. *Journal of Memory and Language, 29*, 566-581.

Healy, A. F., Clawson, D. M., McNamara, D. S., Marmie, W. R., Schneider, V. I., Rickard, T. C., Crutcher, R. J., King, C., Ericsson, K. A., & Bourne, L. E., Jr. (1993). The long-term retention of knowledge and skills. In D. Medin (Ed.), *The psychology of learning and motivation* (pp. 135-164). New York: Academic Press.

Healy, A. F., & Sinclair, G. P. (in press). Long-term retention of training and knowledge. In E. Bjork & R. Bjork (Eds.), *The handbook of perception and cognition, Volume 10* (Memory). New York: Academic Press.

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

Keefe, D. E., & McDaniel, M. A. (in press). The time course and durability of predictive inferences. *Journal of Memory and Language.*

Keenan, J. M., Baillet, S. D., & Brown, P. (1984). The effects of causal cohesion on comprehension and memory. *Journal of Verbal Learning and Verbal Behavior, 23*, 115-126.

Kintsch, E. (1990). Macroprocesses and microprocesses in the development of summarization skill. *Cognition and Instruction, 7*, 161-195.

Kintsch, E., McNamara, D. S., Songer, N. B, & Kintsch, W. (1992). *Revising the coherence of science texts to improve comprehension and learning I: Traits of mammals*. ICS Technical Report No. 92-3, Boulder, CO.

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

Kintsch, W. (1992). A cognitive architecture for comprehension. In H. L. Pick, P. van den Broek, & D. C. Knill (Eds.), *The study of cognition: Conceptual and methodological issues* (pp. 143-164). Washington, DC: American Psychological Association.

Kintsch, W. (1994). Text comprehension, memory, and learning. *American Psychologist, 49*, 292-303.

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

Kintsch, W., & Vipond, D. (1979). Reading comprehension and readability in educational practice and psychological theory. In L. G. Nilsson (Ed.), *Perspectives of memory research* (pp. 325-366). Hillsdale, NJ: Erlbaum.

Mannes, S., & Kintsch, W. (1987). Knowledge organization and text organization. *Cognition and Instruction, 4*, 91-115.

McDaniel, M. A., Einstein, G. O., Dunay, P. K., & Cobb, R. E. (1986). Encoding difficulty and memory: Toward a unifying theory. *Journal of Memory and Language, 25*, 645-656.

McNamara, D. S., & Healy, A. F. (in press). A generation advantage for multiplication skill and nonword vocabulary acquisition. In A. F Healy & L. E. Bourne, Jr. (Eds.), *Learning and memory of knowledge and skills*. Newbury Park, CA: Sage.

McKeown, M. G., Beck, I. L., Sinatra, G. M., & Loxterman, J. A. (1992). The contribution of prior knowledge and coherent text to comprehension. *Reading Research Quarterly, 27*, 79-93.

Miller, J., & Kintsch, W. (1980). Readability and recall of short term prose passages: A theoretical analysis. *Journal of Experimental Psychology: Human Learning and Memory, 6*, 335-354.

Myers, J. L., Shinjo, M, & Duffy, S. A. (1987). Degree of causal relatedness and memory. *Journal of Memory and Language, 26*, 453-465.

O'Brien, E. J., & Myers, J. L. (1985). When comprehension difficulty improves memory for text. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*, 12-21.

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

Raintree Illustrated Science Encyclopedia. (1984). Milwaukee, Wisconsin: Raintree Publishers Inc. (vol. 8, pp. 767-768).

Rauenbusch, F., & Bereiter, C. (1991). Making reading more difficult: A degraded text microworld for teaching reading comprehension strategies. *Cognition and Instruction*, *8*(2), 181-206.

Roller, C. M. (1990). Commentary: The interaction of knowledge and structure variables in the processing of expository prose. *Reading Research Quarterly*, *25*(2), 79-89.

Sanford, A. J., & Garrod, S. C. (1981). *Understanding written language*. London: Wiley.

Schmalhofer, F., & Glavanov, D. (1986). Three components of understanding a programmer's manual: Verbatim, propositional, and situational representations. *Journal of Memory and Language*, *25*, 279-294.

Silver Burdett & Ginn. (1987). *Life Science* (pp. 300-302). Morristown, NJ: Author.

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning & Memory*, *4*, 592-604.

Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science, 3*, 207-217.

Smolensky, P. (1986). Information processing in dynamical systems: Foundations of harmony theory. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing* (vol. 1, pp. 194-281). Cambridge, MA: MIT Press.

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

Author Notes

This study was funded by a grant to Walter Kintsch from the A. W. Mellon Foundation and a grant to Danielle McNamara from the J. S. McDonnell Foundation. We thank Isabel Beck, Bruce Britton, Art Graesser, Grant Sinclair, Murray Singer, and an anonymous reviewer for helpful comments on earlier versions of this manuscript.

Footnotes

1 The sort score is the sum of the inner product between a subject's sorting matrix (s_{ij}) and a weight matrix (w_{ij}) divided by the sum of the positive weights in the weight matrix ($\Sigma\, |w_{ij}| / 2$): $((\Sigma(s_{ij} \bullet w_{ij})) / (\Sigma\, |w_{ij}| / 2))$. Specifically, a subject's sorting matrix consists of 1s and 0s, whereby a 1 indicates a pair of items sorted together and a 0 indicates a pair not sorted together. The weight matrix, indicative of an ideal sorting, was constructed by assigning a weight $w_{ij}=1.0$ to cells of the matrix representing pairings between mammal characteristics, pairings between non-mammal traits, and pairings between shared characteristics. A weight $w_{ij}=0.5$ was assigned to pairings between mammal characteristics and shared characteristics. This yields a total of 102 points for the positive values in the matrix (i.e., 51 * 2). The remaining cells of the weight matrix were assigned negative values in such a way that the sum of the matrix became 0. Specifically, the negative weight was set equal to $-((\Sigma\, |w_{ij}| / 2)/(m-k))$, where $(\Sigma\, |w_{ij}| / 2)$ is the sum of the positive values in the matrix, m is the total number of non-diagonal cells in the matrix, (n^2-n), and k is the number of cells containing positive values. Diagonal cells of the weight matrix were assigned a weight of 0.

Table 1. Proportion of propositions recalled for the four text versions in Experiment 2 for propositions occurring in all texts.

	CM	Cm	cM	cm
Proportion of Propositions Recalled	.43	.43	.42	.35
Micropropositions	.40	.36	.41	.39
Macropropositions	.47	.49	.46	.32

Table 2. Proportion correct responses on the postreading questions for the four text conditions in Experiment 2 by Knowledge and Question Type.

		CM	Cm	cM	cm
High Knowledge					
	Text-based	0.64	0.37	0.47	0.51
	Bridging	0.48	0.43	0.40	0.68
	Elaborative	0.60	0.48	0.44	0.62
	Problem Solving	0.46	0.34	0.27	0.61
	Average	0.55	0.41	0.40	0.61
Low Knowledge					
	Text-based	0.47	0.33	0.38	0.26
	Bridging	0.40	0.38	0.43	0.31
	Elaborative	0.47	0.42	0.42	0.35
	Problem Solving	0.30	0.29	0.28	0.17
	Average	0.41	0.36	0.38	0.27

Table 3. Sorting task scores in Experiment 2 by Text, Knowledge, Item Type, and Test.

			CM	Cm	cM	cm
High Knowledge						
	Text Items					
		Pretest	0.13	0.12	0.14	0.16
		Posttest	0.22	0.23	0.12	0.37
	Non-Text Items					
		Pretest	0.07	0.04	0.12	0.15
		Posttest	0.09	0.04	0.12	0.16
	Overall Change Score		0.11	0.10	-0.01	0.23
Low Knowledge						
	Text Items					
		Pretest	0.10	0.02	0.15	0.09
		Posttest	0.25	0.27	0.18	0.11
	Non-Text Items					
		Pretest	0.06	0.02	0.09	0.08
		Posttest	0.07	0.08	0.10	0.07
	Overall Change Score		0.16	0.32	0.04	0.00

Figure Captions

Figure 1. Propositional and network representations of the two complex propositions in the sentence "Many mammals hibernate because food is often scarce in winter."

Figure 2. The change in the frequency from pretest to posttest with which three different types of items were sorted together with other mammal characteristics for three texts in Experiment 1.

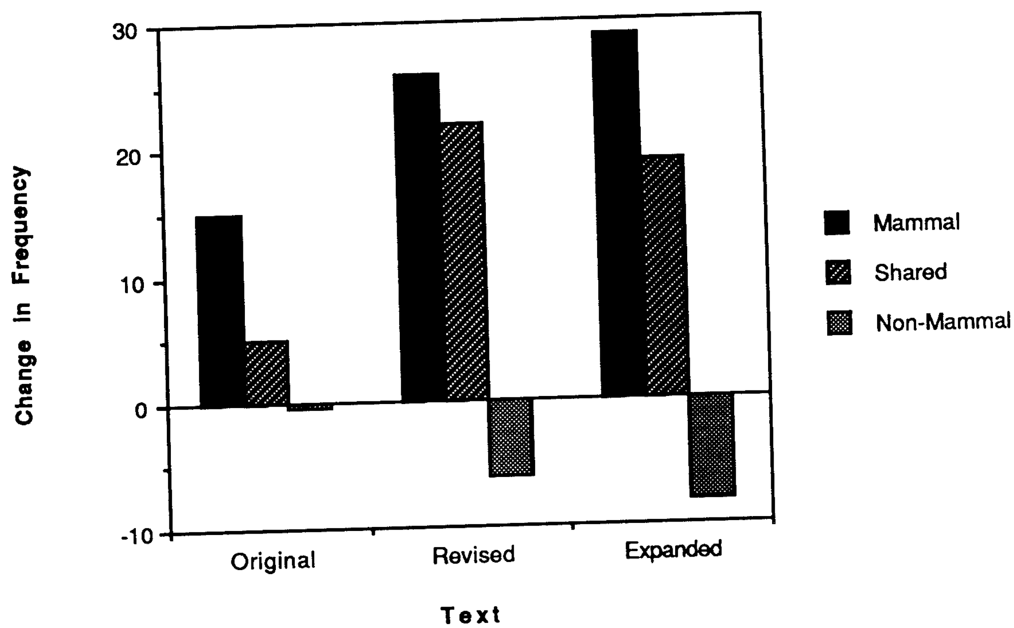
Figure 3. Proportional recall of the text propositions common to all four texts as a function of readers' knowledge in Experiment 2.

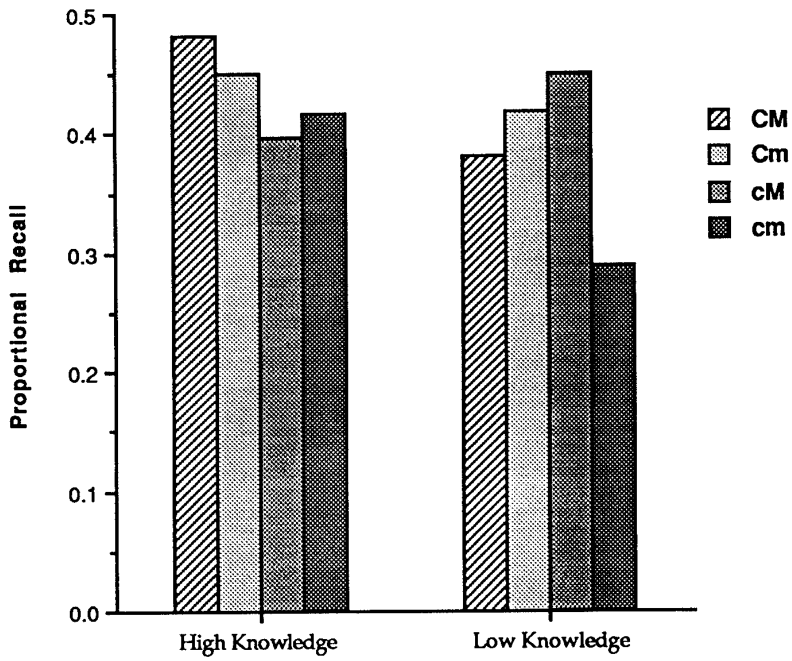
Figure 4. Proportion correct answers to the four types of questions on the postreading test for the maximally (CM) and minimally (cm) coherent texts as a function of readers' background knowledge in Experiment 2.

Figure 5. Change from pretest to posttest in sorting scores for the maximally (CM) and minimally (cm) coherent texts as a function of readers' background knowledge in Experiment 2.

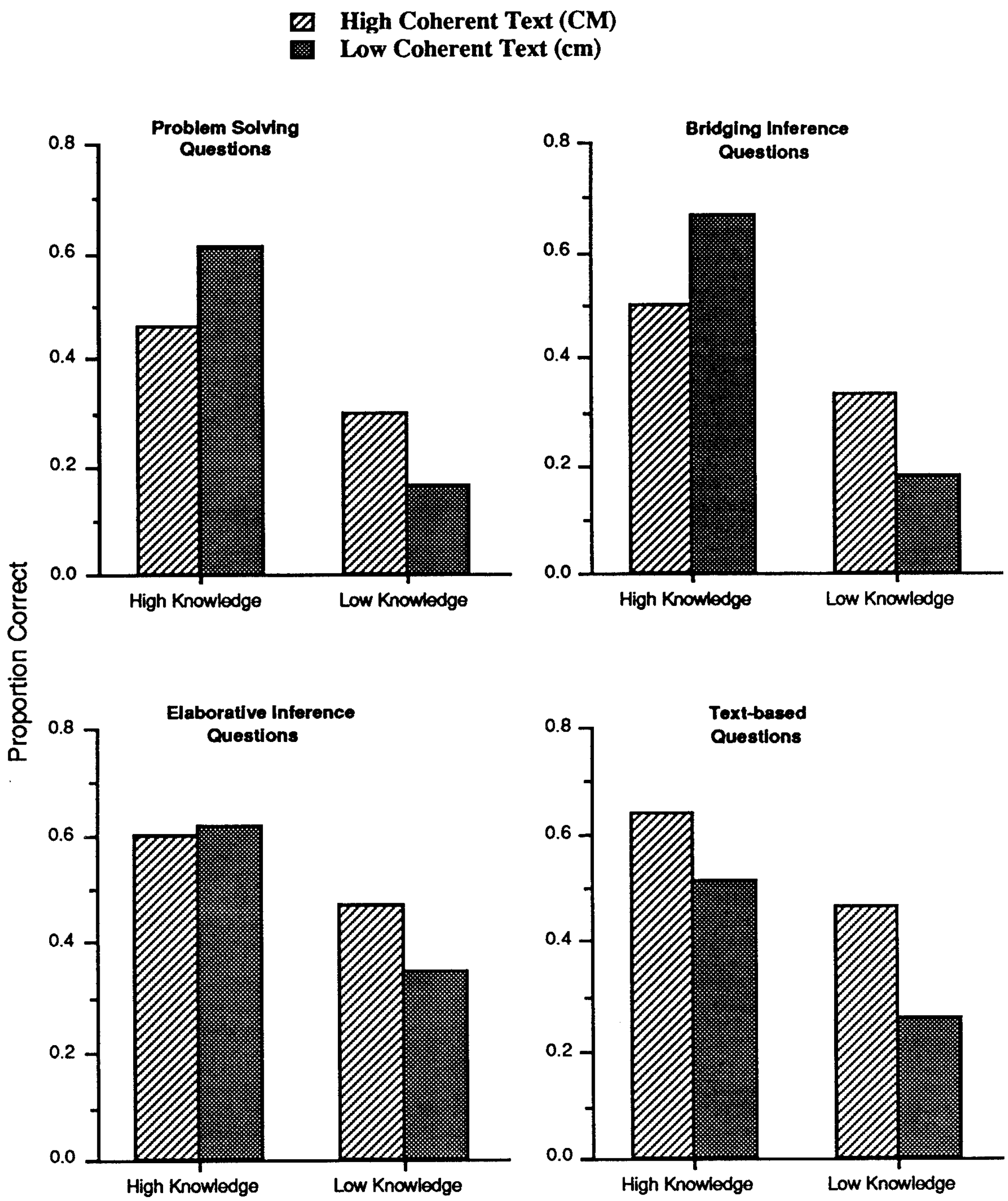
Propositional Representation:

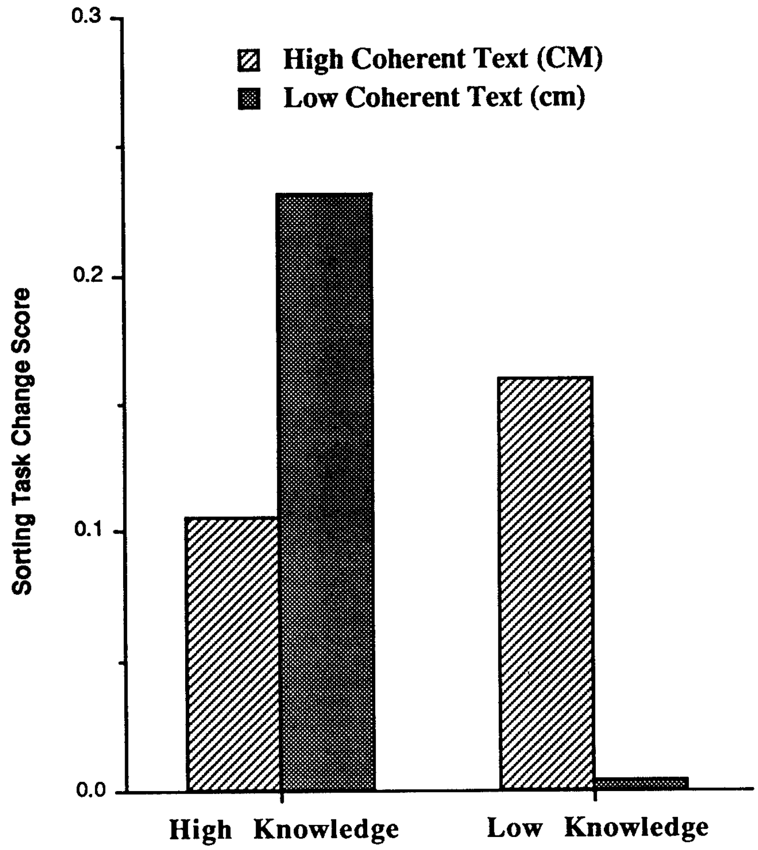
- Complex Proposition
  - Proposition
    - Predicate: **HIBERNATE**
    - Argument
      - Agent: **MAMMALS**
        - Mod: **MANY**
  - Circumstance: --
- **BECAUSE**
- Complex Proposition
  - Proposition
    - Predicate: **SCARCE**
      - Mod: **OFTEN**
    - Argument
      - Object: **FOOD**
  - Circumstance
    - Time: **WINTER**


30
20
10
0
-10
Change in Frequency
Original
Revised
Expanded
Text
Mammal
Shared
Non-Mammal



0.5
0.4
0.3
0.2
0.1
0.0
Proportional Recall
High Knowledge
Low Knowledge
CM
Cm
cM
cm


High Coherent Text (CM)
Low Coherent Text (cm)
Proportion Correct
Problem Solving Questions
Bridging Inference Questions
Elaborative Inference Questions
Text-based Questions
0.0
0.2
0.4
0.6
0.8
High Knowledge
Low Knowledge


0.3
0.2
0.1
0.0
Sorting Task Change Score
High Coherent Text (CM)
Low Coherent Text (cm)
High Knowledge
Low Knowledge

Appendix A

Text excerpts from the original, revised, and expanded versions of the texts used in Experiment 1. The text excerpts are each from the section about mammals' teeth. Text additions are underlined.

Traits of Mammals -- Original Version

Mammals have very specialized teeth. There are four types of teeth in mammals: incisors, canines, premolars, and molars. The number and shape of each of these types of teeth are related to the kind of food the mammal eats. Meat-eating mammals, such as wolves and lions, have long, pointed canine teeth that are used for tearing. Their incisors are chisel-shaped and are used for cutting. Plant-eating mammals, such as horses and cows, have large, flat premolars and molars. These teeth are used for grinding plant materials.

Traits of Mammals -- Revised Version

Another physical trait of mammals is that they can eat many different kinds of food because they have very specialized teeth. This trait also helps them to live in different kinds of environments. There are four types of teeth in mammals: incisors, canines, premolars, and molars. The number and shape of each of these types of teeth are related to the kind of food the mammal eats. Meat-eating mammals, such as wolves and lions, have long, pointed canine teeth, that are used for tearing. Their incisors are chisel-shaped and are used for cutting. Plant-eating mammals, such as horses and cows, have large, flat premolars and molars. These teeth are used for grinding plant materials.

Traits of Mammals -- Expanded Version

Mammals are also successful because over many, many years, they have developed different kinds of specialized teeth. These different teeth allow mammals to eat many different kinds of food. If mammals can eat many different kinds of food, then they will be less likely to die of starvation and become extinct when a change occurs to one of their food sources.

There are four types of teeth in mammals. The number and shape of each of these types of teeth are related to the kind of food the mammal eats. Meat-eating mammals, such as wolves and lions, have long, pointed canine teeth, that are used for cutting. Plant-eating mammals, such as horses and cows, have large, flat premolars and molars. These teeth are used for grinding plant materials. Mammals such as humans have many different kinds of teeth. These help them eat the many different kinds of food in their diets.

Appendix B

Fully coherent and minimally coherent versions of the texts used in Experiment 2. Text additions are underlined or italicized in the fully coherent version: Text which was added to increase coherence at the local level is italicized, and text which was added to increase coherence at the macro-level is underlined.

Heart Disease -- Fully Coherent Version

The heart is the hardest-working organ in the body. We rely on *it to* supply blood regularly *to the body* every moment of every day. Any disorder that stops *the heart from* supplying blood *to the body* is a threat to life. Heart disease *is such a disorder. It* is very common. More people are killed every year in the U.S. by heart disease than by any other disease.

There are many kinds of heart disease, some of which are present at birth and some of which are acquired later.

1. Congenital heart disease

A congenital heart disease is a *defect* that a *baby* is born with. Most babies are born with perfect hearts. *But* one in every 200 *babies is born with a bad heart. For example, hearts have flaps, called valves, that control the blood flow between its chambers.* Sometimes a valve develops the wrong shape. It may be too tight, or fail to close properly, *resulting in congenital heart disease.* Sometimes a gap is left in the wall, *or septum,* between the two sides of the heart. This *congenital heart disease* is often called a "septal defect". When a baby's heart is badly *shaped*, it cannot work efficiently. *It* cannot *pump enough* blood *through the lungs so that it* receives enough oxygen. *As a result*, the baby becomes breathless. The blood *also* cannot get rid of carbon dioxide through the lungs. *Therefore, the blood* becomes purplish, *which causes* the baby's skin *to* look blue. *Fortunately*, it is now possible to save the lives of many "blue babies".

2. Acquired heart disease

Some heart diseases are acquired after the baby is born. Rheumatic fever is *an example of* an *acquired* disease *that* may cause *damage* to the heart. This disease usually follows a sore throat caused by bacteria *known* as streptococci. *This is often called "strep throat". When strep throat causes rheumatic fever*, the tissues of the heart become inflamed. If *the heart* is badly affected, it fails *very soon*. Usually, *however*, it recovers, and the results of the damage are seen only years later. *This is because the rheumatic fever* leaves scars in the valves of the heart. *Therefore*, they cannot work properly. This puts a strain on the heart *so that* eventually it may fail. The effects of the rheumatic fever may take up to twenty or thirty years to appear.

Coronary disease is another example of an acquired heart disease. *This disease affects* the coronary arteries. *These* are the blood vessels that extend across the heart and supply it with blood *from the lungs*. They are very important *because* they give the heart *muscle* the oxygen it needs to carry on working. *In coronary disease* the *coronary arteries* become blocked, *causing* parts of the heart muscle to die *because of the lack of oxygen. When this happens*, the patient has a heart attack, which can be fatal. The blockage of a coronary artery is usually caused by a clot of blood, *called* a "thrombus". When a clot forms in a coronary artery, *this is called* "coronary thrombosis." That is the correct name for a heart attack. In normal arteries, blood does not form clots. But in coronary disease, the walls of the *arteries are not normal. They* become lumpy, rough, and narrow. *The lumps break off and form clots that stop the flow of blood to the heart.*

Other examples of acquired heart disease are arrhythmia, angina, and high blood pressure. Arrhythmia, *which means "lack of rhythm"*, is an interruption of the heart's normal beat. Angina is a sharp pain in the chest which is very similar to that caused by *a heart attack, or* thrombosis. High blood pressure is *one of the most* common *heart diseases. It places a heavy strain on the heart and other organs. Therefore*, if it is not treated, high blood pressure may lead to heart attacks, kidney failure, or other serious problems. High blood pressure *is a disease which* has no symptoms. *Thus*, a person may not be aware of having it unless the blood pressure is measured.

3. Treatment and prevention of heart disease

Since the mid-1960's, medical science has made tremendous progress in the treatment and prevention of heart disease. Both new drugs and new surgical methods have been developed. Among the new drugs for treating heart disease are *chemicals* called "beta blockers". The *beta-blockers* lessen the after-effects of heart attacks; *they* can prevent second attacks; and *they* can lower the blood pressure of people who have high blood pressure. Other drugs dissolve the lumps which break off the walls of arteries *so that they do not stop the flow of blood to the heart.*

Surgical techniques for treating heart disease range from repairing or replacing damaged parts, such as valves or arteries, to replacement of the entire heart. *If a heart has been so damaged that it can no longer function, it can be replaced by* a mechanical heart, or, more often, by a heart transplant. In transplant surgery, the healthy heart of someone who has died replaces the *diseased* heart of the patient. Mechanical devices can be implanted in people's bodies to keep their hearts functioning. The pacemaker *is the most* common *of these devices. It* does not heal the diseased heart, but it relieves the symptoms of an irregular heart *beat* and *maintains* the steady beat *needed* for normal living. When a heart cannot pump enough blood through the lungs because of poorly functioning valves, the valves can be replaced with artificial ones of plastic and metal. *For patients with coronary disease*, "by-pass surgery" is often used to repair clogged or damaged *arteries*. Doctors use pieces of a patient's own veins, often from the leg, to replace the damaged portions of arteries.

Preventive care is also getting better as scientists learn more and more about the causes of heart disease. They have shown that diet can be an important means of controlling heart disease. *For example*, a substance called cholesterol is known to cause a build-up of fatty substances in the blood vessels, which can cause *blood clots to form in the arteries. Therefore*, doctors stress the importance of a diet low in *cholesterol. Similarly,* salt is known to increase blood pressure, so *doctors* recommend a low-salt diet *for patients with high blood pressure.*

Heart Disease -- Minimally Coherent Version

The heart is the hardest-working organ in the body. We rely on a regular blood supply every moment of every day. Any disorder that stops the blood supply is a threat to life. Heart disease is very common. More people are killed every year in the U.S. by heart disease than by any other disease.

A congenital disease is one that a person is born with. Most babies are born with perfect hearts. In about one in every 200 cases something goes wrong. Sometimes a valve develops the wrong shape. It may be too tight, or fail to close properly. Sometimes a gap is left in the septal wall between the two sides of the heart. This is often called a septal defect. When a baby's heart is badly formed, it cannot work efficiently. The blood does not receive enough oxygen. The baby becomes breathless. The blood cannot get rid of carbon dioxide through the lungs. It becomes purplish, and the baby's skin looks blue. It is now possible to save the lives of many blue babies.

The disease called rheumatic fever may cause harm to the heart. The disease usually follows a sore throat caused by bacteria called streptococci. The tissues of the heart become inflamed. If it is badly affected, it fails. Usually it recovers, and the results of the damage are seen only years later. The valves of the heart are left with scars. They cannot work properly. This puts a strain on the heart. Eventually it may fail. The effects of the rheumatic fever may take up to twenty or thirty years to appear.

The blood vessels that extend across the heart and supply it with blood are called the coronary arteries. They are very important. They give the heart the oxygen it needs to carry on working. If they become blocked, parts of the heart muscle will die. The patient has a heart attack, which can be fatal. The blockage of a coronary artery is usually caused by a thrombus, or blood clot. Coronary thrombosis happens when a clot forms in a coronary artery. That is the correct name for a heart attack. In normal arteries, blood does not form clots. In coronary disease, the walls of the blood vessels become lumpy, rough, and narrow.

Arrhythmia is an interruption of the heart's normal beat Angina is a sharp pain in the chest which is very similar to that caused by thrombosis. High blood pressure is very common. If untreated, high blood pressure may lead to heart attacks, kidney failure, or other serious problems. High blood pressure may have no symptoms. A person may not be aware of having it unless the blood pressure is measured.

Among the new drugs for treating heart disease are a family of compounds called beta blocking drugs, or simply, beta blockers. They lessen the after-effects of heart attacks, can prevent second attacks, and can lower the blood pressure of people who have high blood pressure. Other drugs dissolve the lumps which break off the walls of veins and arteries.

Heart transplants are used more often than mechanical hearts. In transplant surgery, the healthy heart of someone who has died replaces the heart of the patient. Mechanical devices can be implanted in people's bodies to keep their hearts functioning. The commonly used pacemaker does not heal the diseased heart, but it relieves the symptoms of an irregular heart and keeps a steady beat for normal living. When a heart cannot pump enough blood through the lungs because of poorly functioning valves, the valves can be replaced with artificial ones of plastic and metal. By-pass surgery is used to repair clogged or damaged blood vessels. Doctors use pieces of a patient's own veins, often from the leg, to replace the damaged portions of arteries.

A substance called cholesterol is known to cause a build-up of fatty substances in the blood vessels, which can lead to heart disease, so doctors stress the importance of a diet low in fats. Salt is known to increase the blood pressure, so a low-salt diet is recommended.

Technical Report 93-04

University of Colorado at Boulder

Institute of Cognitive Science

Muenzinger Psychology Building
Campus Box 344
Boulder, Colorado 80309-0344
(303) 492-5063
FAX: (303) 492-7177
email: cogsci@psych.colorado.edu
URL: http://psychwww.colorado.edu/ics/