UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION: Unclassified	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT: Approved for public release; distribution unlimited.
	5. MONITORING ORGANIZATION REPORT NUMBER(S): AFOSR-TR- 90 1083
6a. NAME OF PERFORMING ORGANIZATION: University of Wisconsin; 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION: Air Force Office of Scientific Research/NL
6c. ADDRESS (City, State and ZIP Code): Department of Psychology, 1202 W. Johnson Street, Madison, WI 53706	7b. ADDRESS (City, State and ZIP Code): Building 410, Bolling Air Force Base, DC 20332-6448
8a. NAME OF FUNDING/SPONSORING ORGANIZATION: Air Force Office of Scientific Research/NL; 8b. OFFICE SYMBOL (If applicable): NL	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER: AFOSR-89-0367
8c. ADDRESS (City, State and ZIP Code): Building 410, Bolling Air Force Base, DC 20332-6448	10. SOURCE OF FUNDING NOS.

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT NO
61102F	2313	A4	

11. TITLE (Include Security Classification): Comprehension of Illustrated Text

12. PERSONAL AUTHOR(S): Arthur M. Glenberg and William E. Langston

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Yr., Mo., Day)	15. PAGE COUNT
Annual Technical Report	FROM 7/1/89 TO 6/30/90	9/18/90	41

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB. GR.	
05	09		Reading; Comprehension; Pictures; Mental Models

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

Pictures help people to comprehend and remember texts. We report two experiments designed to test among several accounts of this facilitation. Students read texts describing four-step procedures in which the middle steps were described as occurring at the same time, although the verbal description of the steps was sequential. A mental representation of the *procedure* would have the middle steps equally strongly related to the preceding and succeeding steps (because the steps are performed simultaneously), whereas a mental representation of the *text* would have the middle step that was described first more closely related to the preceding step than the middle step described second. After reading, strengths of the represented relationships between the steps were assessed. When the texts were accompanied by appropriate pictures, subjects tended to mentally represent the procedure. When the texts were presented alone or with pictures illustrating the order in which the steps were described in the text, subjects tended to mentally represent the text. We argue that these results disconfirm motivational, repetition, and dual code explanations

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT: UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☒ DTIC USERS ☐	21. ABSTRACT SECURITY CLASSIFICATION: Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL: Dr. Alfred R. Fregly	22b. TELEPHONE NUMBER (Include Area Code): (202)767-5021; 22c. OFFICE SYMBOL: NL

DD FORM 1473, 83 APR — EDITION OF 1 JAN 73 IS OBSOLETE.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

3 OCT 1990

of the facilitative effects of pictures. The results are consistent with a version of mental model theory that proposes that pictures help to build mental models of what the text is about. Keywords: Memory psychology; Reading comprehension; Teaching methods/pictures. (EMN)

Comprehension of illustrated text: Pictures help to build mental models

Arthur M. Glenberg and William E. Langston

University of Wisconsin-Madison

Short title: Pictures build models

Subject Index

Comprehension
    expository text
Mental models in comprehension of expository text
Pictures
    effects on comprehension of expository text

Send all correspondence to:

Arthur Glenberg
Department of Psychology
University of Wisonsin
1202 West Johnson Street
Madison, WI 53711

(608) 262-8992

Bitnet: Glenberg@wiscmacc

Abstract

Pictures help people to comprehend and remember texts. We report two experiments designed to test among several accounts of this facilitation. Students read texts describing four-step procedures in which the middle steps were described as occurring at the same time, although the verbal description of the steps was sequential. A mental representation of the *procedure* would have the middle steps equally strongly related to the preceding and succeeding steps (because the steps are performed simultaneously), whereas a mental representation of the *text* would have the middle step that was described first more closely related to the preceding step than the middle step described second. After reading, strengths of the represented relationships between the steps were assessed. When the texts were accompanied by appropriate pictures, subjects tended to mentally represent the procedure. When the texts were presented alone or with pictures illustrating the order in which the steps were described in the text, subjects tended to mentally represent the text. We argue that these results disconfirm motivational, repetition, and dual code explanations of the facilitative effects of pictures. The results are consistent with a version of mental model theory that proposes that pictures help to build mental models of what the text is about.

The literature is overflowing with work investigating the facilitative effects of pictures on text comprehension.[1] And yet, no one has a clear idea of the cognitive processes underlying these effects. In this paper we describe results of research designed to uncover those processes. To foreshadow our conclusions, we find that pictures help to generate (or reinforce) important inferences, and that the probable mechanism responsible for inference generation is a type of mental model (Glenberg, Meyer, and Lindem, 1987; Johnson-Laird, 1983). Importantly, in contrast to much of the previous research on mental models, we will demonstrate effects of mental models in domains of discourse that are not explicitly spatial, thus showing the generality of the mental model construct.

We begin with a consideration of some possible explanations for the facilitative effect of pictures, many of which have been laid out by Levin (1981) and Levie and Lenz (1982). Of course, pictures may facilitate comprehension by providing information that is not available from the text. We will not consider that option here; and, as will be clear shortly, our pictures add no new information to that which is explicitly stated in the text, and yet the pictures demonstrably improve retention. The first hypothesis under consideration is that pictures have a motivating effect. Because texts with pictures may be more enjoyable to read, the reader works harder at understanding the text. This hypothesis predicts that pictures will facilitate performance on all aspects of the text, not just those illustrated.

Pictures may also serve to repeat important information. Just as explicit repetitions have large effects on memory (Glenberg, 1979; Greene, 1989), processing the information twice, once as text and once as a picture, may facilitate comprehension and memory. This hypothesis predicts that pictures will facilitate performance on tests of information explicitly portrayed in the picture (that is, repeated), but that the picture should have relatively little effect on information not represented in the picture.

There are also a number of more sophisticated hypotheses regarding the effects of pictures on the representations derived from reading. One is the dual code theory (e.g., Paivio, 1986). According to this theory, text and pictures result in two different kinds of conceptual representations. These representations may allow independent access to information, and hence benefit retention. We will demonstrate that some forms of this hypothesis can be ruled out. For example, we will show that some intuitively reasonable pictures can hurt performance, an effect not predicted by a simple dual coding hypothesis. However, the hypothesis we adopt, based on the mental model construct, can also be viewed as a type of dual code theory.

Comprehension of a text appears to result in multiple representations (e.g., Carpenter and Just, 1987; van Dijk and Kintsch, 1983), one of which may be a propositional representation of the text itself (a representation of the words and sentences). Another representation may be a mental model, which is a representation of what the text is about, rather than a representation of the text. Different versions of mental model theory specify that mental models are propositional representations (van Dijk and Kintsch, 1983; Carpenter and Just, 1987), representations based on perceptual abilities (Franklin and Tversky, 1990), as well as other alternatives (Johnson-Laird, 1983). For us, a mental model derived from a text has the following characteristics. First, it is a representation of what (a portion of) the text is about. Second, it is a representation that makes use of working memory (Baddeley, 1986), in particular the visuo-spatial scratchpad, and hence has a very limited capacity.

Third, the mental model consists of representational elements arrayed in a spatial medium of the visuo-spatial scratchpad. The representational elements represent objects and ideas derived from the text (or from pictures). The *representational elements* act as pointers to propositional and perceptual information in LTS that describes the objects represented.[2] Although the spatial medium is ordinarily used to represent space, we propose that readers can elect to map one or more of the spatial dimensions onto dimensions suitable for representing the text. Thus, if the text describes particles that differ in energy, one (spatial) dimension can be used to represent energy, and the representational elements corresponding to the particles will be arrayed along that dimension.

Fourth, the mental model reflects the reader's current understanding of the text, and the model is updated as the text progresses. This sort of updating is accomplished by adding and deleting representational elements to reflect the current focus of the text (e.g., Sidner, 1982), as well as changing the locations of the representational elements within the spatial medium as the text describes how an object moves along a represented dimension. For example, as the text describes a change in location of a character, the model is updated (e.g., Glenberg et al., 1987, Morrow, Greenspan, and Bower, 1987). Equivalently, if the text describes how a sub-atomic particle loses energy, the representational element corresponding to the particle is moved along the dimension being used to represent energy.

Because representational elements in a mental model can point to both propositional and perceptual information, they serve to integrate information derived from these separate domains. Thus a text could describe features of an object (e.g., its mass and density), a picture could indicate the object's location in space, and the representational element in the mental model could link the information sources so that they are conceived of as pertaining to the same object.

Mental models allow a particular sort of computation, which we call *noticing*, that enhances comprehension and retention. We propose that whenever a mental model is updated (by adding, deleting, or moving a representational element), attention is focused on the updated element. Following the "spotlight" metaphor of attention, we propose that other representational elements in the spatial vicinity of the updated elements are noticed. When this occurs, the relationship between the updated element and those noticed is encoded and stored (propositionally) along with other propositions from the text. In this way, the mental model acts as an inference generator to assist in the encoding of relationships that are implicit in the text, as well as recoding and reinforcing some relationships that are explicit in the text.

Consider again a text that describes the energies of sub-atomic particles, and how these energies change. In the mental model the representational elements corresponding to the different particles are arrayed along one (spatial) dimension which the reader uses to represent energy. Now, suppose that the text continues with a description of a change in energy of one particle. Updating the mental model consists of adjusting the location of the appropriate representational element, and this adjustment may bring that element into contiguity with a different element representing a different particle. The relationship between the two particles is noticed, and an inference is stored that encodes the fact that the two particles now have the same energy. Given other information from the text, the reader might also be able to infer that the two particles are now of the same class, etc.

. Within this mental model framework, pictures help the comprehension and retention of text in a variety of ways that we group under the term *working memory management*. For now, we will describe the one aspect of working memory management relevant to the experimental work reported here. Recall that mental models are representations of situations described by a text, rather than representations of the text itself. Pictures are also typically representations of situations. Thus, a picture can assist in the construction of a mental model because the structure of the picture (the relations between the parts) are often identical to the required structure of the mental model.

In summary, we propose that pictures assist in the construction and management of mental models in working memory. Furthermore, mental models support the noticing of relationships that are implicit in the text, thus assisting in the creation of a representation that is "richer" or more "elaborate" than would ordinarily be available from a representation of the text itself.

The experiments that follow test a number of predictions of this conceptualization. First, pictures should (often) facilitate the comprehension and retention of text. Second, the facilitation should be greatest for information that is "noticed" when a mental model is formed, but that is left implicit in the text or, because of the structure of the text, is difficult to encode. Third, pictures that seem to be intuitively reasonable adjuncts to a text but that encourage the noticing of inappropriate relations (inappropriate in that the structure of the picture does not reflect the structure of the situation) may reduce comprehension and retention. Finally, these effects should be clearly attributable to a level of representation different from the representation of the text alone.

These predictions contrast in several ways with predictions derived from the motivation, repetition, and simple dual code hypotheses. The motivation hypothesis predicts that pictures will facilitate memory for all aspects of the text, whereas the mental model hypothesis predicts specific facilitation for noticed relations. The repetition hypothesis predicts facilitation only for information directly represented in the picture, whereas the mental model hypothesis predicts facilitation for representational elements entered into the mental model and for which relations are noticed. The simple dual code hypothesis predicts that (reasonable) pictures will always facilitate retention, whereas the mental model hypothesis predicts that pictures that encourage noticing of inappropriate relations will hinder comprehension.

The experiments we report trade on the notion of multiple levels of representation, in particular, a representation of the text, and a representation of the situation portrayed by the text, the mental model. We designed the texts so that the two representations would have different structures, and then probed the subjects to uncover the structures actually developed.

Insert Table 1

As an example, consider the text in Table 1. The text describes a four-step procedure for which the second and third steps are performed at the same time. Note that the text is explicit about the temporal ordering of the steps. Nonetheless, by virtue of the nature of text, one of the co-temporaneous steps ("consider the structure") is described before the the other ("address the audience"). Because of the order of description, a representation of the text is likely to have a stronger connection between the first step ("write a first draft") and the second step ("consider the structure") than between the first step ("write a first draft") and the third step ("address the audience"). In contrast, a representation of the procedure being described (a mental model) should have the second and third steps equally well related to the first (and fourth) steps. Figure 1 illustrates the structure of a mental model for this text.

Insert Figure 1

In the experiments, subjects in the no-picture condition read texts structured much like that in Table 1. That is, all of the texts described four-step procedures, and the second and third steps were always described as occurring at the same time. Subjects in the with-picture condition read the identical texts, and the texts were accompanied by pictures much like that in Figure 1. We then probed for two types of structural relations, near pairs and far pairs. Near pairs are pairs of steps whose descriptions are literally near one another in the text. These are pair (of steps) 1 and 2 and pair 3 and 4. Far pairs are pairs of steps whose descriptions are literally farther in the text, pair 1 and 3 and pair 2 and 4.

Subjects in the no-picture condition should respond primarily on the basis of a representation of the text. Thus there should be evidence for a stronger relation between members of near pairs than far pairs. Suppose that subjects in the with-picture condition are, with the support of the picture, more likely to form a mental model of the procedure. In this case, relations between the steps in the far pairs should be noticed, and there should be evidence for an equally strong relationship between members of near and far pairs. Thus the mental model hypothesis predicts an interaction between distance (near and far) and picture condition (no-picture and with-picture). The motivation and repetition hypotheses predict a main effect of picture condition, but no interaction between picture condition and distance.

As illustrated in Table 1, the text contains step names and facts associated with each step. The motivation and the mental model hypotheses make the same predictions for facts as for step names. The repetition hypothesis makes a different prediction, however. Because the facts are not represented in the pictures, they are not repeated, and the pictures should not facilitate responding to facts. Thus, the repetition hypothesis predicts an interaction between picture condition and name/fact, whereas the motivation and mental model hypotheses do not.

There are several methods available for assessing the degree of relationship between members of a pair. A method that has been employed successfully uses a variant of the priming methodology (Glenberg, et al. 1987; McKoon and Ratcliff, 1980; McNamara, Ratcliff, and McKoon, 1984). After reading, subjects are presented one member of the pair (the prime) and then asked to make a speeded recognition decision to the other member of the pair (the target). Faster responding indicates a stronger (or structurally closer) relationship. We did not use the speeded recognition task for the following two reasons. If there are functionally separate representations (one of the text and one of the situation described by the text), the speeded recognition task may well be based on the representation of the text, and hence would not reveal anything about the mental model. In fact, given that the response requires recognition of the exact words, this seems likely (see Clayton and Chattin, 1989; McNamara Altarriba, Bendele, Johnson, and Clayton, 1989). Second, forming and maintaining mental models requires some cognitive effort. Given that the speeded recognition task can be performed without constructing a mental model, there is little encouragement to do so.

We took three steps to encourage the construction of mental models. First, subjects were explicitly told to learn the order of the steps required to execute the procedures, not the order of the steps in the texts. Second, some texts (referred to as "non-sequential") presented the steps out of order. An example of a non-sequential text is given in Table 2. These texts were included to further illustrate (to the subjects) the difference between the order of the steps in the procedure and the order in the text. Results from the non-sequential texts were not analyzed. Third, a comprehension task was designed that required (for accurate performance) a representation of the order of the steps in the procedure. After reading a text, subjects were probed with a series of pairs of step names or pairs of facts from different steps. The members of the pair were presented one under the other on a computer monitor. The subject was to respond "yes" if the member on the top came from a step performed (when the procedure is executed) immediately before the step from which the other member was taken. Thus, subjects were to respond "yes" to near pairs formed from steps 1 and 2, and steps 3 and 4, and they were to respond "yes" to the far pairs formed from steps 1 and 3 and steps 2 and 4. Subjects were to respond "no" to pairs formed from steps 1 and 4 and steps 2 and 3. We measured both speed and accuracy of responding.

Insert Table 2

Experiment 1

Method

Subjects. Forty-eight subjects from Introductory Psychology courses at the University of Wisconsin--Madison participated in this experiment in exchange for course credit. Three additional subjects were dropped from the experiment; two due to equipment failure, and one who failed to complete the experiment in the time allowed. Subjects were randomly assigned to the two picture conditions.

Materials. Thirty-two texts were constructed, each describing a four-step procedure (see Table 1 for an example text). Four pieces of information were presented for each step: its order in the procedure, the name of the step, and two facts associated with that step. The two facts were short phrases associated with the step. The descriptions of the four steps were presented in a uniform text frame. This text frame was of the form: "There are four steps to be taken when [the procedure name]. The first step is to [the description of the first step]. The next two steps are performed at the same time. One of these is to [the description of the second step]. The other step is to [the description of the third step]. The final step is to [the description of the fourth step]." All of the texts described steps two and three in the procedure as being performed at the same time. These two steps were written so that they could be exchanged without disrupting the flow of the text, and they were randomly ordered each time a procedure was presented.

A pair of pictures was designed to correspond to each text. All of the pictures followed the same diagrammatic outline, and the step names for each procedure were added to this outline. One member of each pair of pictures presented steps two and three in one (left-right) order, and the other member presented these steps in reverse order. Pictures were then selected so that the left-right order of the steps in the diagram matched the order in the text.[3] A sample picture appears in Figure 1.

Each text could also be presented in non-sequential order (see Table 2 for an example non-sequential text). For this order of presentation, steps two and three were presented first, and then the phrase "However, the very first step is to . . ." preceded the description of step one.

Six speeded tests followed each text. Each test consisted of two phrases taken from different steps in the text. Four of the tests required a "yes" response, and two required a "no" response. For "yes" tests, the two phrases were taken from steps which were in sequential order when the procedure was performed. For "no" tests, the pairs of phrases came from either steps 2 and 3 or steps 1 and 4. In both of these instances the correct answer would be "no", either because the steps would be performed at the same time, or because a step (or steps) would be performed between them.

As described previously, the "yes" tests were further classified into near tests (steps 1 and 2, steps 3 and 4), and far tests (steps 1 and 3, steps 2 and 4). Two of the tests for each text were name tests, and four of the tests for each text were fact tests. Name tests contained two step names, whereas fact tests consisted of two facts, one from each of two different steps. There were two facts associated with each step. To counteract order effects between particular pairs of facts, the fact from a step that was used first was randomly determined each time a text was presented. The six tests for each text were presented to the subject in random order with the restriction that a fact test always appeared first.

As there were only four steps in each text, there were only four step names available to use in creating tests. This meant that once two of the names from a particular text were used to construct a far test the only other test possible was another far test. The same was true for near tests and no tests. To solve this problem, six test groups were constructed with six test templates in each. Each test template contained information about the step that a phrase was to come from and whether the test was to be a name or fact test. The six test groups contained different combinations of name near, far, and "no" tests with fact near, far, and "no" tests. For each text, one of these groups of test templates would be selected. Six orderings of these groups of templates were constructed, and these orderings were counterbalanced to ensure that for each text each test type would appear an equal number of times.

Two additional texts and four additional pictures were designed to be used for practice. The experiment was controlled using an Apple Macintosh II computer.

Design. A 2 (with-picture vs. no-picture) X 2 (name vs. fact tests) X 3 (near vs. far vs. "no" tests) mixed factorial design was used for this experiment. Pictures were manipulated between subjects, with half of the subjects seeing pictures while reading the texts, and half of the subjects not seeing pictures. Whether or not a text was presented in sequential order, whether a particular test was composed of name or fact phrases, and whether a particular test represented a near test, a far test, or a "no" test were all manipulated within subjects. One non-sequential text was presented in each block of four texts, with its position in the block being randomly determined. A counterbalancing strategy was used to ensure that once all subjects were run all of the texts would be presented in non-sequential format equally often.

Procedure. All subjects were given detailed instructions prior to beginning the experiment, and subjects were also "walked through" a practice text to ensure that they fully understood the task. Extensive feedback, including explanations of correct answers, was given to the subjects during the first practice trial. Subjects were also given a second practice trial and were encouraged to ask questions.

All subjects saw all thirty-two texts; the order of presentation of the texts randomized for each subject. Subjects performed the same set of operations for each text. Each text was presented one step at a time, reading was self-paced, and total reading time was recorded for each text. In the with-pictures condition, the title and the picture were displayed until the first keypress from the subject. At that time, the title was erased and replaced by the description of the first step of the text. The picture remained on the screen throughout the reading process. In the no-pictures condition, the title and the four step names were presented first, and remained on the screen until the first keypress. At that time, they were erased and replaced by the description of the first step of the text. In both conditions, subjects advanced through the text by pushing either the '/' or 'z' keys on the computer's keyboard. As each step was revealed, the entire text remained on the screen, and new steps began at the end of the preceding steps, in the correct location on the screen. When all of the steps were displayed, they appeared to form a single paragraph.

After reading each text, subjects were warned to expect a series of speeded tests. Next, they were shown fixation points for 1.5 seconds to allow them to position their fingers on the '/' and 'z' keys on the keyboard and focus on the appropriate location on the computer monitor. Tests consisted of two phrases presented one on top of the other in the center of the screen next to the fixation points. The subjects' task was to respond "yes" or "no" to the question: "Would the step containing the phrase on the top immediately precede the step containing the phrase on the bottom, if you were to actually perform the procedure?" Reaction time and response choice were recorded for each test. The key on the keyboard that corresponded to "yes" was always on the side of the subject's dominant hand. Left-handed subjects used the 'z' key for "yes", and other-handed subjects used the '/' key. Subjects were instructed to respond to these tests as quickly as possible without making mistakes.

After responding to the six speeded tests, subjects were given a true/false comprehension question. This question usually asked about the sequence of steps in the procedure described by the text, and was the only test whose intent was clear to the subjects Feedback was given on this question.

After the true/false question, subjects were warned to expect the next text, and the process was repeated. After sixteen texts had been presented, subjects were allowed to stop for a ten minute break.

Results

Because the reaction times were long (median reaction times on the order of 2-5 seconds) and the error rates were high (ranging between 15-50 percent), we decided to focus on proportion correct as the main dependent variable. Data from the texts presented in non-sequential format were not used in the analyses. Two Analyses of Variance were run on the data, one using subjects as the random factor (reported as $F1$), and another using texts as the random factor (reported as $F2$). The significance level was set at .05.

"Yes" Pairs. We focus first on those pairs for which the correct response was "yes". Both near and far pairs are "yes" pairs.

The main effect for pictures (with-picture vs. no-picture) was significant, $F1(1,46) = 7.79$, $MSE = .08$, and $F2(1,31) = 107.15$, $MSE = .011$. Subjects in the with-pictures condition responded more accurately ($M = .82$) than did subjects in the no-pictures condition ($M = .70$). This result replicates many previous studies. Our finding is notable, however, because the pictures are so simple. This makes it difficult to attribute the improvement to additional information present in the pictures that was not present in the texts. Now, how is it that the pictures enhance performance?

The repetition hypothesis predicts an interaction between picture condition and name/fact. Pictures should facilitate performance on the name tests (because the names are repeated in the pictures), but not the fact tests. Contrary to the prediction, the interaction was not significant, $F1(1,46) = 1.12$, $MSE = .011$, and $F2(1,31) = .662$, $MSE = .012$. Thus we can rule out the repetition hypothesis.

The mental model hypothesis predicts a picture condition X distance interaction. Pictures should facilitate the "noticing" of a relationship between far pairs (the relationship being that these steps are sequential). Thus, with pictures, performance on near and far pairs should be similar, whereas without pictures performance should be less accurate for the far pairs. The data relevant to this prediction are presented in Figure 2. The picture condition X distance interaction was significant, $F1(1,46) = 5.67$, $MSE = .037$, and $F2(1,31) = 42.03$, $MSE = .01$. Because the motivation hypothesis fails to predict this interaction, it can be ruled out.

Insert Figure 2

A number of other main effects and interactions were also significant, but were of less theoretical importance. There was a significant main effect for name/fact, $F1(1,46) = 101.80$, $MSE = .011$, and $F2(1,31) = 91.76$, $MSE = .016$. Subjects responded to name tests ($M = .84$) more accurately than fact tests ($M = .68$).

There was also a significant main effect for distance, $F1(1,46) = 8.73$, $MSE = .037$, and

$F2(1,31) = 21.54$, $MSE = .012$. Subjects responded to near tests ($M = .80$) more accurately than far tests ($M = .72$).

The name/fact X distance interaction was also significant, $F1(1,46) = 5.80$, $MSE = .013$, and $F2(1,31) = 5.71$, $MSE = .011$. This interaction occurred because the difference between name and fact near tests was greater than the difference between name and fact far tests.

The picture condition X name/fact X distance interaction was also significant in the analysis by texts, $F2(1,31) = 7.60$, $MSE = .009$.

<u>"No" Pairs</u>. The data discussed so far were for the proportion correct on those tests requiring positive responses (near and far pairs). The data illustrated in Figure 2 serve to disconfirm the repetition and motivation hypotheses and support the mental model hypothesis. Those data are also consistent with a form of dual code theory. Suppose that the picture results in the storage of a long-term pictorial representation which has parts associated with appropriate verbal descriptions. Thus, the pictorial representation of "step 1" is associated with the verbal description of the step. Furthermore, assume that access to the pictorial representation may be obtained from the verbal description. In this case, a subject who saw the picture would do as well on far pairs as on near pairs. The reasoning is that the phrases presented on the test would provide access to the pictorial representation, and the pictorial representation indicates equally well the sequential relationship between members of near pairs and members of far pairs.

The mental model (noticing) account and this dual-coding account make different predictions in regard to the tests requiring a "no" response (that is, the members of the pair are from steps 1 and 4 or steps 2 and 3, which are not sequential steps in the execution of the procedure). Consider first predictions from the dual code approach. When faced with a test, subjects access the verbal description of the pair members and, when available, the pictorial representation. Given access to the pictorial representation, there is no reason to predict differential responding to pair 1 and 4 relative to pair 2 and 3; it is evident from the picture that neither pair involves sequential steps. Thus, the dual code approach predicts a main effect of picture condition (more accurate responding with pictures than without), but no main effect of pair type (pair 1 and 4 versus pair 2 and 3) and no interaction between picture condition and pair type.

Now, consider the noticing account. Subjects in the with-picture condition should notice that steps 2 and 3 occur at the same time (because the representational elements are located at the same point on the temporal dimension in working memory), and infer this proposition or reinforce the proposition derived from the text. Subjects should not notice (as frequently) a relationship between steps 1 and 4 because these steps are not contiguous in the mental model. Thus, this account predicts an interaction between picture condition and pair type: For the pair 2 and 3, performance should be better in the with-picture condition relative to the no-picture condition, but for pair 1 and 4, there should be little difference between the picture conditions.

Insert Figure 3

The relevant data are in Figure 3. As predicted by the noticing account, there was a large difference between the picture conditions for pair 2 and 3, but little difference for pair 1 and 4. This was true for both names and facts. The interaction of picture condition and pair type was significant, $F1\ (1,46)=11.09$, $MSE=.11$, $F2(1,31)=137.48$, $MSE=.01$.

A number of other effects were also significant in the analysis of the "no" data. There were main effects of picture condition, $F1(1,46)=6.55$, $MSE=.116$, $F2(1,31)=56.90$, $MSE=.018$, and name/fact, $F1(1,46)=19.28$, $MSE=.055$, $F2(1,31)=33.16$, $MSE=.026$. There were also significant interactions between name/fact and pair type, $F1(1,46)=11.902$, $MSE=.022$, $F2(1,31)=34.38$, $MSE=.01$, and picture condition, name/fact, and pair type, $F1(1,46)=9.55$, $MSE=.022$, $F2(1,31)=15.57$, $MSE=.01$. This last interaction was produced by the extremely poor performance on tests involving the names of step 2 and 3 when no pictures were involved. We have no reasoned explanation for this poor performance.

Reading time for each text was also collected. Comparison of reading time in the with-picture condition ($M = 45.9$ sec.) and the no-picture condition ($M = 44.0$ sec.) yielded no significant differences for the analysis by subjects, but the difference was significant in the analysis by texts, $F2(1,31) = 6.78$, $MSE = 8040061.09$.

Discussion

The major results are straightforward. First, pictures facilitate comprehension and memory for texts, even when the pictures add no new information. This is true both for information that is repeated in the pictures and for information that is not. Second, we have shown that pictures can affect the representation of the information. Without pictures, it appears that the information is organized much like the text: the relationship between members of near pairs is more strongly represented than that for far pairs. With pictures, the strength of the relationship is about the same for near and far pairs. This pattern of results is consistent with the predictions derived from our mental model approach to comprehension. Namely, the spatial arrangement of components of pictures induces a similar arrangement of representational elements in working memory (the mental model). Representational elements that are close in the mental model may be noticed, so that the relationship between the elements is encoded. This may occur even when the descriptions of the elements are separated in the text. Thus, with pictures the relationship between far pairs is encoded as well as the relationship between near pairs.

A similar conclusion about how pictures affect comprehension of text was drawn by Mayer (1989). Subjects read a passage describing the operation of hydraulic brakes. The passage was either accompanied by an illustration or not. After reading, subjects were given recall tests, verbatim recognition tests, and problem solving tests (answering questions such as "Why do brakes get hot?"). There was little effect of illustrations on the recall and recognition tests, but the illustrations greatly facilitated performance on the problem solving test. Although there are many differences in procedure, Mayer interpreted his findings much as we have interpreted ours, "illustrations can help readers to focus their attention on explanative information in text and to reorganize the information into useful mental models" (Mayer, 1989, page 240).

Although the mental model and dual code approaches make similar predictions for the "yes" data, the predictions diverge for the "no" data. If the advantage of pictures was simply the availability of a veridical pictorial representation, then subjects should respond "no" equally well to pair 2 and 3 and pair 1 and 4. However, if the advantage of pictures derives from noticing particular relationships in a mental model, then responding should be more accurate to pairs 2 and 3 when pictures are presented. The data in Figure 3 strongly support the mental model approach over the dual code approach.

Experiment 2

Experiment 2 was designed to address three goals. The first was to discover those aspects of the with-picture condition that facilitated performance. There are at least three possibilities. In the with-picture condition, the names of the steps were continuously on the screen while the text was presented, but the names were absent in the no-picture condition. Thus, the difference between the conditions may have had little to do with the pictorial aspects, and simply reflected availability of the step names. A second possibility, related to Hegarty and Just's (1988) idea of "formation" is that the boxes provided a method for mentally representing the steps, even when the steps were abstract. A third possibility, consistent with our idea of noticing, is that the spatial arrangement of the boxes was important. That is, the arrangement of the boxes in the picture produced an analogous spatial arrangement in the mental model, and this enhanced noticing of the relationship between members of the far pairs.

To test among these possibilities, we introduced a third condition, the linear-picture condition. This condition was identical to the with-picture condition, except that the boxes in the picture were arrayed vertically, one box under the next, and there were no lines connecting the boxes (see Figure 4 for an example). Thus if the "availability of step names" or the "formation" explanations are correct, the linear-picture condition should give results identical to the with-picture condition. However, if the noticing explanation is correct, then the linear-picture condition should produce worse performance than the with-picture condition. That is, in the linear picture, the members of far pairs are not depicted as spatially close so that there is little support for noticing the correct relationship. The linear pictures may cause some subjects to notice and encode inappropriate relations such as "step 3 follows step 2". This is an inappropriate relationship in that the steps are simultaneous in the procedure as executed. Thus, the linear-picture condition may result in performance worse than in the no-picture condition.

Insert Figure 4

The second goal was to provide another test of the dual code hypothesis, and the linear-picture condition does just that. According to dual code theory, pictures facilitate performance by providing a second representation that can be consulted during retrieval. There is little reason to suspect that the linear pictures would be less helpful than the standard pictures in this regard, hence the dual code theory predicts equivalent performance in the with-picture and linear-picture conditions.

The third goal was to address concerns about specific components of the methodology used in Experiment 1. One concern stems from the complex nature of the performance test, that is, requiring subjects to respond positively only when members of a pair immediately follow one another when the procedure is executed. Might it be that subjects do not really understand all the conditions that must be met for a correct positive response? This is particularly worrisome in the no-picture condition where subjects do not have the opportunity to literally see that steps 2 and 3 are performed at the same time. To overcome this possibility, we provided extensive feedback after the order test associated with each text. For each incorrect response, the subject was shown (on the computer monitor) the pair of statements, the subject's response, the correct response, and a written explanation of the correct response. For example, if a subject responded "no" to a pair taken from steps 1 and 3, the explanation would be, "Step are in sequential order. Correct response is 'yes'," and if the subject responded "yes" to a pair taken from steps 2 and 3, the explanation would be, "Steps happen at the same time. Correct response is 'no'." In addition, there were some "no" pairs ("backward no") in which the steps were performed one after the other, but their order of presentation at the test was reversed.

This extensive feedback might have additional effects that would work against some predictions. Consider the following. Our proposal is that pictures help to build mental models. However, subjects can form mental models from text without pictures (e.g., Glenberg, et al., 1987). Given the encouragement provided by the extensive feedback, some subjects, particularly those in the no-picture condition, might form mental models and thus reduce the interaction seen in Figure 2.

Method

<u>Subjects</u>. Thirty-six subjects from the Madison, Wisconsin area were paid in exchange for participation in this experiment. Two additional subjects were run who failed to complete the experiment in the time allowed, one additional subject's data were lost due to a computer

error, and three additional subjects' data were lost due to experimenter error. Subjects were randomly assigned to the picture conditions.

Materials. The materials used in Experiment 2 were the same as those used in Experiment 1 with the following two exceptions. First, two linear pictures (as in Figure 4) were constructed for each text to be used in the linear-picture condition. One linear picture for each text presented steps two and three in one (top-bottom) order, while the other linear picture presented these steps in the other (top-bottom) order. The linear picture used for each text had the same ordering of steps two and three as the text itself.

The other change to the materials was the creation of "backward no" tests. In these tests, steps were presented in reverse order (ie. a phrase from step two on top of a phrase from step one).

Design. A 3 (with-picture vs. linear-picture vs. no-picture) X 2 (name vs. fact test) X 3 (near vs. far vs. "no" test) mixed factorial design was used for Experiment 2. Picture condition was manipulated between subjects, whereas name/fact and distance were manipulated within subjects. A counterbalancing scheme more efficient than the one used in Experiment 1 was used for this experiment. This scheme insured that after twelve subjects were run in each condition all texts would appear in non-sequential format equally often, and all test types would be equally represented for each text.

Procedure. The procedure used for Experiment 2 was the same as that used for Experiment 1 with three exceptions. First, a third picture condition, the linear picture condition, was added. This condition was the same as the with-picture condition in Experiment 1, but linear pictures were shown to the subjects instead of pictures which model the situation described by the text.

The second change was to add exhaustive feedback following a subject's responses to the six speeded tests that followed each text. This feedback showed a subject how many errors were made, the phrases from the tests responded to incorrectly, the steps those phrases came from, the subject's response to that test, an explanation of why that response was incorrect, and what the correct response should have been. Each error was shown to subjects individually, and subjects pressed the "return" key to advance through the list of errors. A counter on the screen kept track of where subjects were in the error list to prevent confusion. For texts on which subjects made no errors, the message "No errors" appeared on the screen, and subjects pressed the "return" key to continue. The feedback was carefully explained to subjects during the practice texts.

The third change was to add "backward no" tests to the six speeded tests for the non-sequential texts. These tests presented steps in reverse order. Each non-sequential text contained two of these "backward no" tests. These "backward no" tests were included to prevent subjects from adopting the strategy of memorizing only steps 1 and 4 and then responding "yes" whenever one (but not both) of these steps were included in a test. These tests were included only in the non-sequential texts for two reasons. First, because the "backward no" tests were added to prevent subjects from using a particular strategy, there was no theoretically driven reason to analyze these data. Second, because there were already so few name "no" tests in the analyzed texts, we were hesitant to replace any of them.

Results

As in Experiment 1, reaction times to the speeded tests were collected, but for the same reasons as in Experiment 1, we decided to focus on proportion correct as the main dependent variable. Data from the non-sequential texts were not used in the analyses.

"Yes" Pairs. The main effect for picture condition was significant, $F1(2,33) = 7.18$, $MSE = .039$, and $F2(2,62) = 25.98$, $MSE = .029$. Subjects in the with-pictures condition responded most accurately ($M = .81$), followed by subjects in the no-pictures condition ($M = .76$), followed by subjects in the linear-pictures condition ($M = .66$). The results of two planned comparisons showed that performance in the linear-picture condition was significantly poorer than performance in the with-picture condition, $t1(33)=3.72$, $t2(62)=7.73$, and performance in the linear-picture condition was significantly worse than performance in the no-picture condition, $t1(33)=2.47$, $t2(62)=4.11$.

The fact that subjects in the with-pictures condition were most accurate replicates previous research in which pictures were shown to improve performance, but the low performance of subjects in the linear-pictures condition is a new finding. Apparently, "reasonable" pictures do not invariably help comprehension. Because the linear-pictures condition produced the poorest level of performance, we can disconfirm the "availability of step names" and "formation" hypotheses, as well as the dual code hypothesis; all predicted that performance in the linear-pictures condition would be equivalent to performance in the with-pictures condition. The ordering of the three conditions does support the noticing hypothesis. The hypothesis predicted that performance in the linear-pictures condition would be poorer than performance in the with-pictures condition, and, when inappropriate relations are noticed, the hypothesis predicted that performance in the linerar-picture condition would be worse than performance in the no-picture condition.

The noticing hypothesis predicts a pictures X distance interaction. Subjects in the with-pictures condition should notice the sequential relationship between far pairs, whereas subjects in the linear-pictures condition should not. In other words, subjects in the with-pictures condition should show roughly the same level of performance on near and far pairs, whereas subjects in the linear-pictures condition should show better performance on the near pairs than on the far pairs. The data relevant to this prediction are presented in Figure 5. The picture X distance interaction was marginally significant by subjects, $F1(2,33) = 2.84$, $MSE = .02$, $p = .07$, and significant by texts, $F2(2,62) = 3.09$, $MSE = .034$, providing some support for the noticing hypothesis.

Insert Figure 5

Subjects in the no-pictures condition performed better than we would have expected, given the results from Experiment 1. This could have been caused by the explicit feedback which subjects received following the speeded tests. That is, the feedback might have encouraged subjects in the no-picture condition to form a mental model of the procedure described by the text even though they did not receive any picture[4]. If this reasoning is correct, then performance in the first part of the experiment (before too much encouragement) should replicate the picture condition X distance interaction found in Experiment 1 (see Figure 2). After sufficient encouragement to form mental models, there should be little difference between the with-picture and no-picture conditions. In fact, we found just this pattern for the name tests. Examining those texts subjects read in the first third of the experimental session, the difference between near and far pairs was .08 in the with-pictures condition, and .15 in the no-pictures condition. In the final two thirds of the experimental session, the difference was .08 in the with-pictures condition, and .07 in the no-pictures condition. A similar analysis for the fact tests was not as informative because performance on the fact tests in the no-picture condition was close to chance (.58) in the first third of the experiment.

Several other main effects and interactions were also significant, but were of less theoretical importance. First, there was a significant main effect for name/fact, $F1(1,33) = 72.23$, $MSE = .007$, and $F2(1,31) = 101.20$, $MSE = .023$. Subjects responded to name tests ($M = .82$) more accurately than to fact tests ($M = .67$).

The main effect for distance was also significant, $F1(1,33) = 15.1$, $MSE = .02$, and $F2(1,31) = 23.76$, $MSE = .03$. Subjects responded to near tests ($M = .79$) more accurately than to far tests ($M = .70$).

The name/fact X distance interaction was also significant, $F1(1,33) = 5.93$, $MSE = .011$, and $F2(1,31) = 5.68$, $MSE = .026$. The interaction occurred because subjects responded better to near name tests than to far name tests, whereas there was less of a difference between near and far fact tests.

"No" Pairs. As with the "Yes" pairs, the major prediction was for better performance in the with-picture condition than in the linear-picture condition. In fact, subjects were more accurate in the with-picture condition (.76) than in the linear-picture condition (.65); accuracy in the no-picture condition was between these two (.73). An analysis including just the with-picture and linear picture conditions produced a marginally significant effect of pictures in the subjects analysis, $F1(1,33)=3.57$, $MSE=.079$, $p=.07$, and a significant effect in the text analysis, $F2(1,31)=21.20$, $MSE=.04$. Planned comparisons showed that performance in the linear-picture condition was worse than in the with-picture condition, $t2(62)=4.60$, and worse than the no-picture condition, $t2(62)=3.25$, but neither comparison reached standard levels of significance in the analysis by subjects.

For two reasons, other predictions cannot be made as confidently. First, as we noted for the "Yes" data, it seems likely that the extensive feedback encouraged subjects in the no-picture condition to form an accurate mental model, precluding differences between the with-picture and the no-picture conditions. Second, predictions depend on the details of how the noticing process is instantiated. Some subjects might treat the adjacency (in the linear picture) of steps 2 and 3 as an opportunity to encode a "sequential" relationship. This would tend to produce incorrect responding to the pair 2 and 3 (because the steps as being performed simultaneously). On the other hand, if some subjects recall the wording in the text, they might might treat the adjacency of steps 2 and 3 (in the picture) as an opportunity to encode a "simultaneous" relationship, which would tend to produce correct responding. Thus, the only safe prediction is that the conditions will be ordered as with-picture, no-picture, linear-picture. In any event, except for the main effect of names ($\underline{M}$=.76) versus facts ($\underline{M}$=.66), $\underline{F1}(1,33)=29.82$, $\underline{MSE}$=.39, $\underline{F2}(1,31)=16.86$, $\underline{MSE}$=.062, no other source of variance was significant in both the analysis by subjects and the analysis by texts.

Total reading time for each text was also collected. The difference between reading times in the three picture conditions was not significant for the analysis by subjects, $\underline{F1}(2,33) = .256$, $\underline{MSE} = 188594464.84$, but was significant by texts, $\underline{F2}(2,62) = 4.89$, $\underline{MSE} = 26338672.59$. Subjects in the no-pictures condition read the slowest ($\underline{M}$ = 49.7 sec), followed by subjects in the with-pictures condition ($\underline{M}$ = 48.5 sec), followed by subjects in the linear-pictures condition ($\underline{M}$ = 45.8 sec).

Discussion

Three aspects of the results bear emphasizing. First, we successfully replicated the finding from Experiment 1 of similar performance on the near and far pairs when the text is supported by an appropriate picture. Second, not all pictures are appropriate: the linear pictures did not support noticing correct temporal relations between the steps in the procedures, and hence performance in the linear-picture condition was worse than in the with-picture condition and worse than in the no-picture condition. This result also disconfirms the dual code hypothesis. Third, the linear pictures did provide continuous availability of the step names and a concrete image for representing the steps. Thus the relatively poor performance in this condition contradicts the predictions derived from the "availability of step names" and "formation" accounts of the results of Experiment 1.

General Discussion

Our results point to the use of mental models in the integration of information from pictures and texts during comprehension. Before discussing how we view these processes occurring, we will briefly describe three important features of our methodology.

First, our experiments used many texts (although all of the same structure), thus it is unlikely that the results are peculiar to sampling one or just a few content areas. This can be contrasted with other work investigating mental models such as Hegarty and Just (1988), Kieras and Bovair (1984), Morrow, Bower, and Greenspan (1989), Morrow and Greenspan (1987), Perrig and Kintsch, (1985), and Schmalhofer and Glavanov (1986).

Second, our experiments did not involve the pre-memorization of a picture, as in Morrow and Greenspan (1987) and Morrow et al. (1989), so that our reading situation is close to that found in many natural situations.

Third, neither the structure nor the contents of the texts were explicitly spatial. In fact, the structure was temporal in that the texts described the order in which the steps in the procedures were to be performed. This is in contrast to most previous work investigating mental models and text comprehension. Our finding of robust effects in non-spatial domains illustrates the generalizability of the mental model construct.

The results contradict a number of hypotheses describing how pictures facilitate text comprehension. In particular, we have presented evidence contrary to the motivation, repetition, and some versions of the dual code model. Because facilitating effects of pictures were not across the board, we can confidently rule out the motivation hypothesis. Because facilitation could be found for information repeated in the pictures as well as information not repeated in the pictures, we can rule out the repetition hypothesis. Finally, we adduced two pieces of data contrary to the dual code hypothesis. First, in Experiment 1, we found greater facilitation due to pictures when responding "no" to the pair 2 and 3 than to the pair 1 and 4 (see Figure 3). These pairs should be equally well represented in the pictorial representation, and hence on the dual code approach there is little reason to expect a difference. Second, in Experiment 2, a picture (the linear picture) that should have provided access to the steps and facilitated correct responding actually reduced correct responding.

We do not wish to claim that there is no long-term representation of pictures. In fact, it seems quite likely that our subjects could reproduce from memory at least some of the pictures they saw. Similarly, we do not wish to claim that a long-term pictorial representation is never beneficial. Because our pictures were so simple and so similar across texts, we probably decreased any benefit derivable from a pictorial representation. Nonetheless, we did show large effects of pictures even under these constrained conditions. We turn now to a discussion of our mental model explanation of those results.

Our version of mental model theory has a number of attractive features, not the least of which is that it does a credible job of accounting for much of our data. In addition, because we propose that mental models are constructions in working memory, we immediately get the benefit of research on the contributions of working memory to comprehension. Perhaps more importantly, this proposal supplies constraints (e.g., capacity constraints) needed for formal modeling. We also propose that readers can choose how to use the (normally) spatial dimensions of working memory to represent other dimensions and relations. When this is combined with the process of "noticing," we can turn the mental model into a powerful inference generator, but one which has multiple constraints, such as when noticing is done, and the capacity of working memory. These inferences enhance comprehension and restructure the representation of information derived from the text, thereby giving mental models their functional power. Furthermore, because pictures help to build mental models, these constructs allow us to explain how pictures improve comprehension.

We also propose that the long-term effects of mental models are mediated by a propositional representation derived partially from the text, partially from pictures, and partially from the model (the inferences generated). In this manner we need not propose a new type of long-term representational format (or even a separate long-term representation of the mental model), and we can take advantage of the the tremendous literature supporting propositional representational formats.

Finally, we wish to be clear that like Johnson-Laird (1983), our ideas do not necessitate that mental models be imagistic. Representational elements in working memory may point to information that can be used to construct mental images, but they need not. Thus we are not embarrassed by data showing mental-model-like effects with difficult to image material.

These ideas are open to development in a number of directions. Consider first the integration of pictorial (or more generally, spatial) and linguistic information. Clearly this is an important skill that we exercise repeatedly in watching television or when engaged in conversation (Massaro and Cohen, 1990; McGurk and MacDonald, 1976). The results we have presented here clearly demonstrate integration of a sort: information from the text, such as the facts pertinent to each step, is integrated with information from the picture, such as the temporal relationships among the steps. In addition, we have provided a mechanism, representational elements in working memory, to account for the integration. To reiterate, the representational elements are pointers to descriptive propositional information derived from the text and descriptive propositional and perceptual information derived from the picture. Thus the mental model integrates the two sources of information. To be sure, there is other important work on cross-modal integration. However, some of that work deals with pre-memorized pictures and verbal information (e.g., Altariba and McNamara, 1988; McNamara, Halprin, and Hardy, manuscript in preparation), and the methodology in other work (e.g., Loftus, Miller, and Burns, 1978; Pezdek, 1977; Pezdek and Miceli, 1982) has been criticized (McCloskey and Zaragoza, 1985).

A second direction for these ideas is application to problems in development. We envision the construction of a mental model from text as an active, attention-demanding process, not one that occurs automatically with reading. Furthermore, the ability to arbitrarily assign a new meaning to one of the spatial dimensions in working memory, is a skill that almost surely requires learning. This learning may be a precursor to effective use of mental models in abstract reasoning tasks (Johnson-Laird, Byrne, and Tabossi,1989).

Finally, we think that the sort of mental model we are proposing can serve important functions in discourse understanding. For example, the representational elements are very similar to Carpenter and Just's (1977) discourse pointer, and Sidner's (1982) focus. Thus the mental model keeps track of the topic of the sentence and discourse to facilitate inference making (e.g., what a pronoun refers to) and integration of ideas. If we combine with the mental model a compound cue theory of retrieval and priming (Ratcliff and McKoon. 1989), the mechanism becomes a powerful device for comprehension. For example, suppose that retrieval of information from LTS is prompted by representational elements in working memory (or equivalently, the propositional information pointed to) as well as other contents of working memory. A focusing rule can be used to pick out information that is highly related to the conjunction of elements in working memory (e.g., Gillund and Shiffrin, 1984; Hintzman, 1986), so that only the contextually most appropriate information in LTS is primed (or retrieved). Deleting a representational element from working memory terminates the priming, much as Sharkey and Mitchell (1985) found that "exiting" a script reduces priming of script-related concepts. Similarly, two findings reported by MacDonald can be accommodated within this framework. MacDonald and Just (1989) demonstrated that negation (e.g., "Almost every weekend, Elizabeth bakes some bread but no cookies for the children") slows time to recognize the negated noun (as if it has been deleted from the mental model). MacDonald and MacWhinney (1990) and Gernsbacher (1989) have demonstrated that pronominal reference facilitates later access to the antecedent concept and inhibits reference to a similar concept that is not the antecedent concept. Apparently, pronominal reference ensures that the representational element of the antecedent is maintained in working memory (providing later facilitation), while other representational elements may be deleted to recoup capacity for further processing (producing later inhibition).

Earlier, we introduced the terminology "working memory management." Although the terminology may be new, the idea that comprehension requires control over working memory is part of most theories of comprehension (e.g., Van Dijk and Kintsch, 1983; also see Fletcher, 1986 for experimental investigation of strategies of working memory management). We mean by this concept the introduction, maintenance, and deletion of information in working memory. We propose that many beneficial effects of pictures come about through the effect of pictures on working memory management. Here we have demonstrated how pictures can assist in building an accurate mental model (a type of working memory management) that facilitates inference making. Pictures may also ease the search for referents of terms (Larkin and Simon, 1987) and the introduction of those referents into working memory. In a similar vein, pictures can serve as a type of external memory. That is, comprehension of some ideas may require the simultaneous co-occurrence of multiple representational elements, too many to ordinarily hold in working memory at one time. Direct viewing of a picture may provide relatively effortless maintenance of some of the representational elements corresponding to parts in the picture, freeing up capacity for inference generation.

References

Altarriba, J., & McNamara, T. P. (1988). Spatial priming in memory for a pre-experimentally-learned environment. Paper presented at the 60th Meeting of the Midwestern Psychological Association, May, 1988, Chicago.

Baddeley, A. D. (1986). Working memory. Oxford: Oxford University Press.

Carpenter, P. A., & Just, M. A. (1977). Integrative processes in comprehension: In D. Laberge & S. J. Samuels (Eds.), Basic processes in reading: Perception and comprehension. Hillsdale, NJ: Erlbaum.

Clayton, K., & Chattin, D. (1989). Spatial and semantic priming effects in tests of spatial knowledge. Journal of Experimental Psychology: Learning, Memory, and Cognition, 15, 495-506.

Fletcher, C. R. (1986). Strategies for the allocation of short-term memory during comprehension. Journal of Memory and Learning, 25, 43-58.

Franklin, N., & Tversky, B. (1990). Searching imagined environments. Journal of Experimental Psychology: General, 119, 63-76.

Gernsbacher, M. A. (1988). Mechanisms that improve referential access. Cognition, 32, 99-156.

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. Psychological Review, 91, 1-67.

Glenberg, A. M. (1979). Component-levels theory of the effects of spacing of repetitions on recall and recognition. Memory & Cognition, 7, 95-112.

Glenberg, A. M., Meyer, M., & Lindem, K. (1987). Mental models contribute to foregrounding during text comprehension. Journal of Memory and Learning, 26, 69-83.

Greene, R. G. (1989). Spacing effects in memory: Evidence for a two-process account. Journal of Experimental Psychology: Learning, Memory, and Cognition, 15, 371-377.

Hegarty, M., & Just, M. A. (1988). Understanding machines from text diagrams. In A. Mandle & J. Levin (Eds.), Knowledge acquisition from text and pictures. Amsterdam: North-Holland.

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. Psychological Review, 93, 411-428.

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

Johnson-Laird, P. N., Byrne, R. M. J., & Tabossi, P. (1989). Reasoning by model: The case of multiple quantification. Psychological Review, 96, 658-673.

Kieras, D. E., & Bovair, S. (1984). The role of a mental model in learning to operate a device.

Cognitive Science, 8, 255-273.

Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. Cognitive Science, 11, 65-99.

Levie, W. H. (1987). Research on pictures: A guide to the literature. In D. M. Willows & H. A. Houghton (Eds.), The psychology of illustration: Vol I. Basic research. New York: Springer-Verlag.

Levie, W. H., & Lentz, R. (1982). Effects of text illustrations: A review of research. Educational Communication and Technology Journal, 30, 195-232.

Levin, J. R., Anglin, G. J., & Carney, R. N. (1987). On empirically validating functions of pictures in prose. In D. M. Willows & H. A. Houghton (Eds.), The psychology of illustration: Vol I. Basic research. New York: Springer-Verlag.

Levin, J. R., & Lesgold, A. M. (1978). On pictures and prose. Educational Communication and Technology Journal, 26, 233-243.

Loftus, E. F., Miller, D. G., & Burns, H. J. (1978). Semantic integration of verbal information into a visual memory. Journal of Experimental Psychology: Human Learning and Memory, 4, 19-31.

MacDonald, M. C., & Just, M. A. (1989). Changes in activation level with negation. Journal of Experimental Psychology: Learning, Memory, and Cognition, 15, 633-642.

MacDonald, M. C., & MacWhinney, B. (1990). Measuring inhibition and facilitation from pronouns. Journal of Memory and Learning, 29, 469-492.

Massaro, D. W., & Cohen, M. M. (1990). Perception of synthesized audible and visible speech. Psychological Science, 1, 55-63.

Mayer, R. E. (1989). Systematic thinking fostered by illustrations in sceintific text. Journal of Education Psychology, 81, 240-246.

McCloskey, M., & Zaragoza, M. (1985). Misleading postevent information and memory for events: Arguments and evidence against memory impairment hypotheses. Journal of Experimental Psychology: General, 114, 1-16.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. Nature, 264, 746-748.

McKoon, G., & Ratcliff, R. (1980). Priming in item recognition: The organization of propositions in memory for text. Journal of Verbal Learning and Verbal Behavior, 19, 369-386.

McNamara, T. P., Altarriba, J., Bendele, M., Johnson, S. C., & Clayton, K. C. (1989). Constraints on priming in spatial memory: Naturally learned versus experimentally-learned environments. Memory & Cognition, 17, 444-453.

McNamara, T. P., Halpin, J. A., & Hardy, J. K. (manuscript in preparation). The representation and integration in memory of spatial and non-spatial information.

McNamara, T. P., Ratcliff, R., & McKoon, G. (1984). The mental representation of knowledge acquired from maps. Journal of Experimental Psychology: Learning, Memory, and Cognition, 10, 723-732.

Morrow, D. G., Bower, G. H., & Greenspan, S. L. (1989). Updating situation models during narrative comprehension. Journal of Verbal Learning and Verbal Behavior, 28, 292-312.

Morrow, D. G., Greenspan, S. L., & Bower, G. H. (1987). Accessibility and situation models in narrative comprehension. Journal of Memory and Learning, 26, 165-187.

Paivio, A. (1986). Mental representations: A dual coding approach. New York: Oxford University Press.

Peeck, J. (1987). The role of illustrations in processing and remembering illustrated text. In D. M. Willows & H. A. Houghton (Eds.), The psychology of illustration, Vol I: Basic research. New York: Springer-Verlag.

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

Pezdek, K. (1977). Cross-modality semantic integration of sentences and picture memory. Journal of Experimental Psychology: Human Learning and Memory, 3, 515-524.

Pezdek, K., & Miceli, L. (1982). Life-span differences in memory integration as a function of processing time. Developmental Psychology, 18, 485-490.

Pressley, M., & Miller, G. E. (1987). Effects of illustrations on children's listening comprehension. In D. M. Willows & H. A. Houghton (Eds.), The psychology of illustration, Vol. I: Basic research. New York: Springer-Verlag.

Ratcliff, R., & McKoon, G. (1988). A retrieval theory of priming in memory. Psychological Review, 95, 385-408.

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

Sharkey, N. E., & Mitchell, D. C. (1985). Word recognition in a functional context: The use of scripts in reading. Journal of Memory and Learning, 24, 253-270.

Sidner, C. L. (1983). Focusing and discourse. Discourse Processes, 6, 107-130.

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

Willows, D. M., & Houghton, H. A. (1987). The psychology of illustrations, Vol. I: Basic research. New York: Springer-Verlag.

Winn, B. (1987). Charts, graphs, and diagrams in educational material. In D. M. Willows &

H. A. Houghton (Eds.), The psychology of illustration. Vol. I: Basic research. New York: Springer-Verlag.

Author note

This research was funded in part by an Air Force Office of Scientific Research grant (89-0367) to Arthur Glenberg. We thank David Hess who helped to write many of the texts and whose senior honors research provided an important foundation for the work reported here. We also thank Peter Kruley who invented the term "working memory management" and who contributed to the development of the theoretical ideas we tested. Requests for reprints should be sent to Arthur Glenberg, Department of Psychology, University of Wisconsin, 1202 West Johnson Street, Madison, WI 53706.

[1] Consider the following as evidence for this claim. Willows and Houghton (1987) contains five chapters surveying the literature on the psychology of illustration. The chapter by Levie (1987) lists 152 citations on the topic of pictures and learning and cognition; the chapter by Levin, Anglin, and Carney (1987) presents a meta-analysis on the functions of pictures in prose based on 100 experiments contained in 87 reports; the chapter by Pressley and Miller (1987) on illustration and oral prose memory lists 83 citations; Peeck's chapter (1987) on the role of illustrations in processing and remembering text lists 136 citations; and Winn's chapter (1987) on charts, graphs, and diagrams lists 126 references. Of course, there is overlap among the bibliographies. Nonetheless, the immensity of the literature can be appreciated given Levie's claim (1987) that Dwyer and his associates (e.g., Dwyer, 1982-83) have published over 200 studies of a single text and set of illustrations.

[2] Because representational elements are pointers, they may represent quite complex objects by pointing to propositions with many embeddings. Thus an element may represent a single sub-atomic particle, a group or class of particles, an atom, molecule, or whatever. The limit is on the number of separate representational elements (or chunks), not what they represent.

[3] When the left-right order of mention in the diagram did not match the second-third order of description in the text, there was a sense of mismatch between the picture and the text. Thanks to Rebecca Glenberg for this pointing this out to us.

[4] The feedback will also encourage the construction of mental models in the linear-picture condition. However, we assume that the perceptual support provided by the picture overrides any tendency to create a model based solely on the text.

Table 1

Example Sequential Text Used in Experiments 1 and 2

Writing a paper

There are four steps to be taken when writing a paper. The first step is to **write a first draft**. To do this, you must *follow an outline* and *disregard style*.

The next two steps should be taken at the same time. One of these steps is to **consider the structure**. You must correct *flaws in logic* and *gaps between main points*.

The other step is to **address the audience**. You should *explain novel terms* adequately and *support bold statements*.

The final step is to **proof the paper** for grammar, punctuation, and style. It is a good idea to have *someone else do this* for you since you may not notice such *surface details*.

Note. Words in boldface are step names, italicized words are facts.

Table 2

Example Non-Sequential Text Used in Experiment 2

Writing a paper

There are four steps to be taken when writing a paper. The next two steps should be taken at the same time. One of these steps is to **consider the structure**. You must correct *flaws in logic* and *gaps between main points.*

The other step is to **address the audience**. You should *explain novel terms* adequately and *support bold statements.*

However, the very first step is to **write a first draft**. To do this, you must *follow an outline* and *disregard style.*

The final step is to **proof the paper** for grammar, punctuation, and style. It is a good idea to have *someone else do this* for you since you may not notice such *surface details.*

Note. Words in boldface are step names, italicized words are facts.

Figure Captions

1. Example of a picture used in the with-picture condition of Experiments 1 and 2.

2. Proportion correct responding to the "Yes" pairs in Experiment 1.

3. Proportion correct responding to the "No" pairs in Experiment 1.

4. Example of a picture used in the linear-picture condition of Experiment 2.

5. Proportion correct responding to the "Yes" pairs in Experiment 2.

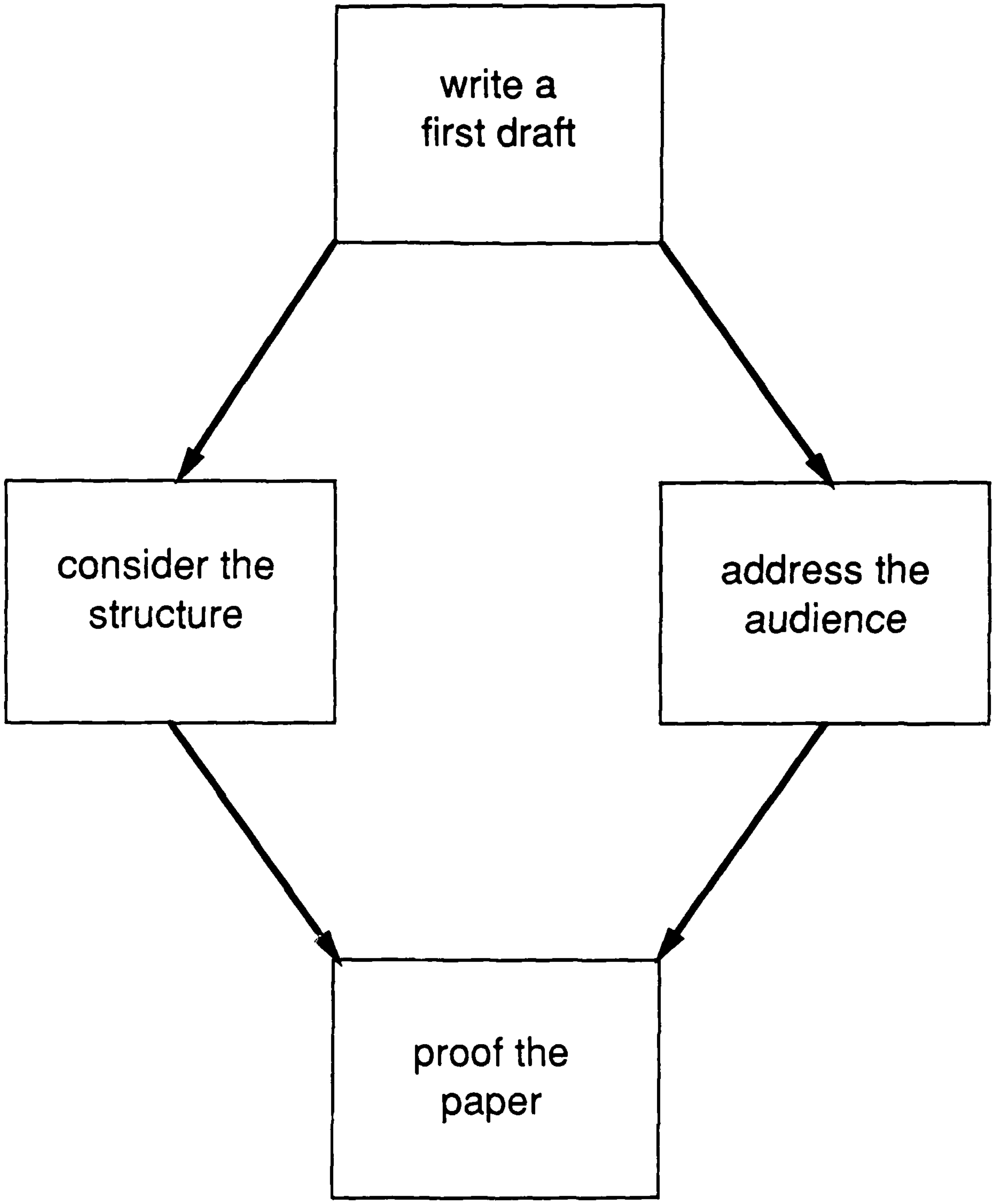
write a first draft
consider the structure
address the audience
proof the paper

Fig 1

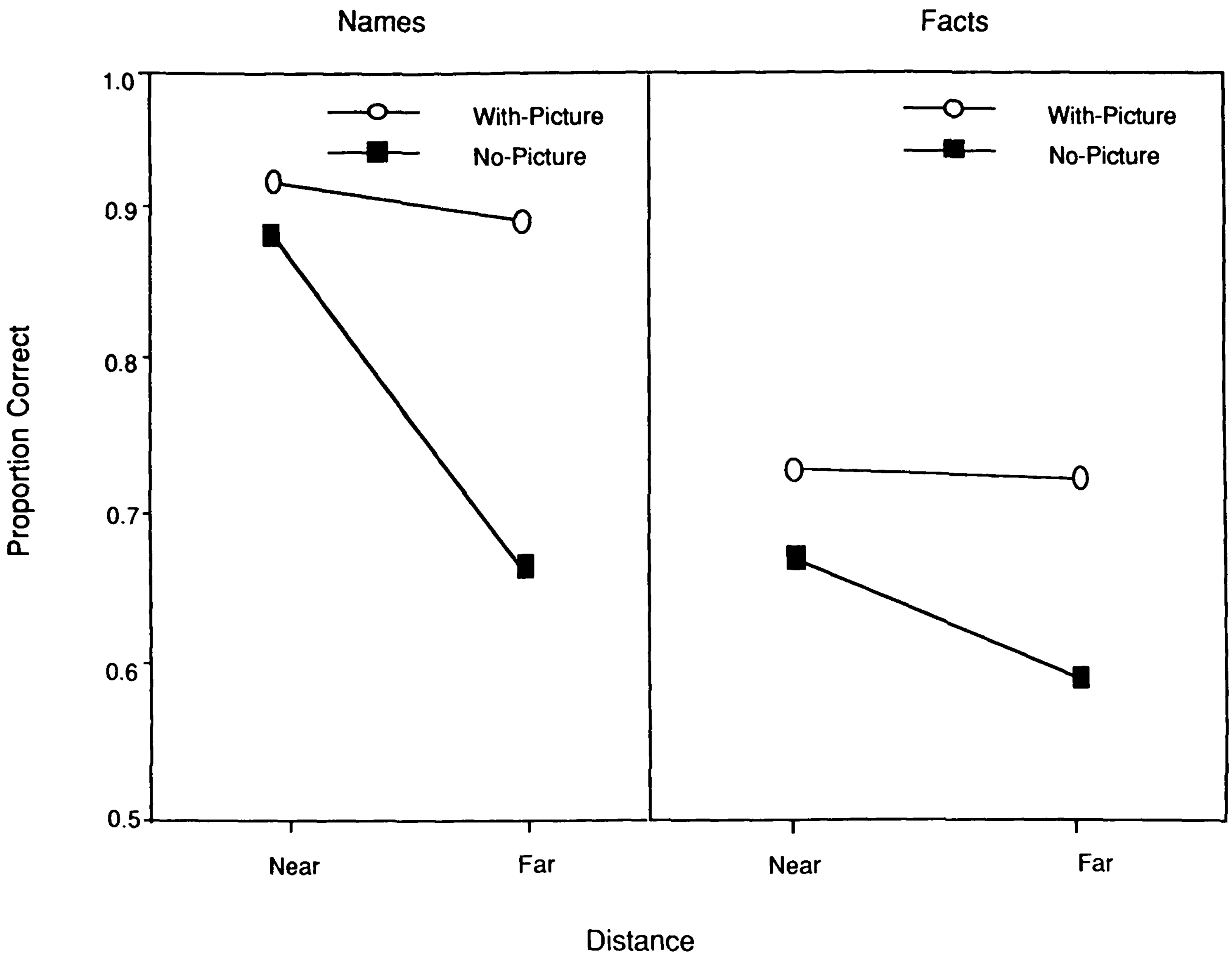
Names
Facts
With-Picture
No-Picture
With-Picture
No-Picture
1.0
0.9
0.8
0.7
0.6
0.5
Proportion Correct
Near
Far
Near
Far
Distance

Fig 2

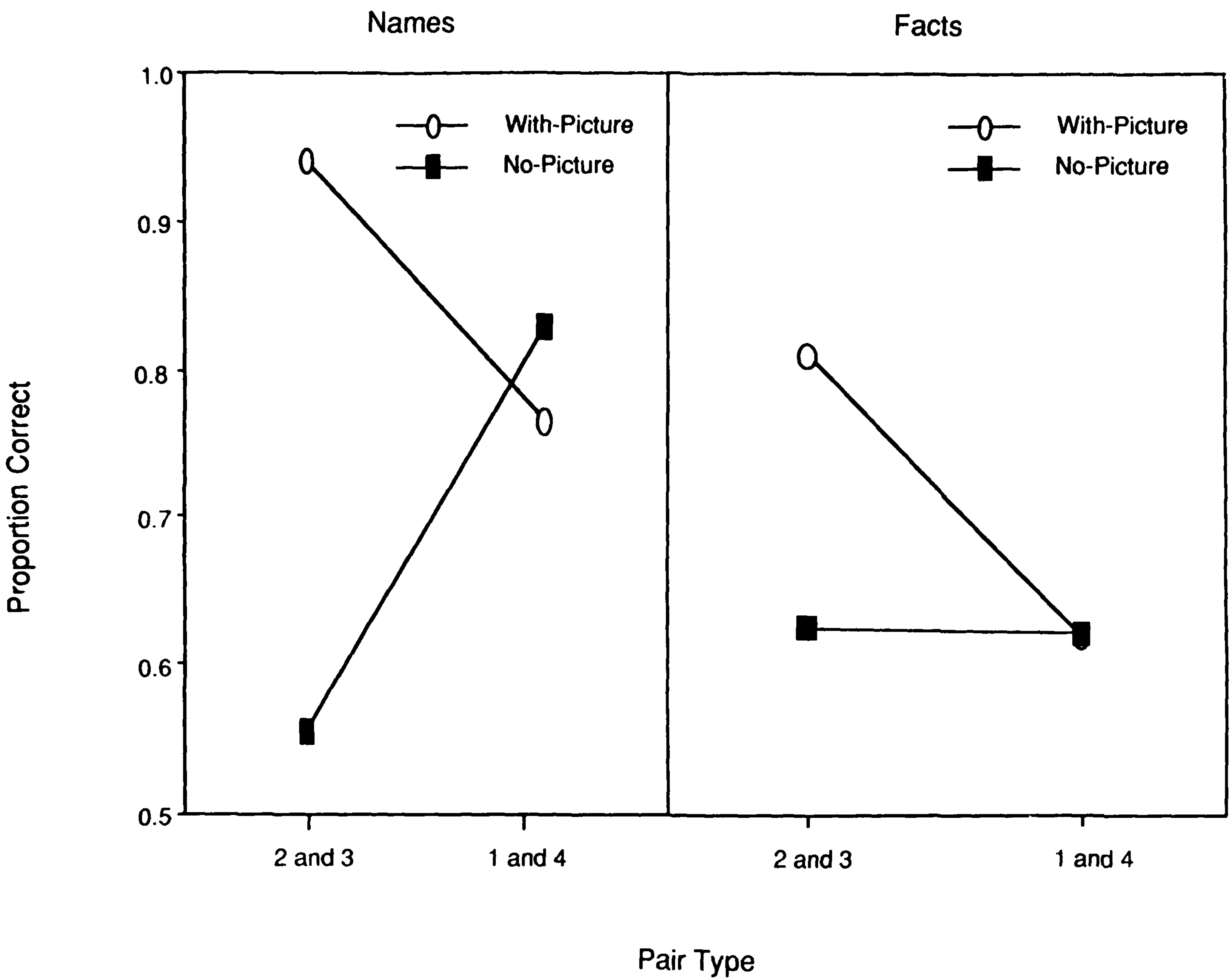
Names
Facts
Proportion Correct
1.0
0.9
0.8
0.7
0.6
0.5
With-Picture
No-Picture
With-Picture
No-Picture
2 and 3
1 and 4
2 and 3
1 and 4
Pair Type

Fig 3

write a
first draft

consider the
structure

address the
audience

proof the
paper

Fig 4

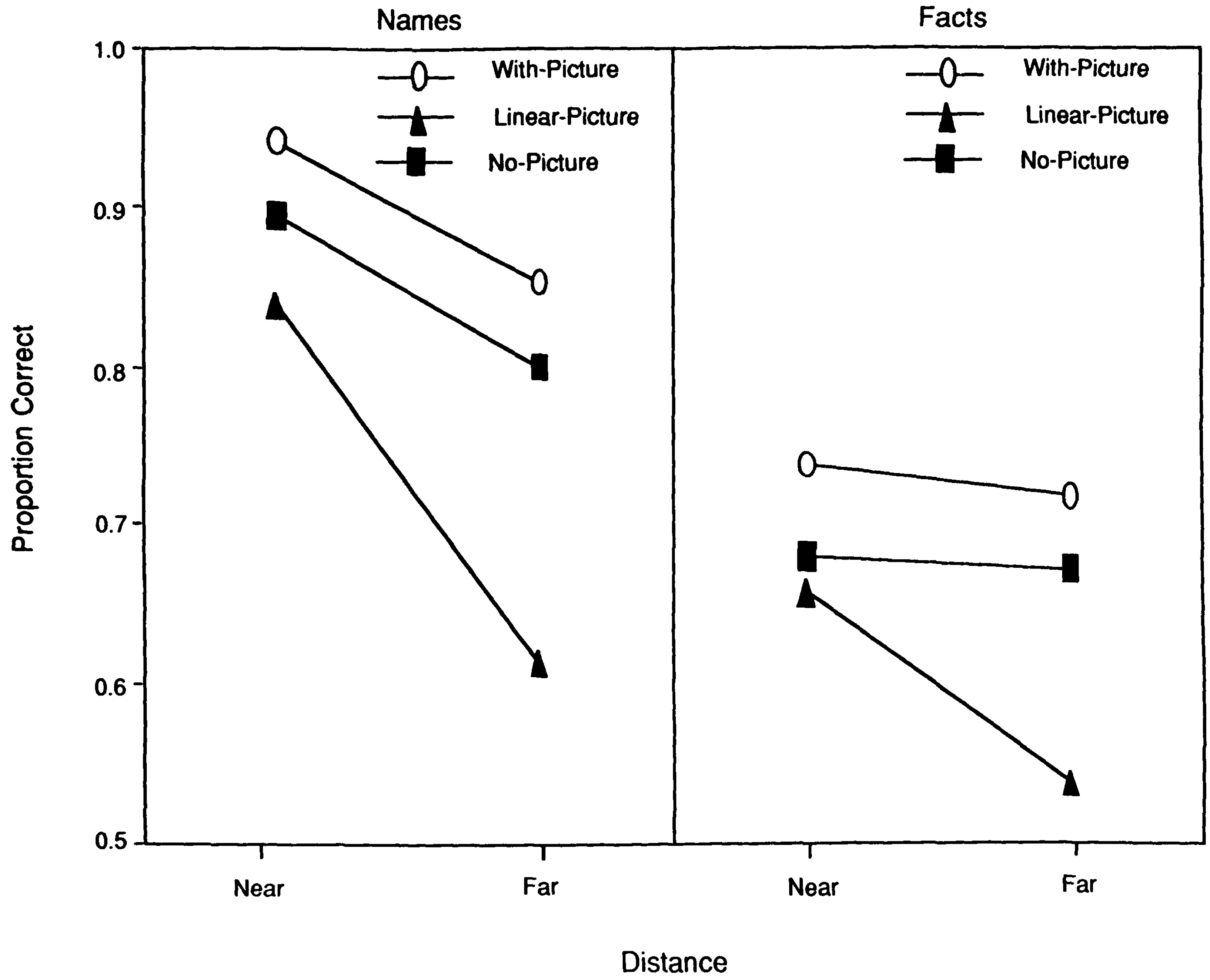
Names
Facts
With-Picture
Linear-Picture
No-Picture
Proportion Correct
1.0
0.9
0.8
0.7
0.6
0.5
Near
Far
Distance

Fig. 5