Constructing Mental Models of Machines from Text and Diagrams

MARY HEGARTY

University of California at Santa Barbara

AND

MARCEL ADAM JUST

Carnegie-Mellon University

Readers' comprehension and eye-fixations are monitored as they read descriptions of simple machines, pulley systems. The comprehension data indicate that readers' comprehension depends on both the medium of instruction and the ability of the reader. The conjunction of text and diagrams particularly facilitated the understanding of how the pulley system moved, whereas either medium alone was sufficient for conveying the system configuration. The eye-fixation data indicate that subjects integrate the information in the text and diagram at the level of individual pulley-system components or groups of connected components. They read the text in increments, often rereading the information about a component or group of components before constructing a spatial mental model of these components with the aid of the diagram. Subjects' diagram inspections vary from local inspections concerned with encoding the relations between two or three components to global inspections concerned with integrating the relations between many components.

People often consult technical and scientific texts to understand how a machine functions, for example, when they want to assemble, operate, or repair the machine. This comprehension process involves constructing an internal representation of the operation of the machine described in the text. The inputs to the comprehension process are the text, the diagram and the reader's prior knowledge, including prior knowledge of the subject matter and strategies for reading technical text. The output is a representation of the referent of the text, which in this case is a mental model of a pulley system. The comprehension processes include reading, diagram interpretation, and coordination of the intake of information from the text and diagram.

This article provides a preliminary account of the process of constructing a mental model of a simple machine, a pulley system, from text and diagrams. We address two basic issues. First, we examine the accuracy of the representations of simple machines that people construct from text, diagrams, and displays combining both media. Second, we examine the comprehension process itself, i.e., how people coordinate their processing of the information in a text and diagram to incrementally construct a mental model of a complex system.

When people read text, they construct representations on several levels, including (a) a representation of the text itself, essen-
tially the propositional content of the text and (b) a representation of the situation or object described in the text (e.g., Just & Carpenter, 1987; Johnson-Laird, 1983; van Dijk & Kintsch, 1983). We refer to these as the text-based representation and the mental model of the referent, respectively. Consistent with previous research, we propose that a mental model is (1) an object-based representation in that elements of the representation are representations of objects or parts of objects and (2) a representation in which information about the configuration of objects is represented directly (Bransford, Barclay, & Franks, 1972; Ehrlich & Johnson-Laird, 1982; Mani & Johnson-Laird, 1982). Mental models are constructed in working memory as a person reads a text by integrating the information presented at different points in the text about the same objects and by elaborating this information using world knowledge or knowledge of spatial constraints.

A growing body of research suggests that constructing a mental model (as opposed to a text-based representation), is facilitated when the text is accompanied by a diagram (Glenberg & Langston, 1992; Hegarty, Carpenter, & Just, 1990; Levie & Lentz, 1982; Mandl & Levin, 1989; Mayer, 1989). Text-and-diagram combinations are particularly effective when the diagrams contain important relational information (Kozma, 1991; Levie & Lentz, 1982; Schallert, 1980), and the learning outcomes require subjects to make inferences or solve problems using a mental model. Text-and-diagram combinations do not facilitate verbatim recall of the text (Mayer, 1989). These studies suggest that people integrate the information in a text and diagram to construct a mental model of their common referent, rather than constructing separate representations of the two media, e.g., a text-based representation and a mental image of the diagram.

Although researchers have theorized about the roles that a diagram might play in a text and diagram combination (Glenberg & Langston, 1992; Hegarty & Just, 1989; Levie & Lentz, 1982; Mayer, 1989), there has been little evaluation of how each of the media contribute individually and conjointly to technical comprehension. Authors of textbooks and technical manuals apparently think that a text-diagram combination is an effective way to communicate how mechanical devices work, because they frequently use this combination. For example, in a previous study, we analyzed a typical introductory manual on basic machines (Bureau of Naval Personnel, 1971) and found that there was a diagram for about every 230 words in the book, and the diagrams occupied approximately 45% of the space in the book (Hegarty & Just, 1989). But would the comprehension of a given device have suffered if the diagram had been omitted, and the readers were required to construct a mental model of the device by visualizing the configurational and kinematic relations between its components? Alternatively, could the readers have figured out how the device works from a diagram alone, without the benefit of the text? These issues are addressed in Experiment 1.

Aside from the question of which medium leads to better understanding, it is also fair to ask how quickly a given level of understanding can be attained from a given medium. It has been theorized that text and diagrams containing the same information are not necessarily equivalent in terms of the processing required to extract the information from the two media (Larkin & Simon, 1987). Thus, from a theoretical perspective, it is interesting to contrast the speed of information acquisition from text versus diagrams in the same spirit that information acquisition has been contrasted in words and pictures (Kroll & Potter, 1984; Potter & Faulconer, 1979; Potter, Kroll, Yachzel, Carpenter, & Sherman, 1986). From a practical standpoint, it is also useful to know whether a text or a diagram would be a more time-effective information source in an emergency situation. Experiment 1 compares the acquisition time for each me-
diurn alone and for the two in combination, to permit such comparisons.

It is plausible that the contributions of text and diagrams to technical understand-
ing depend on additional factors, such as the nature of the information to be ex-
tracted, the mechanical ability of the reader, and the complexity of the device
(Hegarty, Carpenter, & Just, 1990). Consider the issue of what kind of information
is to be extracted. In mechanics, the domain of interest in this article, mental mod-
els contain several different types of information, including information about the
configuration of a machine, i.e., the spatial relations between its components, and in-
formation about kinematics, i.e., the motions of the machine’s components (de
Kleer & Brown, 1984; Hegarty, 1992; Ki-
eras, 1992; Kieras & Bovair, 1984; Metz,
1985). Inspecting a diagram might be suffi-
cient for constructing a representation of
the configuration of a mechanical system,
because a diagram is essentially a static
spatial representation of the system. How-
ever, a static unlabeled diagram does not
directly contain information about the sys-
tem kinematics. Similarly, a text unac-
companyed by a diagram can verbally describe
kinematic relations, but they may be diffi-
cult to visualize in the absence of a dia-
gram. Experiment 1 examines whether the
coordinated processing of a text and dia-
gram might be necessary for constructing a
mental model that includes kinematic infor-

Integration across media. To construct a
mental model from a text and diagram, as
opposed to constructing separate text
based and diagram representations, people
must integrate the information in the two
media, that is, make referential links be-
tween the text and the diagram (Mayer &
Anderson, 1991, 1992; Paivio, 1990) and
combine the information extracted from the
two media. The scanning of the text and the
diagram reflects these integrative pro-
cesses. Although considerable research has
examined how a text is scanned (e.g., Just
& Carpenter, 1987; Rayner and Pollatsek,
1989) and how pictorial stimuli are scanned
(e.g., Friedman, 1979; Loftus & Mack-
worth, 1978; Pollatsek, Rayner, & Hender-
song, 1990; Rayner, 1992) very little is
known about the way text and diagrams are
scanned when they are being compre-
hended in conjunction with each other. In
Experiment 2, we monitored subjects’ eye
fixations to investigate how subjects inte-
grate the information in text and diagrams
to construct mental models of pulley sys-
tems.

The most notable integrative behavior is
the decision to interrupt the reading of the
text to inspect the diagram. First, the fact
that subjects interrupt their reading at all,
suggests that they are constructing an inte-
grated representation of the information in
the text and diagram rather than separate
representations of the two media. Second,
the points at which such interruptions oc-
cur are particularly informative, because
they indicate the approximate size of the
text units that are being related to the dia-
gram. At one extreme, a reader could look
away to the diagram after reading each new
noun phrase that refers to some component
of the pulley system, in order to add the
figural and configural properties of the re-
ferent to his or her mental model. At the
other extreme, a reader could first read the
entire paragraph either to construct a sepa-
rate text-based representation or to con-
struct a preliminary mental model before
inspecting the diagram. The precise pattern
of where the diagram inspections occur,
relative to the reading, should discriminate
between a highly interleaved construction
of a mental model from text and diagrams
versus the construction of separate text and
diagram representations or a text-first con-
struction of a preliminary mental model.

Another question about the coordination
of text and diagram processing concerns
the relation of what is read in the text to
what is inspected in the diagram. Suppose
that in the midst of reading the text, the
reader looks over to the diagram. Since a
diagram represents the configuration of a mechanical system directly, and since this information is a central component of a mental model of a system, we might expect diagram inspections to aid the construction of mental models. One possibility is that upon encountering information in a sentence that is particularly difficult to incorporate into the current mental model, the subject looks over to the diagram for clarification. In this case, one would expect the subjects to inspect the referents of the sentence that was just read. Another possibility is that after reading two or three sentences, a subject attempts to mentally integrate the referential information they impart, and looks at the diagram in aid of the integration. In this latter case, one would expect him or her to inspect components mentioned in these two or three sentences and not just components mentioned in the most recently read sentence. The results of Experiment 2 discriminate between these alternative interpretations of diagram inspections.

Integration within a medium. In addition to integrating the information from the text and diagram, a reader also needs to integrate information within each of the media, i.e., integrate information presented in different parts of the text and also integrate information in different parts of the diagram. First, consider processing of the text. We will refer to an uninterrupted period of scanning the text as a reading episode. Online measures of text processing, such as eye fixations, have revealed that when readers have difficulty constructing an integrated representation of the referent of a text, they often make regressive fixations between the parts of the text that they are attempting to integrate (Just & Carpenter, 1987; Daneman & Carpenter 1983; Ehrlich & Rayner, 1983; Goldman & Saul, 1990). If subjects make regressions primarily within reading episodes, it suggests that they integrate information at the level of the text base before inspecting the diagram. If subjects make regressions primarily across reading episodes, i.e., looking over at the diagram between their first and second reading of a part of the text, it suggests that their integration is not medium-specific. We investigate these possibilities in Experiment 2.

Similarly, the scanning of the diagram can reveal the extent to which subjects integrate information within this medium alone. We refer to a period of uninterrupted scanning of the diagram as a diagram inspection. At one extreme, some diagram inspections might be highly integrative in nature, i.e., have the goal of verifying and elaborating all of the intercomponent relations at the level of the diagram. These diagram inspections would include gazes on all of the diagram components. They are also likely to contain repeated gazes on a given component (as that component is being integrated with other components in the reader's mental model and those other components are also looked at). At the other extreme, inspections of the diagram could have the goal of simply encoding the individual diagram components, analogous to reading words to encode their physical pattern and access their meaning (word encoding and lexical access). These should entail no more than one gaze per component. In Experiment 2, we classify inspections of the diagram along a continuum from simple encoding to global integration.

In summary, we studied the comprehension of pulley-system descriptions in two experiments. In Experiment 1 we investigated the roles of the text, the diagram, and the combined media in constructing mental models of pulley systems. In Experiment 2, we monitored subjects’ eye fixations as they read text accompanied by diagrams to study how they integrated the information in the two media.

Experiment 1

The main purpose of Experiment 1 was to provide baseline data about the roles that text and diagrams play individually and conjointly in the construction of the refer-
ential representation. The effects of different media on comprehension can depend on the ability of the reader, the type of information communicated, and the desired learning outcome (Hegarty & Just, 1989; Kieras & Bovair, 1984; Kozma, 1991; Levie & Lentz, 1982; Mandl & Levin, 1989; Mayer, 1989; Schallert, 1980; Winn, 1987). Therefore it is important to determine how text and diagrams individually affect comprehension of the specific domain and population of interest in this study. We assessed these effects separately for comprehension of the configuration and of the kinematics.

In Experiment 1, subjects studied either text alone, diagrams alone, or text and diagrams conjointly describing the configuration and kinematics of pulley systems. We modeled our stimuli on typical text-and-diagram descriptions of mechanical systems from a technical manual (e.g., Bureau of Naval Personnel, 1971). Figure 1 presents an example of one of the stimuli. Although the text and diagram describe the same object, they are not integrated, e.g., the correspondences between expressions in the text and their referents in the diagram is not made explicit by labeling. It should be pointed out that in no way do we suggest that the stimuli in this experiment are optimal for explaining how mechanical systems work. Instead, we propose that by studying the processes that occur in reading typical texts, we can make prescriptions about what text-and-diagram combinations might facilitate comprehension.

The text and diagram both convey information about configuration, with some redundancy between the two media (see Fig. 1). Thus, we might expect either medium to be sufficient for communicating configuration information. However, subjects who receive the text alone have to visualize the spatial relations between the system components in order to construct a mental model of the device whereas subjects who receive a diagram alone must encode all the relevant spatial relations depicted in the diagram to construct a complete mental model. Thus the required comprehension processes are very different for the two media. By measuring subjects' comprehension, we assessed how well they can accomplish these processes, and by measuring their study times, we assessed how

This pulley system consists of three pulleys, two ropes and one weight. The upper pulley is attached to the ceiling. The other pulleys are free to move up and down. The upper rope is attached to the ceiling at one end, goes under the middle pulley and over the upper pulley, and is free at the other end. The lower rope is attached to the ceiling at one end. It goes under the lower pulley and is attached to the middle pulley at the other end. The crate is suspended from the lower pulley. When the free end of the upper rope is pulled, the rope moves over the upper pulley and under the middle pulley and pulls up the middle pulley. This causes the lower rope to move under the lower pulley and to pull up the crate.

Fig. 1. Sample text-and-diagram description presented to subjects in the experiments (the description of the most complex pulley system).
quickly they can accomplish these processes. Information about kinematics was given only in the text and not in the diagram (see the sample text in Fig. 1). Thus, the kinematic understanding of subjects who received the diagram alone depends on their ability to infer the movement of system components from the static diagram. It is likely that this understanding would be poorer than that of subjects who received the description of kinematics in the text along with the diagram. It is also likely that subjects who received text alone would have poorer kinematic understanding than subjects who received text and diagrams. Previous research has suggested that when verifying sentences describing the kinematics of components in a mechanical system, subjects rely heavily on a diagram of the system as an external memory aid (Hegarty, 1992). Subjects might also need such a memory aid to construct a kinematic mental model from the description of kinematics in the text. If this is the case, we would expect subjects who read text accompanied by diagrams to have better comprehension of kinematics than subjects who read text alone.

In Experiment 1, we also contrasted the comprehension of subjects with high and low mechanical ability. Mechanical ability is highly related to experience with mechanical systems and to physics instruction, suggesting that it probably reflects knowledge of the domain in addition to general reasoning ability (Hegarty, Just, & Morrison, 1988). Readers with more background knowledge have better comprehension of text (Dee-Lucas & Larkin, 1988; Goldman & Duran, 1988; Spilich, Vesonder, Chiesi, & Voss, 1979) and can encode more information from diagrams (Hegarty et al., 1988; Chase & Simon, 1973; Egan & Schwartz, 1979). Thus, we expected that high-mechanical subjects would have better comprehension of each medium.

Method

Materials

Reading materials. Subjects were presented with either text alone, diagrams alone, or text and diagrams describing three different pulley systems. Each text-and-diagram description consisted of an unlabeled diagram of a pulley system and a paragraph listing the components of the pulley system, describing their configuration, and describing the kinematics of the pulley system (see the example in Fig. 1). Each description was presented on a single page.

The text-only group was presented with the same texts but no diagrams. The diagram-only group was presented with the diagrams and no text, except for the statement "the pulley system is operated by pulling on the free end of the rope." This statement was included to cue subjects to consider the kinematics of the pulley system in addition to the configuration shown in the diagram.

The texts described three different pulley systems that differed in complexity, i.e., in the number of components. As is typical in textbooks and technical manuals (e.g., Bureau of Naval Personnel, 1971) the pulley systems were described in order from least complex to most complex for all subjects. The first description was of a pulley system with two pulleys and one rope, the second description was of a pulley system with two pulleys and two ropes, and the third description was of a pulley system with three pulleys and two ropes. The description of the most complex pulley system is presented in Fig. 1. The text length was proportional to the complexity of the pulley system described.

Comprehension measures. Subjects were asked eight comprehension questions after reading or viewing the description of each pulley system. The questions were written so that they could not be answered directly from the text base, in order to assess how
well subjects constructed mental models of the systems. The questions asked about the configuration of components in the pulley system, and the system kinematics. For example, a question about configuration asked “Name all the objects that the lower rope touches” and a question about kinematics asked “How does the lower pulley move when the rope is pulled? (i.e., does it move up, down, or stay in the same place and does it rotate clockwise, counterclockwise, or not rotate).”

Each comprehension question was worth between 1 and 3 points, depending on the amount of information required to answer the question. For example, suppose that after reading the passage presented in Fig. 1, a subject was asked “Name all the objects that the lower rope touches.” He or she would score a point for each of the three correct components listed (i.e., the ceiling, the lower pulley, and the middle pulley). Subjects’ total comprehension score was the sum of their scores on all of the questions. In addition, two comprehension subscores were computed, assessing comprehension of the configuration and kinematics, respectively. The maximum possible score for the configuration and kinematics subscores were 15 and 26, respectively. Two raters scored the comprehension tests independently, and the interrater reliability measures for the two subscores were .94 and .99, respectively.

Individual difference measures. Mechanical ability was measured using an abbreviated form of the Bennett Mechanical Comprehension Test (Bennett, 1969). Spatial ability was measured using the Vandenberg Mental Rotations Test (Vandenberg & Kuse, 1978). Spatial and mechanical ability were highly correlated ($r = .73$), and the effects of these two abilities on the dependent measures were very similar (i.e., there was no case in which there was a significant effect of one ability and not the other). Therefore, only the effects of mechanical ability will be reported.

Subjects and Design

The subjects were 47 students at the University of California at Santa Barbara who participated in the experiment for course credit. Sixteen subjects were assigned to the text-plus-diagram condition, 16 to the text-only condition, and 15 to the diagram-only condition.

Median splits on the Bennett Mechanical Comprehension test scores defined the boundaries between high and low mechanical ability. The median score on the Bennett test was 35 correct out of a possible total score of 48 ($mean = 34.0, SD = 7.9$). There were eight high- and eight low-mechanical subjects in the text-and-diagrams condition, seven high- and nine low-mechanical subjects in the diagram-only condition and eight high- and seven low-mechanical subjects in the text-only condition.

Procedure

Subjects participated in groups of four to six. They were first administered the Bennett Mechanical Comprehension Test and the Vandenberg Mental Rotations Test and then administered a booklet containing the descriptions of the pulley systems and the comprehension questions and a stopwatch to time their reading. The pages of the booklet were arranged so that the page with the comprehension questions about each pulley system followed the description of the pulley system. Each group of subjects was assigned to one condition (receiving either text only, diagrams only, or text and diagrams).

Subjects were instructed to study the description of each pulley system and to try to understand the system so that they would be able to explain to another person how it worked. They were informed that they would later be asked comprehension questions about the systems and that they would be asked to draw the systems. They were allowed to spend as much time as they re-
quired to study each description, but not to look back at a description after they had read it. After studying the description of each pulley system they answered the comprehension questions about that pulley system before reading the next description.

To provide an approximate measure of the time spent studying the pulley system descriptions, the subjects were requested to time themselves using the stopwatch. They were instructed to start the stopwatch as they turned the page to a new pulley-system description and to stop it when they were finished studying the description and were about to turn the page to answer the comprehension questions.

Results and Discussion

Comprehension

Data for the comprehension of configuration and kinematics are presented in Table 1. As might be expected, subjects performed better on questions about configuration than on questions about kinematics, $F(1,41) = 174.42, p < .001$, $MSe = .01$. There was also an interaction between the medium of description and the question type, $F(2,41) = 6.44, p < .01$, $MSe = .01$, indicating that the medium of the descriptions had a greater effect on questions about kinematics than questions about configuration. Specifically the conjoint presentation of text and diagrams was of greater advantage over either medium alone for the comprehension of kinematics than for the comprehension of configuration.

Comprehension of configuration. As Table 1 shows, subjects who studied text and diagrams outscores subjects who read text alone and subjects who viewed diagrams alone on the measure of comprehension of configuration. Planned comparisons indicated that the difference between the text and diagram group and the diagram only group was marginally significant ($t(30) = 2.30, p = .03$), indicating that subjects might not encode all the relevant information from a diagram without direction to the relevant information by an accompanying text. The difference between the text-and-diagram group and the text-only group was not statistically significant ($t(29) = 1.31$) suggesting that subjects can visualize the spatial configuration of the system from the text description without viewing a diagram. Although the required comprehension processes for text alone and diagrams alone are very different, they did not result in different levels of comprehension for the two groups ($t(29) = 1.14$).

We predicted that subjects with high me-

<p>| TABLE 1 |
| MEANS AND STANDARD DEVIATIONS FOR COMPREHENSION SCORES OF SUBJECTS IN EXPERIMENT 1 (EXPRESSED AS PERCENTAGES OF THE TOTAL POSSIBLE COMPREHENSION SCORE) |</p>
<table>
<thead>
<tr>
<th>Medium studied</th>
<th>Text</th>
<th>Diagram</th>
<th>Text and diagrams</th>
<th>All media</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Configuration subscore</strong></td>
<td>Mean (SD)</td>
<td>Mean (SD)</td>
<td>Mean (SD)</td>
<td>Mean (SD)</td>
</tr>
<tr>
<td>High ability</td>
<td>87.5 (10.9)</td>
<td>81.9 (17.9)</td>
<td>96.7 (5.0)</td>
<td>89.0 (13.1)</td>
</tr>
<tr>
<td>Low ability</td>
<td>76.2 (10.8)</td>
<td>72.6 (11.3)</td>
<td>80.9 (15.1)</td>
<td>76.1 (12.4)</td>
</tr>
<tr>
<td>All subjects</td>
<td>82.2 (12.0)</td>
<td>76.7 (14.8)</td>
<td>88.3 (13.9)</td>
<td></td>
</tr>
<tr>
<td><strong>Kinematics subscore</strong></td>
<td>Mean (SD)</td>
<td>Mean (SD)</td>
<td>Mean (SD)</td>
<td>Mean (SD)</td>
</tr>
<tr>
<td>High ability</td>
<td>50.5 (8.3)</td>
<td>58.8 (24.7)</td>
<td>77.4 (13.9)</td>
<td>62.4 (19.7)</td>
</tr>
<tr>
<td>Low ability</td>
<td>28.0 (9.3)</td>
<td>34.1 (24.1)</td>
<td>56.2 (13.6)</td>
<td>39.7 (20.8)</td>
</tr>
<tr>
<td>All subjects</td>
<td>40.0 (14.4)</td>
<td>45.0 (26.7)</td>
<td>66.8 (17.2)</td>
<td></td>
</tr>
</tbody>
</table>
mechanical ability would have better comprehension scores in all conditions than subjects with low mechanical ability. Although Table 1 shows trends in this direction for all groups for the comprehension of configuration, the difference between high- and low-ability subjects was significant only for the text-and-diagram condition \( t(14) = 2.96, p = .01 \).

**Comprehension of kinematics.** Responses to questions about the kinematics of the pulley-systems suggested that subjects needed to study both text and diagrams in order to construct accurate kinematic models. As predicted, subjects who received text and diagrams had higher scores than subjects who received either text alone \( t(29) = 4.70, p < .001 \) or diagrams alone \( t(30) = 2.75, p < .01 \). It is not surprising that subjects who studied text and diagrams had superior comprehension of kinematics than subjects who studied diagrams alone, because information about kinematics was given only in the text. Without the text description of motion, subjects had difficulty inferring the motion of the pulley system components from the static description in the diagram.

It is more remarkable that subjects who studied text and diagrams had superior comprehension of kinematics than subjects who studied text alone, since the kinematic information was given in the text. This result suggests that subjects use the diagram as an external memory aid when constructing a representation of the system kinematics. Without this memory aid, subjects may not have sufficient working memory capacity to both store a spatial representation of the pulley system configuration and transform this representation to imagine the motion of the pulley system. This allows the kinematic information to be mapped onto a spatial representation of the pulley system.

The difference between the comprehension of the text-only group and the diagram-only group was not significant \( t(29) = 0.64 \). Moreover, subjects who studied the two different media understood similar aspects of the pulley systems. The scores of these two groups were compared for each of the individual comprehension questions, producing a significant difference between the text-only group and the diagram-only group on only 1 of the 22 comprehension questions.\(^1\) So although the comprehension processes required for text and diagrams are very different, the two media in isolation seem to have some equipotentiality, as far as the comprehension scores indicate.

Table 1 shows that for all three descriptions, high-ability subjects had better comprehension of kinematics than low-ability subjects. These differences were statistically significant for subjects who studied text and diagrams \( t(14) = 3.08, p < .01 \) and subjects who studied text alone \( t(13) = 4.92, p < .001 \), indicating that high-ability subjects can encode more information about kinematics from the text. The difference in comprehension between high- and low-ability subjects failed to reach statistical significance for subjects who studied diagrams alone \( t(14) = 2.00, p = .06 \).

**Study Times**

We made no specific predictions about study times, so our analyses of these data were more exploratory. As the data in Fig. 2 indicate, the study times varied substantially with medium of description, \( F(2,82) = 31.67, p < .001, MS_e = 2720 \). Post hoc tests indicated that subjects spent less time studying a diagram alone than a text-and-diagram description \( p < .001 \), Neuman–Keuls test) and more time studying a text alone than a text-and-diagram description \( p < .001 \), Neuman–Keuls test). Not surprisingly, the time spent studying a descrip-

\(^1\) The following question about the most complex pulley system, "To what was the lower pulley attached?" All of the subjects who received diagrams answered this question correctly while only 4 of the 16 subjects who received text alone answered it correctly, \( t(29) = 2.18, p < .05 \).
tion increased from the first to the third pulley system (i.e., with the complexity of the pulley system described and the length of the text), $F(2,82) = 39.10, p < .001$, $MS_e = 1292.9$, and the interaction of complexity and medium was statistically significant, $F(4,82) = 8.45, p < .001$, $MS_e = 1292.9$, suggesting that the differences in time spent studying the descriptions increased as the pulley systems became more complex. The effects of mechanical ability on study time were not statistically significant ($F(1,41) = 1.25$).

The study times indicate that subjects spend less time studying a diagram alone than a text alone, although they attain equivalent levels of comprehension. This result is consistent with the view that a mental model of a pulley system is a representation in which the elementary units are representations of the basic components of the pulley system and these elementary units are organized spatially. This is also true of the diagrams which are iconic representations of the pulley systems, depicting their basic components and the spatial relations between these components. The information about each pulley system component is localized to one region of the diagram, facilitating the representation of these components as units (Larkin & Simon, 1987). In contrast, the information about each component of the pulley system is distributed among several different clauses of the text so that subjects have to integrate the information in these different clauses in order to construct the representation of each component. This integration process is investigated in Experiment 2.

Conclusions

Experiment 1 indicated that comprehension of machines from text and diagrams is influenced by both the medium of the description, and the ability of the reader. Subjects did better on the comprehension test if they studied text and diagrams than if they studied either medium alone, and these media effects were largely in the understanding of kinematics. Since the comprehension questions could not be answered directly from the text base, our results support the views that subjects are constructing mental models and that the coordination of text and diagram processing is particularly important to constructing a usable mental model of the motion of components in a pulley system.

Experiment 2

Because readers clearly benefit from the information in both the text and the diagram when they read about mechanical systems, it is meaningful to ask how they integrate the information in the two media to construct mental models of pulley systems. In Experiment 2, we investigated this question by presenting subjects with the combined text-and-diagram descriptions used in Experiment 1, and monitoring subjects’ eye fixations as they read these descriptions.

Experiment 2 examined the level of detail at which the information in a text and diagram is integrated. First, it is possible that information in the text and diagram are not integrated, i.e., that separate representations of the two media are constructed, although the results of Experiment 1 and previous research (Glenberg & Langston, 1992; Mayer, 1989) suggest that this is un-
likely. To distinguish a highly interleaved construction of a mental model from a text-first construction, we analyzed subjects' scanning of the text, how often they looked at the diagram, the points in the text at which they looked at the diagram, and the number of objects fixated during each diagram inspection. If subjects construct a preliminary mental model from the text alone before inspecting the diagram, they should read the entire text before inspecting the diagram. Furthermore, since the information about a given component is presented in different sentences in the text, subjects might need to make regressions while reading the text to combine the information in several different clauses and construct a coherent representation of the referent (cf. Daneman & Carpenter, 1983; Ehrlich & Rayner, 1983; Goldman & Saul, 1990). When they look to the diagram after this reading episode, they should inspect it globally, i.e., gaze at most if not all of the diagram components.

If subjects integrate the information in the text and diagram on a local level, they should inspect the diagram more often, their inspections should occur in midtext and not just at the end of the text, and their inspections should be focused on those particular regions of the diagram that are related to the most recently read section of text. Furthermore, the specific text locations at which subjects inspect the diagram, whether at the end of noun phrases, clauses, sentences or paragraphs, informs us of the size of the text unit that is being related to the diagram.

The differences in comprehension between high- and low-mechanical individuals in Experiment 1 suggested possible variation in how these subject groups process text and diagrams. Experiment 1 suggested that high-mechanical subjects have better comprehension of both a text alone and a diagram alone. Furthermore, high mechanical subjects appear to be better able to locate the relevant information in a diagram so that they are less dependent on an accompanying text to direct its processing (Hegarty et al., 1988; Hegarty & Just, 1989). We propose that the choice between a text-first construction of a preliminary mental model versus a highly interleaved construction of a mental model from text and diagrams is influenced by the cognitive effort required for each. Since high-mechanical subjects might require less effort to construct a mental model from either medium alone, and since switching between processing a text and diagram requires considerable cognitive effort (Sweller, Chandler, Tierney, & Cooper, 1990; Tarzini & Sweller, 1988; Ward & Sweller, 1990) we might expect that high-mechanical subjects would switch less often between the two media than low-mechanical subjects.

Method

Materials

Subjects were presented with the same three text-and-diagram descriptions of different pulley system configurations as were used in Experiment 1. Each description was presented on a single screen, so that the text and diagram were visible at all times during a trial.

A subset of 12 of the comprehension questions from Experiment 1 (4 questions about each pulley system) was used to assess subjects' comprehension of the pulley system descriptions. A reduced set of comprehension questions was used because the question were asked between trials and a long delay between trials can affect the calibration of the eye-tracking equipment. However, this reduced set provided a reliable estimate of subjects comprehension, given that the subset of 12 questions used in both experiments had a correlation of .95 with the total comprehension score in Experiment 1. Unlike Experiment 1 in which some of the questions were multiple choice, all of the questions in this study were open
ended. The maximum possible total comprehension score was 24, the maximum possible scores for questions about configuration was 10 and the maximum possible score for questions about kinematics was 14.

The abbreviated form of the Bennett Mechanical Comprehension Test used in Experiment 1, the Vandenberg Mental Rotations test, and two tests of mental imagery were used to measure subjects' mechanical, spatial, and imagery abilities, respectively. The Reading Span Test (Daneman & Carpenter, 1980) was used to measure reading ability. Because reading span was not significantly correlated with any of the measures of interest in this study, and because the spatial and imagery abilities had very similar effects on the dependent measures to the effects of mechanical ability, only the effects of mechanical ability will be reported.

Apparatus

The stimuli were presented on a DEC VR 260 Monochrome Video Monitor, situated approximately 3 feet from the subject. While subjects studied the text and diagrams, their eye fixations were monitored by a Gulf & Western corneal-reflectance and pupil-center eye tracker and were recorded both digitally and on videotape. The position of the subjects' eye fixation on the stimulus array was measured once every 16.7 ms and output to a Vaxstation II, which also displayed the stimuli on a workstation screen.

Procedure

Before reading the pulley-system descriptions, subjects were administered the Bennett Test of Mechanical Comprehension, the Vandenberg Mental Rotations Test, the Reading Span Test, and the two mental imagery tasks.

They were then given the instructions for the reading task, which consisted of directions for initiating and terminating a reading trial. They were also given a sample text and diagram followed by comprehension questions so that subjects knew in advance what types of information they were expected to acquire from the text. After the eye-tracking equipment was calibrated, the experiment commenced.

Subjects initiated a reading trial by fixating an asterisk on the display and pressing a "start" button and they terminated a trial by pressing the same button when they had finished reading the description. After each trial, subjects were asked comprehension questions about the pulley system described in that trial, before going on to read the description of the next pulley system. As in Experiment 1, all subjects read the descriptions in the same order, beginning with the least complex pulley system and ending with the most complex system.

Subjects and Design

Ten members of the university community at Carnegie-Mellon University were paid to take part in the experiment. Their median score on the Bennett Mechanical Comprehension Test defined the boundary between groups of five high- and five low-ability subjects. The median was 35.5 correct out of a possible total score of 48 ($M = 35.00, SD = 6.14$), which is comparable with the median score of the sample in Experiment 1. Because of equipment failure and measurement error, it was not possible to analyze the eye fixations of one low-ability subject, so this subject's data are omitted from the analyses.

Results and Discussion

The data were analyzed on two levels. At the most aggregate level, analyses of subjects' comprehension scores and study times allowed us to compare the results of Experiments 1 and 2. At the level of process tracing, we identified basic characteristics of subjects' eye-fixation protocols and how these are affected by the complexity of the referent and the ability of the subject.
Aggregate Measures

Comprehension. The comprehension scores replicated the trends observed in Experiment 1 (see Table 2). That is, comprehension scores were higher for questions about configuration and for high-ability subjects, although only the effect of information type (configuration or kinematics) was significant in this experiment, \( F(1,7) = 72.92, p < .001, MS_e = .04. \) The lack of a significant effect of mechanical ability in this study is probably due to the smaller number of subjects, although it may also be due to the open-ended nature of the comprehension questions. Lower scores on comprehension questions were often due to incomplete rather than incorrect answers. For example, when asked to describe the motion of the lower pulley in Fig. 1, a subject who answered “it moves up” received a lower comprehension score than a subject who answered “it moves up and turns counterclockwise.” Thus, comprehension scores reflect subjects’ interpretation of the level of detail that they were being asked to report in addition to the accuracy and completeness of their mental representations.

Study times. As Fig. 2 shows, the study times were comparable to those observed in the text-and-diagram condition in Experiment 1, indicating that the eye-tracking situation did not affect the time spent processing the text and diagrams. As in Experiment 1, study time increased with the complexity of the pulley system, \( F(2,14) = 13.91, p < .001, MS_e = 124.9. \) There was also a marginally significant effect of ability on study times in this experiment, \( F(1,7) = 5.02, p = .06, MS_e = 349.5, \) indicating that low-ability subjects spent more time studying a text-and-diagram description (\( M = 60.71 \) seconds, \( SD = 13.7 \)) than high-ability subjects (\( M = 44.52 \) seconds, \( SD = 8.0 \)).

Process Measures

In order to assess more precisely subjects’ comprehension processes for text and diagrams, their eye fixations were analyzed. At the most detailed level, fixations were aggregated into gazes, i.e., uninterrupted sequences of fixations on a word of the text or a component of the diagram. The components of the diagram were the pulleys, the weight, the floor or ceiling to which the pulley system was attached, and the rope strands running between these points. Gazes on a word of the text were further aggregated into gazes on a clause. Gazes on a diagram component with a duration of less than 100 ms were not included in the analysis because previous research (Loftus, 1981) suggested that people are unable to encode new information from a visual display in less than that time. Gazes on a clause with a duration of less than 250 ms were not included in the analysis because that is the approximate time to read a word (Just & Carpenter, 1987).

To convey the nature of the eye-fixation behavior, Fig. 3 presents a protocol of a typical subject. Each line of this protocol represents a gaze on a clause of the text or a component of the diagram, and the numbers in parentheses at the end of each line indicate the time, in milliseconds, spent on that gaze. In the diagrams on the left side of the figure, the blackened regions show parts of the diagram that the subject fixated during that diagram inspection. Several measures based on such data were used to examine how subjects integrated the information in the text and diagram to construct mental models of the pulley systems. We use the protocol in Fig. 3 to illustrate our discussion of these measures.

Integration across media. The number of

<table>
<thead>
<tr>
<th></th>
<th>High ability</th>
<th>Low ability</th>
<th>All subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Configuration</strong></td>
<td>96.0 (5.5)</td>
<td>90.0 (11.5)</td>
<td>93.3 (8.6)</td>
</tr>
<tr>
<td><strong>Kinematics</strong></td>
<td>58.6 (19.5)</td>
<td>52.5 (17.9)</td>
<td>55.7 (17.9)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>71.2 (14.6)</td>
<td>77.2 (12.4)</td>
<td></td>
</tr>
</tbody>
</table>
Clauses Read:  

**Diagram Component(s) Inspected**

**Reading Episode 1**

1. This pulley system consists of three pulleys, (1763).
2. two ropes, (814).
3. and one weight, (1746).

1. (This pulley system consists of) three pulleys, (666).
4. (The upper) pulley is attached to the ceiling, (1230).

**Diagram Inspection 1**

![Diagram](image)

**Reading Episode 2**

4. (The upper pulley is) attached to the (ceiling), (549).
5. The other pulleys are free to move up and down, (1956).
6. The upper rope is attached to the ceiling at one end, (3524).
7. goes under the middle (pulley), (1081).
8. (The upper rope is attached to the ceiling) at one end, (382).

**Diagram Inspection 2**

![Diagram](image)

**Reading Episode 3**

6. The upper rope is attached to the ceiling at one end, (1797).
7. goes under the middle (pulley), (1031).
8. and over (the upper pulley), (499).

**Diagram Inspection 3**

![Diagram](image)

**Fig. 3.** Protocol of a low-ability subject reading the description of the most complex pulley system (the description in Fig. 1). Each line represents a gaze on a clause of text or component of the diagram. Clauses that are reread are underlined and words in parentheses indicate words that were not directly fixated during a gaze. The diagram components that the reader fixated during each inspection are listed in upper case and are highlighted in the accompanying diagrams.

Times that subjects switch between reading the text and inspecting the diagram provides information about the level of detail at which they integrate information in a text and diagram. In the protocol in Fig. 3, the subject interrupted her reading 8 times to inspect the diagram. On average, subjects interrupted their reading to inspect the diagram 6.07 times per page ($SD = 2.5$), suggesting a highly interleaved construction of the mental model from the text and diagram. Moreover, Fig. 4 shows that the number of diagram inspections increased from the first to the third pulley system description. $F(2,14) = 5.48, p < .01, MS_e = 6.8$. A possible explanation of this effect is that the later descriptions were of more complex pulley systems, so that when there are more components in the pulley system to be represented in a subject’s mental model, subjects switch more often between reading and viewing the diagram.

The text locations at which subjects interrupted their reading of the text to inspect the diagram indicate the approximate size of the text units that are being related to the diagram. Subjects inspected the diagram primarily at the ends of clauses and sen-
Clauses Read: Reading Episode 4 7 goes (under the middle pulley) (648) 6 (the upper rope is attached to) the ceiling at one end (931) 7 goes under the (middle pulley) (831) 6 (the upper rope is attached to) the ceiling at one end (299) 7 goes under the middle pulley, (1148) 8 and over the upper pulley, (1213) 9 and is free at the other end (2362) 10 The lower rope is attached to the ceiling at one end (3111) 11 it goes (under the lower pulley) (299) 12 (and) is attached to the middle pulley (at the other end) (1546)

Diagram Component(s) Inspected:

Diagram Inspection 4 LOWER PULLEY (383)

Reading Episode 5 11 (it goes under) the lower pulley (1946) 12 (and) is attached (to the middle pulley at the other end) (382)

Diagram Inspection 5 RIGHT LOWER ROPE (150) MIDDLE PULLEY (332)

Reading Episode 6 12 (and is attached to the middle) pulley at the other end (1031) 13 The crate is suspended from the lower pulley, (2710)

Diagram Inspection 6 CRATE (316)

FIG. 3—Continued

tences. Whereas 12.3% of the words in the text were at the end of a sentence or a clause, a significantly greater proportion of diagram inspections ($M = 55.6\%, SD = 13.0\%$) occurred at these points ($t(8) = 10.01, p < .001$). Furthermore, 81.8% of diagram inspections ($SD = 14.6\%$) occurred within two words of the end of a sentence or clause, and again this is significantly greater than the proportion of words occurring at these points (36.9%, $t(8) = 9.23, p < .001$). These data replicate previous research (Hegarty & Just, 1989) and indicate that people tend to fully encode a clause or sentence before checking or elaborating the representation of this clause by inspecting the diagram. A clause typically stated a configural or kinematic relation between two components, so the data suggest that subjects inspected the diagram to encode relations between components rather than characteristics of individual components.

The number of diagram inspections increased with the number of clauses in the text, such that the ratio of clauses to diagram inspections did not vary much. The ratio of the number of clauses in the text to the number of diagram inspections was 3.42 ($SD = 1.8$), 3.09 ($SD = 1.9$), and 2.56 ($SD = 0.9$) for the three descriptions, respectively, $F(2,16) = 1.33$, ns. The time spent inspecting the diagram increased from the first to the third pulley system, $F(2,14) = 5.68, p < .05$, $MS_{e} = 24.3$, probably because the diagrams increased in complexity, with subjects spending on average 7.7 s
Reading Episode 7
14 When the free end of the rope is pulled, (2644)
15 the rope moves over the upper pulley, (2345)
16 and under the lower pulley, (865)
17 and pulls up the middle pulley, (1680)
18 This causes the lower rope to move under the lower pulley (3674)

Diagram Inspection 7
LOWER PULLEY (349)

Reading Episode 8
18 (This causes the lower) rope to move under the lower pulley
19 and to pull up the crate. (1064)

Diagram Inspection 8
LOWER PULLEY (200)
MIDDLE PULLEY (964)
UPPER PULLEY (1231)
PULL ROPE (232)
UPPER PULLEY (2095)
RIGHT UPPER ROPE (217)
UPPER PULLEY (366)
RIGHT LOWER ROPE (349)
LEFT UPPER ROPE (316)
LEFT LOWER ROPE (117)
LEFT UPPER ROPE (183)
MIDDLE PULLEY (4073)
RIGHT LOWER ROPE (366)
MIDDLE PULLEY (556)
RIGHT LOWER ROPE (177)
MIDDLE PULLEY (981)
LEFT UPPER ROPE (299)

Fig. 3—Continued

(SD = 6.4), 12.6 s (SD = 8.9), and 15.9 s (SD = 5.05) inspecting the diagram for the three increasingly complex pulley systems.

Diagram inspections. The scanning of the diagram within a diagram inspection can reveal how subjects integrate information using this medium. We identified two different types of diagram inspections, (1) inspections that were short and local to a small number of components of the pulley system and (2) inspections that were longer and more global. We operationally defined a local inspection as an inspection in which three or fewer components of the pulley system were fixated (M = 1.98 components, SD = .36) and a global inspection as

![Diagram of pulley system]

**Fig. 4.** Mean number of diagram inspections for subjects in Experiment 2.
an inspection in which four or more components were fixated ($M = 4.9$ components, $SD = .57$). Thus, in the protocol presented in Fig. 3, the first seven diagram inspections were classified as local inspections and the final diagram inspection was classified as a global inspection. The average length of a global inspection was $4068$ ms ($SD = 1615$) and the average length of a local inspection was $829$ ms ($SD = 199$).

Although the definition of local and global inspections implies a dichotomy, we acknowledge that the local–global distinction is probably more like a continuum, so that the labels “more local” and “more global” might be more appropriate. Nevertheless, there were some qualitative differences between local and global inspections as defined. Subjects were likely to reinspect a component during a global inspection ($M = 1.9$ inspections per component, $SD = .30$) whereas during local inspections they typically fixated a component only once ($M = 1.1$ inspections per component, $SD = .10$). Similarly, subjects spent more time on each component that they inspected during global inspections ($M = 827$ ms, $SD = 156$) than during local inspections ($M = 447$ ms, $SD = 116$). These differences allow us to speculate about the different roles that local and global inspections might play in the comprehension process.

Local inspections probably serve the purpose of constructing the detailed representation of a small section of the pulley system, in particular, establishing coreference between expressions in the text and parts of the diagram and converting the text based representation to a mental model of the referents. Consistent with this interpretation, most inspections classified as local involved fixations on either a single component or on connected or adjacent components (see Table 3), where adjacent components are defined as components that touch. This interpretation is similar to the account of formation inspections proposed by Hegarty and Just (1989) and is supported by data on relations between text and diagram processing that are presented later.

Global inspections are probably more concerned with combining the detailed representations produced by local inspections to construct a complete mental model of the system. This is suggested by the results that global inspections involved repeated gazes on a given component and a longer gaze duration on a given component, as that component is being integrated with the reader’s mental model. Consistent with this interpretation, global inspections constituted a greater proportion of inspections at the end of the text ($M = 53.7\%, SD = 23.2$) than in midtext ($M = 29.1\%, SD = 22.5$, $r(8) = 2.58, p < .05$). Both these data and their interpretation conform to the account of reactivation inspections provided by Hegarty and Just (1989).

The integration of information about different components may be particularly important for constructing a mental model of the movement of system components, which was described at the end of the text on each page. A recent study (Hegarty, 1992) showed that to verify the motion of a pulley system component by viewing a static diagram, subjects inspect not just the component in question, but also inspect the diagram more globally, looking at components that precede it in the causal chain of events. In contrast, they inspect the dia-

![Table 3: Location of Fixations during Local Inspections](image)
gram locally when verifying configural rela-
tions between components.

As Fig. 4 shows, subjects made more global inspections when reading about more complex pulley systems, $F(2,14) = 8.49, p < .01, MS_e = 1.1$. This is consistent with our interpretation of global inspections. If subjects first construct detailed representations of components of a pulley system, the combination of these component representations would be more difficult in the case of pulley systems with more components. The inference of the motion of more complex pulley systems from a static diagram is also more difficult for more complex pulley systems. The effects of complexity on the number of local inspections, $F(2,14) = 1.99$, did not reach statistical significance (see Fig. 4).

Text processing. Although the data suggested a highly interleaved construction of the mental model from text and diagrams, this does not exclude the possibility that processing is text-directed at a local level, i.e., that subjects attempt to integrate the information at the level of the text base before inspecting the diagram. The lines in italics in the protocol in Fig. 3 indicate that the subject often regressed to earlier clauses within a reading episode. This re-reading behavior might reflect failure to un-
derstand the text the first time she read it. It might also suggest that the reader first at-
tempts to integrate the information in dif-
ferent clauses with the same referent, be-
fore interrupting her reading to inspect the diagram.

To examine the integrative function of dia-
gram inspections, we classified regres-
sions as either within or across reading ep-
isodes. Regressions within reading ep-
isodes might be interpreted as attempts to in-
tegrate the information in different clauses at the level of the text base before inspecting the diagram. A substantial pro-
portion of regressions (34.5%) were of this type (see Fig. 5), suggesting that subjects did attempt to integrate information in dif-
ferent clauses of the text, presumably to con-
struct a preliminary mental model of a component or group of components before inspecting the diagram. As Fig. 5 shows, regressions within reading episodes increased with the complexity of the pulley system $F(1,7) = 10.32, p = .01, MS_e = 6.9$.

Regressions across reading episodes might serve the purpose of relocating the subject's place in the text. Consistent with this interpretation, most regressions across reading episodes ($M = 72.0\%, SD = 18.0\%$) occurred immediately after a dia-
gram inspection. This place-finding process may reflect some of the cognitive effort in-
volved in integrating the information in a text and diagram (Sweller et al, 1990; Tarmizi & Sweller, 1988; Ward & Sweller, 1990). The number of regressions across reading episodes also increased monotonically from the first to the third description, i.e., with the complexity of the pulley sys-
tem, $F(1,7) = 10.92, p = .01, MS_e = 9.5$.

Scanning of the text accounted for most of the time spent studying the text-and-diagram descriptions ($M = 76.7\%, SD = 8.2\%$). Not surprisingly, subjects spent more time reading the later texts which...
were both longer and described more complex pulley systems, spending 24.9 s (SD = 7.3), 39.9 s (SD = 14.1), and 40.5 s (SD = 6.8) reading about the three increasingly complex pulley systems, $F(1,7) = 15.53, p < .001$, $MS_e = 45.71$.

**Text direction of diagram processing.** The relation of what is read to what is inspected in the diagram provides information about how the text directed the diagram inspection. Within a local diagram inspection, subjects typically fixated components that were referents of the most recent reading episode. For example, in the protocol presented in Fig. 3, the first diagram inspection is on the upper pulley, which is the referent of the last clause the subject read before inspecting the diagram.

We considered two ways in which a reading episode might direct a local inspection of the diagram. One possibility is that subjects inspect the diagram after they read particularly difficult clauses, i.e., clauses that cannot be easily translated to a mental model on the basis of the text alone. If this is the case they should always inspect the referent of the most recently read clause. Another possibility, suggested by the protocol in Fig. 3, is that subjects attempt to integrate the information in a number of recent clauses of text and use the diagram to construct or to check their referential representation of this integrated information. If this is the case they should inspect the referents of a number of recent clauses, and not just the most recent clause.

On average, 36% (SD = 8%) of the components fixated by subjects during local inspections were referents of the most recently read clause of text. This is greater than the average number of components referred to per clause (23%, $t(8) = 4.66, p < .001$), indicating that subjects’ diagram inspections tended to be directed by this most recent clause. However, a further 44% of fixated components (SD = 8%) were referents of the next two most recently read clauses, so that a total of 80% (SD = 7%) were referents of the three most recently read clauses of the text. Again this is greater than the average number of components referred to in any three consecutive clauses (53%, $t(8) = 12.17, p < .001$). These data suggest that subjects do not inspect the diagram just to construct the representation of the single most recently read clause. They support the view that readers first integrate the representations of related clauses of text and use the diagram to construct or check their mental model of this integrated information.

A possible interpretation of this behavior depends on the assumption that the unit of representation for the text base is the clause whereas the unit of representation for the mental model of the referent is the component. It may be necessary to read and integrate several clauses about a component of a pulley system before a determinate mental model of this component can be constructed (cf. Mani & Johnson-Laird, 1982). Thus, if one or two clauses about a component do not specify a single spatial representation of that component, the reader might postpone the construction of the mental model until more information is read.

In summary, the eye-fixation data suggest that readers construct a mental model of a pulley system incrementally by integrating the information in a text and diagram. Their reading behavior suggests that they read the text in increments, often re-reading the information about a component or group of components to integrate this information at the level of the text base before constructing a mental model with the aid of the diagram. Their diagram inspections are largely concerned with elaborating the configurial and kinematic relations between components described in the text and these vary from local inspections concerned with encoding the relations between two or three components and global inspections concerned with integrating the relations between many components. We now
consider how these processes are affected by mechanical ability.

**Effects of Mechanical Ability**

Low-mechanical subjects should have more difficulty comprehending both the text and the diagram than high-mechanical subjects, and consequently they should be more dependent on the integration of information in the two media. Consistent with these predictions, low-mechanical subjects reread more clauses than high-mechanical subjects, both within \((F(1,7) = 10.32, p = .01, MS_e = 6.9)\) and across \((F(1,7) = 10.92, p = .01, MS_e = 9.5)\) reading episodes. The greater number of regressions within a reading episode suggests that low-ability subjects had more difficulty integrating the information in different clauses of the text independently of the diagram. The greater number of regressions across reading episodes partly reflects the fact that these subjects inspected the diagram more often, so that there were more reading episodes for these subjects. Low-mechanical subjects also spent more time reading the descriptions \((M = 40.5 \text{ s per page}, SD = 8.8)\) than high-mechanical subjects \((M = 30.7 \text{ s}, SD = 4.3, F(1,7) = 4.91, p = .06, MS_e = 130.9)\).

As predicted, and as shown in Fig. 4, subjects with low mechanical ability switched between processing the text and diagrams more often than subjects with high ability, \(F(1,7) = 15.04, p < .01, MS_e = 6.8\). There was also a trend for low-ability subjects to spend more time inspecting the diagram \((M = 14.0 \text{ s per diagram}, SD = 5.6)\) than high-ability subjects \((M = 9.8 \text{ s}, SD = 5.1)\), which did not reach statistical significance, \(F = 1.41\). To gain further insight into the differences in the number of diagram inspections, we assessed the effects of mechanical ability separately on local and global inspections.

Subjects with low mechanical ability made more local inspections than high-mechanical subjects, \(F(1,7) = 15.79, p < .01, MS_e = 5.5\), (see Fig. 5). This result is consistent with the result of Experiment 1 that low-ability subjects had more difficulty constructing the representation of a section of the pulley system from the text alone and suggests that they compensate for that difficulty by inspecting the diagram more often. It is also possible, that once the high-ability subjects have seen the diagram, they are better able to hold its representation in working memory, possibly because this information contained fewer chunks for these subjects (cf. Egan & Schwartz, 1979). In contrast, there was no significant effect of mechanical ability on the number of global inspections, \(F(1,7) = .07\). This result suggests that the major difference in the comprehension processes of low- and high-ability subjects is in the construction of local representations from the text description and not in the integration of these local representations or in the construction of the kinematic model.

**General Discussion**

The research reported in this article provides preliminary accounts of (1) the mental models that readers construct from text and diagrams describing pulley systems and (2) how readers integrate information in text and diagrams to construct these mental models. Experiment 1 indicates that readers need to process both the text and diagram to construct a complete mental model of a pulley system. Experiment 2 suggests that readers integrate the information in a text and diagram at a level corresponding to several clauses worth of text. Thus, they construct mental models of the pulley systems in increments, first integrating information about components of the system to construct local representations of these components, and later combining these local representations to construct a global representation of the pulley system. Both experiments indicate that readers with low mechanical ability have more difficulty constructing mental models of pulley systems from text and diagrams, as indicated by their lower comprehension scores, their
longer study times, and their more frequent regressions and diagram inspections.

The Advantage of Text and Diagrams over Either Medium Alone for Constructing Mental Models

Together, the two experiments suggest why a combined text-and-diagram description might be better than either medium alone. First, consider the advantage of text and diagrams over text alone, shown in Experiment 1 and in previous research (Glenberg & Langston, 1992; Levie & Lentz, 1982; Mandl & Levin, 1989; Mayer, 1989). Experiment 1 showed that this advantage is primarily for kinematic information, i.e., information that was presented in the text alone, so that it is not merely due to the fact that diagrams present additional information or present information in a different format. Experiment 2 showed that diagram inspection is largely text directed, so that the diagram inspection following each reading episode was typically focussed on the referents of that episode. Together, the two experiments suggest that a text-and-diagram combination has an advantage over text alone because the diagram acts as an external memory aid, freeing up working memory resources (cf. Glenberg & Langston, 1992; Kieras, 1992). In the mechanical domain of interest in this study, we propose that these resources are needed primarily to visualize the motion of components.

Next, consider the advantage of text and diagrams over diagrams alone. Experiment 1 showed that there is such an advantage for comprehension of both configuration and kinematics. Considering this result together with the text-directed inspection of diagrams shown in Experiment 2, we propose that although the configuration of a system is directly represented in the diagram, subjects do not always encode this information without direction from an accompanying text. This is consistent with other studies showing that subjects with low mechanical ability do not encode all the relevant information in mechanical diagrams (Hegarty et al., 1988; Ferguson & Hegarty, in press) and that the time they spend inspecting a diagram is proportional to the length of its accompanying text (Hegarty & Just, 1989).

A similar interpretation can be made of the advantage of an accompanying text for understanding kinematic information from diagrams. Although other research has suggested that subjects can infer the motion of components of a mechanical system from information about their configuration alone (Hegarty, 1992), it appears that readers do not carry out these inference processes unless they are directed to do so by the text or other embellishments of the diagram. However, readers do construct more useable kinematic representations of mechanical systems from diagrams if the diagrams are labeled with arrows indicating direction of motion and verbal descriptions of the movement of components (Mayer & Gallini, 1990).

The Incremental Construction of Mental Models from Text and Diagrams

Experiment 2 suggested that subjects integrate the information in a text and diagram at a local level corresponding to a few clauses of text describing a single component or group of connected components. A key question emerging from the present study is how the local units are defined. Several factors might determine the extent of these local representations. First, the components in a local representation are spatially contiguous in the diagram. Second, they are functionally related, because in pulley systems, components that directly affect each other’s motions are spatially contiguous. Third, in the texts that subjects read in these experiments, connected components were typically described in consecutive clauses in the text. It is therefore not clear from this study whether subjects’ incremental construction of the referential representation is being guided by spatial or functional contiguity or by the text. This issue could be investigated further by com-
paring the processing of diagrams in different domains, e.g., electronic diagrams, in which there is less correspondence between spatial contiguity and functional relatedness between components. Future studies might also examine the processing of texts in which consecutive clauses do not describe common or connected components.

The incremental model of construction of the referential representation in a text-and-diagram situation resembles the model of text processing proposed by Kintsch and van Dijk (1978). According to that model, propositions must be activated in working memory at the same time in order to be integrated, and the number of propositions activated at a given time is limited. In this theoretical framework, the local integration process can be seen as a consequence of the capacity limitations of working memory. In order for information in a text and diagram to be integrated, a text-based representation and a diagram-based representation must be activated at the same time. So a text should be able to facilitate integration of the elements of a mental model (which is externally represented by a diagram) by referring to the to-be-integrated elements in the same clause or adjacent clauses.

The local integration results imply that in order to promote comprehension of multimedia presentations such as text accompanied by diagrams, both media should be available simultaneously so that readers can make frequent switches between reading the text and inspecting the diagram. Consistent with this implication, students have been found to have better comprehension of animated diagrams (Mayer & Anderson, 1991, 1992) and instructional movies (Baggett, 1984) when a verbal commentary is presented simultaneously or in close synchrony with the visual presentation. A related result is that for static text and diagrams, improved comprehension results if the text and diagram are spatially integrated, i.e., if the text referring to a part of the diagram is presented close to that part (Mayer, 1989; Mayer & Gallini, 1990; Sweller et al., 1990; Tarmizi & Sweller, 1988; Ward & Sweller, 1990; Whalley & Fleming, 1975). Thus physical contiguity of verbal and visual media in time or space can reduce some of the cognitive load (Sweller et al., 1990) associated with mentally integrating the media.

As is typical of textbooks and instructional manuals, the pulley systems were described in order from least complex to most complex. Several effects of complexity support our view of the incremental construction of mental models from text and diagrams, e.g., the result that readers switch more often between the text and diagram when reading about more complex systems. Since complexity covaried with practice, it is possible that the effects we observed were practice effects rather than effects of complexity. However, if anything, we would expect comprehension of the descriptions to become less effortful with practice, i.e., to be characterized by faster reading and inspection times, fewer diagram inspections and regressions, and the opposite results were observed. Therefore, although it would be wise to separate the effects of practice and complexity in future studies, our present knowledge suggests that these effects are most likely to be due to complexity.

Individual Differences

When a text and diagram are read, the spatial representation of the referent can either be constructed from the text-based representation or it can be encoded directly by inspecting the external spatial representation in the diagram. The individual differences data in the two experiments were consistent with our proposal that a reader's choice in this situation depends on the cognitive effort involved for each, with construction from the text being the less effortful choice for high-ability subjects and construction from the diagram being the less effortful choice for low ability subjects. Ex-
periment 1 indicated that high-mechanical readers are more able to construct spatial representations of pulley systems from the text, suggesting that for these subjects, constructing a mental model from the text is relatively effortless. Thus, in Experiment 2, high-ability subjects inspected the diagram relatively seldom. In contrast, Experiment 1 indicated that low-mechanical readers are less able to construct spatial representations of pulley systems from text, suggesting that for these subjects it is less effortful to switch often to the diagram to construct a spatial representation of a pulley system. Thus, in Experiment 2, they inspected the diagrams more often.

Given that working memory appears to be a limiting factor in the integration of information in a text and diagram, it is surprising that the measure of reading span was unrelated to any of the measures of processing in Experiment 2. One would expect that readers with lower reading spans would have more difficulty integrating the information presented in different clauses about a common referent, as shown by previous research (Daneman & Carpenter, 1983). The sample size in this experiment may have been too small to show such differences. Alternatively, the texts in this experiment may not have taxed working memory, since they were generally coherent in the sense that information about common referents was presented in consecutive clauses.

Generalization to Other Domains

The study of the comprehension of technical text is a very rich research area and it is possible that some of the results of these studies will not generalize to the comprehension of some other types of text-graphic combinations. In general, most graphics are more based on conventions than the simple iconic diagrams studied in this research (Bertin, 1983; Ferguson, 1977) and familiarity with these conventions is particularly important for comprehension (Kolers, 1973). For example, cross-sectional diagrams and exploded views are often used to show the configuration of components in systems such as pumps, gears, or brakes, where the functional components are internal to the device. In viewing these more complex diagrams, an important process is that of differentiating the relevant from the irrelevant information (Canelos, Taylor, & Gates, 1980; Parkhurst & Dwyer, 1983). Diagrams of some other systems, e.g., electronic devices, do not display the physical structure of the system that they depict so that it might be much more difficult to construct mental models of such devices from text and static diagrams. Some recent studies have investigated the potential of computer animations for communicating the operation of such systems (Kieras, 1992; Mayer & Anderson, 1991, 1992). Finally, other types of graphics, such as graphs and flow-charts, depict abstract relations, rather than relations between physical components and support different types of inferences to diagrams of physical systems. Therefore, much more research is needed on the general issue of the comprehension of graphics in text.

In contrast, the proposed process of coordinating reading and diagram inspection has the potential of being domain general. Regardless of the domain of study, it is likely that readers integrate information in text and graphics in local units that are manageable within the capacity of working memory, and that they later combine these local units at a more global level. Future research should study this coordination process for different types of text-graphic combinations.

Conclusion

This research provides important insights into the roles that text and diagrams play in the construction of mental models of machines and into the process of integrating these two media. It suggests a preliminary model of technical comprehension in which the information in a text and diagram is in-
integrated first on a local level and later on a global level. The local integration process might be characterized by the following steps. At step 1, subjects read and interpret a new clause of text. At step 2 they check whether the current clause shares a common referent with other recently read clauses. At step 3 they attempt to integrate the clause representation with the previously stored information about these referents. During this local integration process subjects reread earlier clauses or inspect their referents in the diagram if this is necessary to reactivating or integrate the previously stored information. Although subjects presumably begin by constructing a text-based representation of each clause, we propose that the outcome of the local integration process is a mental model of a component or group of connected components. The global integration process involves reactivating these mental models of components and combining them to construct a mental model of the whole pulley system, including a representation of the motion of the system. We propose that during this process, the diagram serves as an external representation of the mental model, freeing up working memory resources for integrating the mental models of components and imagining the motion of the system.

REFERENCES


(Received August 26, 1992)

(Revision received March 11, 1993)