Individual differences in mental animation during mechanical reasoning

MARY HEGARTY and VALERIE K. SIMS
University of California, Santa Barbara, California

In three experiments we tested the effects of spatial visualization ability on performance of a motion-verification task, in which subjects were shown a diagram of a mechanical system and were asked to verify a sentence stating the motion of one of the system components. We propose that this task involves component processes of (1) sentence comprehension, (2) diagram comprehension, (3) text-diagram integration, and (4) mental animation. Subjects with low spatial ability made more errors than did subjects with high spatial ability on this task, and they made more errors on items in which more system components had to be animated to solve the problem. In contrast, the high-spatial subjects were relatively accurate on all trials. These results indicate that spatial visualization is correlated with accuracy on the motion-verification task and suggest that this correlation is primarily due to the mental animation component of the task. Reaction time and eye-fixation data revealed no differences in how the high- and low-spatial subjects decomposed the task. The data of the two groups of subjects were equally consistent with a piecemeal model of mental animation, in which components are animated one by one in order of the causal chain of events in the system.

Suppose that you are shown a diagram of the brake system in your car and are asked to predict what will happen to each of the components of the system when you step on the brake pedal. What cognitive processes are involved in inferring the movement of the system components from the static diagram? Why are some people more able than others to make these inferences? In this article we describe individual differences in how people infer the motion of parts of a mechanical system when they are given a static display of the system. The experiments reported here provide an analysis of the cognitive processes involved in this mechanical-reasoning task and explore the relation between spatial visualization ability and performance on such a task.

The mechanical-inference processes explored in this paper are central to many real-world tasks, such as reading text-and-diagram descriptions of how machines work, operating complex machinery, diagnosing faults in machines, and designing new machines (Hegarty & Just, 1989; 1993; Kieras & Bovair, 1984; Rouse & Morris, 1986; Williams, Hollan, & Stevens, 1983). These processes have previously been studied by psychometricians as a component of mechanical ability (Bennett, 1969; McGee, 1979; Smith, 1964). Our analysis of the most frequently used tests of mechanical ability revealed that 38% of the items in the Bennett Mechanical Comprehension Test (Bennett, 1969) and 47% of the items in the Mechanical Reasoning Test of the Differential Aptitudes Test (Psychological Corporation, 1990) involve inferring the motion of a component of a mechanical system from a static diagram of the system. In a typical item of this type, subjects are shown a static diagram of a mechanical system (such as a system of belts and pulleys or a gear system) and given the direction of motion of one of the system components (Component A). They are asked to determine the direction of motion of another component (Component B), given that Component A is moving in the direction shown.

This study therefore contributes to our understanding of some of the cognitive components of mechanical ability. By understanding these cognitive components, we will be more able to select students for vocational training and compensate for the deficits of students who are less able.

Experimental Task

In this article we analyze the component processes involved in a mechanical-reasoning task, developed by Hegarty (1992a, 1992b), in which people are shown a diagram of a pulley system and are asked to verify a sentence stating the motion of one of the system components. We refer to this as the *motion-verification task*. An example of this task is shown in Figure 1.

This task is typical of items in mechanical-ability tests, in that subjects are shown a system in which there are multiple interacting parts, and they must infer the mo-

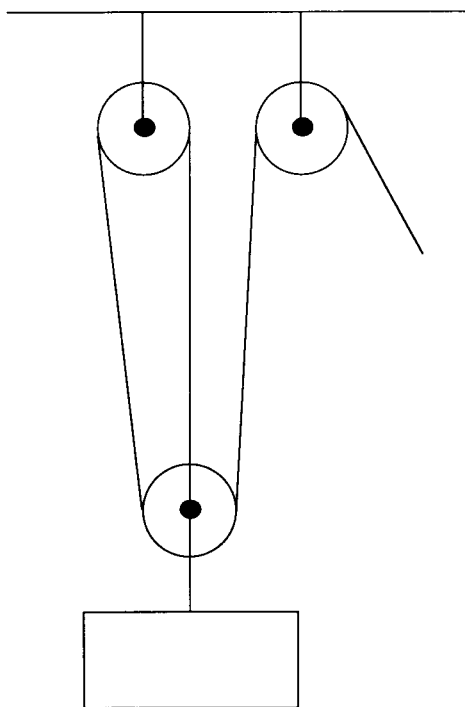The upper left pulley turns counterclockwise

Figure 1. A sample item from the motion-verification task, showing the diagram of Pulley System 1.

tion of one of these parts. Although tests of mechanical reasoning contain items about many types of mechanical systems, including hydraulics, pulleys, gears, and levers, we focused exclusively on pulley problems. This does not compromise the generality of the research, because previous analyses of mechanical-reasoning tests (Cronbach, 1984; Hegarty, Just, & Morrison, 1988) have shown that separate scores for different types of items are highly correlated.

Task analysis. Figure 2 presents a sequence of six steps that a subject might complete in solving the test item presented in Figure 1. These are based on previous models of text–diagram comprehension (Hegarty & Just, 1993; Mayer & Anderson, 1991) and sentence–picture verification (Carpenter & Just, 1975; Clark & Chase, 1972; MacLeod, Hunt, & Mathews, 1978). First, the subject might read the sentence and form an internal representation of its meaning. Second, he/she might inspect the diagram and construct a static spatial representation of the configuration of the pulley system. Third, the subject needs to form referential links between the text and the diagram; that is, it is necessary to identify the component in question in the diagram (e.g., identify which circle in the diagram corresponds to the upper left pulley). Fourth, the subject must infer the motion of the component described in the sentence. Fifth, he/she must compare this inferred motion with the motion described in the sentence—a step that involves converting the two sources

of information to a common representational format. Finally, the subject can make a response.

Thus, we assume that the cognitive processes involved in performing the motion-verification task include *text-comprehension processes*, which occur at Step 1, *diagram-comprehension processes*, which occur at Step 2, *text-diagram integration processes*, which occur at Steps 3 and 5, and *mental animation processes*, which occur at Step 4. In the present research, we examine how each of these component processes contributes to errors and reaction time and how they are related to spatial visualization ability.

The mental animation processes involved in the motion-verification task (Step 4 in Figure 2) have been examined in previous research. Hegarty (1992b) proposed a piece-meal model of mental animation in which subjects infer the motion of components of a pulley system in the order of the causal chain of events in the system. For example, in the problem presented in Figure 1, the subject begins with the knowledge that the free end of the rope is being pulled and infers from this that the rope is moving to the right over the right upper pulley, which then enables him/her to infer that the right upper pulley is turning clockwise. The subject continues to infer the motion of each successive component of the system, until he/she has inferred the motion of the component in question—in this case, the upper left pulley. Note that according to this model, the mental animation component of the task involves two cognitive processes: (1) the process of breaking down the task of animating a whole pulley system into smaller tasks of animating particular "links" in the causal chain, and (2) the process of animating these "links"— that is, inferring the motion of one component from that of an adjacent component.

Consistent with the piecemeal model of mental animation, subjects make few errors and have relatively short response times when asked to verify the motion of components early in the causal chain of events in the pulley system's operation (e.g., the right upper pulley in Figure 1). Their error rates and reaction times increase as a function of the distance of the component in question from the input of the causal chain—that is, the number of pulley system components that they have to animate to solve the problem. Furthermore, eye-fixation data indicate that when verifying the motion of a component, subjects tend to look at that component and earlier components in the causal chain, but not later components (Hegarty 1992a, 1992b). One goal of the present research was to investigate possible individual differences in how subjects break down the mental animation task, by analyzing the reaction times and eye fixations of groups with different abilities.

Spatial Visualization and Mechanical Reasoning

In this study we focused on the differences in performance in the motion-verification test that could be attributed to spatial visualization ability. Previous theoretical accounts of how people infer the operation of a
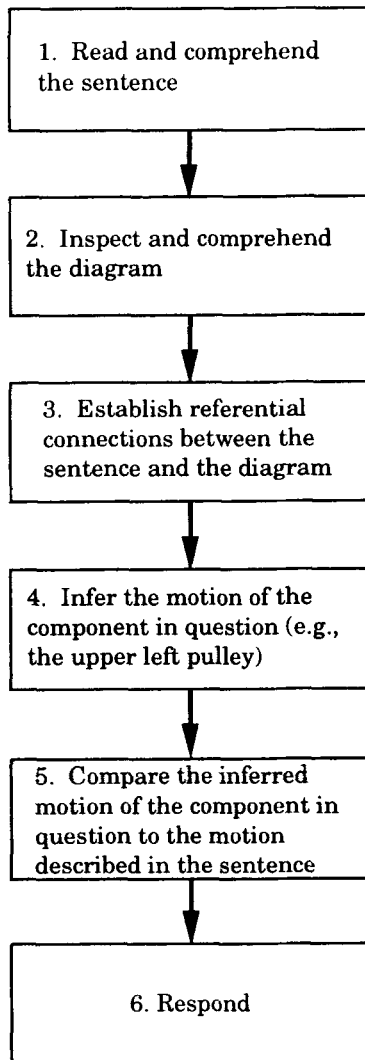
Figure 2. Steps in the solution of an item from the motion-verification task.

physical device have described this as "envisioning," or "running a mental model" (de Kleer & Brown, 1984; Reiger & Grinberg, 1977; Williams et al., 1983). These accounts have been criticized as being vague and involving terminological inconsistencies (Rips, 1986). One of the goals of the present research was to examine whether running a mental model can be linked to the more well-documented theoretical construct of spatial visualization, which was described by Thurstone (1950) as "the ability to visualize a configuration in which there is movement or displacement among the parts of the configuration" (p. 18). If running a mental model of a physical device involves spatial visualization, then individuals who are high in spatial visualization ability should be more successful at such mental-model processing than individuals who are low in spatial visualization ability.

Evidence from the psychometric literature adds support to the view that spatial visualization and inferring the oper-

ation of a physical device share common processes. Performance on spatial visualization tests has a correlation of about .65 with performance on mechanical-ability tests (Bennett, 1969). Tests that load on the spatial visualization factor include surface development tests, paper-folding tests, and form board tests (Ekstrom, French, & Harman, 1976). Our analysis of the motion-verification task and previous task analyses of such spatial tests (Carpenter & Just, 1986; Shepard & Feng, 1972) suggest that the tasks share two abilities. First, they both require the ability to transform spatial representations accurately. Second, they both require the ability to keep track of a series of mental operations in working memory, because both tasks require subjects to make a series of spatial transformations on the given representation; for example, subjects imagine a series of paper folds or the motions of a series of components in a mechanical device. Because the common abilities are specific to the mental animation component of the motion-verification task, we predicted that spatial visualization ability would be particularly predictive of this component process.

Our emphasis on the relation between spatial visualization and performance on the motion-verification task does not imply that we consider spatial visualization to be the only ability that is important in mechanical reasoning. In fact, a previous study stressed the importance of general reasoning ability and domain-specific knowledge in mechanical ability (Hegarty et al., 1988). Rather, our goal in the present study was to identify particular components of mechanical-reasoning ability that are related to spatial visualization ability.

In Experiment 1, we analyze accuracy and reaction time in order to examine individual differences in the motion-verification task and a static version of the task that involves text and diagram comprehension and text–diagram integration, but not mental animation. In Experiment 2, we analyze subjects' eye fixations for a more precise examination of individual differences in the motion-verification task. Finally, in Experiment 3, we analyze performance on the motion-verification task and a version of the task that involves diagram comprehension and mental animation, but not text comprehension or text–diagram integration. In all of the experiments, we measure how spatial ability and other psychometric measures are related to performance.

EXPERIMENT 1

In Experiment 1 we contrast the performance of low-spatial and high-spatial subjects on the motion-verification task and a static version of the task. In this experiment, subjects verified two types of verbal statements about the pulley systems depicted in the diagrams—sentences describing kinematic relations between system components, as described above, and sentences describing static relations between system components, such as "The upper left pulley is attached to the ceiling." To solve the static items, the subjects had to perform Steps 1, 2, 3, and 5 in Figure 2, but not mental animation (Step 4). If spatial

ability is related to the ability to comprehend a sentence, the ability to comprehend a static diagram, or the ability to integrate a text and diagram, then there should be a difference between high- and low-spatial subjects on both the static and the kinematic items. If spatial ability is related to the mental animation component of the task alone, then there should be a difference between high- and low-spatial subjects on the kinematic items only. Furthermore, if there is a difference due to spatial ability in mental animation, then it should be greatest on kinematic items in which the component in question is later in the causal chain, because these items require more mental animation processes and place a greater load on working memory.

We included measures of the three major spatial factors identified in the psychometric literature (Lohman, 1979): spatial visualization, speeded rotation, and spatial orientation. Spatial visualization tests measure the ability to make a series of transformations on a spatial representation. Speeded rotation tasks measure the ability to determine whether a given stimulus is a rotated version of a target stimulus; thus, they require subjects to make a single transformation rather than a series of transformations. Spatial orientation tests measure the ability to determine how an object or scene will appear when viewed from a new perspective. We hypothesized that performance on the motion-verification task would be most highly correlated with spatial visualization.

There are several ways in which spatial ability might affect performance on the motion-verification task. First, low-spatial subjects might make more errors than high-spatial subjects. Second, low-spatial subjects might be slower at the task than high-spatial subjects. Third, the pattern of reaction times might be different, revealing possible differences in how subjects break down the mental animation task. For example, a person with high spatial ability might be able to animate the whole system at once, or to animate a component by just considering the local interactions between that component and the components with which it comes into contact directly. If this is the case, the person's response times to verify motion should be equal at all positions in the causal chain. Thus, the piecemeal model of mental animation (Hegarty, 1992b) might not be equally characteristic of both high- and low-spatial subjects.

Method

Subjects

Fifty-three undergraduate students from the University of California at Santa Barbara participated in the pretesting phase of the experiment. Thirty-seven of these subjects were classified as either "high spatial" or "low spatial" on the basis of their scores on the Paper Folding Test (Ekstrom, French, & Harman, 1976). The high-spatial subjects scored 13 or higher on the test (M = 15.0, SD = 1.50, range = 13.0–19.4), which has a maximum possible score of 20, and the low-spatial subjects scored 11 or lower (M = 9.0, SD = 1.51, range = 6.2–11.0).

Thirty-five of the selected subjects returned to participate in the experiment, but 2 of these failed to follow the instructions correctly, so the final sample consisted of 33 subjects (16 high spatial and 17 low spatial). All of the subjects received course credit for their participation.

Stimuli

Pretests. The subjects were administered the Paper Folding Test-VZ-2 (Ekstrom et al., 1976) as a measure of spatial visualization ability and the Vandenberg Mental Rotation Test (Vandenberg & Kuse, 1978) as a measure of speeded rotation. The Guilford-Zimmerman Spatial Orientation Test (Guilford & Zimmerman, 1953) was used to measure spatial orientation ability. A shortened form of the Bennett Mechanical Comprehension Test (Bennett, 1969), containing 48 of the items in the original test, was used to measure mechanical ability.

Experimental task. Each stimulus was composed of a sentence and a diagram of a pulley system, with the sentence placed above the diagram (see Figure 1). The subjects' task was to verify whether the sentence was true or false with regard to the depicted pulley system. The sentences described either static relations between components of the pulley system or kinematic events that occur when the free end of the pull rope is pulled. An example of a sentence describing a static relation between components in the pulley system depicted in Figure 1 is "The lower pulley is attached to the weight." Examples of sentences describing a kinematic relation are presented in Figure 1 and the Appendix.

The components described in these sentences were components in three different locations in the pulley system, components involved in interactions toward the beginning of the causal chain of events in the system, components in the middle of the causal chain, and components toward the end of the causal chain. Statements about kinematic events either described the motion of a pulley or the motion of a rope over or under a pulley. The example in Figure 1 describes a kinematic relation toward the end of the kinematic chain. The true kinematic statements are listed in the Appendix, in order, from the beginning of the causal chain.

The subjects were presented with the pulley system diagrams, presented in Figures 1 and 3, and the mirror images of these diagrams (producing left and right isomers of the two systems). Half
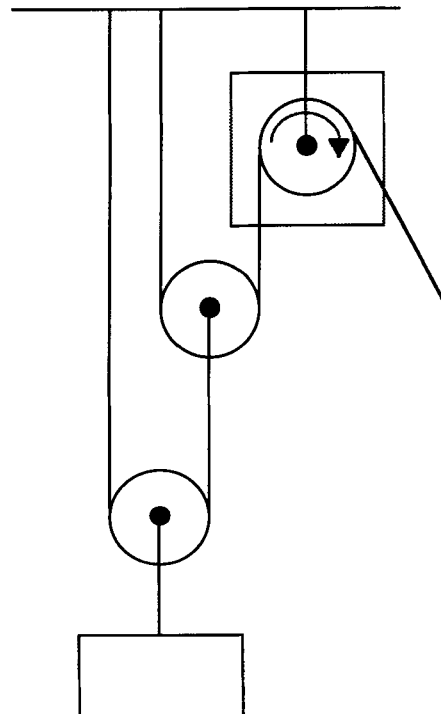

Figure 3. A sample item from the arrow-verification task in Experiment 2, showing the diagram of Pulley System 2.

of the sentences were true and half were false. For each combination of the other factors, there were two sentence-diagram pairs—one describing the motion of a pulley (e.g., "The upper right pulley turns clockwise") and one describing the motion of part of the rope (e.g., "The rope moves to the right over the upper pulley"), yielding a total of 96 unique sentences.

The stimuli were presented in two blocks of 48 trials each. Half of the subjects saw Block 1 first, and half saw Block 2 first. Stimuli were presented in a random order within a block. The false trials were included in order to make the verification task realistic, but it was not possible to standardize the difficulty of the false sentences describing static relations.[1] Thus, only the true trials (48 trials) were analyzed in this experiment.

Apparatus

The stimuli were presented on Macintosh IIci computers using Mindlab software (Meike, Bharucha, Baird, & Stoeckig, 1988). Mindlab is a programming shell that presents stimuli and collects reaction times, with a temporal resolution of 16.6 msec.

Procedure

Pretesting. The subjects were pretested in groups of up to 10. They were administered the standardized tests in the following order: (1) the Guilford–Zimmerman Spatial Orientation Test, (2) the Vandenberg Mental Rotation Test, (3) the Paper-Folding Test, and (4) the Bennett Mechanical Comprehension Test.

Experimental task. The subjects were tested either individually or in groups of 2 or 3. The experiment lasted approximately 30 min. First, the experimenter gave written and verbal instructions as follows. To reduce errors due to failure to understand the diagrams, she showed the subjects diagrams of the two pulley systems that they would see during the experimental trials. Then she introduced them to the terminology used to refer to the different pulley system components ("the upper pulley," "the lower rope," etc.), pointed out each component, and traced the paths of the ropes. She also told them that when a pulley was fixed to the ceiling, a rope, or the weight, they should imagine that there was a metal bar that went through the pulley, on which the pulley rotated, and that this bar was attached to the ceiling, the rope, or the weight by a brace. Thus, it was emphasized that the connection of a pulley to another component did not prevent the pulley from rotating.

The subjects were seated in front of a Macintosh IIci computer. They were told that on each trial they would be asked to view a sentence and a diagram of a pulley system, and that their task was to decide whether the sentence was true or false with regard to the depicted pulley system. They were instructed to work as quickly as possible without sacrificing accuracy.

The subjects were instructed to press the space bar on the computer keyboard to initiate a trial. After the space bar was pressed, the stimulus appeared. The subjects were instructed to read the text, view the diagram, and to indicate their responses by pressing either a *yes* or *no* key (the "k" and "d" keys on the keyboard, respectively). When one of these keys was pressed, the stimulus display was terminated, and a message was presented that instructed the subjects to press the space bar to begin the next trial. The subjects were not given feedback on the accuracy of their performance, because our goal was to measure performance on the motion-verification task, not learning. The subjects were administered 10 practice trials, randomly selected from the 96 experimental trials, and were given an opportunity to ask questions before the experiment commenced. The practice trials were not analyzed.

Results and Discussion

We analyzed the data on two levels. First, we examined the correlations between the ability measures and overall error rates on the experimental task, which are shown in Table 1. Second, we analyzed the contributions

Table 1
Correlations Among the Measures in Experiment 1

Test	Average Error	Test Paper Folding	Bennett	Vandenberg
Paper folding	−.700*			
Bennett	−.493†	.757*		
Vandenberg	−.499†	.549*	.600*	
Spatial orientation	−.486†	.503†	.552*	.665*

*$p < .001$.   †$p < .01$.

of the stimulus factors and the ability measures to the variance in errors and reaction times.

Overall Measures

The overall error rate was .065 ($SD = .056$). As Table 1 shows, this error rate had significant correlations with performance on all of the psychometric measures. This is not surprising, because the measures were also highly intercorrelated. It should be noted that these correlations are probably artificially high, because the pretest subjects in the middle range on spatial visualization ability were not included in the experiment.

Of the spatial ability tests, the Paper-Folding Test was most highly correlated with errors on the motion-verification task. This spatial task has most in common with the experimental task, because it involves making a series of transformations on a spatial representation; thus, it has a working-memory component in addition to the spatial transformation component.

Error Analysis

In a $2 \times 2 \times 3$ analysis of variance (ANOVA), we investigated the contributions of spatial visualization ability (high or low), sentence type (static or kinematic), and location of the referent in the causal chain (beginning, middle, or end) to the variance in errors. The data for the left and right isomers of the pulley system diagrams were collapsed. Because the specific configuration of the two pulley systems interacted with other factors, we analyzed the error rates for the two pulley systems separately. The error data for the different experimental conditions are presented in Figure 4. We first discuss general characteristics of the data and then discuss individual differences in performance.

We predicted that the subjects would make more errors verifying kinematic statements than verifying static statements, because the former items involve the process of mental animation in addition to the processes of text and diagram comprehension and text–diagram integration. Consistent with this prediction, the subjects made more errors on kinematic trials than on static trials [proportion = .01 for static, .11 for kinematic trials, $F(1,31) = 23.82$, $MS_e = .02$, $p < .001$ for Pulley System 1; proportion = .03 for static, .11 for kinematic trials, $F(1,31) = 15.56$, $MS_e = .02$, $p < .001$ for Pulley System 2].

The piecemeal model of mental animation predicts that subjects should make more errors when verifying the mo-
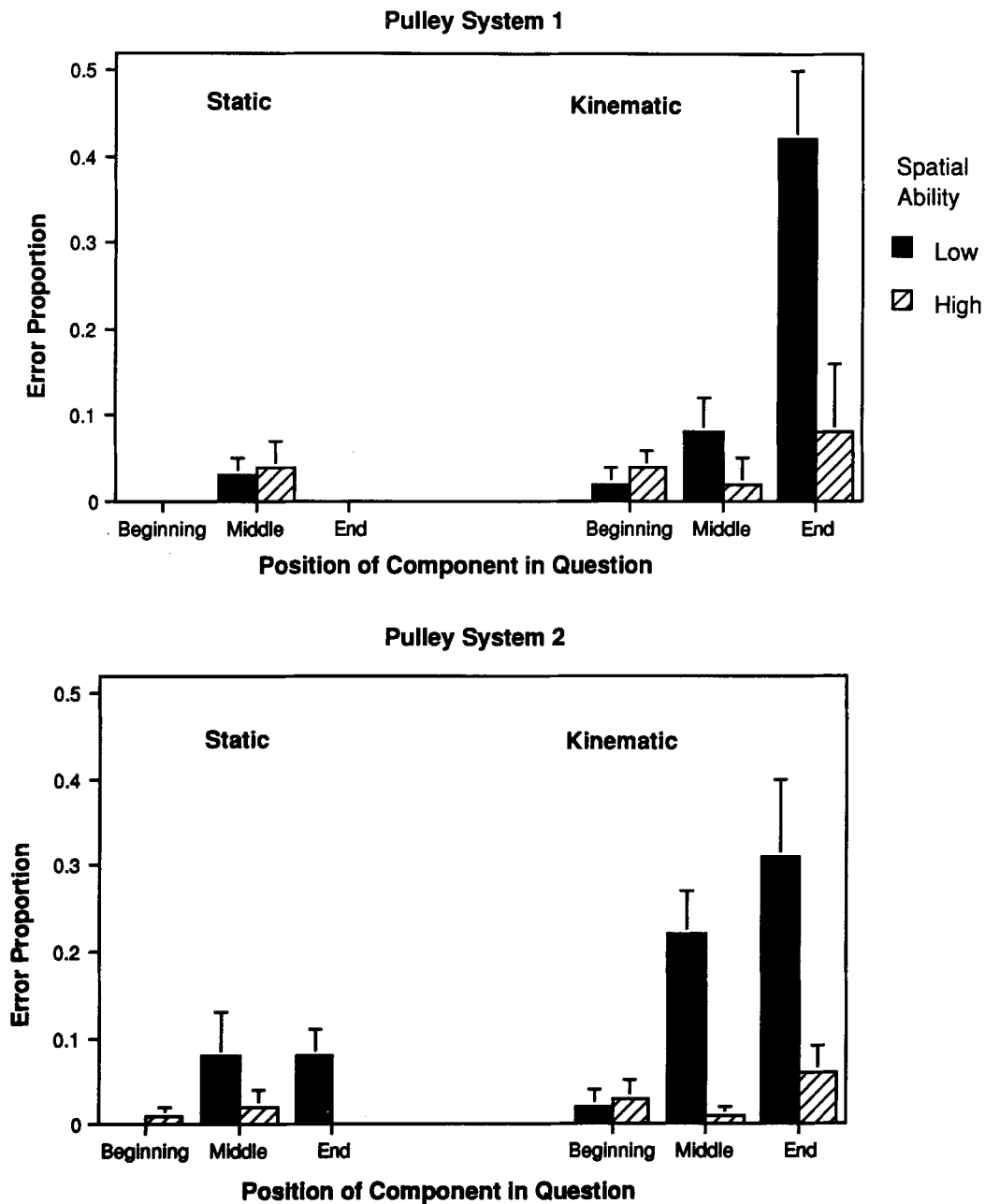
Pulley System 1

Pulley System 2

Figure 4. Mean proportions of errors made by the high- and low-spatial subjects in Experiment 1 in verifying static and kinematic statements describing components at different positions in the causal chain of events in the pulley systems.

tion of components later in the causal chain, because these items require them to animate more components and keep track of more information in working memory. Consistent with this prediction, the subjects made more errors verifying statements about components later in the causal chain of events [$F(2,62) = 11.88, MS_e = .02, p < .001$ for Pulley System 1; $F(2,62) = 7.92, MS_e = .02, p < .001$ for Pulley System 2]. As Figure 4 shows, there was also an interaction between type of statement (static or kinematic) and location of the referent [$F(2,62) = 16.43,$

$MS_e = .02, p < .001$ for Pulley System 1; $F(2,62) = 2.97, MS_e = .02, p = .06$ for Pulley System 2]. Simple effects analyses revealed significant effects of component location on the number of errors in kinematic trials [$F(2,62) = 8.85, MS_e = .03, p < .01$ for Pulley System 1; $F(2,62) = 5.18, MS_e = .03, p < .01$ for Pulley System 2], but not in static trials [$F(2,62) < 1.5$, for both pulley systems].

Individual differences. We predicted that the low-spatial subjects would make more errors on the mental

animation task, and that these errors would be made primarily on trials requiring mental animation—kinematic trials. There was a main effect of spatial ability; low-spatial subjects made a greater proportion of errors than high-spatial subjects [.09 for low spatial and .03 for high spatial, $F(1,31) = 8.27$, $MS_e = .02$, $p < .01$ for Pulley System 1; .11 for low spatial and .03 for high spatial, $F(1,31) = 24.95$, $MS_e = .02$, $p < .001$ for Pulley System 2].

There was a significant interaction between spatial ability and sentence type [$F(1,31) = 9.73$, $MS_e = .02$, $p < .01$ for Pulley System 1; $F(1,31) = 7.50$, $MS_e = .02$, $p = .01$ for Pulley System 2]. This provides information about possible component processes' contributing to the subjects' errors. An effect of spatial visualization ability on both static and kinematic trials suggests that this ability might be related to any of the components in the motion-verification task. An effect of spatial ability on the kinematic trials only suggests that spatial visualization is related to the mental animation processes alone. Simple effects analyses revealed larger effects of spatial ability on performance on the kinematic trials [$F(1,31) = 9.98$, $MS_e = .04$, $p < .01$ for Pulley System 1; $F(1,31) = 19.89$, $MS_e = .03$, $p < .001$ for Pulley System 2] than on the static trials [$F(1,31) < 1$ for Pulley System 1; $F(1,31) = 5.46$, $MS_e = .01$, $p < .05$ for Pulley System 2]. These results suggest that the errors related to spatial visualization were largely in the mental animation component of the task. On static trials, the low-spatial subjects made more errors (.05) than the high-spatial subjects (.02), but only in the case of Pulley System 2. Previous research has suggested that subjects have particular difficulty comprehending the diagram of Pulley System 2, especially the middle pulley of this system, which is both supported by the upper rope and connected to the lower rope (Hegarty, 1992b).

If low-spatial subjects have particular difficulty with the mental animation component of this task, they should make more errors when asked to verify the motion of components later in the causal chain of events in a pulley system. This was the case. For Pulley System 1, their error proportions were .02, .08, and .42, for the beginning, middle, and end of the causal chain, respectively. For Pulley System 2, their error proportions were .02, .22, and .31 for these respective locations. Spatial visualization ability interacted with location of the referent in the causal chain [$F(2,62) = 7.90$, $MS_e = .02$, $p < .001$ for Pulley System 1; $F(2,62) = 7.27$, $MS_e = .02$, $p < .01$ for Pulley System 2]. There was also a significant three-way interaction between ability, sentence type, and pulley location for Pulley System 1 [$F(2,62) = 7.83$, $MS_e = .02$, $p < .001$]. Simple effects revealed a significant interaction of spatial visualization with location of the referent for kinematic trials [$F(2,62) = 8.85$, $MS_e = .03$, $p < .001$ for Pulley System 1; $F(2,62) = 5.18$, $MS_e = .03$, $p < .01$ for Pulley System 2], but not for static trials.

In contrast to the low-spatial subjects, the high-spatial subjects had relatively low error rates (<3% overall). A separate post hoc ANOVA of the errors of these subjects alone revealed no significant effects of sentence type [$F(1,16) = 3.15$, $MS_e = .004$, n.s., for Pulley System 1; $F(1,16) = 2.79$, $MS_e = .01$, n.s., for Pulley System 2], pulley location ($F < 1$, for both pulley systems), or the interaction of these two factors [$F(2,32) = 1.40$, $MS_e = .006$, n.s., for Pulley System 1; $F(2,32) = 1.71$, $MS_e = .02$, n.s., for Pulley System 2]. These results suggest that high-spatial subjects are largely successful at the mental animation task and that the observed effects of sentence type and location were due to errors by the low-spatial subjects.

Reaction Time Analysis

The reaction time data are shown in Figure 5. The data from trials in which the subjects made errors and data points greater than three standard deviations above the mean (3.3% of trials) were omitted from the analyses. Again, the reaction times for the two pulley systems are analyzed separately, because the pulley system configuration interacted with other effects. In a $2 \times 2 \times 3$ ANOVA we assessed the effects of spatial ability, statement type, and component location on reaction times.

Consistent with our analysis of the motion-verification task, the subjects spent more time verifying kinematic statements than verifying static statements [$F(1,31) = 172.96$, $MS_e = 4.1$, $p < .001$ for Pulley System 1; $F(1,31) = 173.46$, $MS_e = 2.8$, $p < .001$ for Pulley System 2]. Presumably, the additional time spent on kinematic items reflects time to mentally animate the system.

The piecemeal model of mental animation predicts an increase in reaction time with distance from the beginning of the causal chain for verification of kinematic but not static items. Consistent with this prediction, there was a significant effect of location of the referent on reaction time [$F(2,62) = 29.21$, $MS_e = 2.9$, $p < .001$ for Pulley System 1; $F(2,62) = 18.58$, $MS_e = 3.13$, $p = .001$ for Pulley System 2]. This factor interacted with the type of statement to be verified (static or kinematic) [$F(2,62) = 14.08$, $MS_e = 3.6$, $p < .001$ for Pulley System 1; $F(2,62) = 24.67$, $MS_e = 3.2$, $p = .001$ for Pulley System 2]. As Figure 5 shows, the reaction times for kinematic trials increased from the beginning to the end of the causal chain [beginning = 7.1 sec, middle = 8.7 sec, end = 11.1 sec, $F(2,62) = 24.99$, $MS_e = 5.27$, $p < .001$ for Pulley System 1; beginning = 7.1 sec, middle = 8.6 sec, end = 10.8 sec, $F(2,62) = 24.03$, $MS_e = 4.87$, $p < .001$ for Pulley System 2]. However, this trend was not observed in the verification times for statements about static relations [beginning = 5.1 sec, middle = 4.9 sec, end = 5.5 sec for Pulley System 1; beginning = 5.3 sec, middle = 6.6 sec, end = 5.2 sec for Pulley System 2].

Individual differences. There were no differences between the high- and low-ability subjects in reaction time, nor did spatial ability interact with any of the other factors in the experiment. Taken together with the error data, these data suggest that high- and low-spatial subjects use similar strategies to break down the motion-verification task, but that high-spatial subjects are more accurate in carrying out these strategies.
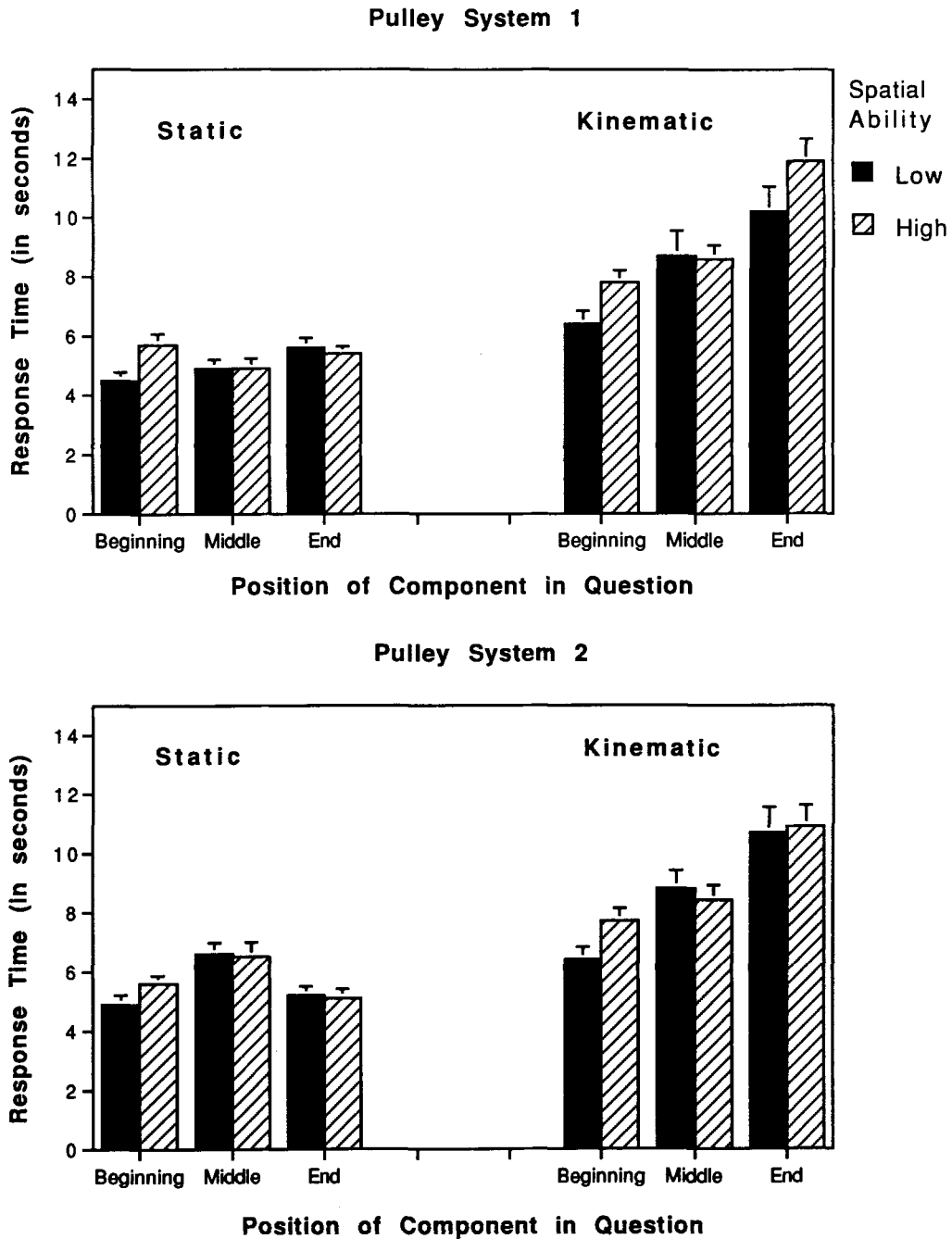
Pulley System 1

Pulley System 2

Figure 5. Mean response times of the high- and low-spatial subjects in Experiment 1 in verifying sentences describing static and kinematic relations between system components at the beginning, middle, and end of the causal chain in the pulley systems.

Conclusions

The results from Experiment 1 supported our prediction that spatial ability is related to performance on the motion-verification task. The differences between the low- and high-spatial subjects were greatest on the items that required the most mental animation processes, suggesting a striking effect of spatial ability on mental animation. Spatial ability was also related to performance on the static items, suggesting that this ability is also related to diagram comprehension.

In contrast, the results provided no evidence that the low- and high-spatial subjects used different strategies to break down the mental animation task. Reaction time data for both groups of subjects were equally consistent with the view (Hegarty, 1992b) that subjects mentally animate pulley systems by inferring the motion of components in

order of the causal chain of events in the machine. Thus, the effects of spatial ability appear to be largely in how accurately subjects mentally animate particular "links" in the causal chain. The observed data may have resulted from subjects' incorrect inferences of the motion of some components, or a failure to keep track of all of the component motions in working memory when the causal chain was relatively long.

The observed pattern of individual differences in accuracy but not in reaction time is consistent with research showing that, for complex spatial visualization tasks, speed of processing is a poor predictor of an individual's accuracy, and that accuracy is the more discriminating measure between high- and low-spatial individuals (Egan, 1976; Lohman, 1979).

EXPERIMENT 2

Experiment 1 showed that high-spatial subjects are more accurate in performing the motion-verification task, although their reaction times provided no evidence that they used different strategies to break down the task. In Experiment 2 we explored further possible evidence for strategy differences by monitoring subjects' eye fixations as they performed the motion-verification task.

It is possible that the poorer performance of low-spatial subjects on diagram comprehension and mental animation occurs because they spend less time on these processes and relatively more time on text comprehension. If this is the case, then low-spatial subjects should spend less time inspecting the diagram and more time inspecting the text than high-spatial subjects. Alternatively, low-spatial subjects might spend more time than high-spatial subjects on the diagram, as they struggle to understand it. The eye-fixation data allowed us to observe how subjects of different abilities allocated their time among these component processes.

The eye-fixation analysis also provided more precise data on how subjects break down the mental animation task. According to the piecemeal model, when asked to infer the motion of a component, subjects infer the motion of components before this component in the causal chain, but not components after it. Therefore, they should inspect the sentence, the referent, and components before the referent, but not components after the referent in the causal chain. Previous research has shown that subjects' eye-fixation data are consistent with this prediction, although the subjects spent a small amount of their time inspecting components after the referent in the causal chain. In Experiment 2, we investigated whether low-spatial subjects would spend more time inspecting these less relevant parts of the display.

Method

Subjects

Twenty-one students from the University of California at Santa Barbara participated in the experiment for course credit. They were selected from a pool of subjects who were pretested on the Paper Folding Test (Ekstrom et al., 1976). Ten of the subjects had paper-folding scores of 13 or above (M = 16.6, SD = 1.6, range = 13.8-19), and 11 had scores of 8 or below (M = 6.0, SD = 2.2, range = 1.6-7.8). One of the low-spatial subjects (with a score of 6.6 on the Paper Folding Test) had below-chance performance on the mental animation task, suggesting that he misunderstood the instructions; his data were not included in the analyses.

Stimuli and Apparatus

The stimuli consisted of the 48 kinematic trials used in Experiment 1. In this experiment the display differed, in that each stimulus was composed of a sentence on the left of the screen and a diagram of a pulley system on the right. As before, the subjects' task was to state whether the sentence was true or false with regard to the depicted pulley system. The subjects were presented with the stimuli twice, in two blocks of 48 stimuli. The stimuli were presented in a random order within a block.

The stimuli were presented on a DEC VR 260 Monochrome Video Monitor, situated approximately 1 m from the subject. The subject's eye fixations were monitored with an Iscan corneal-reflectance and pupil-center eye tracker (Model RK-426), which has a resolution of less than 1° of visual angle. The tracker sampled the position of the subject's gaze every 16 msec, and output the x and y coordinates to a DEC Vaxstation 3200. Further processing of the data is described below.

Procedure

First, the experimenter presented the same written and verbal instructions that were used in Experiment 1. A headrest was then fitted comfortably to the subject's head; the subject was asked to move as little as possible during the experiment. After the eye-tracking equipment had been calibrated, the subject was asked to fixate an asterisk that appeared in the top left corner of the screen and to press a button in order to begin and end each trial. As soon as the button was pressed, a sentence and diagram appeared on the screen. The subject viewed the sentence and diagram and responded either "true" or "false" verbally, and by pressing one of two buttons, marked "true" and "false." These two responses were later checked for consistency, and if they were inconsistent, the verbal response was taken as the subject's response. The subjects were given 10 practice trials before the experiment commenced.

Results

Error Analysis

The overall error proportion was .097, and the error rate had a correlation of .615 (p < .05) with scores on the Paper-Folding Test. We analyzed the contribution of spatial visualization ability (low or high) and location of the referent in the causal chain (beginning, middle, or end) to the variance in errors. As in Experiment 1, the data for the two pulley systems were analyzed separately.

As Figure 6 shows, the low-spatial subjects made significantly more errors than the high-spatial subjects [$F(1,18)$ = 5.61, MS_e = 1.3, p < .05 for Pulley System 1; $F(1,18)$ = 5.43, MS_e = 1.7, p < .05 for Pulley System 2]. These main effects of spatial visualization ability replicate the results of Experiment 1.

Consistent with Experiment 1 and with previous research, the subjects made more errors verifying statements about components later in the causal chain of events [$F(2,36)$ = 11.25, MS_e = 0.7, p < .001 for Pulley System 1; $F(2,36)$ = 6.84, MS_e = 1.2, p < .01 for Pulley System 2]. There was an interaction of spatial visualization ability with position in the causal chain [$F(2,36)$ = 15.17, MS_e = 0.7, p < .001 for Pulley System 1;
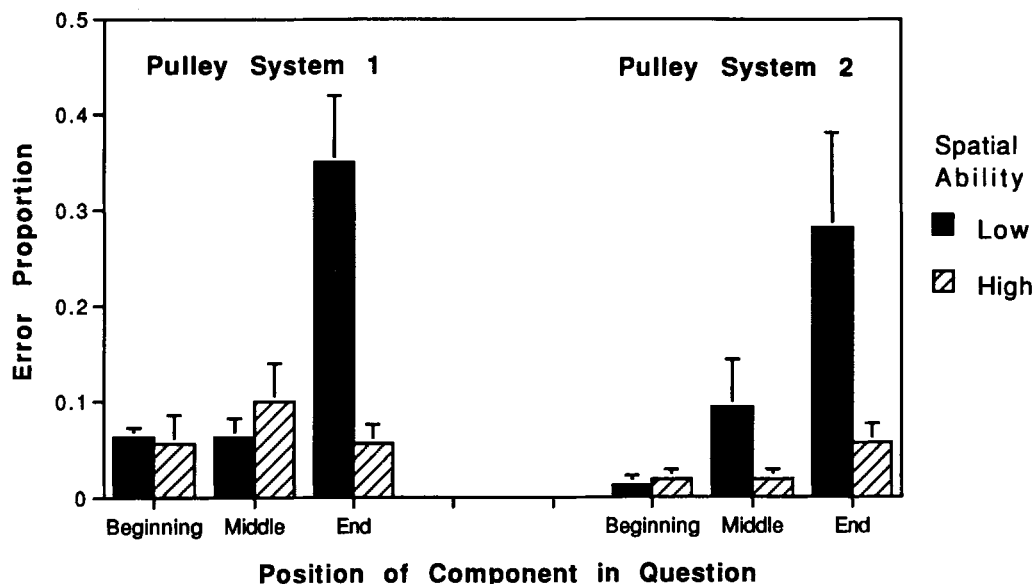
Figure 6. Mean proportions of errors made by the high- and low-spatial subjects in Experiment 2 in verifying kinematic statements describing components at different positions in the causal chain of events in the pulley systems.

$F(2,36) = 3.74$, $MS_e = 1.2$, $p < .05$ for Pulley System 2]. Simple effects analyses indicated that for both pulley systems, the error proportions for the low-ability subjects increased with the distance of a component from the beginning of the causal chain of the device [beginning = .06, middle = .06, end = .35, $F(2,36) = 25.82$, $MS_e = 0.7$, $p < .001$ for Pulley System 1; beginning = .01, middle = .09, end = .28, $F(2,36) = 10.32$, $MS_e = 1.2$, $p < .001$ for Pulley System 2], whereas location had no significant effects on the errors of the high-ability subjects ($F < 1$, in both cases). As Figure 6 shows, the error rates of the high-ability subjects were low (.05 overall). These data are consistent with the interpretation that low-spatial subjects have difficulty with the mental animation process, so they make more errors on trials that require more mental animation steps, whereas high-spatial subjects are largely successful at mental animation.

Reaction Time Analysis

The reaction time data are shown in Figure 7. The data from trials in which the subjects made errors and data points greater than three standard deviations above the mean (< 1% data) were omitted from the analyses. In a 2×3 ANOVA we assessed the effects of spatial visualization ability (high or low) and component location (beginning, middle, or end) on reaction time for the two pulley systems.

The reaction time data were consistent with the results of Experiment 1. Although there was a trend for the low-spatial subjects to have longer reaction times, the effects of spatial ability did not reach statistical significance for either Pulley System 1 [$F(1,18) < 1$] or Pulley System 2 [$F(1,18) = 2.54$, n.s.]. Furthermore, the location of a component in the causal chain did not interact significantly

with spatial ability [$F(2,36) = 1.82$, n.s., for Pulley System 1; $F(2,36) < 1$, n.s., for Pulley System 2], suggesting a similar increment in reaction time for the high- and low-ability subjects with each additional component to be mentally animated. Consistent with the piecemeal model of mental animation and with previous results, reaction time increased with the distance of a component from the beginning of the causal chain of events [beginning = 7.3 sec, middle = 7.8 sec, end = 10.0 sec, $F(2,36) = 51.61$, $MS_e = 1.6$, $p < .0001$ for Pulley System 1; beginning = 7.3 sec, middle = 8.8 sec, end = 10.5 sec, $F(2,36) = 26.29$, $MS_e = 3.9$, $p < .001$ for Pulley System 2]. Taken together with the error data, these results are consistent with the interpretation that high- and low-spatial subjects use similar strategies to break down the motion-verification task, but that high-spatial subjects are more accurate at carrying out these strategies.

Eye-Fixation Analysis

To investigate whether high- and low-spatial subjects would allocate their response time to different parts of the display, we aggregated the subjects' eye fixations to gazes, consisting of a single fixation or group of fixations on a particular sector of the display. The display sectors were rectangles enclosing the text, and the individual pulleys, rope strands,[2] weight, and sections of the ceiling were defined so that each section contained a single connection to a rope or pulley. Each sector enclosed a single pulley system component, with the exception that the rope strands lying above or below a pulley were included in the same sector as that pulley.

Because the sentences differed in length on different trials, the time that the subjects spent reading the text was expressed as a proportion of the number of characters in
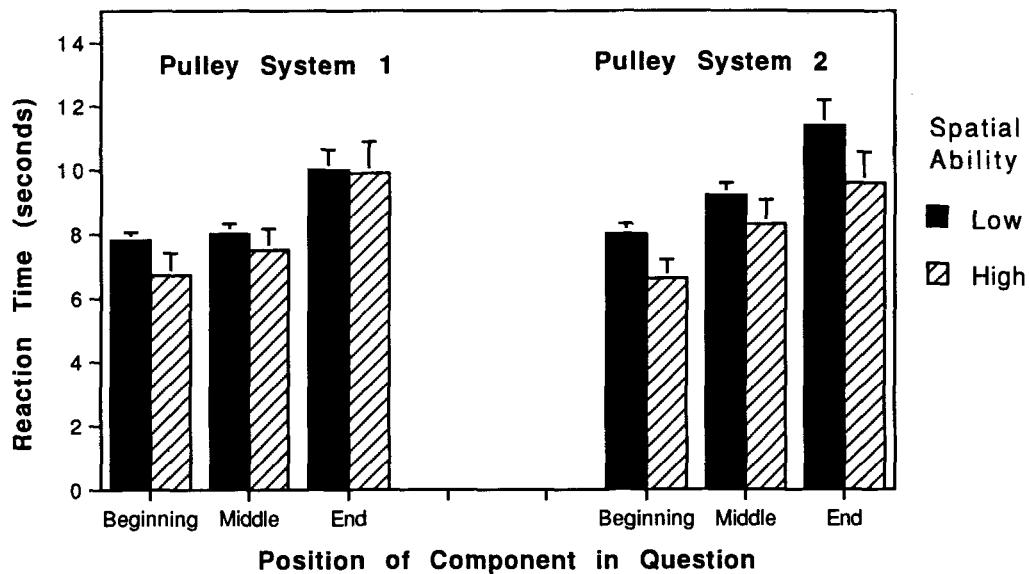
Figure 7. Mean response times of the high- and low-spatial subjects in Experiment 2 in verifying sentences describing kinematic relations between system components at the beginning, middle, and end of the causal chain in the pulley systems.

the text on a given trial (expressed as milliseconds per character of text). The time spent inspecting parts of the diagram was classified further as (1) time spent fixating the component (or components) described in the sentence, (2) time spent fixating pulley system components whose motions occurred before the referent in the causal chain of events, and (3) time spent fixating pulley system components whose motions occurred after the referents in the kinematic chain. The order of events, presented in the Appendix, defined the order of events in the causal chain. The breakdown of reaction time is shown in Figure 8.

Time on text. We expected time on the text to be proportional to the length of the text on a given trial. It is interesting that there was an unpredicted effect of referent location on time spent reading the text, in that the subjects spent more time reading when the referent was at the beginning or end of the causal chain. For Pulley System 1, the mean time per character of text was 102 msec (SD = 23.2) at the beginning of the causal chain, 100 msec (SD = 25.4) at the middle, and 112 msec (SD = 24.9) at the end. For Pulley System 2, the mean time was 110 msec (SD = 26.7) at the beginning of the causal chain, 102 msec (SD = 26.1) at the middle, and 120 msec (SD = 30.1) at the end. The total time spent reading the text, in seconds, is shown in Figure 8.

An investigation of individual differences did not reveal a main effect of spatial ability [$F(1,18) < 1$, for both pulley systems] or an interaction of spatial ability with location of the referent in the causal chain [$F(2,36) < 1$, for both pulley systems]. Furthermore, the proportion of time spent reading the text did not differ for the low- and high-spatial subjects [$F(2,36) < 1$]. Thus, there was no evidence that subjects of different abilities differed in how

they allocated their time to text comprehension versus diagram comprehension and mental animation.

Time on components before the referent. According to the piecemeal model, subjects animate each component before the referent in the causal chain. Therefore, the time spent inspecting components before the referent should increase with the distance of a component from the beginning of the causal chain (i.e., the number of components before the referent). As Figure 8 shows, there was an increase in time spent on components before the referent [$F(2,36) = 58.75$, $MS_e = .34$, $p < .0001$ for Pulley System 1; $F(2,36) = 53.18$, $MS_e = .54$, $p < .0001$ for Pulley System 2].

On this measure, there was again no main effect of spatial ability [$F(1,18) < 1$, for both pulley systems] nor did spatial ability interact with component location [$F(2,36) < 1$, for both pulley systems]. The fact that the high- and low-spatial subjects showed a similar increase on this measure with distance of the referent from the beginning of the causal chain suggests that the piecemeal model of mental animation is equally descriptive of their strategies.

Time on the referent. According to the piecemeal model of mental animation, subjects should spend the same amount of time inspecting the referent at different positions in the causal chain, because the increment in reaction time with distance from the beginning of the causal chain should be due to additional time spent on components before the referent. Location of the pulley system did not affect time on the referent for Pulley System 1 [$F(2,36) = 1.13$, $MS_e = .25$, n.s.]. However, there was an unpredicted effect of pulley location on this variable for Pulley System 2 [$F(2,36) = 3.67$, $MS_e =$
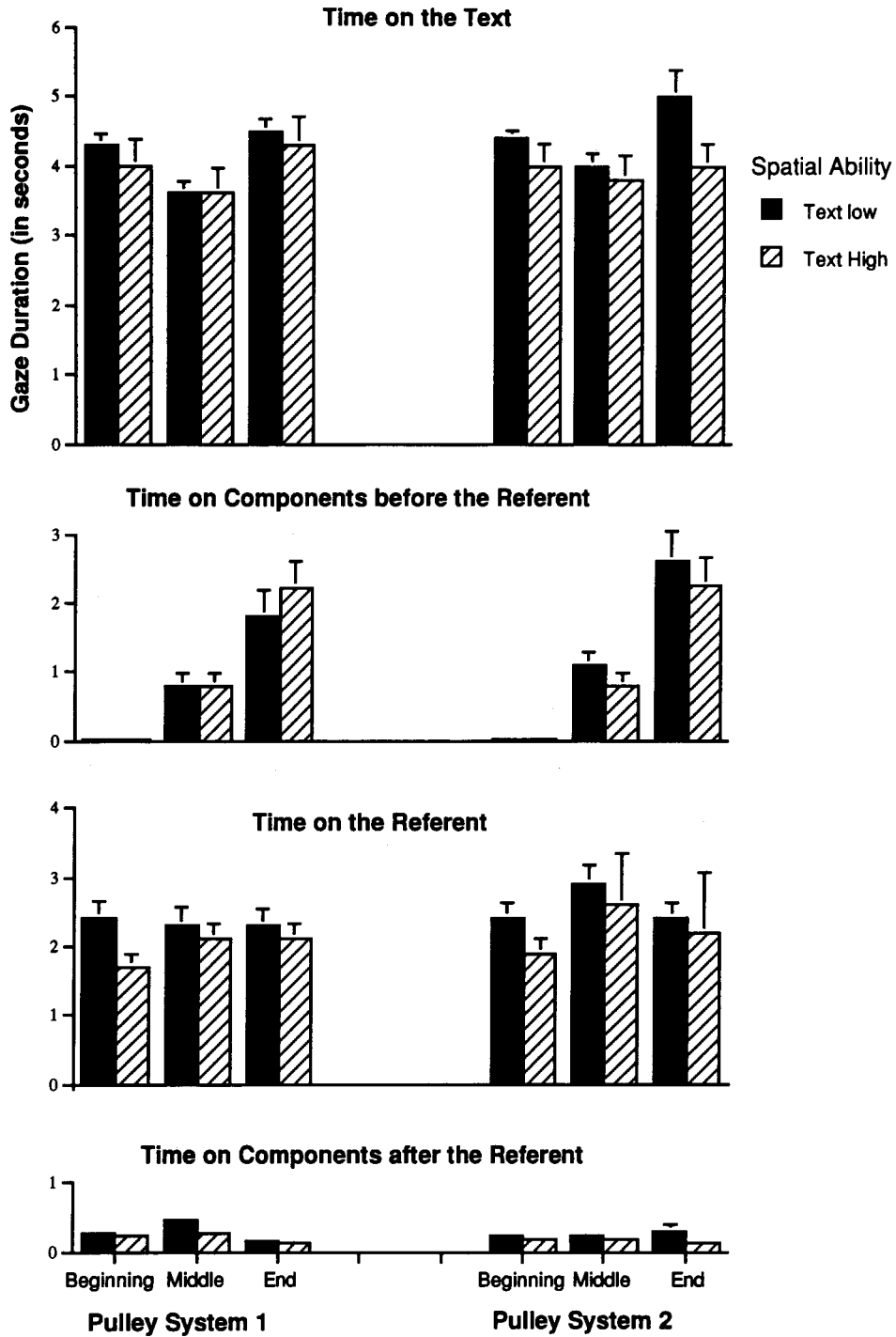
Figure 8. Breakdown of reaction time in Experiment 2 into time spent inspecting the text, components before the referent(s) in the causal chain, the referent(s), and components after the referent(s) in the causal chain.

.43, p < .05]. Inspection of Figure 8 shows that this was due to the subjects' taking a longer time to inspect the middle pulley when this was the referent. The error data in Experiment 1 suggest that people have particular difficulty in understanding the configuration of this pulley, which might explain why the present subjects spent more time inspecting it.

A comparison of the high- and low-spatial subjects revealed no significant main effects of time spent inspecting the referent [$F(1,18)$ = 2.31, MS_e = 1.2, n.s., for Pulley System 1; $F(1,18)$ = 2.50, MS_e = 1.4, n.s., for Pulley System 2] and no significant interactions of location of the referent [$F(2,36)$ = 1.6, MS_e = .25, n.s., for Pulley System 1; $F(2,36)$ < 1, MS_e = .43 for Pulley System 2]. Thus, there was no evidence that spatial ability affected this measure.

Time on components after the referent. Finally, according to the piecemeal strategy, the subjects should not spend any time inspecting components after the referent in the causal chain for any of the pulley locations. The subjects spent very little time inspecting components after the referent in the causal chain (less than 7% of time spent inspecting the diagram), and this variable was unaffected by location in the case of Pulley System 2 [$F(2,36)$ < 1]. For Pulley System 1 there was an unpredicted effect of location, in that the subjects spent the most time on these components when the referent was the lower pulley (the middle of the causal chain) [$F(2,36)$ = 10.01, MS_e = .02, p < .001]. They spent an average of .25 sec (SD = .04) when the referent was at the beginning, .36 sec (SD = .05) when it was in the middle, and .14 sec (SD = .02) when the referent was at the end. The fact that the subjects spent some time inspecting components after the referent in the causal chain suggests that they may not have used the piecemeal strategy exclusively. For example, they might have inferred the motion of a component from the end of the causal chain on some trials, particularly for Pulley System 1. It is also possible that this time reflects initial comprehension of the diagram, search for the relevant components, or errors in calibrating the eye tracker.

Finally, there were no significant main effects of spatial ability on time spent inspecting components after the referent in the causal chain of events in a pulley system [$F(1,18)$ = 2.31, MS_e = .05, n.s., for Pulley System 1; $F(1,18)$ = 2.50, MS_e = .07, n.s., for Pulley System 2], nor did spatial ability interact significantly with location [$F(2,36)$ = 2.23, MS_e = .02, n.s., for Pulley System 1; $F(2,36)$ < 1, MS_e = .03, n.s., for Pulley System 2]. Therefore, we cannot conclude that the groups differed in the time spent inspecting these irrelevant parts of the display.

Conclusions

In conclusion, the results of Experiment 2 replicated both the error and reaction time patterns in Experiment 1. Although there was a strong effect of spatial ability on accuracy, especially for items that involved more mental animation, we failed to find significant differences between the reaction times of the high- and low-spatial subjects. Furthermore, we took a closer look at how the subjects broke down the motion-verification task by monitoring their gaze duration on the text, the referent, and on the components before and after the referent, and again found no differences between the high- and low-spatial subjects on these more precise measures. These data support the view that high- and low-spatial subjects use the same basic strategy to break down the mental animation process, but that low-spatial subjects are less accurate at carrying out this strategy.

EXPERIMENT 3

Experiments 1 and 2 suggested that the effects of spatial visualization ability were primarily in the mental animation component of the motion-verification task. In Experiment 3 we examined further possible sources of error by comparing performance on the motion-verification task with a new task, in which subjects viewed a diagram with an arrow showing the motion of one of the components and verified whether the arrow showed the correct direction of motion of the component (see Figure 3). This arrow-verification task shares the processes of diagram comprehension (Step 2 in Figure 2) and mental animation (Step 4) with the sentence-verification task used in Experiment 1, but does not involve sentence comprehension (Step 1) or text-diagram integration (Steps 3 and 5). If subjects make errors in sentence comprehension or text-diagram integration, then they should be more accurate on the arrow-verification task. If they do not make errors on these processes, then there should be no difference in performance on the two versions of the task. If spatial visualization is specifically related to the mental animation component of the task, then errors on sentence comprehension or text-diagram integration should not be related to low spatial visualization ability.

A comparison of the arrow-verification task with the sentence-verification task also allowed us to examine how the additional processes of text comprehension and text-diagram integration in the sentence-verification task affect mental animation performance. Sentence verification should take longer than arrow verification, because it involves these additional processes. This increment in reaction time can indicate whether the text comprehension and text-diagram integration are independent of diagram comprehension and mental animation, or whether they interfere with these processes. In the sentence-verification task, subjects have to keep the sentence representation in working memory as they mentally animate the diagram, so that they can compare this representation with the results of their animation processes. The additional memory load might have the effect of slowing down the mental animation processes (see Carpenter, Just, & Shell, 1990; Just & Carpenter, 1992), in which case the increment in reaction time should be proportional to the number of components that must be mentally animated to solve the par-

ticular item. It should be smallest at the beginning of the causal chain, greater in the middle of the causal chain, and greatest at the end of the causal chain in a pulley system. The additional memory load might also interfere with the accuracy of the mental animation process. If this occurs, there should be an increment in errors due to text–diagram integration processes. This increment would be smallest for items at the beginning of the causal chain, greater in the middle of the causal chain, and greatest at the end of the causal chain in a pulley system.

In Experiment 3 we also examined the possibility that the performance differences between the low- and high-spatial subjects that we observed in Experiments 1 and 2 were not specifically due to spatial ability. Thus, we included a measure of verbal ability in Experiment 3 in addition to a measure of spatial ability. We predicted that motion verification would be correlated with spatial ability, but not verbal ability.

Method

Subjects

Fifty-one undergraduate students from the University of California at Santa Barbara participated in the experiment. The data of 2 of the subjects (with scores of 6.0 and 11.6, respectively, on the Paper Folding Test) were not included in the analyses because they failed to reach chance performance on the experimental tasks. The subjects who scored 13 or more on the Paper Folding Test ($n = 19$, $M = 15.6$, $SD = 1.8$, range $= 13$–18.8) were classified as high spatial, and those who scored 11 or less ($n = 22$, $M = 7.9$, $SD = 2.1$, range $= 2.4$–10.4) were classified as low spatial.

Stimuli

Pretests. The subjects were administered the Paper Folding Test–VZ-2 (Ekstrom et al., 1976) as a measure of spatial visualization ability and the Advanced Vocabulary Test I-V-4 (Ekstrom et al., 1976) as a measure of verbal ability.

Experimental tasks. The subjects performed two versions of the motion-verification task—the sentence-verification task and the arrow-verification task. All the subjects performed both versions of the task, and the order of presentation of versions was counterbalanced across subjects.

The sentence-verification task consisted of a subset of 24 of the items in the kinematic condition of Experiment 1 (see Figure 1)—that is, those items that described the motion of the pulleys (e.g., "The upper pulley turns clockwise").

On trials in the arrow-verification task, the subjects were presented with a diagram of a pulley system, which was drawn in black on a white screen. On each trial a red box was drawn around one of the pulleys in the system and a red arrow was drawn on the pulley, showing a possible direction of motion of the pulley (see Figure 3). The subjects' task was to verify whether the arrow showed the correct direction of motion of the pulley when the free end of the rope in the pulley system was pulled. The diagrams were identical to those in the sentence-verification task, and there was an item of the arrow-verification task corresponding to each of the items in the sentence-verification task. Thus, the arrows showed the motion of pulleys at the beginning, middle, and end of the causal chain. For each location, there was an item showing the correct direction of motion and an item showing the opposite direction of motion. Again, two replications of the 48 stimuli were presented in a random order in one block of trials.

Apparatus

The stimuli were presented on Macintosh IIci computers using Mindlab software (Meike et al., 1988).

Procedure

The subjects were tested either individually or in groups of 2 or 3; the experiment lasted approximately 45 min. They were first administered the Paper Folding Test, which was followed by the Advanced Vocabulary Test. Then the experimenter presented written and verbal instructions for the first experimental task. The instructions for the sentence-verification task were the same as those presented in Experiment 1. In the instructions for the arrow-verification task, the subjects were first familiarized with the diagrams, as in Experiment 1. They were instructed that on each trial they would see a diagram of a pulley system with a box drawn around one of the pulleys and an arrow showing the motion of that pulley. They were told that their task was to verify whether the arrow showed the correct direction of motion.

After the subjects had received the instructions for the first experimental task, they performed that task.[3] Then they were given instructions for the second task, and they performed that task. In both tasks, the procedure for initiating a trial and responding was identical to that in Experiment 1.

Results and Discussion

We analyzed the data on two levels. First, we computed overall accuracy scores on the basis of the two experimental tasks and examined the correlations of these measures with spatial and verbal ability. These analyses were based on the data from all 49 subjects. Then we analyzed the contributions of the stimulus factors and the ability measures to the variance in errors and reaction times. These analyses were based on the data from the subjects classified as either high or low ability.

Overall Measures

The subjects made a greater proportion of errors on the sentence-verification task ($M = .16$, $SD = .14$) than on the arrow-verification task ($M = .12$, $SD = .13$) [$t(48) = 2.84$, $p < .01$], indicating that the text–diagram integration processes accounted for a significant proportion of errors in the motion-verification task.

Consistent with Experiments 1 and 2, spatial visualization ability had significant negative correlations with errors on both sentence verification ($r = -.37$, $p < .01$) and arrow verification ($r = -.39$, $p < .01$). Furthermore, a measure of errors on text–diagram integration processes (equal to the number of errors on sentence verification minus the number of errors on arrow verification) was not significantly correlated with spatial ability ($r = .004$). These results suggest that although sentence comprehension and text–diagram integration accounted for a significant proportion of errors, the correlation of spatial visualization with the motion-verification task was due to the components shared by the sentence- and arrow-verification tasks—mental animation and diagram comprehension.

Contrary to our predictions, verbal ability was significantly correlated with errors on sentence verification ($r = -.38$, $p < .01$). However, the correlation with arrow verification failed to reach statistical significance ($r = -.23$). The subtraction of the text comprehension and text–diagram integration processes weakened the correlation with verbal ability, suggesting that the correlation with verbal ability was partly due to the text comprehension

and text–diagram integration processes and not just diagram comprehension and mental animation. The correlation of this measure with the difference score (sentence-verification errors minus arrow-verification errors) was −.23.

These data suggest that although both spatial and verbal ability are correlated with accuracy, they are related to different aspects of performance. A multiple regression analysis, in which the measures of spatial and verbal ability were used to predict the overall error rate, was consistent with this interpretation of the data. Both variables made significant contributions to the regression [β = .007, $t(46)$ = 2.03, p < .05 for verbal ability; β = .013, $t(46)$ = 3.00, p < .01 for spatial ability], and together they accounted for 23.3% of the variance in errors [$F(2,46)$ = 5.94, p < .01]. The correlation between the measures of spatial and verbal ability was .07. This analysis suggests that although verbal ability contributes to performance on the motion-verification task, spatial ability makes a unique contribution to performance independently of verbal ability.

Error Analysis

In a 2×3×2 ANOVA, we investigated the contributions of task type (sentence or arrow verification), location of the referent in the causal chain (beginning, middle, or end), and spatial ability (high or low) to the variance in errors. For these analyses, only the data from the subjects who were classified as high spatial (n = 19) and low spatial (n = 22) were included. As in Experiment 1, the data for the two pulley systems were analyzed separately. The data were collapsed over left and right isomers, over order of presentation of the two tasks, and truth value of the item (true or false), because none of these factors had significant interactions with the other experimental factors. Figure 9 shows the mean error rates in the different experimental conditions.

The effects of task type indicate whether sentence comprehension and text–diagram integration were significant sources of error in the motion-verification task. A significant effect in Pulley System 2 [$F(1,38)$ = 9.60, MS_e = .02, p < .01] indicated that, for this pulley system, the subjects made more errors on the sentence task than on the arrow task. There was also a trend in this direction for Pulley System 1, which was not statistically significant [$F(1,38)$ = 1.89, MS_e = .03, n.s.]. These results indicated that errors due to sentence comprehension and/or text–diagram integration processes occurred primarily in items concerning the more difficult pulley system.

As in Experiments 1 and 2, the error rate was highly related to the location of the component in question in the causal chain of events in the pulley system [$F(2,76)$ = 7.11, MS_e = .03, p < .01 for Pulley System 1; $F(2,76)$ = 13.91, MS_e = .05, p < .0001 for Pulley System 2]. If the storage of the sentence representation interferes with the accuracy of mental animation, there should be an interaction between task type and location of the referent. These interactions were not significant

[$F(1,38)$ < 1 for Pulley System 1; $F(1,38)$ = 1.60, n.s., for Pulley System 2], suggesting no such interference.

Individual differences. Consistent with the results of Experiment 1, the high-spatial subjects made fewer errors than the low-spatial subjects [$F(1,38)$ = 4.82, MS_e = .11, p < .05 for Pulley System 1; $F(1,38)$ = 5.82, MS_e = .12, p < .05 for Pulley System 2]. The interaction between task type and spatial ability was not significant [$F(1,38)$ < 1, for both pulley systems], supporting the conclusion that spatial visualization ability is not differentially related to the accuracy of the text–diagram integration processes.

There was an interaction between location and spatial ability [$F(2,76)$ = 3.11, MS_e = .03, p = .05 for Pulley System 1; $F(2,52)$ = 5.55, MS_e = .05, p < .01 for Pulley System 2]. Analysis of simple effects revealed an effect of pulley location on the error proportion of the low-spatial subjects [beginning = .06, middle = .20, end = .26, $F(2,76)$ = 10.21, MS_e = .03, p < .001 for Pulley System 1; beginning = .06, middle = .19, end = .37, $F(2,76)$ = 19.42, MS_e = .05, p < .001 for Pulley System 2], but not high-spatial subjects [M = .08 errors overall, $F(2,76)$ < 1, for both pulley systems]. This replicates the results of Experiments 1 and 2 and supports the view that spatial visualization ability is primarily related to accuracy in the mental animation component of the task.

Reaction Time Analysis

In a 2×3×2 ANOVA, we assessed the contributions of task type, location of the referent in the causal chain, and spatial ability to the variance in reaction times. The mean reaction times are shown in Figure 10. The data from error trials and data points greater than three standard deviations above the mean for all subjects in any trial (1.3% of data) were omitted from the analysis. As in the error analysis, the data were collapsed over left and right isomers, order of presentation of the two tasks, and truth value (true or false), because none of these factors had significant interactions with the other experimental factors. Because of the high error rate in this experiment, when we omitted error trials, some of the subjects had missing data in at least one of the experimental conditions. Because the proportion of missing data was small (1.2%), we replaced the missing data with the mean for all the subjects.

Consistent with the piecemeal model of mental animation and the results of Experiment 1, reaction time increased with the distance of a component from the beginning of the causal chain [beginning = 5.0 sec, middle = 6.8 sec, end = 8.5 sec, $F(2,76)$ = 83.86, MS_e = 2.9, p < .001 for Pulley System 1; beginning = 4.9 sec, middle = 6.4 sec, end = 9.6 sec, $F(2,76)$ = 58.29, MS_e = 7.9, p < .001 for Pulley System 2].

As we predicted, sentence verification took longer than arrow verification [$F(1,38)$ = 74.99, MS_e = 10.4, p < .001 for Pulley System 1; $F(1,38)$ = 28.88, MS_e = 13.3, p < .001 for Pulley System 2]. If the storage of the sentence interferes with the speed of mental animation,
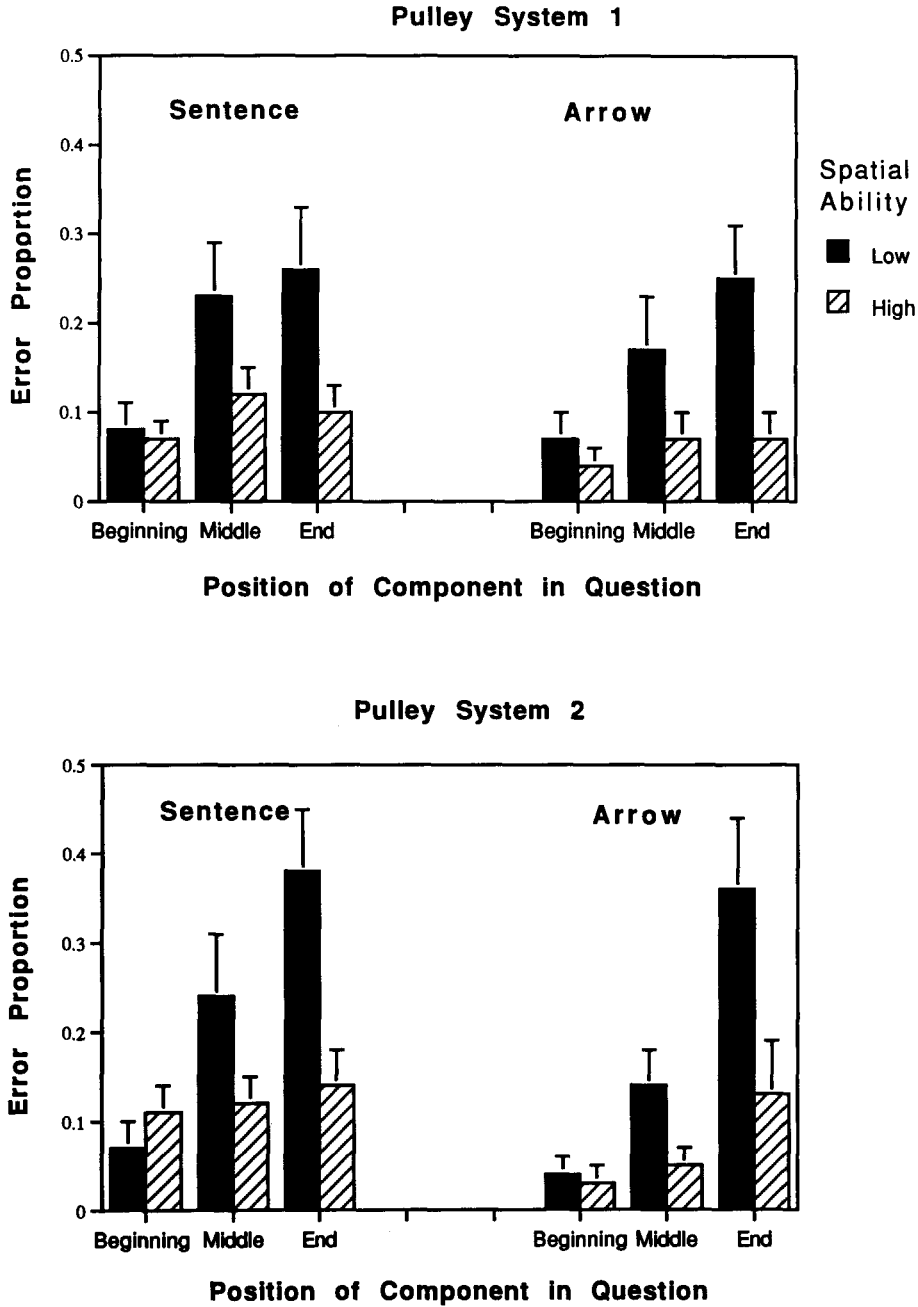
Pulley System 1

Pulley System 2

Figure 9. Mean proportions of errors made by the high- and low-spatial subjects in the sentence-verification and arrow-verification tasks for components at the beginning, middle, and end of the causal chain in the pulley systems.

there should be an interaction between task type and location of the referent in the causal chain, in that the increment in reaction time should increase with distance of the component in question from the beginning of the causal chain. This interaction was not statistically significant [$F(2,76) < 1$, for both pulley systems], indicating that storing a sentence to be verified in working memory does not interfere with the speed of mental animation. This is consistent with previous research showing that a ver-

bal working-memory load is unlikely to interfere with spatial information processing (Baddeley & Lieberman, 1978; Brooks, 1968).

Individual differences. As in Experiments 1 and 2, the main effect of spatial ability on reaction time was not significant [$F(1,38) < 1$ for Pulley System 1; $F(1,38) = 1.83$, n.s., for Pulley System 2]. However, for Pulley System 2, there was a significant interaction between spatial ability and task type [$F(1,38) = 6.11$, $MS_e = 13.3$,
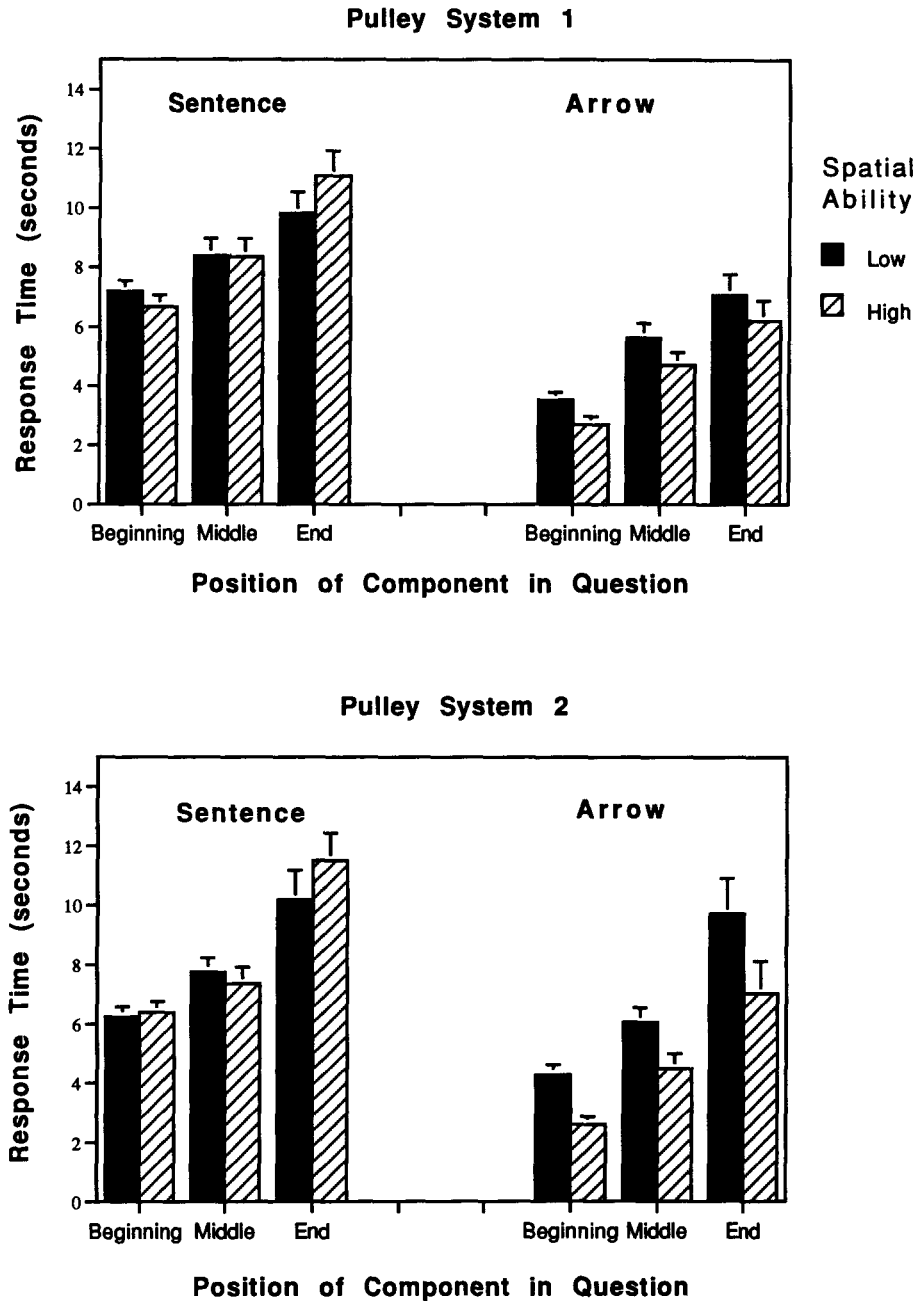
Pulley System 1

Pulley System 2

Figure 10. Mean response times of the high- and low-spatial subjects in verifying sentences and arrows describing static and kinematic relations between system components at the beginning, middle, and end of the causal chain in the pulley systems.

$p < .05$]. This interaction did not reach statistical significance for Pulley System 1 [$F(1,38) = 1.78$, $MS_e = 10.4$, n.s.]. Simple effects analyses revealed that, for Pulley System 2, the high-spatial subjects had shorter reaction times than the low-spatial subjects on the arrow-verification task [$F(1,38) = 5.79$, $MS_e = 20.13$, $p < .05$], but the reaction times of the two groups did not differ on the sentence-verification task [$F(1,38) < 1$]. The mean difference in reaction time between the two tasks was

3.71 sec for the high-spatial subjects and 1.37 sec for the low-spatial subjects. A possible explanation for the arrow-verification results is that when the low-spatial subjects were unburdened by the sentence comprehension and text-diagram integration processes, they spent more time on the diagram-comprehension processes in the case of Pulley System 2. This explanation is consistent with the higher accuracy on the arrow-verification task for this pulley system.

Conclusions

In summary, the results from Experiment 3 showed that although the combined processes of sentence comprehension and text–diagram integration are a significant source of error in the motion-verification task, errors on these processes are not related to spatial visualization ability. The results also showed that although verbal ability is correlated with accuracy in the motion-verification task, spatial visualization has an effect on performance that is independent of the effect of verbal ability on performance.

Finally, Experiment 3 showed that the sentence-comprehension and text–diagram integration processes are independent of mental animation in that they do not affect either accuracy or speed of mental animation. Although the increment in reaction time due to sentence comprehension and text–diagram integration did not vary with position in the causal chain, it did differ for the high- and low-spatial subjects. Our tentative explanation of this result is that the low-spatial subjects spent more time on diagram comprehension in the arrow-verification task—that is, when unburdened by the processing of the sentence.

GENERAL DISCUSSION

This research provides an account of individual differences in a mechanical-reasoning task in which subjects must infer the motion of a component in a mechanical system (i.e., mentally animate the component) from a static diagram of the system. In three experiments, we found a characteristic pattern of errors in this task, in that these errors increase with the distance of the component to be animated from the beginning of the causal chain in the system. These experiments also showed that accuracy in this task is related to spatial visualization ability—only the low-spatial subjects showed this characteristic pattern of errors, whereas the high-spatial subjects were largely successful at the task. Although the high- and low-spatial subjects showed different error patterns, they had similar reaction time patterns. In particular, both groups showed a characteristic increase in reaction time with distance of the component in question from the beginning of the causal chain in the system. Furthermore, the eye-fixation data indicate that the two groups divided their response times similarly among the different parts of the display.

A task analysis suggested four cognitive components of the motion-verification task: sentence comprehension, diagram comprehension, text–diagram integration, and mental animation. We isolated these components experimentally, and we found that spatial visualization predicts accuracy on diagram comprehension and mental animation, but not accuracy on sentence comprehension and text–diagram integration. The reaction times for these components were not completely additive, because the subjects in Experiment 3 appeared to spend more time on diagram comprehension when the task did not include sentence comprehension and text–diagram integration. Thus,

a comparison across experiments indicates that the additional time spent on sentence verification compared with arrow verification in Experiment 3 was considerably less than the gaze duration on the text in Experiment 2.

What might account for the differences in accuracy in the mental animation component of the task? Individual differences in the mental animation process might occur in how subjects break down the process of mentally animating a complex system into the animation of individual components, or in how they mentally animate the individual components. The reaction time and eye-fixation patterns showed no evidence for differences in how subjects break down the animation task; they indicated that both groups of subjects animated the pulley systems in a piecemeal way, and in order of the causal chain of events in the system.[4] Thus, the differences appear to lie in how accurately they carry out this piecemeal strategy.

Either less accurate spatial transformation processes or a spatial working memory with a smaller capacity might account for the poor performance of the low-spatial subjects on mental animation. First, high- and low-spatial subjects might differ in the accuracy of spatial transformation processes; that is, they might differ in their ability to infer the motion of any one component from that of an adjacent component. If there was some probability of error for each "link" in the causal chain, then subjects would make more errors for items later in the causal chain, as we observed. Second, high- and low-spatial subjects might differ in spatial working memory. Assuming that subjects keep track of the motions of each component that they animate, there is a greater load on working memory when a component motion is later in the causal chain of events, and so errors on these later items might result from a loss of information from working memory. This explanation is consistent with the observed correlation between spatial visualization and mental animation, because spatial visualization performance might also be expected to involve spatial transformation processes and working memory. The research presented in the present article does not discriminate between these two accounts of the differential performance of high- and low-spatial subjects. In fact, the two accounts may not be empirically separable, given that previous research has shown a tradeoff between storage of spatial information and spatial processes in working memory (Baddeley & Lieberman, 1978; Carpenter & Just, 1986).

Previous theoretical accounts describe the process of inferring the operation of a physical device as envisioning, or running, a mental model (de Kleer & Brown, 1984; Reiger & Grinberg, 1977; Williams et al., 1983). Our research suggests that running a mental model shares processes with spatial visualization ability. A remaining issue concerns the basic nature of these shared processes. It has been suggested that running a mental model involves transforming mental images (Schwartz & Black, 1992; Spoehr & Horvath, 1989). It has also been suggested that the accurate transformation of visual images underlies spatial visualization ability (Kosslyn, Brunn, Cave, & Wal-

lach, 1985; Poltrock & Agnoli, 1986; Smith, 1964). Thus, one possible account is that people infer the motion of mechanical systems by transforming mental images. An alternative view is that people infer the motion of components in a mechanical system by applying rules of mechanical reasoning. In fact, it is possible that mental imagery and such rule-based reasoning are characteristic of individuals with different abilities or different stages of practice with the task (Schwartz & Black, 1992). This possibility could be tested in a future study, in which subjects of different abilities could be asked to give verbal protocols at different stages of practice at the motion-verification task, or by measuring the effects of spatial and verbal interference tasks on the process of inferring motion in machines.

Although we have emphasized the processes that are common to spatial visualization and motion verification in this paper, there are also some differences between the tasks. First, the Paper-Folding Test is speeded, whereas the motion-verification task is self-paced, and previous research has shown that speeded spatial tasks are not always good predictors of performance in unspeeded tasks (Carpenter & Just, 1986). Also, the motion-verification task is a more complex task, and it involves other skills, such as text and diagram comprehension, text–diagram integration, and the decomposition of the complex task into subproblems—the animation of individual "links" in the causal chain. Thus, the particularly spatial processes in this task may be somewhat overshadowed by verbal and general intelligence processes (Carpenter & Just, 1986; Carpenter et al., 1990), accounting for the moderate correlations between spatial visualization and motion verification.

In conclusion, the results of the present study expand our understanding of mechanical ability. Previous researchers have suggested that mechanical ability involves general reasoning skills and specific knowledge of machines (Cronbach, 1984; Hegarty et al., 1988). In the research presented in this article we describe another cognitive component of mechanical ability—mental animation, or the ability to infer motion from static diagrams—which is related to spatial visualization ability. Further research is required to test how our findings generalize to other inferences about motion. This future research should investigate whether spatial visualization can be employed to infer other aspects of motion in machines (e.g., the displacement of a machine component after a given motion) or the relative rates of motion of different components. These studies will further clarify the functions of spatial visualization in mechanical reasoning.

REFERENCES

BADDELEY, A. D., & LIEBERMAN, K. (1978). Spatial working memory. In R. Nickerson (Ed.), *Attention and performance VIII* (pp. 521-539). London: Academic Press.

BENNETT, C. K. (1969). *Bennett Mechanical Comprehension Test.* New York: The Psychological Corporation.

BROOKS, L. R. (1968). Spatial and verbal components of the act of recall. *Canadian Journal of Psychology, 22,* 349-368.

CARPENTER, P. A., & JUST, M. A. (1975). Sentence comprehension: A psycholinguistic processing model of verification. *Psychological Review, 82,* 45-73.

CARPENTER, P. A., & JUST, M. A. (1986). Spatial ability: An information processing approach to psychometrics. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (pp. 221-253). Hillsdale, NJ: Erlbaum.

CARPENTER, P. A., JUST, M. A., & SHELL, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices Text. *Psychological Review, 97,* 404-431.

CLARK, H. H., & CHASE, W. G. (1972). On the process of comparing sentences against pictures. *Cognitive Psychology, 3,* 472-517.

CRONBACH, L. J. (1984). *Essentials of psychological testing* (4th ed.). New York: Harper & Row.

DE KLEER, J., & BROWN, J. S. (1984). A qualitative physics based on confluences. *Artificial Intelligence, 24,* 7-83.

EGAN, D. E. (1976). *Accuracy and latency scores as measures of spatial information processing* (NAMRL Report No. 1224). Pensacola, FL: Naval Aerospace Medical Research Laboratory.

EKSTROM, R. B., FRENCH, J. W., & HARMAN, H. H. (1976). *Kit of factor-referenced cognitive tests.* Princeton, NJ: Educational Testing Service.

GUILFORD, J. P., & ZIMMERMAN, W. S. (1953). *Guilford-Zimmerman aptitude survey.* Orange, CA: Sheridan Psychological Services.

HEGARTY, M. (1992a). The mechanics of comprehension and comprehension of mechanics. In K. Rayner (Ed.), *Eye movements and visual cognition: Scene perception and reading* (pp. 428-443). New York: Springer-Verlag.

HEGARTY, M. (1992b). Mental animation: Inferring motion from static diagrams of mechanical systems. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 18,* 1084-1102.

HEGARTY, M., & JUST, M. A. (1989). Understanding machines from text and diagrams. In H. Mandl & J. Levin (Eds.), *Knowledge acquisition from text and picture* (pp. 171-194). Amsterdam: North-Holland.

HEGARTY, M., & JUST, M. A. (1993). Constructing mental models of machines from text and diagrams. *Journal of Memory & Language, 32,* 717-742.

HEGARTY, M., JUST, M. A., & MORRISON, I. M. (1988). Mental models of mechanical systems: Individual differences in qualitative and quantitative reasoning. *Cognitive Psychology, 20,* 191-236.

JUST, M. A., & CARPENTER, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99,* 122-149.

KIERAS, D. E., & BOVAIR, S. (1984). The role of a mental model in learning to operate a device. *Cognitive Science, 8,* 255-273.

KOSSLYN, S. M., BRUNN, J., CAVE, K. R., & WALLACH, R. W. (1985). Individual differences in mental imagery: A computational analysis. In S. Pinker (Ed.), *Visual cognition* (pp. 195-243). Cambridge, MA: MIT Press.

LOHMAN, D. F. (1979). *Spatial ability: Individual differences in speed and level* (Tech. Rep. No. 9). Stanford, CA: Stanford University, Aptitude Research Project, School of Education.

MACLEOD, C. M., HUNT, E. B., & MATHEWS, N. N. (1978). Individual differences in the verification of sentence-picture relationships. *Journal of Verbal Learning & Verbal Behavior, 17,* 493-507.

MAYER, R. E., & ANDERSON, R. B. (1991). Animations need narrations: An experimental test of a dual-coding hypothesis. *Journal of Educational Psychology, 83,* 484-490.

MCGEE, M. G. (1979). *Human spatial abilities.* New York: Praeger.

MEIKE, B., BHARUCHA, J., BAIRD, J. C., & STOECKIG, K. (1988). *Mindlab.* Santa Barbara, CA: Intellimation.

POLTROCK, S. E., & AGNOLI, F. (1986). Are spatial visualization ability and visual imagery ability equivalent? In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (pp. 255-296). Hillsdale, NJ: Erlbaum.

THE PSYCHOLOGICAL CORPORATION (1990). *The Mechanical Reasoning Test of the Differential Aptitudes Test.* New York: Author.

REIGER, C., & GRINBERG, M. (1977). The declarative representation and procedural simulation of causality in physical mechanisms. *Proceedings of the Fifth International Joint Conference on Artificial Intelligence, 5,* 250-255.

RIPS, L. J. (1986). Mental muddles. In M. Brand & R. M. Harnish (Eds.), *Problems in the representation of knowledge and belief* (pp. 258-286). Tucson, AZ: University of Arizona Press.

ROUSE, W. B., & MORRIS, N. M. (1986). On looking into the black box: Prospects and limits in the search for mental models. *Psychological Bulletin*, **100**, 349-363.

SCHWARTZ, D. L., & BLACK, J. B. (1992). *Shuttling between depictive models and abstract rules: Induction and fall-back* (Tech. Rep. No. CCT92-2). New York: Teachers College.

SHEPARD, R. N., & FENG, C. (1972). A chronometric study of mental paper folding. *Cognitive Psychology*, **3**, 228-243.

SMITH, I. M. (1964). *Spatial ability: Its educational and social significance.* London: University of London Press.

SPOEHR, K. T., & HORVATH, J. A. (1989, November). *Running a mental model: Evidence from reaction time studies.* Paper presented at the meeting of the Psychonomic Society, Atlanta, GA.

THURSTONE, L. L. (1950). *Some primary abilities in visual thinking* (Psychometric Laboratory Report No. 59). Chicago: University of Chicago.

VANDENBERG, S. G., & KUSE, A. R. (1978). Mental rotations: A group test of three-dimensional spatial visualization. *Perceptual & Motor Skills*, **47**, 599-604.

WILLIAMS, M. D., HOLLAN, J. D., & STEVENS, A. L. (1983). Human reasoning about a simple physical system. In D. Gentner & A. L. Stevens (Eds.), *Mental models* (pp. 131-153). Hillsdale, NJ: Erlbaum.

NOTES

1. The difficulty of the kinematic trials was standardized, because for these trials the false items always stated the direction of motion that was exactly opposite that stated in the true items—for example, "counterclockwise" if the true direction was clockwise. For the static statements, the false items described a component as being attached to some object other than that to which it was actually attached (e.g., "The upper pulley is attached to the weight"). In this example, the false statement might be very implausible to subjects, because "upper" means at the top of the diagram, and the weight was always at the bottom of the diagram. Therefore, this sentence might in fact be somewhat easier to verify than the true statement.

2. A rope strand was a section of rope that lay over or under a pulley, ran between two components (e.g., a pulley and the ceiling), or had a free end. For example, in the pulley system presented in Figure 3, the lower rope has three rope segments: the right side of the rope, which moves up when the rope is pulled, the section lying under the pulley, which moves to the right, and the left section running between the lower pulley and the ceiling, which shortens.

3. The subjects were not given practice trials in this experiment. Note that this did not appear to affect the experimental outcomes, because performance on the sentence-verification task replicated similar conditions in Experiments 1 and 2.

4. It should be noted that the eye-fixation data showed some discrepancies from the pattern predicted by the piecemeal model—for example, in the time spent inspecting components after the referent in the causal chain—but these discrepancies were similar for both the high- and low-spatial subjects.

APPENDIX
Static and Kinematic Statements About the Two Pulley Systems That the Subjects Verified in Experiment 1

Pulley System 1

The upper rope moves to the right over the upper pulley.
The upper pulley turns clockwise.
The upper rope moves to the right under the middle pulley.
The middle pulley turns counterclockwise.
The lower rope moves to the right under the lower pulley.
The lower pulley turns clockwise.

Pulley System 2

The rope moves to the right over the right upper pulley.
The right upper pulley turns clockwise.
The rope moves to the right under the lower pulley.
The lower pulley turns counterclockwise.
The rope moves to the left over the left upper pulley.
The left upper pulley turns counterclockwise.

Note—Only statements about the right isomer of each pulley system are listed.