Reasoning about the referent of a picture versus reasoning about the picture as the referent: An effect of visual realism

DANIEL L. SCHWARTZ
Vanderbilt University, Nashville, Tennessee

Research on picture perception and picture-based problem solving has generally considered the information that enables one to "see" and think about a picture's subject matter. However, people often reason about a picture or representation as the referent itself. The question addressed here is whether pictorial features themselves help determine when one reasons about the referent of an image, as with an engrossing movie, and when one reasons about the image in its own right, as with abstract art. Two experiments tested the hypothesis that pictures with relatively high fidelity to their referents lead people to think about those referents, whereas pictures with relatively low fidelity lead people to think about the picture as a referent. Subjects determined whether marks on the bottom and top boards of an open hinge would meet if the hinge were closed. Accuracy and latency results indicated that subjects who saw realistic displays simulated the physical behavior of the hinge through analog imagery. In contrast, subjects who saw schematic displays tended to reason about static features of the display such as line lengths and angles. The results demonstrate that researchers must be cautious when generalizing from reasoning about diagrammatic materials to reasoning about the referents themselves.

In the psychological study of picture perception, a primary objective has been the delineation of visual forms, cues, relationships, or information that allow one to "see" and think about the referent of a picture (e.g., Buswell, 1935; Gibson & Bridgeman, 1987; Hochberg, 1962; Rock, Halper, & Clayton, 1972; Wertheimer, 1958). A rich set of issues arises when the information in a picture equally specifies different referents. For example, Hochberg (1974) argued for the existence of mental representation on the basis of a one-to-many mapping between stimulus and percepts. He argued that multiple perceptions of the same stimulus occur because people map an underspecified picture onto different representations of prior perceptual experience. For example, if people interpret one face of a Necker cube as being in the foreground, they will enrich the other cube faces so that the overall image will be consistent with their representation of cubes. If people interpret the same face as being in the background, they will enrich the rest of the cube in a different way, again in accordance with their cube representation. While Hochberg's argument revolved around stimuli that create optical illusions, there is a more general one-to-many mapping involving pictorial stimuli. Every picture

has at least two distinct readings—a picture "of" a referent and a picture "as" a referent (cf. Arnheim, 1969; Siegel, 1978). According to Hochberg's argument, a picture is interpreted according to its referential function and can be embellished by reference to perceptual representations of its physical referent. However, one can also interpret a picture (or any representation) as a referent in its own right and embellish it by reference to formal features, conceptual structures, and other things that do not depend on prior perceptions of the distal referent (e.g., Gombrich, 1974; Hanson, 1961). The research task then is to find the factors that help people resolve the one-to-many mapping between representation and reference inherent to visual artifacts. When will one view a picture as a window on a referent, as with an engrossing movie, and when will one view a picture as a referent itself, as with abstract art?

Controlling for cultural history (cf. Deregowski, Muldrow, & Muldrow, 1972; Hochberg & Brooks, 1962), the hypothesis examined here is that, with respect to its referent, the visual impoverishment, or infidelity, of a picture influences the way people "view" and think about that picture. I will call this the fidelity hypothesis: the degree to which a picture violates or omits naturally occurring perceptual information helps to specify the picture's self or other reference.[1] A "realistic" picture leads one to think in terms of the picture's referent. A picture with infidelities leads one to think in terms of the representation. An intuitive example comes from movies in which a particular special effect does not work quite right, and one briefly sees the special effect rather than the action it portrays. The hypothesis is not meant to imply that

people cannot switch between, or simultaneously entertain, two views on a picture. For example, those in the arts often consider the relationship between the formal features of an image and its semantic content (e.g., McLuhan, 1965). Instead, the hypothesis is meant to suggest that, in general, the perceptual features of an image provide information (or lack of information) that supports an initial disambiguation of the image's referential status, a disambiguation which may have functional consequences in subsequent problem solving (cf. Chastain & Burnham, 1975).

Prior research shows that both physical and spatial infidelities influence what people perceive and recall. For example, one class of research has used a violations paradigm in which elements of an image are placed in impossible or unusual spatial and physical relations (e.g., Biederman, Mezzanotte, & Rabinowitz, 1982; Mandler & Stein, 1974). The results generally indicate that identification, recall, recognition, and reconstruction are diminished by violations. A particularly relevant finding comes from Freyd, Pantzer, and Cheng (1988). In one condition, subjects were shown a picture of a pot hanging from a hook. They then saw a second picture in which the hook had been removed. Subsequently, the subjects were shown a third picture and had to judge if it was the same as the second picture they had just seen. As though gravity had been operating on their representations of the image, subjects tended to make "same" judgments when the pot had actually been displaced downward in the third picture. Of importance to the current work, this displacement effect disappeared if the initial picture did not include a hook supporting the pot. One interpretation of this result is that when the original image had an infidelity with respect to physical support, subjects represented an image of a pot rather than a real pot with mass.

In other research, investigators have examined the role of perceptual impoverishment by comparing the effects of diagrams, photographs, films, and physical objects on measures of problem solving, learning, and transfer (e.g., Ferguson & Hegarty, 1995; Joseph & Dwyer, 1984; Kaiser, Proffitt, & Anderson, 1985; Moore & Schwartz, 1994; Murray, 1970; Schwartz & Black, in press). Excepting tasks that require fine discriminations of unfamiliar and visually complex referents (e.g., a photograph of a heart, Parkhurst & Dwyer, 1983), studies with significant results have found that people perform better when working with more visually realistic materials. However, these tasks have generally examined people's understanding of the referents of the images and not the images themselves. One exception is research that examined adolescent cognition about linear relationships using physical winches, computer diagrams of winches, and number tables (Moore, 1993). Although Moore found descriptive evidence that the physical winch yielded superior transfer to other tasks (cf. Ferguson & Hegarty, 1995), she also found that the computer diagram and number table led to earlier numerical activity. Conceivably, the latter result, similar to Freyd's, could be

due to a focus on the properties of the representations themselves rather than a focus on the physical and causal device to which they referred.

Developmental studies provide clear documentation of how behaviors may change depending on the referential status accorded to a stimulus. For example, in Mischel's delay-of-gratification studies, children were shown a reward such as a cookie (Mischel, Shoda, & Rodriguez, 1989). The children were told they could either take the reward immediately or wait to receive double the reward. The results showed that children delayed gratification longer if they were encouraged to think of the physical reward as though it were a picture. Research by DeLoache and Marzolf (1991) shows how perceptual fidelity may influence the child's referential assignment and subsequent reasoning. Young children were shown a representation of a toy hidden in a room. Afterwards, they were told to find the real toy. When the representation was a miniature toy placed in a scale model of the room, $2\frac{1}{2}$-year-olds could not locate the target toy in the larger room. However, if the representation was a photograph of the toy in its hiding place, $2\frac{1}{2}$-year-olds could locate the object. DeLoache contends that children interpreted the miniature toy and scale model as referents in their own right, whereas they interpreted the photograph as a representation of the referents. One may speculate that because the photograph was less resemblant to a real object, young children were more sensitive to its symbolic function.

To conduct an initial examination of the hypothesis that perceptual fidelity helps to specify the referential status of a picture, subjects in the following experiments reasoned about hinges. Their task was to determine whether a mark on each leg of a hinge would meet if the hinge were to close. Figure 1 shows the "realistic" hinge that, according to the hypothesis, should lead people to reason by reference to real hinges, and the "abstract" hinge that should lead people to reason by reference to the representation. Although the realistic image is cartoonish, several cues might lead subjects to "see through" the image to a physical hinge. There were the perceptual cues, including one-point perspective and occlusion, and there were the force cues, including the taut rope supporting the mass of the upper board. The abstract drawing omits these cues. Despite verbal instructions indicating that the abstract image is a hinge that closes, the perceptual impoverishment of this drawing may cause people to think about the image as the referent rather than the physical hinge to which it refers.

If subjects reason in terms of physical hinges, they may be inclined to solve the problem by mentally depicting the hinge's closing behavior (Hegarty, 1992; Schwartz & Black, in press). An implied motion paradigm has shown that people can interpolate a pivoting motion between two flashing boards (Foster, 1975). If the current results show that people can interpolate a pivoting motion from a static display, this would be a useful finding, because there has been relatively little work demonstrating mental imagery for objects that are
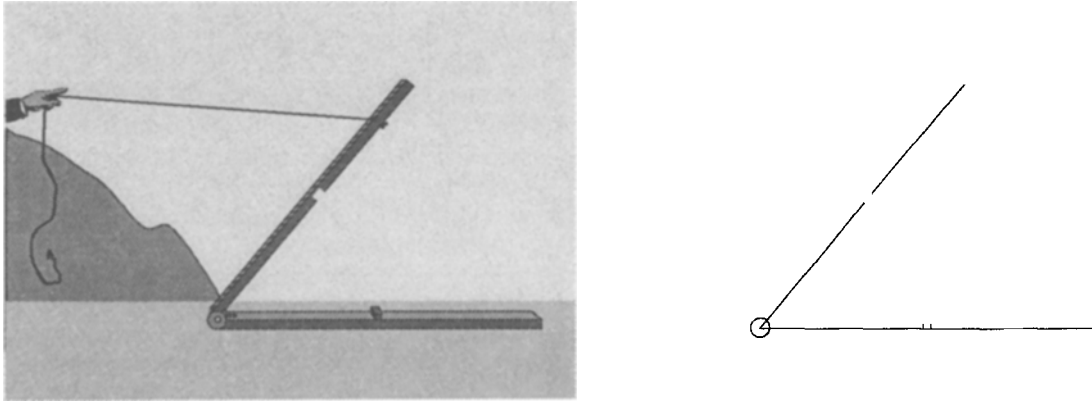
Figure 1. A realistic and abstract version of the hinge problem. *Will the top mark meet with the bottom mark if the hinge closes?* One group of subjects saw the realistic version, and a second group saw the abstract version. The marks will meet. For the realistic picture shown here, it may be easier to determine the match by moving one's head to the left, in line with the perspective shown in the image.

constrained by physical principles other than object rigidity (e.g., pivot joints; Parsons, 1987).

If subjects focus on the lines and angles that constitute the image of the hinge, they may be inclined to compare line lengths or implied angles. I will call this and other similar approaches "feature based" to capture the idea that subjects are reasoning about discrete features of the hinge image rather than about the behavior of a hinge. Evidence of feature-based reasoning is not direct evidence that one is reasoning about a representation rather than its specific referent. For example, one could physically measure and compare the features of a physical hinge. Because referents and resembling images necessarily share properties, it should be possible to find, or cause, crossovers between referent- and image-focused reasoning (e.g., Mischel et al., 1989). Moreover, feature-based reasoning may be peculiar to the abstract images of the current task. Consequently, feature-based reasoning is used as the prediction that arises from a consideration of how people would reason about the static image if they did not enrich it by reference to their dynamic knowledge of physical hinges.

Research by Cooper and others (Cooper, 1976; Cooper & Podgorny, 1976; Kail, Carter, & Pellegrino, 1979) suggests how one might distinguish between simulation and feature-based strategies. Cooper found two latency patterns among subjects who determined the equivalency of two Atteneave figures. In one case, latencies increased with the size of the angular disparity between the two figures. This indicated an analog rotation according to the rationale that larger angular disparities require more processing, because there are more degrees through which to rotate a figure (Shepard, 1968/1982). Thus, the model for this strategy was that subjects rotated the two figures into alignment and then made a holistic comparison of their configurations. Supporting the interpretation of a holistic comparison, the similarity between the two figures did not influence the latencies. In the second case, latencies increased as the figures became more

similar, but did not vary as a function of angular disparity. This indicated a rejection-oriented, feature-based comparison. Subjects chose a feature on one figure and searched for it on the second figure. If they found a matching feature, they then picked a second feature and repeated the process. Nearly identical figures caused long latencies because subjects needed to check numerous features before they chanced on a discriminating feature.

In the hinge task, one might expect a somewhat similar split in data patterns. If subjects imagine a falling hinge through analog imagery, as expected for the realistic hinge, latencies should depend on the opening, or angular disparity, of the hinge. If subjects solve the problem through a feature-based approach, as expected for the abstract image, then the determinant of response time should be the size of the mismatch between the upper and lower board segments, not the angular disparity. Larger mismatches should lead to faster responses, because greater differences in magnitudes make for faster and more accurate discriminations (Johnson, 1939; Moyer, 1973). So, whereas Cooper's research relied on the number of mismatched features to show a feature-based strategy, the current studies relied on the metric size of a mismatch. In the situation where the board segments match, one might expect a two-stage process. Subjects who begin a match problem with a feature-based approach should find it very difficult to determine the answer. Presumably subjects would allow for a margin of error in their length estimations, and it would be difficult to discriminate whether a near mismatch should really be a match. Consequently, they may abandon their rejection strategy in favor of a confirmation strategy. One possible confirmation strategy is to use the preceding analog strategy. If subjects using a feature-based strategy do switch to an analog strategy for the match problems, then the analog transformations should begin later for them than for those subjects who solve the problem with an analog strategy from the outset.

EXPERIMENT 1

To contrast analog and feature-based strategies for the hinge task, I tried to make the difficulty of the mismatch problems increase with the angular disparity if using an analog strategy, but decrease if using a feature-based strategy. This would make it so that analog solutions would take longer and be more error prone at larger angular disparities, whereas feature-based solutions would become quicker and more accurate. The way this was done was to make the size of a mismatch increase with the angular disparity. Figure 2 shows how the position of the gap on the upper leg of the hinge varies for two angular disparities. Notice that for the two types of mismatch problem, undershoot and overshoot, the 80° angular disparity yields larger gap offsets than the 20° angular disparity. For a feature-based strategy, the increase in offset should have a relatively large effect. This strategy may capitalize on the increasing discriminability of the negative problems without being appreciably affected by the increasing angular disparity. Because a feature-based strategy does not require a larger pivoting transformation at larger angles, response times and errors should decrease as the discriminability of the negative problems grows. For the analog strategy, the increase in offset should have a relatively small effect compared with the angular disparity. Latencies should primarily be a function of how far subjects need to imagine pivoting the upper leg. Additionally, the accuracy benefit of a larger offset should be mitigated, because subjects have more opportunity to compound transformation errors over a larger angular disparity.

The relationship between the gap offset and the angular disparity required a discrimination function that was sufficient to benefit a feature-based strategy, but not so generous that subjects could detect an obvious mismatch. Using the observations of pilot subjects, I altered the interior angles of the problems by 10% for the mismatches. In Figure 2, the undershoot and overshoot problems are displaced by 10% of the match interior angles, yielding a larger offset for the larger angular disparity (the match positions are the same in each case). Figure 3 shows how this function translates into the gap's offset across the problems of Experiment 1. The figure shows that the offset increases as a function of the angular disparity and the distance of the bottom mark from the hinge axis.

For the current task, if one associates an analog strategy with thinking about the referent of the picture and a feature-based strategy with thinking about the picture as the referent, there are four predictions that would support the fidelity hypothesis. (1) For subjects who view the realistic hinge, there should be a positive correlation between latency and angular disparity for the match and mismatch problems. (2) For subjects who view the abstract hinge, there should be a negative correlation between latency and angular disparity for the mismatch problems. (3) The intercept of the linear model describing the latency data for the match problems should be greater in the abstract condition than in the realistic condition, because subjects will have first tried to solve the problem using a feature-based strategy. (4) As the angular disparity increases on the mismatch problems, the abstract-condition responses should become more accurate, whereas the realistic-condition responses should become less accurate.

Method

Subjects. Thirty-two undergraduates and graduates from Columbia University (26 females) participated for financial compensation. (One abstract-condition subject was replaced because she physically measured the hinge lengths.) The subjects were assigned randomly to ensure equal gender distributions.

Stimuli and Design. The experiment included one between-subjects and two within-subject factors. The between-subjects fac-
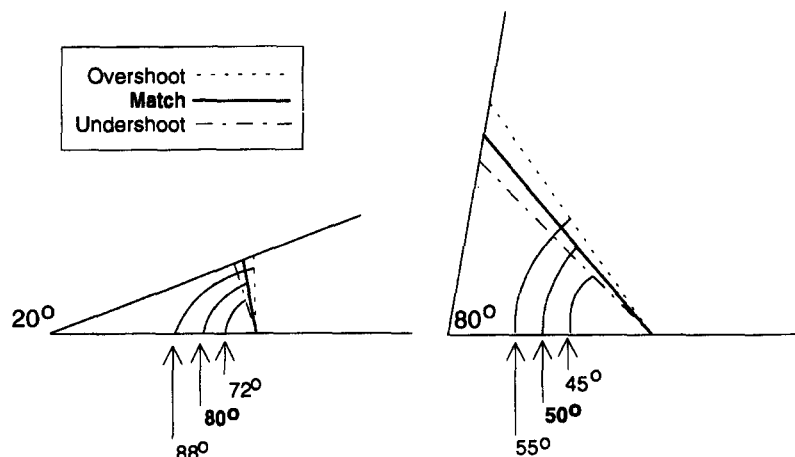
Figure 2. The position of the upper mark for the match and mismatch problems at two angular disparities. The lines intersecting the upper leg of the hinge show the center positions of the gap for the undershoot, match, and overshoot problems. Note that the mismatch offsets along the upper leg are greater for the larger angular disparity. The amount of offset is determined by a 10% change to the interior angle that would put the gap at the match position.
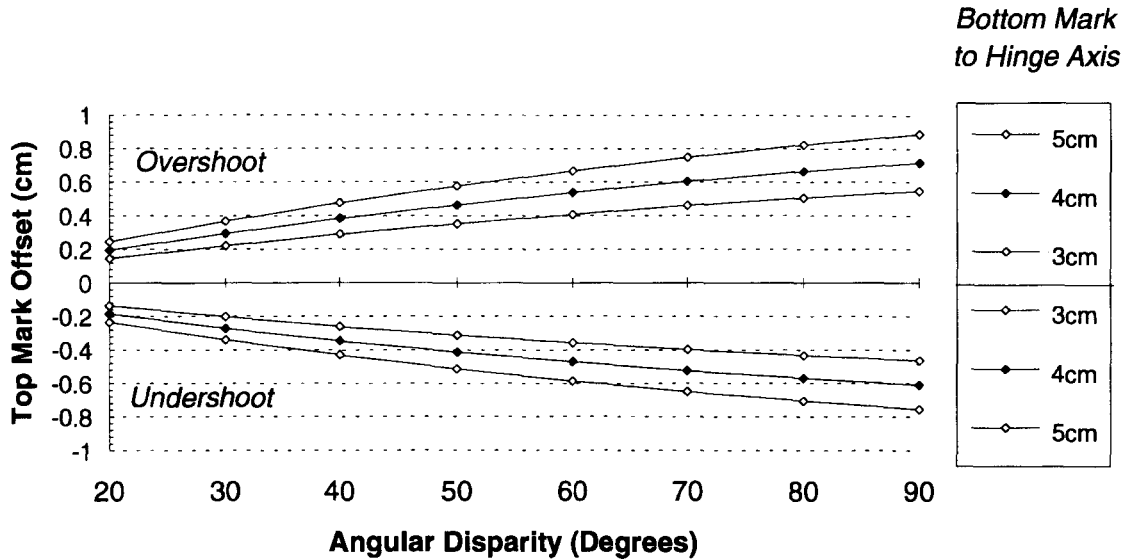
Bottom Mark to Hinge Axis

Figure 3. The size of the undershoot and overshoot offsets plotted by angular disparity and the distance of the lower mark along the bottom leg of the hinge.

tor was the *display condition*, which had the levels of *realistic* and *abstract* drawings (see Figure 1). The *angular disparity* factor determined the openness of the hinge ranging from 20° to 90° in 10° increments. The *problem type* factor determined whether a problem was a *match*, an *undershoot*, or an *overshoot*, as described above. The ratio of problem types was 2 matches to 1 undershoot to 1 overshoot. To prevent memorization, the problems were varied in secondary ways. The length of the top board was either 6.25 or 8.25 cm. (The bottom board was always 8.75 cm.) For the 6.25-cm board, the bottom mark was either 3.25 or 4.25 cm from the hinge axis. For the 8.25-cm board, the distance was either 4.25 or 5.25 cm. These variations were not of theoretical interest to the current investigation and will not be analyzed. All told, each subject saw 128 experimental trials (8 angular disparities × 2 mark distances × 2 board lengths × 4 match/mismatch). Eight random problem sequences used in either direction created 16 problem orders for each condition.

Procedure. Subjects sat approximately 1 m from an IBM 8513 VGA monitor. They placed the forefingers of each hand on the "yes" and "no" keys ("?" and "z," respectively) and used their thumbs to press the spacebar. The subjects worked 20 practice problems that used slightly different sizes. For the first of the practice problems, the experimenter stated, "This is a hinge, and the question is whether the top mark will meet with the bottom mark if it closes. If you think it will, press the 'yes' key, and if you think it won't, press the 'no' key." After the choice was made, the computer provided textual feedback stating whether or not the answer was correct. For the negative problems, additional feedback was provided by a second bottom mark that showed its true match position. The subjects were encouraged to examine the feedback to calibrate their thinking if necessary. The subjects pressed the space bar for the next problem. After two or three initial practice problems, the experimenter explained that the recorded trials would not have any feedback and that both speed and accuracy were important. The subjects finished the practice working alone. The computer notified the subjects that the recorded trials were beginning. Four filler problems were appended to the front of the recorded trials. The subjects debriefed by explaining their strate-

gies on three representative problems, one being from the other condition.

Results

Averaging across trials, the realistic display led to faster and less variable responses (2.08 sec, $SD = 1.49$) than did the abstract display (2.84 sec, $SD = 2.52$).[2] The realistic display also yielded more correct responses (74.5%, $SD = 9.4$% across subjects) than the abstract display (69.4%, $SD = 10.3$%). The primary hypotheses involve the effect of the angular disparity on the latency and accuracy data in each display condition. First, I analyze the correct response latencies represented in the top row of Figure 4. Suggesting an analog transformation, the realistic condition latencies had a positive linear trend across the mismatch and match problems, except for the two smallest angles of the undershoot problems. In contrast, the abstract-condition latencies had a positive trend for the match problems but a negative or flat trend for the mismatch problems. This suggests that subjects may have used a feature-based, rather than analog, solution for the abstract mismatch problems.

To analyze the latency data, regression slopes of latency on angular disparity were fit for each subject separately for the correctly answered undershoot, match, and overshoot problems. The resulting coefficients served as dependent measures in multivariate analyses (cf. Lorch & Myers, 1990). Table 1 shows the means of the individuals' regression coefficients within each condition and the variability across subjects. The first analysis tested whether the realistic-condition data had a sufficient linear trend to support the interpretation of an analog transformation for the match and mismatch problems.
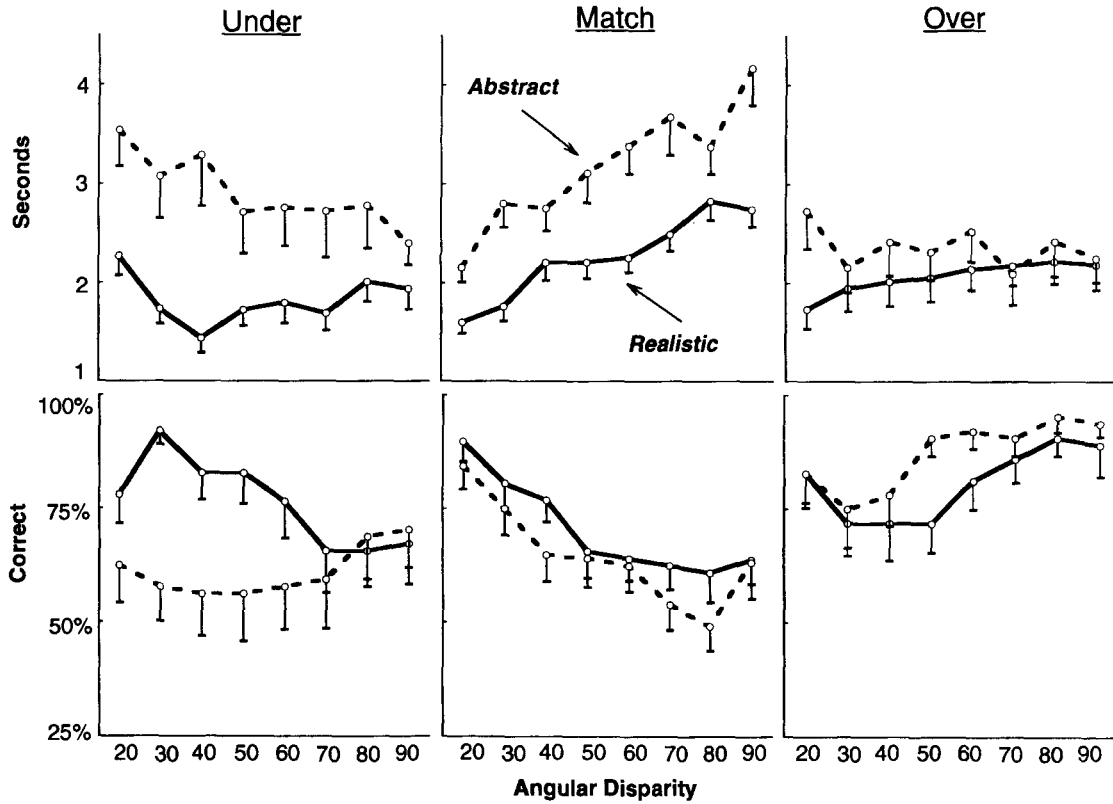
Figure 4. Correct response latencies and percent correct across the angular disparities for each problem type and display condition (Experiment 1). Error bars represent the standard error of the mean across trials for latencies and across subjects for the percents correct.

The slope and intercept coefficients for each individual served as the data in a doubly repeated multivariate analysis with problem type (under, match, over) as the within-subject factor. For the realistic condition, there was an overall positive slope [$F(1,15) = 9.62, p < .01$]. The slopes and intercepts did not reliably differ for the three problem types [$F(2,12) = 2.75, p > .1, HF = .41$, and $F(2,12) = .46, p > .1$, respectively].[3] Table 1 shows that the undershoot problems yielded a descriptively flatter slope than the other two problem types. This flatness was an artifact of the 20° problems which had inflated laten-

cies and errors in both display conditions. The extremely small mismatch offset may have been problematic for the subjects. Excluding the 20° problems, the average undershoot slope became .010 sec/deg instead of .003 sec/deg and the intercept became 1.66 sec instead of 1.75 sec. These coefficients are more in line with the coefficients of the match and overshoot problems which were not appreciably changed by omitting the 20° problems.

For the abstract display (including the 20° problems), the combined mismatch slopes were reliably different from the match slopes [$F(1,15) = 21.07, p < .01$], as were

Table 1
Within-Subject Linearity Measures on Angular Disparity (Experiment 1)

	Latency						Accuracy			
	Slope		Intercept		R^2		$Exp(B)$		% Correct	
Condition	M	SD	M	SD	M	SD	M	SD	M	SD
Realistic										
Under	.003	.013	1.75	.81	.074	.076	.977	.036	76.6	20.3
Match	.015	.013	1.57	1.44	.176	.116	.976	.026	70.8	19.8
Over	.007	.021	1.85	1.87	.110	.095	1.023	.095	80.6	19.8
Abstract										
Under	−.013	.023	3.79	2.34	.081	.108	1.016	.037	61.2	28.0
Match	.027	.027	1.85	1.19	.136	.130	.979	.038	64.7	15.0
Over	−.007	.022	2.96	1.95	.110	.108	1.019	.035	87.3	13.9

Note—Standard deviations represent between-subjects variability. Latency slopes are in seconds/ degree and intercepts are in seconds. $Exp(B)$ represents the change in the odds of a correct answer for each degree of increase in angular disparity.

the intercepts [$F(1,15) = 15.25, p < .01$]. These results may be interpreted as evidence that subjects in the abstract condition used an analog transformation for the match problems but a feature comparison for the mismatch problems.

To test whether the patterns in the two conditions were reliably different, a new statistical design included the between-subjects factor of display condition. The key contrast tested whether the relationships between the match and mismatch coefficients were different in the two conditions. In formulaic terms, the contrast was whether *Abstract*: (*Under* + *Over*)/2 − *Match* = *Realistic*: (*Under* + *Over*)/2 − *Match*. The contrast was reliable [Hotellings $F(4,27) = 4.64, p < .01$] and extended to the slopes [$F(1,16) = 9.87, p < .01, HF = .53$] and intercepts [$F(1,16) = 9.07, p < .01$].

Assuming analog solutions for the match problems in both display conditions, the evidence that subjects in the abstract condition began their analog transformations later was mixed. The contrast of the two conditions did not reveal a reliable difference between the match-problem slopes or intercepts [Hotellings $F(2,29) = 2.05, p > .1$]. This null result is somewhat surprising given the descriptive differences between the conditions. Forgoing the generalizability of a within-subject analysis, I further explored the data with a multiple regression. Using the correctly answered match problems, the dependent measure was response time and the independent measures were the display condition, angular disparity, and the condition × disparity interaction. The best regression model [$F(2,1380) = 72.84, p < .01, R = .31, SE = 2.04$] included the display condition variable [$t(1380) = 8.23, p < .01, B = .90, SE = .11$], the angular disparity variable [$t(1380) = 8.83, p < .01, B = .021, SE = .00$], and the constant term [$t(1380) = 7.88, p < .01, B = 1.13, SE = .14$]. The condition × disparity interaction was not significant [$t(1379) = 1.58, p > .1$]. Interpreting these results in terms of an analog transformation, both conditions exhibited an average rotation rate of 48.2°/sec for the match problems. The intercept was 1.13 sec for the realistic condition and 2.04 sec for the abstract condition. The aim of Experiment 2 was to discover whether individual differences could explain why this analysis revealed a significant difference between the intercepts of the two conditions whereas the within-subject analysis did not.

The predictions made for the accuracy data received mixed support from the data represented in the bottom row of Figure 4. The realistic condition was expected to invite analog solutions which become more error prone for larger angular disparities for all three problem types. The data showed that the undershoot and match problems did yield decreasing accuracy (except for the 20° undershoot problem). However, the overshoot problems yielded increasing accuracy past 50°. The abstract condition showed the predicted drop in accuracy for the match problems and a modest increase in accuracy for the undershoot and overshoot problems, as would be expected if a feature-based strategy was capitalizing on the larger mismatches.

The bottom half of Table 1 contains the means and between-subjects variability of the individuals' percent correct and the *Exp(B)* coefficients of accuracy on angular disparity. The *Exp(B)* coefficients represent how the odds of a correct answer change as a function of the angular disparity. Values greater than 1 indicate that accuracy increases with the angular disparity, whereas values less than 1 indicate that accuracy declines with the angular disparity. A multivariate analysis with display condition as a between-subjects factor and problem type as a within-subject factor was performed on the percent correct and *Exp(B)* values. There was a reliable difference in the pattern of results between each condition [Hotellings $F(4,27) = 3.46, p < .05$]. The univariate statistics locate this effect in the *Exp(B)* values [$F(2,24) = 3.98, p < .05; HF = .41$] but not the percent correct [$F(2,24) = 2.49, p > .1$].[4]

In the abstract condition there was a strong tendency to judge a problem as an overshoot. This bias undermines the latency analyses; subjects may have guessed for many of the correct answers that were used in the latency analyses. Nonetheless, even if subjects were ultimately guessing, the reliable differences in the accuracy and latency trends within the abstract condition and between conditions suggests that there were real strategic differences.

Discussion

Latencies, variances, errors, and linear trends differed between the two conditions. These results document effects of different graphical displays in the hinge task. Unfortunately, the low correlations, the high between-subjects variability, and the imperfect fit of the data with the original predictions mitigate strong process claims. After I interpret the main results from the current study, I describe possible explanations for the noisy nature of the data.

For the realistic drawing of the hinge, the leading hypothesis was that subjects would simulate the hinge's pivoting behavior. Based on analog imagery research, latencies and error levels should increase with the size of the pivot. To a large extent, the data supported these predictions. Even for the undershoot and overshoot problems in which the size of a mismatch increased with the size of the required pivot, latencies and errors generally increased with angular disparity. This is a compelling finding, because the increasing mismatch discriminability would presumably counteract some of the effects of having to conduct a longer transformation. Two exceptions to the predicted linear patterns were found for the 20° undershoot problems and the larger disparities of the overshoot problems.

For the abstract condition, the leading hypothesis was that subjects would initially try to solve the problems by relying on static features of the image. For the mismatch problems, errors and latencies should drop as the angular disparity increases, because the size and discriminability of a mismatch also increases. For the match problems, the feature-based comparison should be dif-

ficult because subjects must decide whether a small difference is within their margin of judgment error. Consequently, subjects might abandon the feature-based, rejection strategy for the analog strategy. To a large extent, the data supported this process model. For the mismatch problems, latencies and errors tended to decline with increases in angular disparity. For the match problems, latencies and errors tended to rise with increases in angular disparity. Moreover, the intercept for the match problems was larger in the abstract condition than in the realistic condition, suggesting that a feature-based effort may have preceded an analog solution in the abstract condition. However, in the current experiment, the latter result was not reliable when the data were nested within subjects. Exceptions to the predicted patterns were that the overshoot latencies were quite flat across the angular disparities, and the errors did not drop as the disparities increased for the undershoot problems.

A notable aspect of the mismatch data for each condition was that either the latency data or the error data followed the predicted pattern, but not both. For example, for the realistic undershoots, accuracy declined with increases in angular disparity (as predicted), but latencies did not rise as much as expected. And, for the abstract undershoots, latencies declined with the increasing angular disparity (as predicted), but accuracy did not rise as much as expected. One possible remediation for this speed–accuracy tradeoff is to increase the rate at which the offset grows over angular disparity for the abstract problems and to reduce the rate for the realistic condition. This might cause the slopes to rotate upward a small amount in each condition. To avoid the confound of this remediation, in the following experiment I only modified the size of the overshoot offsets to remove the overshoot bias found in the abstract condition.

The relatively low correlation between latency and angular disparity may have three sources. The first source is the variability inherent in a difficult task. In traditional analog imagery tasks, subjects make configural comparisons (e.g., two letter shapes). Consequently, the differences between an identical and mirrored image reveal themselves throughout the image after the rotation is completed. In the hinge task, the discrimination between a match and mismatch problem involved fractions of centimeters. The second source of the low correlations may be the small range of usable angular disparities. This small range was necessary because gravity pulls the board in the wrong direction for angles greater than 90°. Because the hinge task is limited to smaller disparity increments over a smaller angular range than most analog imagery tasks, differences in latency across angular disparities will be less pronounced. Consequently, there is less of an opportunity for strong correlations against background noise. The third source of low correlations may be strategy switching. Although the evidence for the two conditions pointed toward distinct strategies, subjects probably mixed a number of different strategies (cf. Lohman & Kyllonen, 1983). The clearest

evidence of mixing comes by comparing the match and mismatch data from the abstract condition. Debriefing also provided evidence of strategy mixing. Subjects from the realistic condition mentioned several strategies, ranging from pivoting to tilting their heads. Subjects from the abstract condition mentioned numerous strategies based on the static features of the problem, including comparing line lengths, projecting right angles, looking for isosceles triangles, bisecting angles, and projecting a circle with the hinge axis as the center point.

Individual differences may have added further variability to the results. For example, in the realistic condition there were 2 subjects who did not appear to have used an analog transformation. One subject had a 4-sec intercept term and 1 had a negative slope. In the following study, a larger sample may help determine if these differences were due to stable individual differences in strategy or were the tail end of a normal distribution.

EXPERIMENT 2

The second experiment tested the stability of the first experiment's results by examining whether the linear patterns replicated under several modifications. This is important, given the high level of unexplained variance. I therefore attempted in this experiment to improve accuracy in the abstract condition so that latency comparisons between conditions would not be potentially obscured by overall accuracy differences.

Several adjustments to improve the profile and interpretability of the data were made in this experiment. First, the number of training trials was doubled to allow subjects sufficient time to calibrate their mismatch tolerances and to settle into a consistent strategy. Second, there was only a single upper-board length to reduce variance associated with two board lengths. Third, there were no 20° or 90° problems. The 20° problems may have required too much precision, and the 90° problems may have invited a more feature-based approach for the realistic condition because of the physical impossibility that a taut rope could hold the board at this angle (cf. Shiffar & Freyd, 1991). Fourth, the number of problems at each angular disparity was increased to yield a more stable estimate for linearity analyses. Fifth, the gap offsets for the undershoot and overshoot problems were made symmetrical. In Experiment 1, the offset of the overshoots increased faster than the undershoots (see Figure 3). This asymmetry may have led to a number of the perplexing accuracy results. Finally, the population sample was increased by 50% to support an analysis of individual differences in strategy use.

Method

Subjects. Forty-eight adults responded to advertisements posted in the Columbia University community offering compensation for participation in a computer display study. (Two subjects had to be replaced due to experimenter error.) The subjects were assigned randomly, with 8 males in the realistic condition and 7 males in the abstract condition.

Design and Procedure. The following minor adjustments were made to Experiment 1. The angular disparity ranged from 30° to 80°. The top board was always 8.25 cm. The bottom mark was set at 3.25, 4.25, and 5.25 cm for each angular disparity. The subjects saw four match, two undershoot, and two overshoot problems for each mark distance at each angular disparity. The gap offsets of the overshoot problems were the same as the equivalent undershoots. All told, there were 144 recorded trials (6 angular disparities × 3 mark distances × 8 match/mismatch). Procedurally, the subjects saw 48 training trials and took a brief break between the training and the recorded trials.

Results

The experimental modifications brought the two conditions closer in terms of average accuracy and latency. The realistic condition had an average correct response latency of 2.44 sec (SD = 2.05) with an accuracy of 70.0% (SD = 14.3% across subjects). The abstract condition had an average correct response latency of 2.94 sec (SD = 2.15) with an accuracy of 71.0% (SD = 12.3%). The top row of Figure 5 indicates that the latency patterns were similar to the first experiment. The match problems in both conditions led to positive linear trends, whereas the no-match problems led to a crossover interaction between the conditions. Table 2 shows that for the undershoots and overshoots, the realistic condition tended to have positive slopes, whereas the abstract condition

tended to have negative slopes. As in Experiment 1, there was a high level of unexplained variance. However, if one covers the 30° problems in Figure 5, the mismatch problems show fairly consistent slopes.

An analysis identical to Experiment 1 used the coefficients found by regressing correct response latencies on angular disparity for each subject for each problem type (see Table 2). Of main interest, the slopes and intercepts from the match and the mismatch problems exhibited the same crossover interaction between the two conditions as they had in Experiment 1 [Hotellings $F(4,43)$ = 3.32, $p < .05$]. When separating out the contrast of the slopes, a subject in the realistic condition had a match intercept of 17 sec and added undue violations to the sphericity assumption needed for retrieving the univariate statistics (see Note 3). When excluding the outlying subject, the display condition × problem type interaction was significant for the slopes [$F(1,24)$ = 4.27, $p < .05$, HF = .53].

Improving upon the marginal evidence of Experiment 1, the match-problem intercepts from each condition were significantly different [$F(1,46)$ = 4.57, $p < .05$]. As before, the match slopes were not reliably different [$F(1,46)$ = .55, $p < .5$]. To facilitate comparison to Experiment 1, I regressed the angular disparity, the display condition, and the condition × disparity interaction against the
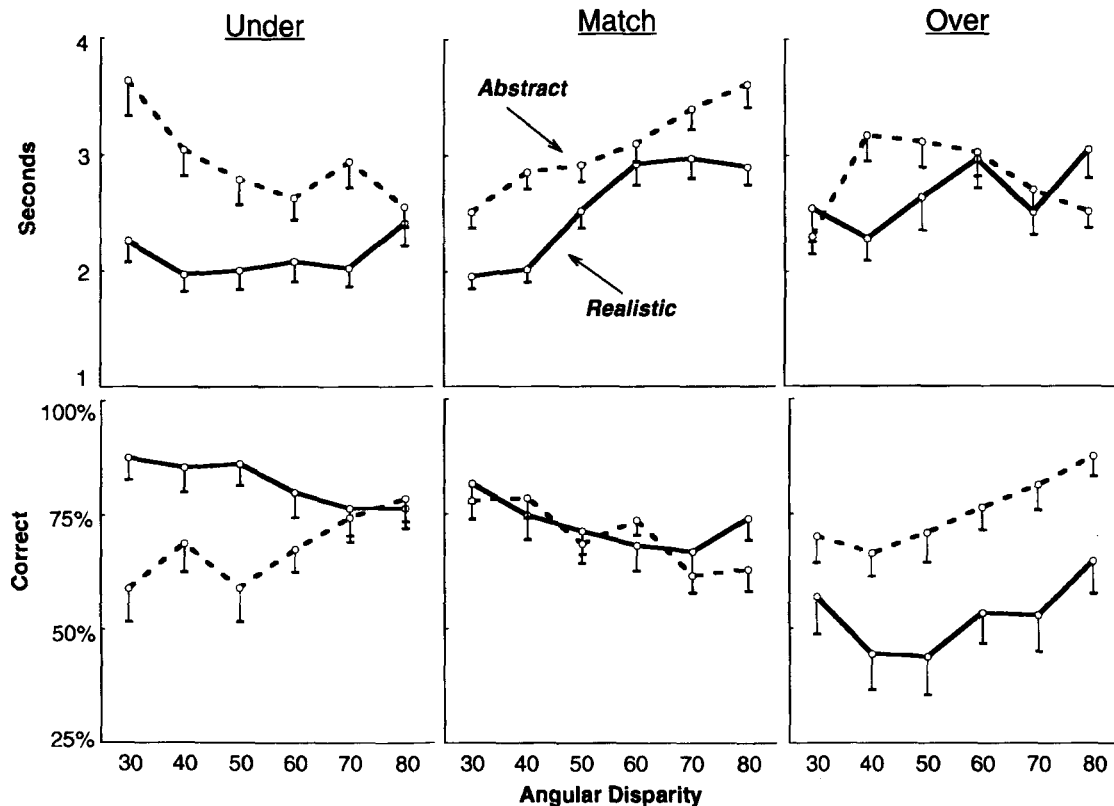

Figure 5. Correct response latencies and percent correct across the angular disparities for each problem type and display condition (Experiment 2). Error bars represent the standard error of the mean across trials for latencies and across subjects for the percents correct.

Table 2
Within-Subject Linearity Measures on Angular Disparity (Experiment 2)

| | Latency | | | | | | Accuracy | | | |
| | Slope | | Intercept | | R^2 | | $Exp(B)$ | | % Correct | |
Condition	M	SD	M	SD	M	SD	M	SD	M	SD
Realistic										
Under	.006	.033	1.90	1.98	.096	.100	.985	.052	81.9	20.4
Match	.026	.023	1.30	1.02	.105	.093	.986	.036	72.7	19.8
Over	.009	.053	2.60	4.14*	.157	.155	1.016	.049	52.7	32.0
Abstract										
Under	−.021	.032	4.23	2.58	.090	.160	1.024	.049	67.8	23.9
Match	.021	.017	1.88	.87	.060	.050	.978	.033	70.5	14.0
Over	−.004	.022	2.98	1.72	.080	.082	1.042	.061	75.4	21.8

*The elevated values are due to one subject who had an intercept of 17 sec. If excluded, the means of the overshoot slopes and intercepts become 0.16 sec/deg and 1.95 sec, respectively.

correct-response latencies (without nesting by subject). The best model [$F(2,2469) = 55.7$, $p < .01$, $R = .21$, $SE = 2.04$] includes the angle disparity term [$t(2469) = 8.73$, $B = .022$, $SE = .00$], the display condition term [$t(2469) = 6.10$, $B = .53$, $SE = .09$], and the constant term [$t(2469) = 9.05$, $B = 1.34$, $SE = .15$]. The condition by disparity interaction was not significant [$t(2468) = −.4$]. In terms of an analog transformation, both groups had a pivoting rate of 47.8°/sec. The intercept was 1.34 sec for the realistic condition and 1.87 sec for the abstract condition. These results are comparable to the first experiment for which the pivoting rate was 48.2°/sec and the intercepts for the realistic and abstract conditions were 1.13 and 2.04 sec, respectively.

The right-hand columns of Table 2 and the bottom row of Figure 5 show that the realistic and abstract conditions had different error trends, as in Experiment 1. A multivariate analysis of the individuals' $Exp(B)$ coefficients (i.e., the slope of accuracy by angular disparity) and percents correct for the three problem types indicated a reliable difference between the two conditions [$F(4,43) = 3.93, p < .01$]. This effect distributed over two contrasts. Due to the poor performances on the overshoot problems in the realistic condition, the percents correct on the overshoot and undershoot problems were in a different relationship in the two conditions [$F(1,22) = 12.01, p < .01$; $HF = .49$]. More importantly, the combined overshoot

and undershoot slopes were in a different relationship to the match slopes in the two conditions [$F(1,22) = 6.34$, $p < .05$].

Figure 6 shows that subjects' overshoot accuracy in the realistic condition was in a bimodal distribution. This result suggests stable individual differences. However, when searching for patterns that might reveal different strategies, there were no systematic correlations. Correlation analyses that attempted to separate the two modes of the distribution on other dependent measures accounted for less than 1% of the variance. Table 3 shows the regression statistics of latency on angular disparity separated by the above- and below-chance performers. The only notable pattern was the inflated overshoot latencies for those who performed below chance. This was due to the combination of a few long latencies and a small sample of correct response trials due to the below-chance performances. The lack of differentiation on the latency measures makes it difficult to assign strategy differences to the two modes of the accuracy distribution.

Discussion

The latency and accuracy trends over angular disparity echoed the findings of Experiment 1. The realistic and abstract displays led to opposing trends on the mismatch problems, and the abstract display led to longer latencies on the match problems. Notably, the linear latency pat-
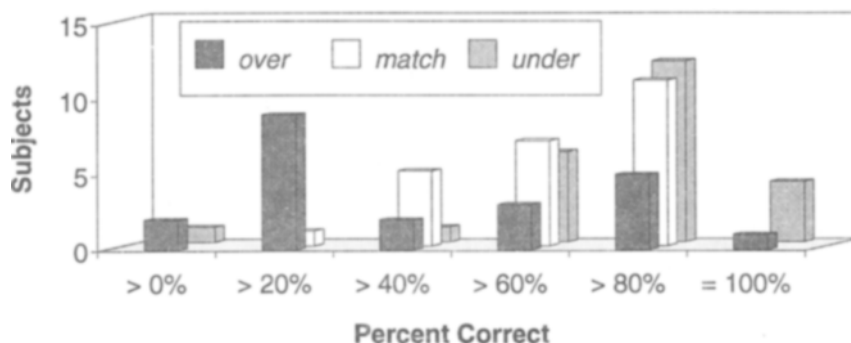
Figure 6. Accuracy distributions for the realistic condition (Experiment 2).

Table 3
Latency Regression Means for the Realistic Condition Separated
by Overshoot Accuracy (Experiment 2)

Overshoot Accuracy	n	Problems	Slope (sec/°)		Intercept (sec)	
			M	SD	M	SD
Below chance	13	Under	.004	.026	1.92	1.76
		Match	.023	.023	1.29	1.29
		Over	.012	.069	3.01	5.44
Above chance	11	Under	.010	.042	1.87	2.09
		Match	.029	.024	1.31	.97
		Over	.005	.027	2.11	1.82

terns replicated Experiment 1, even though the abstract subjects did not show the previous strong tendency to judge a problem as an overshoot, and many subjects in the realistic condition gained a tendency to judge an overshoot problem as a match. The overshoot problems in the realistic condition yielded a bimodal accuracy distribution with about half of the subjects performing below chance. However, no evidence was found that differentiated the subjects in either mode on any of the other dependent measures. Given that the crossover interaction in correct rejection latencies held up in a second experiment that had low correlations, identifiable individual differences, and a shift in response bias (i.e., the overshoot problems), the effect of the graphic display on reasoning appears to be a stable result. To further document the phenomenon, one possible step is to find a better function for creating the mismatch offsets (e.g., a logarithmic function; Buckley & Gillam, 1974). This might yield data more perfectly in accord with the original predictions. However, given the inherent variability of this difficult task, a more appropriate next step may be to generalize these results to other stimuli and tasks.

GENERAL DISCUSSION

The current experiments demonstrated differential performances in both latency and accuracy that result from two levels of perceptual fidelity on the same task. This task was to determine whether the marks on either leg of a hinge would meet if the hinge closed. Process models contrasting analog imagery and feature-based judgments provided the basic framework for describing the empirical results. The model for the analog process was that subjects transformed a mental image of the hinge as they would expect the hinge to behave in the world. Data from the more realistic display supported this model, because there was a positive linear relationship between angular disparity and latency. In its own right, this is a useful finding, because it indicates that people can use analog imagery to reason about physically constrained transformations (i.e., about a pivot point). The interpretation of an analog transformation for the realistic display gains strength when it is contrasted with the data from the abstract condition, data which I interpret as demonstrating the different results that come from the main hypothetical alternative to an analog strategy.

This alternative is a more discrete, feature-based strategy. The model for the feature-based process was that subjects initially followed a rejection-oriented strategy in which they looked for metric differences in static features of the display. The data from the abstract display supported this model, because mismatch latencies dropped as the angular disparity became greater, presumably due to the increasing mismatch discriminability. Further supporting this interpretation, mismatch accuracy increased with the angular disparity for the abstract display but not for the realistic display. Finally, for the match problems of the abstract display, latencies increased with the angular disparity at the same rate as in the realistic condition. Importantly, the intercept for the abstract match problems was over a half second greater than the realistic problems' intercept. This suggests that after subjects in the abstract condition could not detect a difference using a feature-based approach, they switched to an analog strategy.

One problematic aspect of the process models for each condition is that the data had a high level of unexplained variance that makes it difficult to infer stable internal processes. Nonetheless, the crossover interactions from Experiment 1 held up in Experiment 2, although there were several changes in procedure and results. For example, unlike Experiment 1, the overall accuracy and variability of the abstract condition in Experiment 2 approximated that of the realistic condition. Thus, the different patterns from the two conditions cannot be ascribed to more general performance issues such as not being able to see the abstract display clearly. Given the stability of the overall patterns in the midst of high variability, a prudent conclusion might be that the two process accounts are descriptions of the modal strategies for the two displays.

Given the effect of pictorial realism, it becomes useful to consider what might cause this effect. According to the fidelity hypothesis, the integrity of the visual information determines whether one reasons about the referent or the image. Schwartz and Black's (in press) computer simulation of depictive reasoning provides one possible account of the psychological mechanisms. In this simulation, a mental model was constructed around the information provided in the visual array. Of particular importance was whether the provided information was sufficient for constructing a referential model that

included constraints for resolving a physical or spatial relationship without excessive memorial enrichment (cf. Farah, Rochlin, & Klein, 1994; Marr, 1982). For example, the program simulated how people might imagine the coordinated rotations of two meshing gears. In one case, the simulation received gear pictures that included surface information about the gear circumferences. With this information, the computer simulation constructed a model that used the friction at the surface contact points to coordinate the rotations of the two gears. In a second case, the picture of the gears was a simple line drawing. Consequently, the program could not use perceptual information about the gear surfaces to support a simulation that depended on surface friction. Instead, the program extracted metric information about the radii of the circles and applied more general geometric principles. For the current experiments, the solidity and implied mass of the realistic hinge drawing may have supported the construction of a representation that could depict the behavior of the hinge. In the abstract display, there was less physically relevant information available for the construction of a depictive model.

A competing account is that subjects reasoned about dynamic behaviors with the realistic display because the extra pictorial detail obscured the precise metrics and made it difficult to reason in a more feature-based manner. Although this might be the case for other pictures with distracting detail, it seems unlikely for the hinge task, given that subjects in the realistic condition were able to use size information with an accuracy that matched the abstract condition. Although anecdotal, it is informative to note the subjects' introspection on the task. During debriefing, subjects were shown displays from each condition and were asked to predict which condition they thought would yield the better results. Notably, many of the subjects felt that the realistic display was easier to work with even though they noted that there were more distracting details. They felt that the realistic display made them think of concrete objects and gave them something to "hold on to." This supports the idea that the informational completeness or fidelity of an image supports the construction of a referential model (cf. Cooper, 1989), not that the complexity of the image prevents a feature-based solution. Further research using more variations in pictorial complexity could address this question more fully.

A potential alternative to the fidelity hypothesis might be termed the "similarity hypothesis." This hypothesis posits that the reasoning for a specific display is determined by the similarity of that display to prior experiences. The hypothesis might be framed in terms of priming so that each display primed its own body of prior knowledge. For example, the schematic drawing of the abstract condition may have primed the feature-based reasoning associated with similar graphic conventions. The similarity hypothesis is akin to Hochberg's (1974) suggestion that ambiguous visual information is enriched by relying on mental representations of similar, previously experienced information. The similarity hypothesis could complement the fidelity hypothesis, if one assumes that priming effects are secondary to the initial referential disambiguation. For example, reasoning about an image or representation requires access to some prior knowledge, and priming provides a general account for the knowledge retrieval. Indeed, evidence of consistent priming comes from the many subjects in the abstract condition who explicitly mentioned trying to use various geometric principles. However, the similarity and fidelity hypotheses do differ. With the similarity hypothesis, an impoverished image primes memories for similar impoverished images. With the fidelity hypothesis, the impoverished information initially leads one to think about the representation rather than its referent, regardless of the similarity to the previous images one has experienced. This leads to at least two operational differences that can help experimentally to tease apart the relative contributions of fidelity and priming.

According to the fidelity hypothesis, impoverished images should become the referent even if one has had no prior experience that could be primed by the particular type of infidelity shown in the image. For example, painters that preceded the Renaissance, such as Giotto, experimented with depth cues (e.g., Guillaud & Guillaud, 1987). Their solutions, which predated the development of current perspective techniques, should not remind many people of any other perceptual experience. In this situation, the fidelity hypothesis predicts that one initially reasons about the painting as the referent rather than the scene it portrays, whereas the similarity hypothesis has no direct prediction. The second operational difference is that the fidelity hypothesis posits that the effect is based in perceptual information and is not simply a demonstration of general priming effects. This means that it should be difficult to undermine the initial effects of perceptual fidelity by priming conceptual knowledge or knowledge of social expectations. So, according to the fidelity hypothesis, inducing a "feature-based" set should have a minimal influence on how one solves the hinge problem. For example, short of directives to the contrary, subjects should tend to use analog imagery for the realistic image even if they receive problems that alternate between realistic and abstract presentations or if they have just previously solved multiple geometry problems about triangles.

For the current purposes of showing that an image provides information that helps resolve the one-to-many mapping between picture and reference, I have operationalized two levels of reference by contrasting analog and feature-based reasoning for two levels of perceptual fidelity. However, there are many different ways people can construe the relationship between pictorial representation and reference. For example, one might view a picture as representing a class of referents rather than a specific referent, as is often the case with young children's drawings (Gardner, 1980). In the current case, given the lack of individuating perceptual information in the abstract display, subjects may have seen the abstract image as the representation of the idea of a hinge, not

any existing hinge. For the realistic display, subjects may have seen a specific physical hinge. It would be informative to develop tasks that could differentiate the reasoning for the many possible referents of a visual representation (e.g., an existing referent, a possibly existing referent, a class of referents, a picture of a picture of a referent, the picture itself, etc.). Perhaps, using real hinges, photographs of hinges, cartoons of hinges, pictures of pictures of hinges, and diagrams of hinges could provide some leverage on the issue. However, for the current task, determining what features or lack of features are sufficient to lead people to reason about different levels of reference may be impossible given the narrow bandwidth of possibly discriminating evidence. More generally, it may be very difficult to use a single task to delineate a discrete list of pictorial features and accompanying referential interpretations, given the interactions of different visual features, cultural and perceptual experiences, and spatial and semantic knowledge (Hochberg, 1962).

Regardless of the ultimate explanation of the current effects of graphic realism, the implications of the experiments are fairly clear with respect to research that employs diagrams or pictures as surrogates for real physical objects. One cannot assume that research conclusions based on diagrammatic materials will generalize to physical materials. In particular, diagrams may yield more static, feature-based reasoning than the simulative reasoning one might find with a picture or physical object.

The current studies may also have implications for education, although the effects are diminutive compared with other influences in the classroom. As computers become a more prevalent part of the classroom experience, the designer may want to make strategic decisions about when to use realistic or schematic graphics. One possible guideline is that if one wants the student to imagine the referent, one should use photorealism. However, if one wants the student to reflect on more symbolic or formal descriptions of the object, one should use schematic diagrams. For example, one might use an actual image of a person sliding down an inclined plane to help the student think in terms of forces as they are perceptually experienced. After these perceptual notions are brought to mind and reflected upon, one may want to switch to more abstract drawings of two dimensional blocks on oblique lines. Perhaps, the student will then be in a position to connect intuitive and analytic treatments of mass, friction, and gravity.

REFERENCES

ARNHEIM, R. (1969). *Visual thinking*. London: Faber & Faber.

BIEDERMAN, I., MEZZANOTTE, R. J., & RABINOWITZ, J. C. (1982). Scene perception: Detecting and judging objects undergoing relational violations. *Cognitive Psychology*, 14, 143-177.

BUCKLEY, P. B., & GILLMAN, C. B. (1974). Comparison of digits and dot patterns. *Journal of Experimental Psychology*, 103, 1131-1136.

BUSWELL, G. T. (1935). *How people look at pictures*. Chicago: University of Chicago Press.

CHASTAIN, G., & BURNHAM, C. A. (1975). The first glimpse determines the perception of an ambiguous figure. *Perception & Psychophysics*, 17, 221-224.

CLIFF, N. (1987). *Analyzing multivariate data*. New York: Harcourt Brace Janovich.

COOPER, L. A. (1976). Individual differences in visual comparison processes. *Perception & Psychophysics*, 19, 433-444.

COOPER, L. A. (1989). Mental models of the structure of visual objects. In B. E. Shepp & S. Ballesteros (Eds.), *Object perception: Structure and process* (pp. 91-119). Hillsdale, NJ: Erlbaum.

COOPER, L. A., & PODGORNY, P. (1976). Mental transformations and visual comparison process: Effects of complexity and similarity. *Journal of Experimental Psychology: Human Perception & Performance*, 2, 503-514.

DELOACHE, J. S., & MARZOLF, D. P. (1992). When a picture is not worth a thousand words: Young children's understanding of pictures and models. *Cognitive Development*, 7, 317-329.

DEREGOWSKI, J. B., MULDROW, E. S., & MULDROW, W. F. (1972). Pictorial recognition in a remote Ethiopian population. *Perception*, 1, 417-425.

FARAH, M. J., ROCHLIN, R., & KLEIN, K. L. (1994). Orientation invariance and geometric primitives in shape recognition. *Cognitive Science*, 18, 325-344.

FERGUSON, E. L., & HEGARTY, M. (1995). Learning with real machines or diagrams: Application of knowledge to real-world problems. *Cognition & Instruction*, 13, 129-160.

FOSTER, D. H. (1975). Visual apparent motion and some preferred paths in rotation group SO(3). *Biological Cybernetics*, 18, 81-89.

FREYD, J. J., PANTZER, T. M., & CHENG, J. L. (1988). Representing statics as forces in equilibrium. *Journal of Experimental Psychology: General*, 117, 395-407.

GARDNER, H. (1980). *Artful scribbles*. New York: Basic Books.

GIBSON, J. J., & BRIDGEMAN, B. (1987). The visual perception of surface texture in photographs. *Psychological Research*, 49, 1-5.

GOMBRICH, E. H. (1974). Projective transformations as determining visual space perception. In R. B. MacLeod & H. L. Pick, Jr. (Eds.), *Perception: Essays in honor of James J. Gibson* (pp. 84-94). Ithaca, NY: Cornell University Press.

GUILLAUD, J., & GUILLAUD, M. (1987). *Giotto: Architect of color and form*. New York: Clarkson N. Potter.

HANSON, N. R. (1961). *Patterns of discovery*. Cambridge: Cambridge University Press.

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

HOCHBERG, J. (1962). The psychophysics of pictorial perception. *Audio-Visual Communication Review*, 10, 22-54.

HOCHBERG, J. (1974). Higher-order stimuli and inter-response coupling in the perception of the visual world. In R. B. MacLeod & H. L. Pick, Jr. (Eds.), *Perception: Essays in honor of James J. Gibson* (pp. 17-39). Ithaca, NY: Cornell University Press.

HOCHBERG, J., & BROOKS, V. (1962). Pictorial recognition as an unlearned ability: A study of one child's performance. *American Journal of Psychology*, 75, 624-628.

JOHNSON, D. M. (1939). Confidence and speed in the two-category judgment. *Archives of Psychology*, 34, 1-52.

JOSEPH, J. J., & DWYER, F. M. (1984). The effects of prior knowledge, presentation mode and visual realism on student achievement. *Journal of Experimental Education*, 52, 110-121.

KAIL, R., CARTER, P., & PELLEGRINO, J. (1979). The locus of sex differences in spatial ability. *Perception & Psychophysics*, 26, 182-186.

KAISER, M., PROFFITT, D., & ANDERSON, K. (1985). Judgments of natural and anomalous trajectories in the presence and absence of motion. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 11, 795-803.

KIRK, R. E. (1982). *Experimental design: Procedures for the behavioral sciences* (2nd ed.). Pacific Grove, CA: Brooks/Cole.

LOHMAN, D. F., & KYLLONEN, P. C. (1983). Individual differences in solution strategy on spatial tasks. In R. F. Dillon & R. R. Schmeck (Eds.), *Individual differences in cognition* (Vol. 1, pp. 105-135). New York: Academic Press.

SCHWARTZ

Lorch, R. F., Jr., & Myers, J. L. (1990). Regression analyses of repeated measures data in cognitive research. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 16*, 149-157.

Mandler, J. M., & Stein, N. L. (1974). Recall and recognition of pictures by children as a function of organization and distractor similarity. *Journal of Experimental Psychology, 102*, 657-669.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information.* San Francisco: W. H. Freeman.

McLuhan, M. (1965). *Understanding media: The extensions of man.* New York: McGraw-Hill.

Mischel, W., Shoda, Y., & Rodriguez, M. L. (1989). Delay of gratification in children. *Science, 244*, 933-938.

Moore, J. L. (1993). *Comparisons of a physical model and computer representations in reasoning and learning about linear functions.* Unpublished doctoral dissertation, Stanford University.

Moore, J. L., & Schwartz, D. L. (1994). Mental models for proportional reasoning. In *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society* (pp. 640-645). Hillsdale, NJ: Erlbaum.

Moyer, R. S. (1973). Comparing objects in memory: Evidence suggesting an internal psychophysics. *Perception & Psychophysics, 13*, 180-184.

Murray, F. B. (1970). Stimulus mode and the conservation of weight and number. *Journal of Educational Psychology, 61*, 287-291.

Parkhurst, P. E., & Dwyer, F. M. (1983). An experimental assessment of students' IQ level and their ability to profit from visualized instruction. *Journal of Instructional Psychology, 10*, 9-20.

Parsons, L. M. (1987). Imagined spatial transformations of one's hands and feet. *Cognitive Psychology, 19*, 178-241.

Rock, I., Halper, F., & Clayton, T. (1972). The perception and recognition of complex figures. *Cognitive Psychology, 3*, 655-673.

Schwartz, D. L., & Black, J. B. (in press). Analog imagery in mental model reasoning: Depictive models. *Cognitive Psychology.*

Shepard, R. N. (1982). On turning something over in one's mind. In R. N. Shepard & L. A. Cooper (Eds.), *Mental images and their transformation* (pp. 19-23). Cambridge, MA: MIT Press. (Original work published 1968)

Shiffar, M., & Freyd, J. J. (1991). Apparent motion of the human body. *Psychological Science, 1*, 257-264.

Siegel, I. E. (1978). The development of pictorial comprehension. In B. S. Randhawa & W. E. Coffman (Eds.), *Visual learning, thinking, and communication* (pp. 93-111). New York: Academic Press.

Wertheimer, M. (1958). Principles of perceptual organization. In D. Beardslee & M. Wertheimer (Eds.), *Readings in perception* (pp. 115-135). Princeton: Van Nostrand.

NOTES

1. In the current work, I do not distinguish between informational impoverishment and infidelity. In general, they are highly correlated. Except for pictures that intentionally violate certain constraints (e.g., interposition), infidelities generally arise from informational omissions as in the case of omitting reflectance and supporting surfaces in the diagram of a device.

2. In both experiments, outliers greater than 3 SDs from the condition mean were replaced with the condition mean plus 3 SDs. This was less than 2% of the cases in all conditions.

3. To retrieve the univariate statistics, the degrees of freedom have been reduced from 30 to 12 according to the Huynh-Feldt correction (HF) for sphericity violations (Cliff, 1987). Correction values are noted throughout the text when they were required.

4. The more appropriate, but less interpretable, contrast involves an arcsine transformation of the ratio data (Kirk, 1982). This analysis yielded the same results as the analysis of the percents themselves.