Journal of Experimental Psychology:
Human Perception and Performance
1987, Vol. 13, No. 3, 395–404

0096-1523/87/$00.75

Effects of Stimulus Complexity on Mental Rotation Rate of Polygons

Mark D. Folk
Law School, Stanford University

R. Duncan Luce
Harvard University

Both spatial and propositional theories of imagery predict that the rate at which mental images can be rotated is slower the more complex the stimulus. Four experiments (three published and one unpublished) testing that hypothesis found no effect of complexity on rotation rate. It is argued that despite continued methodological improvements, subjects in the conditions of greater complexity may have found it sufficient to rotate only partial images, thereby vitiating the prediction. The two experiments reported here are based on the idea of making the discriminative response sufficiently difficult so as to force the rotation of complete images. The first one scaled the similarity between standard polygons and certain systematically mutated versions. From the ratings so obtained, two levels of perceived similarity, high and low, were defined and served as separate conditions in a response-time, image rotation experiment. The second experiment tested the complexity hypothesis by examining the effect of similarity on rotation rates and its interaction with levels of complexity. The results support the complexity hypothesis, but only for the highly similar stimuli. Rotation times were also generally slower for high as compared with low similarity. It is argued that these results arise because subjects rotate incomplete images when the stimuli are not very similar.

Shepard and Metzler (1971) showed that the time subjects require to determine whether two three-dimensional angular shapes are identical or different is an increasing linear function of the angular disparity between the two patterns. Since this initial demonstration, numerous studies have shown that the monotonic (often linear) relation between transformation time and distance (or angle) is robust. (See Shepard & Podgorny, 1978, and Kosslyn, 1980, for reviews.) One common, although disputed, interpretation of this empirical relation is that mental images are transformed in small steps, thus requiring them to pass through intermediate stages before the desired transformation is achieved (cf. Cooper & Shepard, 1973a). Although little doubt exists that the processing of visual representations is responsible for this phenomenon, considerable debate has emerged regarding the representational structures on which processing mechanisms operate (see Kosslyn & Pomerantz, 1977; Pylyshyn, 1973).

Theoretical accounts of visual imagery can be divided into two broad camps. Supporters of propositional theories (e.g., Anderson, 1978; Palmer, 1975) argue that visual information is represented in an abstract propositional format, not unlike the networks that support semantic information in artificial intelligence programs (e.g., Minsky & Papert, 1972) and visual representations in computer vision systems (e.g., Riseman & Hansen, 1978). Supporters of spatial or analog theories (e.g., Cooper & Shepard, 1973a; Kosslyn, Pinker, Smith, & Schwartz, 1979) counter that visual information is represented in an analog medium that at least partially preserves the metric and spatial information inherent in original real-world percepts.

Spatial image theories can themselves also be divided into two competing views about the way in which images are manipulated. The "holistic" spatial theory posits that images are transformed in a unitary process—that is, the entire image is mentally rotated[1] all at once. The "piecemeal" version, in contrast, suggests that image rotations are accomplished, first, by parsing the image into units and then by mentally rotating the individual pieces sequentially. (See Kosslyn, 1981, for a further account of this distinction.) Much recent research on image transformations has pitted differing predictions against each other from the several theoretical accounts in an attempt to allow behavioral data to determine the adequacy of each type of model. (For an alternative viewpoint, see Anderson, 1978).

In the course of this debate, the issue of whether stimulus complexity affects the rate of mental image rotation has generated a curious puzzle. Although both propositional and piecemeal theories predict that increasing internal complexity should slow rotation rate, four separate tests—two in Cooper (1975), Cooper and Podgorny (1976), and Schwartz (1979)—failed to find evidence for such an effect, which, if taken at face value, provides support for the holistic point of view. All of these studies used the same definition of stimulus complexity,

---

This work is extracted from the Harvard University senior honors thesis of the first author, which was awarded a 1983 Hoopes prize for an undergraduate paper. The research was supported by National Science Foundation Grant IST 79-24019, awarded to R. Duncan Luce and Louis Narens at the University of California, Irvine, and by Harvard University Research Grant HU 33-672-5411 awarded to R. Duncan Luce.

We wish to thank Stephen Kosslyn, Miriam Schustack, Robert Rosenthal, and William Vaughan for critical comments and helpful suggestions. Comments from two anonymous referees and Roger Shepard have led to many improvements and clarifications.

Correspondence concerning this article should be addressed to R. Duncan Luce, William James Hall 930, Harvard University, Cambridge, Massachusetts 02138.

[1] We use the term *mental rotation* throughout as a description of the phenomenon and in reference to the instructions given the subjects. It is not intended to prejudge the actual brain processes involved.

namely, the number of vertices determining inflections on the perimeter of Attneave and Arnoult (1956) Method 1 random polygons. (See Attneave, 1957, for a further account of this definition of complexity.)

Should these studies be so interpreted? Are both the propositional and piecemeal accounts of visual imagery largely misguided? Is the empirical definition of complexity incorrect? Or, despite continued refinements in previous research, are the methodologies and stimulus materials the culprits? We argue the latter, specifically claiming that the design of previous studies was not sufficiently sensitive to assess adequately the effect of complexity.

Our basic methodological point is simple: In order effectively to test for complexity effects, the experimental test must force subjects mentally to rotate structurally complete stimuli (in all cases, polygons) the full rotational angles at all levels of complexity. We first argue that this criterion was probably not met in any of the four studies mentioned. Second, we propose a design intended to increase the likelihood of its being met. And, finally, we report data that support the predictions of the effect of complexity on rotation rate.

Critique of Previous Studies on Stimulus Complexity

Cooper (1975) performed the first study bearing on the issue of stimulus complexity. She used Attneave and Arnoult's (1956) Method 1 random polygons at five levels of complexity consisting of 6, 8, 12, 16, and 24 points. Subjects first learned to discriminate "standard" versions of the figures from their reflected mirror-image versions. On test trials, polygons were presented to subjects in the upright position and at orientations that were 60°, 120°, 180°, 240°, and 300° from upright. The task was to determine whether a presentation was normal or mirror-reversed. Cooper's main finding was that there were no differences among times to evaluate figures of different complexity.

A possible reason why complexity failed to matter is that Cooper used as distractor items mirror-reversed polygons, and they were visible to subjects during the entire comparison task. Reflection is a global property of a stimulus, one that subjects can easily detect without maintaining highly detailed images of standard forms. For example, Hochberg and Gellman (1977) reported that images of stimuli containing highly perceptible "landmark" features were mentally rotated more quickly than images in which such identity cues were absent. It seems likely that Cooper's subjects found it unnecessary mentally to rotate all of the local, feature-specific information the full angle to succeed in the discrimination task.

Cooper's (1975) second experiment focused on the time required for subjects to "prepare" for the presentation of a stimulus in a particular orientation. She used the preparation task described by Cooper and Shepard (1973b) in which subjects were given both identity and orientation information and performed mental rotations to the indicated positions before the stimulus was presented. In Cooper's second experiment, subjects discriminated between the same set of standard and mirror-reversed polygons in the same orientations used in the first study. The results replicated Cooper's previous finding in that rotation rates were unaffected by stimulus complexity.

Why did complexity effects again fail to appear? It seems likely, for two reasons, that Cooper's preparation task did not require subjects mentally to rotate structurally complete polygons. The first reason relates to the extensive practice Cooper's subjects received. These subjects—the same ones as in the first experiment—practiced discriminating between the same set of normal and reflected stimuli for the second time. The second reason relates to the ease of discriminating between mirror-reversed and standard polygons (discussed above). In the context of the preparation task, if subjects anticipate that the upcoming discrimination will be easy, it seems unlikely that they will mentally rotate complete mental images when structurally partial images suffice. Both of these reasons raise the possibility that Cooper's subjects performed the task with incomplete "test form" representations and that, consequently, complexity variations in the polygons may not have been incorporated in images of these forms.

To eliminate that possibility, Cooper and Podgorny (1976) designed an experiment in which subjects discriminated not only between standard and reflected versions of random figures but also between standards and "mutated" versions—that is, patterns that were systematically related to standard forms. They selected one polygon from each of the five levels of complexity that Cooper (1975) had tested earlier. For each standard, the experimenters generated approximately 40 different mutants that varied in the number of perturbed points as well as in the extent of the perturbations.

The mutants included in the rotation task were chosen on the basis of a study that scaled the similarity of the sets of mutated polygons. Subjects rated the similarity of each mutant to its standard on a 7-point scale. They were encouraged to select ratings from the entire 7-point scale and to maintain a fixed criterion throughout. The final stimulus set associated with each standard consisted of six of the rated mutants together with the respective mirror-reflective patterns. All test trials required a preparatory rotation before the test form was presented. The main result of interest in this experiment was that, once again, stimulus complexity had no effect on rotation times.

The failure to detect complexity effects may, we argue, still result from a failure to induce subjects to rotate complete images. The use of perturbed stimuli was surely a step in the right direction, because discriminating them from the standard required more detailed internal representations than did discriminating mirror-reflected patterns. Still, there may have been a problem. Recall that raters were encouraged to maintain a fixed criterion for judging all of the figures. It is therefore plausible that more and larger perturbations were required for a complex mutant to be rated as equally similar to its standard than for a less complex one. Anderson (1978) pointed out that, consequently, subjects in the rotation test actually needed to remember a smaller proportion of points in complex images in order to obtain the same probability of discriminating a perturbation. Therefore, he claimed, it is still quite likely that, despite Cooper and Podgorny's attempt to eliminate the problem, subjects were performing the tasks with only partial images of complex polygons. In order to force subjects to maintain all the available structural information in their image of the original stimulus, Anderson suggested that all mutated stimuli should be constructed by mutating the same number of points by the same amount for all levels of complexity.

This experiment was performed by Schwartz (1979). He used Attneave and Arnoult's random polygons in a preparatory rotation task similar to that of Cooper and Podgorny. The polygons had 6 and 10 points, and mutants were formed by perturbing a single point by the same amount in all cases. The results showed that rotation rates for the 10-point figures were slightly slower than for the 6-point figures; however, increases in time to rotate more complex images through larger angles were not significant.

In addition to these four studies, two—Pylyshyn (1979) and Yuille and Steiger (1982)—reported evidence that the authors interpreted as an effect of complexity on rotation rate of mental images. Both studies, however, suffered from a design flaw, pointed out by Kosslyn (1980, p. 301) in his discussion of the Pylyshyn study, namely, that the two figures to be compared were presented simultaneously and thus there was no guarantee that subjects relied on mental rotation to discriminate between the patterns. For example, in one of the Yuille and Steiger experiments, two block figures were presented, one of which was the standard and the other was a rotated version of either its mirror image or the mirror image with a subset of the blocks twisted 90° out of alignment. They interpreted the fact that the subjects were slower in the mirror-twisted condition than in the mirror-alone condition as evidence that subjects carried out the discrimination task by making a piecemeal comparison of figure segments rather than a holistic image rotation. They argued that had holistic rotation been used, there should have been no difference between the two conditions. On the other hand, the authors noted that the more complex, twisted forms contained sufficient distinguishing features so that it was unnecessary to compare the entire figure with the standard to make a correct discrimination. In other words, with the figures presented simultaneously, it was sufficient to focus on the distinguishing features, and it was unclear that any, let alone complete, rotation was needed.

This conclusion sheds little light on the factors that influence performance when stimuli are presented sequentially and when preparatory rotations are required before test items are presented. In such designs, subjects cannot rely on distinguishing features because the figures cannot be compared directly. These results, as well as those of the other four studies, are consistent with the view that when discrimination is easily made by mentally rotating only a portion of the figure, that is exactly what subjects do.

Overcoming the Tendency to Rotate Partial Images

Despite continued methodological improvements, even in the later studies subjects may still have mentally rotated incomplete images. This difficulty must be corrected if the issue of complexity effects is to be resolved. It seems to us that the best way to ensure that subjects maintain complete images is to design a task in which it is to their advantage to do so. For example, if the required discrimination between an image and a test form is made sufficiently difficult, subjects may be forced to maintain more structural information in the rotated image—else they will be unsuccessful at the discrimination task. Because increasing the similarity of any two patterns increases the difficulty of discriminating between them (e.g., Moyer, 1973), that may be a way to induce the subjects to perform complete mental rotations with fully detailed images. That idea underlies our design.

Two experiments reported here investigated the complexity hypothesis by using stimuli with two levels of rated complexity, namely, polygons of the type used by Schwartz (1979) incorporating 6 and 10 points. To avoid possible artifacts arising from peculiar properties of individual stimuli, three standard forms were generated at each of the two levels. Mutants were constructed by perturbing a single point by the same amount in all mutated forms. In addition to testing for complexity effects, the question was raised whether the similarity of patterns in the discrimination task influenced the amount of structural information subjects maintain in rotated images (and thus affects rotation rate).

The idea was, first, to have subjects rate the similarity of mutated polygons. These ratings were used to define high- and low-similarity groups of mutants for each shape at each level of complexity. These stimuli were then used in a preparatory image rotation experiment, similar to that carried out by Cooper and Podgorny.

If the above reasoning about subjects' possible strategies in a preparatory rotation paradigm is correct, the two similarity conditions should affect rotation rate differently: In the condition of high similarity, where the required discrimination is intentionally difficult, subjects should be compelled to rotate mentally a considerable amount of structural information, whereas in the condition of low similarity, where the discrimination is easier, rotation of less complete images should suffice. For each level of complexity, rotation rates for highly similar mutants should be slower than for those of low similarity. In other words, similarity conditions should produce a main effect on rotation time. In addition, because more detailed images should be mentally rotated in the condition of high similarity than in that of low similarity, the following is predicted: Complexity should affect rotation rates in the high-similarity condition, as predicted by both propositional and piecemeal image theories (Anderson, 1978; Kosslyn, 1980). In contrast, because complete structural information is not needed to maintain discrimination success in the low-similarity condition, complexity should not affect rotation rates in that case, which would replicate previous results (particularly Schwartz, 1979).

Experiment 1

The standard polygons used in this study were three 6-sided random forms and three 10-sided figures. Twenty-four mutants were generated from each standard by perturbing a single point by a fixed amount in all cases; thus, one standard form and its 24 mutated versions can be thought of as a "family" of closely related shapes. One group of subjects was requested to rate on a 9-point scale the similarity of each mutant to its respective standard, thereby providing the data used to select mutants for the rotation distractor set. Aside from the perturbation procedure and the number of rating categories employed, Experiment 1 closely resembled the task described by Cooper and Podgorny (1976).

In addition to providing a basis for choosing distractor items for the preparatory rotation task, this experiment investigated the following hypothesis. In his critique of Schwartz's study, Anderson (1978) suggested that one-point perturbations might yield higher average similarity ratings for more complex figures. Pylyshyn (1979) corroborated this claim by arguing that similarity ratings were generally correlated with comparison difficulty. In other words, these authors anticipated that the one-point perturbations would produce a confounding of similarity with the complexity of the stimulus.

Method

Subjects. Twelve Harvard–Radcliffe undergraduates volunteered to participate as paid subjects in a single experimental session that lasted between 1½ and 2 hr.

Stimuli and apparatus. Six standard polygons, three with 6 sides and three with 10 sides, were constructed according to Attneave and Arnoult's (1956) Method 1 by plotting and connecting points on Cartesian graph paper (100 × 100 squares, 5 squares to the centimeter). The six standard polygons were assigned alphabetic names: A, B, and C to the 6-sided figures, and D, E, and F to the 10-sided forms.

Twenty-four mutants complying with Anderson's (1978) one-point perturbations were generated from each standard according to Attneave and Arnoult's (1956) Method 8. Specifically, one point on the perimeter of a polygon was selected, and its coordinates were moved five squares (1 cm) along either the x- or y-axis. The lines determining the perimeter of the new shape were redrawn as before, but now including the relocated point. The points to be mutated and the direction of the perturbation were such that the following constraints applied. For 6-sided figures, each vertex was perturbed four times, once in each possible direction (+5x, −5x, +5y, −5y). For 10-sided polygons, each vertex was perturbed twice, once along the x-axis and once along the y-axis, with the directions selected at random; the 4 remaining mutants were constructed by randomly selecting four points for perturbation. Thus, four of the vertices were perturbed three times, and the remaining six were perturbed twice. Randomly assigning the direction of perturbation was also such that each direction was sampled six times. All mutants thus generated from each standard were arbitrarily numbered from 1 to 24.

All polygons were cut out of flat black poster board and were centered and mounted on 8½" × 11" pieces of white construction paper. Figures were given a standard orientation on these sheets so that the longest axis was aligned vertically. Each sheet was labeled with the figure's corresponding letter and number designation (e.g., Standard C for one of the standard polygons and A14 for one of the mutants).

Three sets of the 144 mutants were constructed, each of which contained the entire set of mutants in a different across-family randomization order. Each order was such that the following conditions were met: (a) No more than 2 consecutive mutants were from the same stimulus family; (b) no more than 2 consecutive mutants involved perturbations of the same corresponding point (e.g., only 2 of any 3 mutants included perturbations at the uppermost point in the standard orientation); (c) 12 mutants from each family appeared in each half of the set of patterns.

Procedure. Subjects participated individually and were randomly selected to receive one of the three randomization orders. They were given the set of 144 patterns, copies of the six standard figures, a typewritten set of instructions, and sheets for recording similarity ratings. The instructions requested them to compare each mutant with its corresponding standard form and to record a similarity rating on a 9-point scale (where 1 was *extremely similar* and 9 was *barely similar*) before moving on to the next mutated pattern.

Results

The results of interest in this study are not so much the test statistics that are calculated from the rating data but rather how the ratings were used to select a set of rotation distractor stimuli from the set of 144 mutants. First, however, it seems wise to determine how reliable the ratings were. Do the ratings really vary according to changes in the stimuli, or are they independent of stimulus parameters and vary according to individual subjects' perceptions? This question was addressed by calculating a version of Cronbach's (1951) coefficient α that was adjusted to determine rater reliability rather than item reliability, its original appliction. Apparently, changes in the stimuli were responsible for the variance in the ratings subjects reported, $\alpha = .85$. This result is important because it indicates that for a large proportion of rating items, subjects assessed similarity by focusing on similar attributes of the stimulus. Had the coefficient indicated that subjects focused on different stimulus parameters, basing the rotation distractor set on these ratings would be less reliable in isolating similar and dissimilar mutants.

The hypothesis suggested by Anderson (1978) and Pylyshyn (1979) that using one-point perturbations across levels of complexity would produce a confounding of similarity with complexity was tested by considering the average rating (across subjects) provided to each mutant in the Mann-Whitney U test for large samples. This hypothesis was not supported in the present study, $z = -.162$, $p > .1$.

Finally, the mutants within each family were rank ordered on the basis of mean ratings from *most similar* (1) to *least similar* (24). Ties were broken by assigning the lower rank to the one with the smaller standard deviation of ratings. If both the mean ratings and standard deviations were equal, ties were broken randomly. In terms of these rankings, patterns ranked 1 through 8 were defined to be the set of *highly similar* distractors, and those ranked 17 through 24 defined the set of *least similar* distractors. Due to this choice, ratings for the least similar member of the highly similar set were significantly lower than ratings for the most similar member of the least similar patterns.

Discussion

Two important results emerged from this study, each of which suggests additional research. First, Cooper and Podgorny's (1976) argument that certain properties of their random polygons seemed more perceptually salient than others in determining the similarity between standards and mutants can be generalized to the one-point perturbations considered here. This result is interesting for three reasons. (a) The population of shapes in the present study was certainly more homogeneous than were Cooper and Podgorny's greatly perturbed stimuli. (b) Despite this homogeneity, similarity ratings varied appreciably because each subject selected similarity ratings from all nine categories. (c) Subjects' ratings to the perturbed stimuli were highly reliable. Apparently, changes in certain stimulus parameters incorporated in a mutant determine the perceived similarity between the pattern and its standard version. Identifying these parameters is an issue that has perplexed psychologists since Attneave and Arnoult (1956) introduced the notion of quantitative random polygons.

The second noteworthy result from this study was that similarity ratings were not confounded with stimulus complexity. One possible way this could come about is for subjects to locate the perturbed vertex in each mutant before estimating the similarity of the mutant to the standard. Complexity variations were clearly irrelevant to judging similarity in such a scheme.

It is important to note that the present study does not fully test Anderson's and Pylyshyn's hypothesis for the following reason. If subjects identified the perturbed point in each mutant, it seemed reasonable to assume, all else being equal, that more time was required to compare complex figures than simple patterns with their standards. The hypothesis could be more adequately tested if subjects saw mutated stimuli for only a short time and were required to estimate similarity immediately. Such a paradigm would not provide sufficient time for identifying the locus of perturbation, and, hence, subjects would need to rate similarity based on gestalts rather than by comparing the magnitude of identified perturbations.

Experiment 2

On each trial of the preparatory image rotation experiment, a subject studied a standard polygon in the standard orientation for 3 s. The instructed goal was to recognize the pattern and to form a clear mental image of it. After this, an orientation cue pointed to one of six positions, and the subject immediately began mentally rotating the image of the standard into congruence with the cue. When the rotation was completed, the subject pressed a button, and a test pattern was immediately presented at the same orientation as the cue. The task was to determine whether the test pattern was a standard or a mutant. The main dependent variable was the time subjects required to complete the preparatory rotation. The independent variables of interest were complexity of polygons and the similarity of mutants to their standards. Secondary dependent variables were the time and accuracy of the discrimination.

Using the data of Experiment 1, 48 mutants (8 from each family of shapes) were selected as highly similar and 48 as least similar to their respective standards. These two groups of stimuli were used to define two similarity conditions, high and low. A within-subjects design was used in which each subject was run in each similarity condition separately, but in counterbalanced orders. At no time during the experimental sessions was the subject told about the identity of a particular similarity condition or how it was defined.

Results from 4 pilot subjects indicated that the manipulation of similarity conditions worked entirely too well. Subjects simply were unable to discriminate mutants from standards in the condition of high similarity. Error rates were astoundingly high (higher, in fact, than expected on the basis of chance alone), reflecting the fact that subjects responded *same* (i.e., standard) to nearly all mutants. Therefore, a learning task was introduced into the experimental procedure. On trials of this task, subjects were required to discriminate standards from mutants in the standard orientation, and they continued until a criterion was satisfied. The mutants used in the learning task were not used in the experimental trials. This learning task differed considerably from Cooper's (1975) and Cooper and Podgorny's (1976) learning trials in that our subjects practiced discriminating polygons exclusively in the standard orientation. In the previous learning tasks they received extensive practice at all orientations.

Results from 2 additional pilot subjects indicated that the learning task did, in fact, facilitate discriminating polygons on the experimental trials. However, despite successful completion of the discrimination practice trials, error rates for these pilot subjects were still quite high. Aside from increasing the number of learning trials or requiring a higher criterion of accuracy, one possible way to decrease error rates is to make it advantageous for subjects to provide correct discrimination responses. The last refinement made on the experimental procedure was to increase subjects' motivation in the task by changing the payment scheme. The 2 pilot subjects were paid $4.00 per hour plus 3 cents for every correct response. Subsequent subjects were paid $30.00 for completing the entire experimental task and were penalized 10 cents for every incorrect discrimination response.[2] As will be argued later, changing the remuneration scale from a wage plus bonuses to a salary minus penalties profoundly increased subjects' motivation in the task. Thus, the *Method* section of this experiment concerns only the final version of the study.

Method

Subjects. Eight undergraduate volunteers from Harvard–Radcliffe college were paid $30.00 for completing the experimental task and were penalized 10 cents for each incorrect discrimination response. Subjects participated in experimental sessions that lasted between 2 and 2½ hr on 2 consecutive days (1 subject required three sessions). None had participated in Experiment 1, and all were naive regarding the hypotheses and purposes of the experiment.

Stimuli and apparatus. The standard polygons were the same standards used in Experiment 1. The mutated stimuli were those defined as high and low similarity, using the rankings from the first study. These 48 patterns consisted of 8 mutants from each family of shapes. The experimental distractor set for the condition of high similarity consisted of mutants ranked 1, 2, 4, 5, 7, and 8 in each family. Patterns ranked 3 and 6 in each family were used as distractor items for practice trials in this condition. Mutants ranked 17, 18, 20, 21, 23, and 24 in each family were included in the experimental distractor set for the condition of low similarity, and mutants ranked 19 and 22 were used in practice trials. Thus, the experimental distractor set for each condition of similarity consisted of 36 mutants (6 from each family), and the practice distractor sets consisted of 12 mutants (2 from each family). The method for selecting mutants for the practice distractor sets was chosen because it seemed to provide the most representative sampling of the similarity variations present in each set of mutants for each family. Figure 1 shows a sample of the mutants from each family used in each condition of similarity.

Subjects were asked mentally to rotate images of standard polygons to the following six orientations—0°, 60°, 120°, 180°, 240°, and 300°. The discrimination task was such that the test item was a standard on half of the trials and a mutant on the other half. Mutants were assigned to the orientations according to a 6 × 6 permuted Latin square (after Fisher & Yates, 1953). That is, each of the 6 mutants within each family was assigned to a different orientation, and mutants occurring at each orientation were such that all six ordinal levels of similarity were represented across families. The orientation cue was a black arrow positioned outside a black outline circle (11.5 cm in diameter). Subjects were requested mentally to rotate their image of the standard to that orientation.

Photographic slides were made of all stimuli and were presented to subjects on a rear-projection screen. Subjects were seated 71 cm in front

---

[2] This idea arose in a discussion with Miriam Schustack.

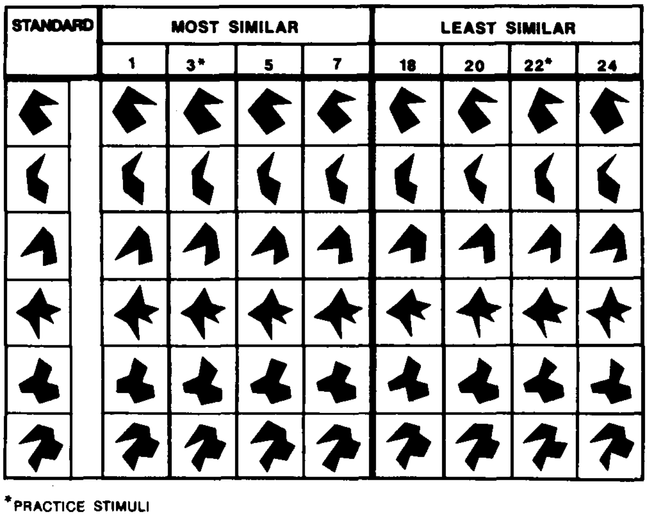

Figure 1. Sample of random polygons from each family of shapes used in the image rotation experiment.

of the screen, and the polygons subtended a visual angle both horizontally and vertically of about 7.15°; orientation cues subtended 9.19°. Within each condition of similarity the slides were randomized across families in such a way that the following constraints were met: (a) No more than two consecutive stimuli came from the same family; (b) neither the same orientation nor the same discrimination responses were required on more than two consecutive trials; (c) each half of the stimulus items in each condition was composed of equal numbers of same and different stimuli from each family in each orientation. A computer was programmed to control the presentation of the stimuli, operate the slide projectors, and record responses and response times.

Procedure. The subject was seated alone facing the rear-projection screen in front of a response box with three buttons arranged horizontally and labeled *same, advance,* and *different.* For half the subjects, the *same* response button was on the right, and for half it was on the left. The *advance* button was always in the middle. Subjects were informed of the payment scheme, and then they read a typewritten set of instructions. They were allowed approximately 30 min to study and learn the shapes of the six standard polygons, which were given to them on separate pieces of paper. Subjects were also allowed to study a black outline drawing of each standard. They were instructed to practice forming mental images of the standard and to compare their images to the original figures. All but 2 subjects considered the time allotted adequate for completing this task.

When subjects reported being able to form accurate images, they began the learning task, which consisted of two blocks of 72 trials. One half of the stimuli were standard polygons; the other half was composed of mutants of high and low similarity. The stimuli were randomized across families according to the constraints listed above with the exception that all polygons were presented in the standard orientation. On each trial, a polygon was projected onto the screen, and subjects were required to respond whether it was a standard (*same* button) or a mutant (*different* button). Subjects were informed that discrimination accuracy, not speed, was their primary objective, for they could not proceed to the rotation trials until they discriminated one set of 144 patterns with no more than 10 mistakes (93% accuracy criterion). Subjects not reaching the accuracy criterion again studied and practiced forming images of the standard shapes for approximately 10 min and then repeated the learning task with the same set of items in a different random order. This procedure was repeated until each subject reached the criterion.

That done, subjects moved on to the experimental trials. Half were randomly assigned to receive the condition of high similarity first, the other half receiving the low-similarity condition first. The similarity conditions defined the two blocks of trials in this experiment; each block consisted of 24 practice trials and 72 experimental trials.

The temporal sequence of events on each trial is illustrated in Figure 2.

Each trial began with a standard polygon presented in the standard orientation. That pattern remained on the screen for 3 s, during which time subjects were instructed both to recognize the standard and to form an accurate mental image of it. The standard was then replaced

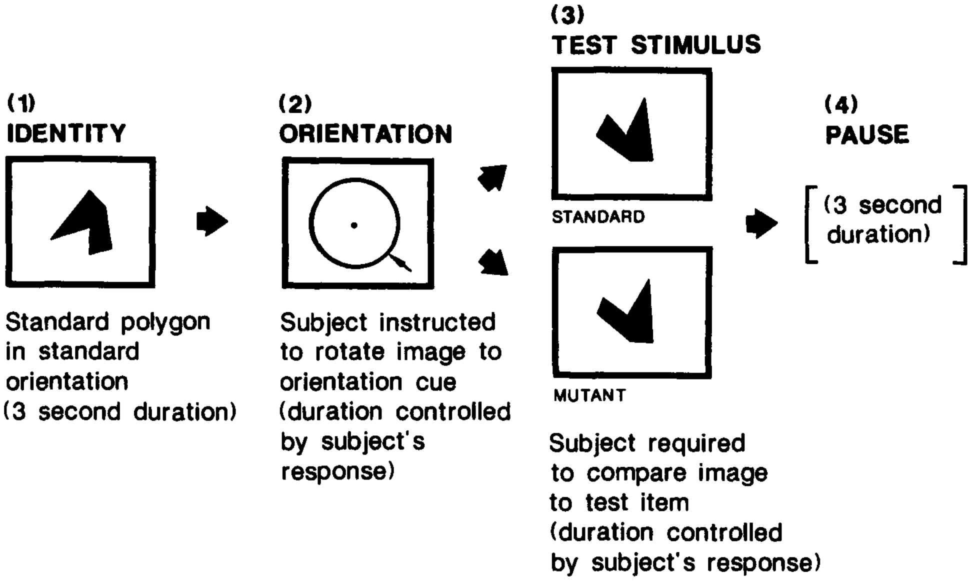

Figure 2. Sequence of events on a trial of the image rotation experiment.

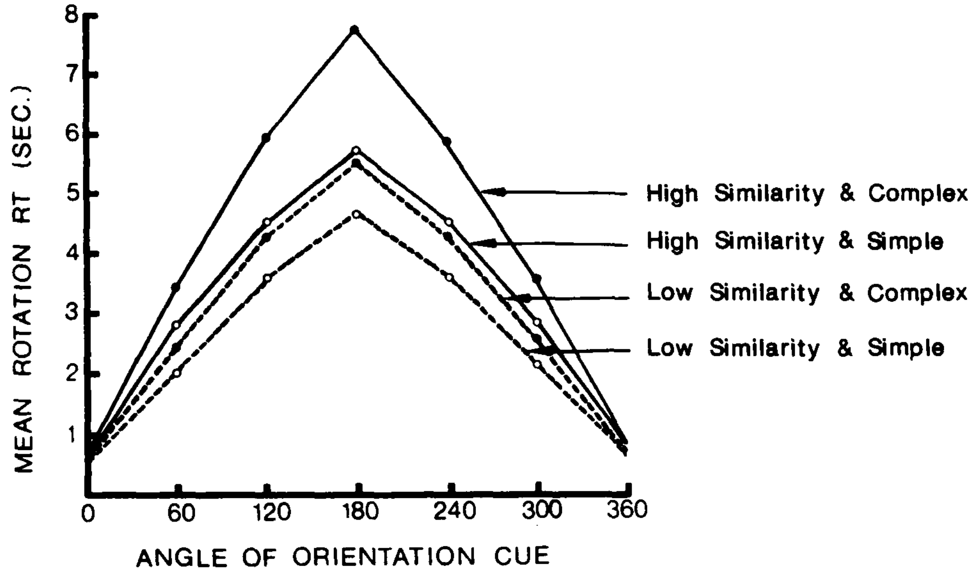

Figure 3. Mean times, in seconds, required mentally to rotate images of simple and complex polygons to six orientations in the two similarity conditions.

by an orientation cue. Subjects were instructed to begin mental rotation of their image as soon as the cue was presented and to press the *advance* button when the rotation was completed. The cue remained on the screen throughout the rotation. The lapsed time (i.e., the total duration of the cue) was recorded and was the dependent measure of primary interest. Subjects were instructed to rotate mentally their images in the direction of the shortest distance to the specified orientation. For example, if the cue pointed to 240°, subjects were told to rotate images 120° counterclockwise rather than 240° clockwise. Subjects were allowed to choose either direction for rotations of 180°. Of course, when the cue pointed to 0°, no rotation was required, and they were told to press the *advance* button as quickly as possible. Finally, they were instructed, "Rotate your image *as quickly as possible*" and "Rotate your image *completely* to the orientation designated by the arrow [orientation cue]."

As soon as subjects pressed the *advance* button, a test stimulus was presented in the orientation that had been specified by the cue. It was either a standard or a mutant. Subjects were instructed to make the *same/different* comparison as quickly and accurately as possible and to press the *same* or *different* button as appropriate. The test stimulus remained on the screen until the subject responded to it. This was followed by a 3-s intertrial interval. At the conclusion of the study, the experimenter tabulated the incorrect *same/different* judgments, subtracted the 10-cent fine for each error from the $30.00 salary, and then described the purpose of the study.

Results

Rotation times. Figure 3 shows the mean time to rotate images in the four conditions (Similarity × Complexity) presented in the experiment. Two observations are noteworthy. First, although the pattern of increased time with distance is nearly linear, some deviation exists. In particular, it could arise because either the 0° times or the 180° ones are too fast. Second, the effects of similarity and complexity appear as a "fan," that is, they modify the rate of rotation.

We discuss the results in greater detail in terms of a seven-way analysis of variance: Block Order (of similarity condition) × Subjects × Similarity × Complexity × Families (three families nested within each level of complexity) × Distance × Replications (first or second rotation of a pattern to a given orientation). The variable *replications* was replaced by *judgment* (*same* or *different*) in the analysis of discrimination data. We first report the analysis based on an additive model for the dependence of rotation time on the several independent variables. Considering the three most important independent variables tested in this experiment—similarity (s), complexity (c), and rotation distance (d)—the usual specification of the simplest additive model is

$$T = T_0 + f(s) + g(c) + h(d),$$

where T = rotation time to any nonzero orientation; T_0 = response time to identify the orientation cue at 0° (base rate); $f(s)$ = a function of similarity; $g(c)$ = a function of complexity; and $h(d)$ = a function of distance.

The analysis is reported in Table 1. For the three major variables—d, c, and s—all of the possible interactions were significant, as would be expected in a model such as Kosslyn's in which the image was mentally rotated by parts.

The simple fan pattern exhibited in Figure 3, however, strongly suggests that these interactions may be described far more simply as a single triple interaction, that is, as a multiplicative model: Similarity and complexity simply modify the rate of rotation but do not introduce any additive terms. Certainly, the Kosslyn model suggests such a rate effect rather than the introduction of additive stages. So, consider the following model:[3]

---

[3] This multiplicative model may seem inconsistent with the analysis of variance reported in Table 1 because it exhibits only a triple-order interaction. Actually, it is not in the following sense: Define

Table 1
Analysis of Variance of Rotation Times, T

Experimental variable	*df*	*F*	*p*<
Main effects			
Distance (D)	5, 30	31.48	.001
Complexity (C)	1, 6	25.56	.005
Similarity (S)	1, 6	45.30	.005
Second order			
C × D	5, 30	18.89	.0001
S × D	5, 30	13.66	.0001
C × S	1, 6	10.94	.02
Third order			
C × S × D	5, 30	3.55	.02
Other variables			
D × R	5, 30	3.25	.02
B × D × R	5, 30	2.21	.08
B × D × F	20, 120	2.54	.001
B × S × F	20, 120	2.54	.001
B × C × D × R	5, 20	2.15	.09
S × F × D × R	20, 120	1.59	.07

Note. B = block order; F = families within complexity; R = replication. Interactions not reported were all *ps* > .1.

$$T = T_0 + f(s)g(c)h(d).$$

Indeed, as is approximately true of Figure 3, if $h(d) = d$, then $f(s)g(c)$ is simply the rate of rotation.

This model may be transformed into the additive form required for the analysis of variance by subtracting times at 0°, T_0, from both sides of the equation and by taking the logarithm of the terms, yielding the following log-linear model:

$$\log(T - T_0) = \log f(s) + \log g(c) + \log h(d).$$

To determine the extent to which this model describes the data, each subject's mean time at 0° (across the four conditions of similarity and complexity) was subtracted from the times at all nonzero orientations. The logarithms of the differences were then subjected to the seven-way analysis of variance mentioned earlier, with the results shown in Table 2.

The results show that the log-linear model provides a good description of the primary experimental variables, except for the significant interaction of similarity and distance. The main effects confirm that subjects required more time to perform the larger rotations in the condition of high similarity than in the condition of low similarity. Mental rotations were slower in the high-similarity condition; presumably this was in order to maintain more complete structural information in their images for the difficult discrimination task. The significant Similarity × Distance interaction suggests that the following modification of the log-linear model more accurately describes the data:

$$\log(T - T_0) = \log f(s) + \log g(c) + \log h(d) + \log j(d, s).$$

Perhaps the simplest model that introduces this $s \times d$ interaction is

$$T = T_0 + f(s)g(c)d^{1+m(s)}.$$

The exponent, with $m(s) < 0$, introduces the slight nonlinearity exhibited in Figure 3. The remaining three interactions, which are barely significant, seem largely of methodological interest.

It should be mentioned that rotation times were *three* times slower than those reported by Schwartz (1979) and nearly *five* times slower than those reported by Cooper (1975) and Cooper and Podgorny (1976). Such large differences suggest either that our subjects may have mentally rotated complete images, thereby decreasing rotation rates in order to facilitate discriminations, or that practice in the earlier studies may have allowed subjects to achieve more fully integrated representations of the shapes. We cannot tell which from the present data.

Discrimination data. Discrimination data were considered in a seven-way analysis of variance of the same factors, except replication was replaced by judgment (*same/different*). *Same/different* response times provide an indication of how well subjects followed the instructions for the rotation task. Recall that subjects were specifically instructed, "Rotate your images *completely* to the orientation designated by the arrow." If subjects indeed followed this instruction, the orientation at which test forms appeared should not have affected discrimination times. The analysis, however, did not support this hypothesis: Discrimination times were sharply affected by orientation, $F(5, 30) = 7.93$, $p = .0001$. Cell means indicated that times at 0° were primarily responsible for this effect; average response times at this orientation were 400 ms faster than at all other positions. This result is evident in Figure 4, which shows the mean time to discriminate patterns in the four experimental conditions.

To determine whether the 0° orientation was indeed responsi-

Table 2
Analysis of Variance of Rotation Rates, log(T − T₀)

Experimental variables	*df*	*F*	*p*<
Main effects			
Distance (D)	4, 24	24.31	.0001
Complexity (C)	1, 6	29.77	.002
Similarity (S)	1, 6	20.43	.004
Second order			
S × D	4, 24	2.87	.05
Other variables			
B × D	4, 24	2.69	.06
C × D × R	4, 24	2.81	.05
B × C × D × R	4, 24	3.25	.03

Note. B = block order; R = replications. Interactions not reported were all *ps* > .1.

$$f^*(s) = f(s) - 1,\ g^*(c) = g(c) - 1,$$

and

$$h^*(d) = h(d) - 1;$$

then the model is simply

$$\begin{aligned} T &= T_0 + [1 + f^*(s)][1 + g^*(c)][1 + h^*(d)] \\ &= T_0 + 1 + f^*(s) + g^*(c) + h^*(d) \\ &\quad + f^*(s)g^*(c) + f^*(s)h^*(d) + g^*(c)h^*(d) + f^*(s)g^*(c)h^*(d). \end{aligned}$$

This illustrates how very careful one must be in interpreting the meanings of significant interactions in an analysis of variance for a model of the underlying process.

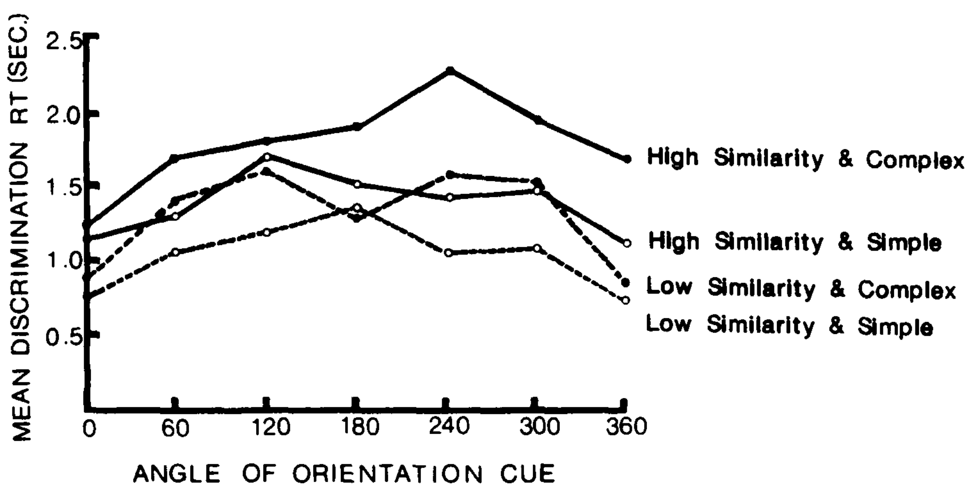

Figure 4. Mean times, in seconds, required to make *same/different* discriminations to simple and complex polygons at six orientations in the two similarity conditions.

ble for this effect, the discrimination time data were considered in an additional analysis of variance that excluded the 0° data. Because the data in Figure 4 do not suggest that the multiplicative model is better than the additive one, we did the analysis in terms of T rather than $\log(T - T_0)$. The result was that effects of distance now were only marginally significant, $F(4, 24) = 2.38$, $.08 < p < .09$. Thus, it is safe to conclude that orientations did not strongly affect response times. Considering that subjects received extensive training in discriminating patterns at 0°, it is possible that including this point biased the analysis. Because the second analysis was nearly identical to the original (except for the distance effect), all results henceforth are reported from the second analysis of variance that excluded times at 0°.

Unlike rotation times, similarity conditions only moderately affected discrimination times, with highly similar test forms requiring more time to discriminate, $F(1, 6) = 5.76$, $.05 < p < .06$; complex patterns required more time than simple forms, $F(1, 6) = 13.27$, $p < .02$. These two results are readily interpreted within the context of the spatial image model described by Kosslyn (1980). According to that model, images are generated and maintained as a collection of distinct parts, the number depending upon the complexity of the image. That is why complexity is expected to affect rotation rate. And because images fade and become degraded over time, the probability of losing structural information from the image is greater for complex images. Therefore, it is not surprising that complex images required more time to be discriminated and that highly similar patterns required more time than less similar ones. This interpretation is further supported by the marginally significant result that complex patterns required more time to discriminate at more disparate orientations from upright, $F(4, 24) = 2.39$, $.07 < p < .08$. However, if complex patterns indeed become more degraded over time than simple ones, then they should require more time both for highly similar test items in general and for highly similar test items at more exaggerated orientations. Neither prediction was supported by the data—$F(1, 6) = 0.59$, $p > .4$, and $F(4, 24) = .61$, $p > .6$, respectively. It appears that subjects adopted a rotation strategy that allowed them not only mentally to rotate detailed images but also to maintain complete structural information for the discrimination task. This accords with the slow rotation rates noted earlier.

Error rates. Table 3 shows the percentage of errors (to both *same* and *different* items, as well as to all items) in each of the four experimental conditions, two levels of complexity by two levels of similarity. All 8 subjects exhibited similar error distributions.

The only striking result was that the pattern of errors was as expected both in terms of the hypotheses tested in this experiment and by appealing to Kosslyn's hypothesis of image decay. It is interesting that error data in this study provided the only *strong* support for the degraded-images interpretation. Note that all previous support of this hypothesis was marginally sig-

Table 3
Percentage of Errors in Various Experimental Conditions

	All items[a]		*Different* items[b]		*Same* items[b]	
Condition	Simple	Complex	Simple	Complex	Simple	Complex
Low similarity	7.3	9.4	4.2	6.9	10.4	11.8
High similarity	22.6	24.0	22.2	29.2	22.9	18.8

[a] $N = 36$ per cell, per subject. [b] $N = 18$ per cell, per subject.

nificant except for Schwartz's data, which were highly significant. As mentioned earlier, the long rotation times found here suggest that subjects maintained more complete images than did Schwartz's subjects. It is not surprising, then, that subjects performed the discrimination task with only slightly degraded images. Error rates in the present study were quite low, considering the difficulty of the discriminations. (Compare Schwartz's overall error rate of 33.6% with the rates obtained here.)

General Discussion and Conclusions

The data presented suggest three conclusions. First, stimulus complexity does affect the rate at which images are mentally rotated, but only if subjects transform complete structural representations. Second, the level of similarity affects the strategies that subjects follow when mentally rotating images. Apparently, they mentally rotated incomplete images when the similarity was low, but were compelled mentally to rotate complete ones when the similarity was great. Not only were rotation times generally longer in the latter case, but complexity effects were detected only in this condition. Third, subjects appeared to be highly motivated in this experiment and provided rotation data that were very clean and systematic and error rates that were reasonably low.

These results are important for three reasons. First, they fully support, for the first time, predictions regarding stimulus complexity on mental rotation rate. Second, they make clear that great care must be taken in selecting stimulus material. The predicted effects of complexity are based on models in which processing mechanisms operate on *complete images*, and the complexity hypothesis was supported in this study only when we were reasonably certain that subjects mentally rotated detailed images. The four earlier experiments considering this issue may have failed to detect complexity effects because the stimuli were such that subjects did not need mentally to rotate complete images to make the *same/different* comparisons. Third, although no systematic study was carried out, we suspect that subjects' level of motivation greatly influenced the results. Considering the fact the subjects were (relative to previous studies) unpracticed and mentally rotated each pattern only twice to each orientation, the results were surprisingly clean. One possible explanation is that it was in the subjects' best interests to rotate complete images the full angle. Without adopting such a mental rotation strategy, subjects might not have been successful in the discrimination task, and poor performance would have decreased their monetary reward. The payment scheme in which subjects were penalized for incorrect responses seemed to motivate them to a greater extent than did the original payment plan that provided bonuses for correct answers. This is consistent with the general finding in the decision theory literature that subjects tend to be asymmetric in their evaluations of gains and losses (Kahneman & Tversky, 1979).

References

Anderson, J. R. (1978). Arguments concerning representations for mental imagery. *Psychological Review, 85,* 249–277.

Attneave, F. (1957). Physical determinants of the judged complexity of shapes. *Journal of Experimental Psychology, 53,* 221–227.

Attneave, F., & Arnoult, M. D. (1956). The quantitative study of shape and pattern perception. *Psychological Bulletin, 53,* 452–471.

Cooper, L. A. (1975). Mental rotation of random two-dimensional shapes. *Cognitive Psychology, 7,* 20–43.

Cooper, L. A., & Podgorny, P. (1976). Mental transformations and visual comparison processes: Effects of complexity and similarity. *Journal of Experimental Psychology: Human Perception and Performance, 2,* 503–514.

Cooper, L. A., & Shepard, R. N. (1973a). Chronometric studies of the rotation of mental images. In W. G. Chase (Ed.), *Visual information processing* (pp. 75–176). New York: Academic Press.

Cooper, L. A., & Shepard, R. N. (1973b). The time required to prepare for a rotated stimulus. *Memory & Cognition, 1,* 246–250.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16,* 297–334.

Fisher, R. A., & Yates, F. (1953). *Statistical tables for biological, agricultural and medical research* (4th ed.). London: Oliver & Boyd.

Hochberg, J. H., & Gellman, L. (1977). The effect of landmark features on "mental rotation" times. *Memory & Cognition, 5,* 23–26.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47,* 263–291.

Kosslyn, S. M. (1980). *Image and mind.* Cambridge, MA: Harvard University Press.

Kosslyn, S. M. (1981). The medium and the message in mental imagery: A theory. *Psychological Review, 88,* 46–66.

Kosslyn, S. M., Pinker, S., Smith, G. E., & Schwartz, S. P. (1979). On the demystification of mental imagery. *The Behavioral and Brain Sciences, 2,* 535–581.

Kosslyn, S. M., & Pomerantz, J. R. (1977). Imagery, propositions, and the form of internal representations. *Cognitive Psychology, 9,* 52–76.

Minsky, M., & Papert, S. (1972). *Artificial intelligence progress report.* Cambridge, MA: MIT. (Project MAC, Artificial Intelligence Laboratory Memo 252)

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

Palmer, S. E. (1975). Visual perception and world knowledge: Notes on a model of sensory-cognitive interaction. In D. A. Norman & D. E. Rumelhart (Eds.), *Explorations in cognition* (pp. 179–237). San Francisco: Freeman.

Pylyshyn, Z. W. (1973). What the mind's eye tells the mind's brain: A critique of mental imagery. *Psychological Bulletin, 80,* 1–24.

Pylyshyn, Z. W. (1979). The rate of "mental rotation" of images: A test of a holistic analogue hypothesis. *Memory & Cognition, 7,* 19–28.

Riseman, D., & Hansen, E. (1978). *Computer vision systems.* New York: Academic Press.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science, 171,* 701–703.

Shepard, R. N., & Podgorny, P. (1978). Cognitive processes that resemble perceptual processes. In W. K. Estes (Ed.), *Handbook of learning and cognitive processes: Human information processing* (Vol. 5, pp. 279–307). Hillsdale, NJ: Erlbaum.

Schwartz, S. P. (1979). Studies of mental image rotation: Implications for a computer simulation of visual imagery (Doctoral dissertation, The Johns Hopkins University). *Dissertation Abstracts International, 40,* 2413B. (University Microfilms No. 79-24, 639)

Yuille, J. C., & Steiger, J. H. (1982). Nonholistic processing in mental rotation: Some suggestive evidence. *Perception & Psychophysics, 31,* 201–209.

Received June 25, 1986
Revision received December 16, 1986
Accepted January 8, 1987 ■