COGNITIVE PSYCHOLOGY 26, 1–32 (1994)

The Representation of Location in Visual Images

KYLE R. CAVE

Vanderbilt University

STEVEN PINKER

Massachusetts Institute of Technology

LIANA GIORGI

Cambridge University

CATHERINE E. THOMAS

Dartmouth College

LAURIE M. HELLER

Massachusetts Institute of Technology

JEREMY M. WOLFE

Harvard Medical School and Brigham and Women's Hospital

AND

HELEN LIN

Massachusetts Institute of Technology

By definition, visual image representations are organized around spatial properties. However, we know very little about how these representations use information about location, one of the most important spatial properties. Three exper-

Thanks to David Small, Joe Garrison, and Shelene Chang for their help in implementing and executing the experiments. Thanks also to Martha Farah, Steve Kosslyn, Dave Irwin, Geoff Loftus, and Mike Tarr for many useful suggestions and comments. This work was supported by a National Science Foundation Graduate Fellowship and by a Whitaker Health Sciences Graduate Fellowship to Kyle Cave, by NSF Grant 85-18774 and a grant from the Sloan Foundation to the M.I.T. Center for Cognitive Science, and by National Institute of Mental Health Training Grant MH-14268, awarded to the Center for Human Information Processing, University of California, San Diego. Correspondence and reprint requests should be addressed to Kyle Cave, Vanderbilt University, Department of Psychology, 301 Wilson Hall, Nashville, TN 37240.

0010-0285/94 $9.00

iments explored how location information is incorporated into image representations. All of these experiments used a mental rotation task in which the location of the stimulus varied from trial to trial. If images are location-specific, these changes should affect the way images are used. The effects from image representations were separated from those of general spatial attention mechanisms by comparing performance with and without advance knowledge of the stimulus shape. With shape information, subjects could use an image as a template, and they recognized the stimulus more quickly when it was at the same location as the image. Experiment 1 demonstrated that subjects were able to use visual image representations effectively without knowing where the stimulus would appear, but left open the possibility that image location must be adjusted before use. In Experiment 2, distance between the stimulus location and the image location was varied systematically, and response time increased with distance. Therefore image representations appear to be location-specific, though the represented location can be adjusted easily. In Experiment 3, a saccade was introduced between the image cue and the test stimulus, in order to test whether subjects responded more quickly when the test stimulus appeared at the same retinotopic location or same spatiotopic location as the cue. The results suggest that location is coded retinotopically in image representations. This finding has implications not only for visual imagery but also for visual processing in general, because it suggests that there is no spatiotopic transform in the early stages of visual processing.

The input provided by the retina is organized spatially: shape information is intertwined with information about spatial properties such as location, size, and orientation. The visual system must identify objects by comparing patterns within the input with patterns stored in memory, but variations in location, orientation, and size are generally irrelevant to an object's identity. If the final product of visual processing is to be used in higher-level reasoning and problem solving, then information about the identity of the represented objects and their spatial properties must be factored apart.

However, the findings from a number of visual imagery experiments indicate that some visual processing tasks rely on representations in which information about spatial properties has not been factored apart from shape information. Spatial properties play an important role in the organization of these image representations; this is one of the reasons that such representations are called "images." Many experiments have documented the spatial organization of images using the following logic: if the time to perform a particular shape discrimination task depends on the stimulus orientation (or size), then the shape representations used in this discrimination task must vary in some important way when the stimulus orientation (or size) varies. In other words, the representation of shape is intertwined with the representation of orientation or size: This is not an abstract representation in which shape, size, orientation and location are all factored apart and represented as separate propositions (Pinker, 1984).

Well-known examples of experiments measuring the importance of a

spatial property come from mental rotation studies by Shepard and his colleagues (Shepard & Cooper, 1982). In these experiments, subjects were asked to determine whether a visual stimulus matched another stimulus that was either presented simultaneously or remembered by the subject. On different trials, the stimuli appeared at different orientations, and over the course of the experiment the difference between the orientations of the two shapes to be compared was systematically varied. The time necessary to compare the shapes increased as the orientation differences increased. Thus, even though orientation was irrelevant to the shape matching task, it exerted a strong effect on the response time. Other experiments have used similar methods to demonstrate that the time to compare two shapes can vary with the difference between their sizes (Bundesen & Larsen, 1975; Bundesen, Larsen, & Farrell, 1981; Cave & Kosslyn, 1989; Kubovy & Podgorny, 1981; Larsen, 1985; Larsen & Bundesen, 1978; Sekuler & Nash, 1972). Clearly, the representations used in these tasks must be orientation- and size-specific: the same shape at different orientations and sizes must be represented differently.

Cooper and Shepard (1973) found that although extra time was necessary to process stimuli at nonstandard orientations, subjects could save most of this extra time if they knew the shape and the orientation of the upcoming stimulus before it appeared. Apparently subjects prepared for a stimulus by creating an image representation of the shape at the cued orientation and matched it against the stimulus when it appeared. In contrast, if they knew only the shape or only the orientation, response times once again rose sharply with orientation difference. The fact that subjects cannot prepare effectively when they know only the shape or only the orientation suggests that orientation is represented integrally with shape. For the representations used in the mirror-reversal task, it is apparently not possible to represent a particular orientation without representing a specific shape at that orientation or vice-versa (but see Hinton & Parsons, 1981).[1]

Given all these results indicating that visual image representations record a particular orientation and size, we can ask whether they also record a particular location. It is possible that they do not; that they are at a level of representation that is location-independent but orientation-specific. This type of representation might be useful, given that as we move or as objects in the environment move, the locations of the objects relative to

[1] Although Cooper and Shepard's subjects discriminated handedness, not shape per se, Tarr and Pinker (1989, 1990) showed that the same effects occur when subjects make pure shape identifications, ignoring handedness (as long as stimuli were not overlearned or trivially discriminable). Thus we will assume that mental rotation effects reveal properties of the representations underlying shape processing.

us can vary over a wide range, and we must be able to recognize them no matter where they appear. However, as we move we generally maintain a constant orientation relative to the environment, and most of the other objects we encounter do the same. Therefore we generally view objects at a standard orientation, and rarely find it necessary to recognize them at other orientations. Presumably this is why an object's location has little impact on recognition (barring differences in surrounding and acuity), whereas the same is not true for orientation: some objects can be very difficult to recognize when they are upside-down (Rock, 1983). The visual system may factor out location differences early in processing, either by normalizing all object representations so that they represent a standard location or by transforming the input in a way that removes location information entirely. In fact, neuroanatomical and neurophysiological studies indicate that one region of visual cortex is dedicated to processing location and a separate region is dedicated to processing shape (Ungerleider & Mishkin, 1982).

Because the representations used in mental rotation must include shape information, they might reside in the region that specializes in shape, and thus location information might be factored out of them. However, other evidence from imagery experiments suggests that location is important in image representations, and thus is probably not factored out. Farah (1985) asked subjects to image a large shape while watching for a much smaller stimulus to appear somewhere on a display. She found that subjects were better at detecting this stimulus if it appeared within the area of the imaged shape. If the image location affects the perception of a stimulus, then location must be part of the image representation. However, Farah instructed her subjects to use imagery, and they may have focused visual attention on the area covered by the shape, thus enhancing their responses to stimuli that fell within this area. Thus her effect does not prove that location is integrally represented with the information regarding shape.

Another possible demonstration of the use of location in image representations comes from experiments demonstrating that the time to scan from one location to another increases with the distance scanned (Kosslyn, 1973; Kosslyn, Ball, & Reiser, 1978; Finke & Pinker, 1982, 1983; Pinker, Choate, & Finke, 1984). However, these tasks specifically *required* the use of location information, because the location of each dot was crucial in determining the correct response: in Kosslyn's tasks, subjects had to visualize a dot moving to the correct location; in Pinker and Finke's, they had to indicate whether an arrow pointed to any dot. Thus, even if subjects had available a representation in which shape was coded independently of location, they would not have been able to use it. The image representation of a visual object includes information about the

relative locations of its different parts, but the location of the object as a whole is not necessarily represented. Subjects in these experiments might have represented the entire configuration of dots or landmarks as a single object, so that the relative locations of each would be preserved.

We need an experiment in which a shape discrimination is required, but in which the stimulus location is not relevant to the response. Experiment 1 will test the importance of location in the task used by Cooper and Shepard. Experiment 2 will use a different approach that tests more directly how location information is encoded in the representations used in mental rotation.

EXPERIMENT 1

Cooper and Shepard concluded that subjects can only generate the necessary mental representation to compare with the stimulus if they know all the spatial properties of the stimulus. Their logic underlies Experiment 1. In Cooper and Shepard's experiments, the stimulus always appeared in the center of the display, so subjects always knew its location before it appeared. In this experiment, subjects will not know the stimulus location before it appears. If they are nonetheless able to prepare the appropriate representation, Cooper and Shepard's logic would imply that location is not encoded in these representations.

Another alternative is that location information is included, but that unlike orientation and size information, it can be easily adjusted. In this case, subjects might prepare a representation without knowing the location and then adjust the represented location after the stimulus appears. This possibility will be addressed in Experiment 2.

Method

Subjects. Seventeen volunteer subjects from the M.I.T. Department of Brain and Cognitive Sciences subject pool were tested and were paid for their services. Most were M.I.T. undergraduates, and their vision was normal or corrected to normal. One subject was rejected for failing to follow the instructions.

Apparatus. The experiment was controlled by an IBM PC/XT computer. Stimuli were displayed using EGA graphics on an NEC Multisync monitor. Subjects' responses were recorded with two microswitches, one for each hand, and a foot pedal.

Stimuli. Each test stimulus consisted of a single character, either an uppercase J (with a serif on top), an uppercase R, or the numeral 4 (in its open version, with right-angle segments). Examples can be found in Fig. 1. The stimuli appeared in eight different orientations, 0, 45, 90, 135, 180, 225, 270, and 315°, and in both normal and mirror-reversed forms. Each character was 1.4 cm in height (1.6° of visual angle).

Each test stimulus was preceded by a cue appearing at the center of the screen. The cue was one of two types, as depicted in Fig. 1. One type consisted of one of the three characters at one of the eight possible orientations. This type of cue informed the subject of both the shape and the orientation of the upcoming test stimulus. These cues were never mirror-reversed and gave the subject no information about whether or not the test stimulus would be mirror-reversed. The second type of cue consisted of an arrow at one of the eight

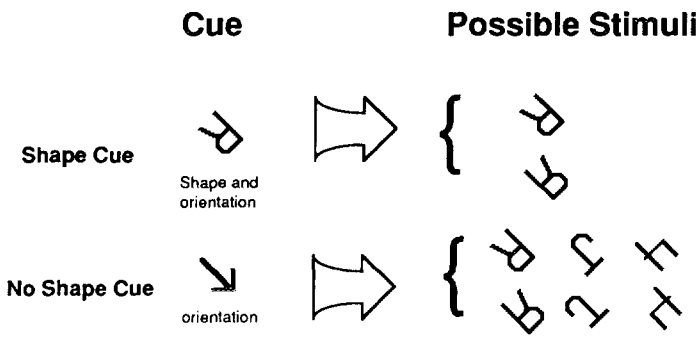

FIG. 1. Examples of the two types of cue used in Experiment 1, and the possible stimuli that could appear after them.

orientations. It revealed the orientation of the upcoming stimulus, but not its shape. The information provided by either cue type was always correct: the orientation and shape never deviated from the values cued.

Procedure. The subject was seated in front of the CRT in a dimly lit room. A chin rest was used to maintain a constant viewing distance of 50 cm. The subject was instructed to keep one hand on each of the response keys.

Each trial began with the presentation of a cue at the center of the screen. The subject was asked to study the cue and to press the foot pedal when ready to proceed. When the subject pressed the pedal, the cue immediately disappeared and the test stimulus appeared at one of four locations, which defined the corners of an imaginary square. Each of the four locations was 5.6 cm (6.4°) from the center of the screen. The test stimulus was one of the three characters, either normal or mirror-reversed, at one of the eight possible orientations. The computer synchronized the onset of the test stimulus with the beginning of a video cycle, waited 120 ms, and removed it at the beginning of the next video cycle. This display time was too short to allow eye movements to the stimulus before it disappeared. The stimulus was a black figure on a white background, and when it disappeared the display was completely white, so that no residual image of the stimulus would persist on the display after it had been removed.

The subject's task was to press the key under the dominant hand if the test stimulus was normal and to press the other key if it was mirror-reversed. If the subject gave an incorrect response, the screen briefly flashed red and a buzzer sounded. The computer recorded the time interval between the presentation of the test stimulus and the subject's response. It also made a record of the error trials and repeated each of them once at the end of the session.

Each possible combination of the two cue conditions, three shapes, eight orientations, four locations, and two response conditions (normal and mirror-reversed) was used, giving a total of 384 different types of trials. There were four instances of each type for each subject, not including those that were repeated because of errors. For each subject, the 1536 trials were arranged in a different random order, with all the different trial types intermixed. Each subject required three different testing sessions to complete all the trials, with each session lasting between 40 and 60 min.

Results

The response time data were submitted to an analysis of variance (ANOVA) with cue type (shape or no-shape), orientation, handedness

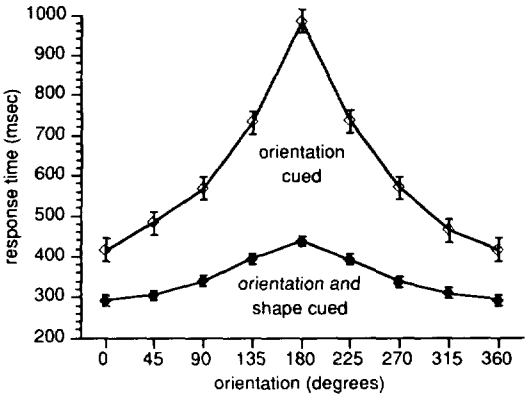

FIG. 2. Response times from the Shape Cue and No-Shape Cue conditions of Experiment 1.[2] (Points for 0° are replotted at 360°.)

(normal or mirror-reversed), location, and shape (R, J, or 4) as factors. Subjects made incorrect responses on 4% of the trials (including those repeated because of earlier errors), and all of these trials were excluded from the analysis. The response times from both cue conditions are presented in Fig. 2 as a function of orientation. In general, subjects responded much more quickly when the shape was cued, $F(1,15) = 31.8, p < .001$. If subjects are able to prepare images only when they know the shape beforehand, then moving the orientation further from upright should not increase response times in the Shape Cue condition in the way that it does in the No-Shape Cue condition. A contrast analysis revealed that the linear increase with orientation was indeed much stronger in the No-Shape Cue condition than in the Shape Cue condition, $F(1,105) = 135.9, p < .001$. The difference between the two conditions is clear in Fig. 2 and is very similar to that found by Cooper and Shepard. Thus subjects appear to be able to generate and use images in this task, even without knowing the correct location in advance.

Because subjects were apparently able to prepare an image in the Shape Cue condition and not in the No-Shape Cue condition, we used two other ANOVAs to examine the results of the two conditions separately. Not surprisingly, a contrast in the no-shape cue analysis established a strong linear increase in response time with the deviation of the orientation from

[2] Error bars were calculated separately for the Shape Cue and No-Shape Cue conditions, using a procedure suggested by Geoff Loftus (personal communication). In each case, an ANOVA was computed for just that condition, and the length of the error bar was calculated as $\sqrt{MSI/n}$, where MSI is the mean square for the subject × orientation interaction (error term for the orientation factor), and n is the number of values contributing to each mean. These error bars are appropriate for comparisons within each condition, but not across the two conditions. A similar procedure is used for error bars in later graphs.

upright, $F(1,105) = 285.9$, $p < .001$. The linear trend accounted for 92% of the variance among the means. The residual from linearity was also significant, $F(6,105) = 4.0$, $p < .005$, as might be expected given Cooper and Shepard's data.

The overall analysis showed that the linear increase in response time with orientation difference was much smaller when the shape was cued than when it was not. Nevertheless, a contrast in the shape cue ANOVA showed that the increase was still significant, $F(1,105) = 113.8$, $p < .001$. This rise is apparent at the bottom of Fig. 2. The linear trend accounted for 96% of the variance among the means, and the residual from linearity was not significant, $F < 1$. As with the no-shape cues, shape cue responses were slower to mirror-reversed stimuli, $F(1,15) = 59.3$, $p < .001$.

The main analysis produced other significant effects that were not directly relevant to the questions at hand and hence are not a priori tests; they will not be considered here. Analysis of the error rates revealed that none of the critical results described above could be attributed to speed/accuracy trade-offs.

Discussion

The results from this experiment are highly similar to those from Cooper and Shepard's experiment, which implies that location knowledge is not necessary for preparing an image representation. Perhaps this conclusion should not be surprising, because this property makes these representations much more useful in a world in which objects can often appear at any visual field location.

The large orientation-effect differences between the Shape Cue and No-Shape Cue conditions makes it clear that the stimuli are processed very differently in the two conditions. Nevertheless, the response time increase with orientation difference in the Shape Cue condition is also significant, even though it is much smaller than in the No-Shape Cue condition. Subjects are likely to rotate stimuli occasionally in this condition even if they know the shape, either because they ignore or forget the cue or because they want to perform an additional test to be sure of their response. An inspection of Shepard and Cooper's (1982) graphs (their figures 4.5, 4.11, and 4.13) reveals that advance information about shape and orientation did not completely prevent response time from rising somewhat for orientations near 180°.

These results show that location information is treated differently from orientation or size information in visual images, because subjects prepare an image without knowing the location. However, this experiment does not give a definitive answer about the location-specificity in the mental representations that are used in this task. Location differences could be handled in two different ways. In the Shape Cue condition, the stimulus

might be normalized or otherwise recoded into a location-independent form and then compared against a location-independent image representation. On the other hand, the stimulus might be represented as occupying a particular location, probably at the center of the display. Once the stimulus appears, the location of the image could then be adjusted so that it matches the stimulus location and it could be compared with the stimulus. We cannot compare our results and those of Cooper and Shepard with enough precision to measure whether our subjects required extra time to adjust the represented location. Experiment 2 will clarify location's role in these representations.

EXPERIMENT 2

Experiment 2 is designed to determine whether the representations used in mental rotation include location information that must be adjusted to match a stimulus location. To measure location adjustment, we induced subjects to create an appropriate image by telling them the shape and orientation of the upcoming stimulus and by leading them to believe that it would appear at a particular location. We could then measure the "movement" of the image to another location. As in Experiment 1, all stimuli appeared the same distance from the fixation cross so that there were no substantial acuity differences. Unfortunately, by presenting stimuli at different locations, we ran the risk of introducing eye movements, which would add to the response times and obscure the important effects. Eye movements were not a problem in Experiment 1, because the cue always appeared at the center of the screen. The subjects did not know where the stimulus would appear, and it disappeared too quickly for them to make a saccade. In this experiment, subjects had to know in advance where the stimulus was likely to appear; therefore we had to prevent them from saccading to that location.

Cooper and Shepard showed that the amount of time necessary to adjust orientation varied with the amount of adjustment. The current experiment tested for a similar pattern in location adjustments by varying the amount of adjustment necessary in different trials. Looking for processing differences at different locations is complicated by the allocation of visual attention. Subjects respond more quickly to stimuli that appear at an expected location (Posner, Nissen, & Ogden, 1978; Posner, Snyder, & Davidson, 1980), and, at least in some circumstances, response time tends to increase with larger distances between the expected location and the actual location (Downing & Pinker, 1985; Rizzolatti, Riggio, Dascola, & Umiltá, 1987). To ensure that any distance effect is due to the image adjustment and not to general attentional factors, we must compare response times with and without image representations. Because subjects cannot create an image without knowing the shape (Cooper & Shepard,

1973), Experiment 2 had two different conditions, with the same sort of shape cues and no-shape cues used in Experiment 1. For clarity, the shape cue condition will now be referred to as the Image condition, and the no-shape cue condition will be called the No-Image condition.

Method

Subjects. This experiment required subjects to fixate at one location while attending to a stimulus at a distant location. Additionally, because a pupil-tracker was monitoring eye position, subjects were required to keep their eyes open from the beginning to the end of each trial. When they began the first session, most subjects were not able to successfully complete many trials. Most of them, however, improved with practice, and a total of 15 completed the entire experiment. A similar number of subjects began the experiment but quit after practice, either on their own or at the suggestion of the experimenter. All subjects were from the M.I.T. Department of Brain and Cognitive Sciences subject pool, and all were paid for their time, whether or not they completed the experiment. Most were M.I.T. undergraduates, and their vision was normal or corrected to normal. These subjects did not participate in the previous experiment.

Apparatus. The computer, video display, and response keys used in this experiment were the same types as those used in Experiment 1. An ISCAN Model RK-416 pupil-tracker monitored eye movements. The eye-tracker received an image of the subject's left eye from an RCA TC2000 video camera with a close-focus lens and an infrared filter. A table light fitted with an infrared filter illuminated the subject's eye, and a chin rest and forehead restraint held the subject's head in place.

The pupil-tracker received a video image from the camera 60 times each second. In each image, the pupil tracker hardware located the pupil by identifying a large dark region. It then calculated the center of this region, and transmitted the horizontal and vertical coordinates of the center's location within the video image to the computer. When the subject steadily fixated on a single location, there was a small amount of variation in the coordinates from the pupil tracker, due mainly to small changes in the video image from cycle to cycle. The program controlling the experiment recorded the eye position once at the beginning of each trial and then monitored the eye position continuously until the response. If the distance between the original recorded position and the current eye position ever exceeded a threshold, the trial was aborted. Before subjects were tested, we determined the lowest level at which we could set the threshold without producing an inordinate number of false alarms. The threshold value we used generally allowed for the detection of eye movements of 2.5° of visual angle or greater.

Stimuli. As in Experiment 1, each stimulus consisted of a single character, the letter J, the letter R, or the numeral 4. As before, the characters could be normal or mirror-reversed. This experiment required that a number of different factors be varied and thus required a large number of different types of trials. Therefore, only four orientations were used, 45, 135, 225, and 315°.

As in Experiment 1, each test stimulus was preceded by either a shape cue or a no-shape cue. Each shape cue consisted of a character at a particular orientation. Each no-shape cue was a rectangle the same size as the characters, with a small arrow indicating its top, as shown in Fig. 3. As with the shape cue, the orientation of the box indicated the orientation of the upcoming test stimulus.

For the duration of each trial, a small fixation cross occupied the center of the screen. Each of the cue and test stimuli was positioned on an imaginary circle that was centered on the fixation cross, so that they were all 5.6 cm (6.4°) from the fixation cross. The cue could

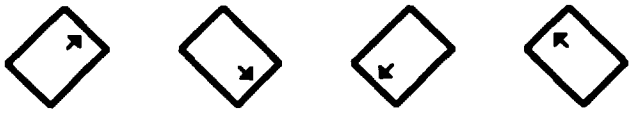

FIG. 3. The four no-shape cues used in Experiment 2.

appear at one of two locations, either to the far left or the far right of the display. The cue was not aligned with the fixation cross; instead, it was slightly higher, as shown at the top of Fig. 4. (A line connecting the cue and the fixation cross would intersect a horizontal line with an angle of 12°.) This displacement was to ensure that test stimuli appearing at the cued location would not receive any attentional benefit or cost that might come from being aligned with the fixation cross. The test stimulus could occur at the same location as that of the cue or at one of the locations 20, 95, or 160° around the imaginary circle in either direction from the cued location, yielding distances of 1.9, 8.3, and 11.0 cm between cue and test stimulus.

Procedure. As in Experiment 1, the subject was seated in front of the CRT in a dimly lit room. The subject was instructed to keep one hand on each of the response keys.

Each trial began with the presentation of the small fixation cross at the center of the display. The cross remained on the screen until the end of the trial, and subjects were instructed to keep their eyes fixed on it until then. The eye position was recorded 1500 ms after the fixation cross appeared and it was monitored for the rest of the trial. If the

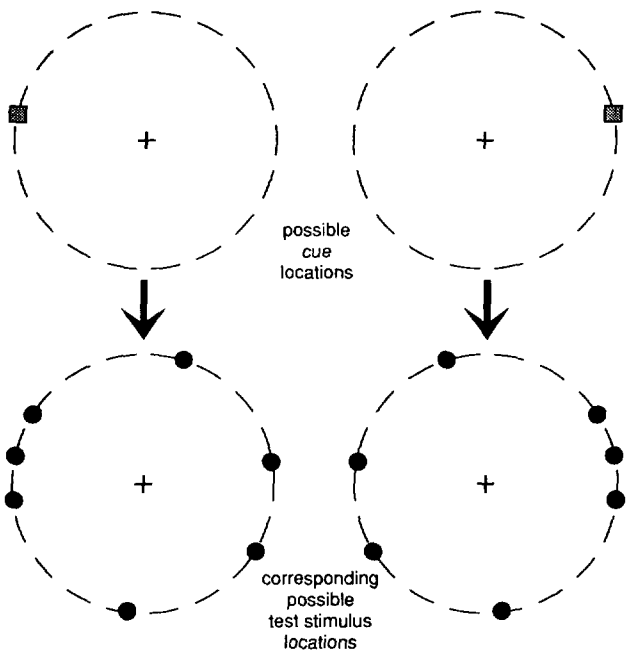

FIG. 4. The top two figures illustrate the two possible cue locations relative to the fixation cross. The bottom two figures illustrate the seven possible test stimulus locations that correspond to each of the cue locations.

computer detected an eye movement at any time during the trial, a buzzer sounded, the screen flashed red, and the trial was aborted.

The cue appeared just after the eye position was recorded. Because subjects did not know the cue location until after the eye monitoring began, they could not begin the trial with their eyes fixed on the cue. The cue remained on the screen for 700 ms. In the Image condition, the cue provided information about the shape, orientation, and location of the test stimulus that would appear soon afterward; in the No-Image condition, the cue provided only orientation and location information. The shape information (when it was available) and the orientation information were always accurate. The location information, however, was only accurate on half the trials. In the other half, the location was evenly distributed over the other six possible locations. Because of the disproportionately large number of stimuli at the cued location, subjects were better off to position images at the cued location than at any other location. The instructions stated that the test stimulus would often occur at the same location as the cue, but there was no mention of imagery or image position.

The test stimulus appeared 1500 ms after the cue disappeared. As in Experiment 1, the computer waited for the beginning of a video cycle, presented the test stimulus, waited for an interval of 120 ms, and then removed it at the beginning of the next video cycle. As before, subjects pressed the key under the dominant hand if the test stimulus was normal, and the other key if it was mirror-reversed. If the subject pressed the wrong key, the screen flashed blue and a buzzer sounded (using a different tone than that used to indicate an eye movement). Error and eye movement trials were saved and were repeated at the end of each block of 128 trials, and this process continued until all trials in that block had been completed correctly. After every 32 trials, the computer stopped presenting trials and the subject was allowed to take a break.

With two cue types, three shapes, four orientations, two responses (normal and mirror-reversed), there was a total of 48 trial types for each location. For each cue location, six stimuli occurred at the same location as the cue, and one appeared at each of the six possible uncued locations. Because there were two different cue locations, there were a total of 1152 trials for each subject, not including those that were repeated because of errors. A new random order was generated for each subject, with all the different trial types intermixed. Each subject usually required about four testing sessions, each lasting for an hour or less.

Results

The response time data were submitted to an ANOVA with cue type, orientation, handedness (normal or mirror-reversed), shape (R, J, or 4), distance, cue hemifield (left or right), and placement around the circle (clockwise or counterclockwise, which was irrelevant for stimuli at the cued location) as factors. Subjects made incorrect responses on 2% of the trials. As described above, incorrect trials were repeated at the end of the block, and only response times from correct trials were included in the analysis.

The main purpose of Experiment 2 was to measure the effects of distance on the use of images. A linear contrast revealed that response times generally increased with the distance between cue and test stimulus, $F(1,42) = 96.6$, $p < .001$. The linear trend accounted for almost 99% of the variance among the means, and the residual from linearity was not significant, $F < 1$. More importantly, a second contrast showed that this increase was greater in the Image (Shape Cue) condition than in the

No-Image (No-Shape Cue) condition, $F(1,42) = 12.3$, $p < .005$. Figure 5 shows the data from these two conditions, along with the best-fitting regression line for each. In the No-Image condition, the increase with distance probably reflects the attentional effects described earlier. The slope in the Image condition is much larger, suggesting that subjects are "moving" the image from the cued location to the stimulus location.

As in Experiment 1, subjects responded more quickly when they knew the shape of the stimulus beforehand, $F(1,14) = 129.0$, $p < .001$. As expected, contrasts showed that responses were faster for the 45 and 315° (−45°) orientations than for the 135 and 225° (−135°) orientations, $F(1,42) = 160.8$, $p < .001$, and that the disadvantage for the larger rotations was greater when the shape was unknown, $F(1,42) = 54.2$, $p < .001$. These results suggest that subjects were using the shape information to generate images, and that these images helped them with the more misoriented stimuli, just as in Experiment 1.

This pattern is apparent in Fig. 6, which shows response times for each orientation. In each graph, the upper line represents the No-Image condition and in each case the response time for the smaller rotations (45 and 315°) are much lower than those for the larger rotations (135 and 225°). The lower line in each graph, from the Image condition, also shows an advantage for the smaller rotations, but because of the use of the image,

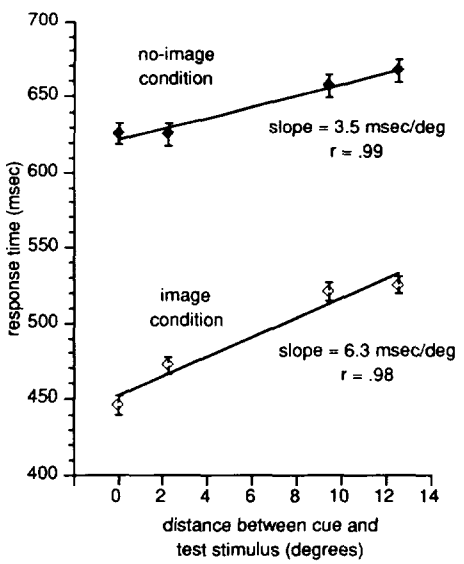

FIG. 5. Response time as a function of (straight line) distance between cue and test stimulus for both Image and No-Image conditions. The r values give the correlation between mean response time and distance for each of the two conditions.

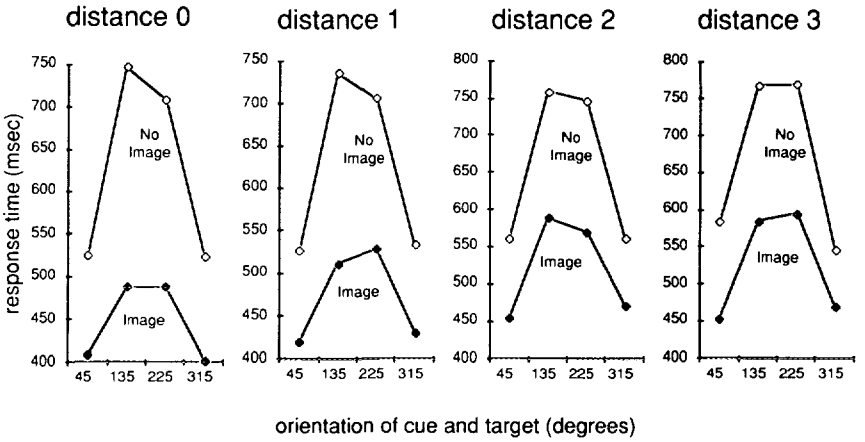

FIG. 6. Response times for each of the four orientations, plotted separately for each distance between cued location and stimulus location.

the advantage is smaller. This pattern is consistent with the claim that subjects move images to the stimulus location, because even at the greatest distance, there is still a much smaller orientation effect in the Image condition than in the No-Image condition. The presence of a small but significant orientation effect in the Shape Cue condition of Experiment 1 indicates that even when subjects had enough information to create an image in advance, they sometimes rotated the stimulus, presumably because they occasionally failed to attend to the cue or lost the image before they could compare it against the stimulus.

An interesting and unexpected result emerges when the four graphs in Fig. 6 are compared. Each represents the data from one of the distances between image and test stimulus. Once the image has been moved to the test stimulus location, its usefulness should not depend on the distance it traveled. In other words, the advantage for small rotations over large rotations should not vary with distance. However, the lower lines in Fig. 6 show that in the Image condition, the gap between small and large rotations grows as the distance increases.

To illustrate this pattern more clearly, we calculated the response time difference between small and large rotations for each distance and displayed these differences in Fig. 7. The figure shows that there is no linear increase in the orientation difference when there are no images, and a contrast confirms that the linear increase rises when images are used, $F(1,126) = 4.3$, $p < .05$. The longer the distance from the cue to the stimulus, the farther the image must be moved before it can be compared with the stimulus. Thus with longer distances, the image must be maintained longer and is more likely to be lost. Additionally, the longer dis-

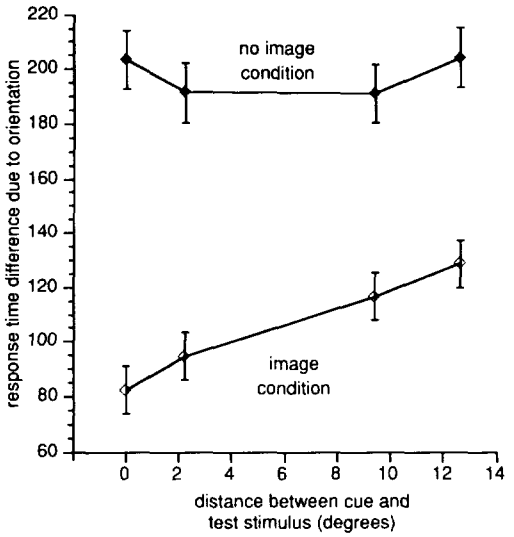

FIG. 7. The difference between the mean response times for the two large rotations (135 and 215°) and the two small rotations (45 and 315°).

tance means that more work must be invested in the image before it can be used, making it more feasible to abandon the image and rotate the stimulus instead. Therefore, subjects are likely to rotate the stimulus more often with longer distances, and when the response times are averaged there will be a larger orientation effect with longer distances.

Once again, subjects responded more slowly to mirror-reversed stimuli, $F(1,14) = 29.3$, $p < .001$. Neither cue hemifield nor placement direction (clockwise or counterclockwise) exerted a measurable influence on response time, $F < 1$ in both cases.

As in Experiment 1, the main analysis produced other unpredicted significant effects that appeared to be irrelevant and will not be considered here. Analysis of the error rates revealed no evidence of a speed/accuracy trade-off.

Discussion

Although Experiment 2 used a different procedure from Cooper and Shepard's, it leads to similar conclusions. In both tasks, differences in irrelevant spatial properties between the stimulus and the image require an amount of time proportional to the size of the difference, and in both experiments the need for this adjustment time varies depending on whether or not subjects know the shape of the stimulus in advance. The results from both experiments seem to reflect a shape-specific representation in which spatial properties are represented integrally with shape and in

which these spatial properties are adjusted gradually. Otherwise, it would be difficult to explain why the time necessary to compensate for a change in one of these spatial properties should depend on knowledge of the shape.

The data presented in Fig. 5 illustrate this point particularly well because of the counterintuitive pattern that they exhibit: the condition with the faster response times also exhibits the larger slope. Overall, subjects were faster when they could prepare an image. When they did, however, they had relatively more trouble with stimuli that were farther from the cued location. Cave and Kosslyn (1989) found a similar pattern in a size scaling experiment. In that case, responses were generally faster when subjects prepared for the correct shape to appear than when they prepared for the incorrect shape. With the correct shape, however, response time increased more sharply with the ratio between expected and actual size than it did with the incorrect shape. Together these two results suggest that images can generally be used in visual perception to speed certain shape processing tasks, but that in doing so they make visual processing more susceptible to irrelevant differences in spatial properties. For an image representation to be optimally useful in a comparison with a visual stimulus, it must be encoded at the right location. By demonstrating this fact, Experiment 2 underscores once again that the organization of the image representation corresponds in many ways to the organization of the retinal projection.

The No-Image condition of Experiment 2 also shows an increase in response time with distance, presumably reflecting differences in the allocation of attention to different locations in the visual field. The effect of distance seems to be very different in the Image condition than in the No-Image condition, both because it is about twice as large in the Image condition and also because the orientation effect increases with distance in the Image condition but not in the No-Image condition. For these reasons, we can be confident that the distance effect in the Image condition reflects something more than just a general attentional "window" or "spotlight."

Although image location adjustments exhibit the same analog properties as orientation adjustments, Experiment 2 shows that location adjustments are fast, compared to mental rotation. Moving an image is easy enough that the subjects in Experiment 1 apparently chose to form the image in advance, and then move it after they knew the correct location. This easy translation of images makes them much more useful in a world full of translating objects and translational eye movements.

EXPERIMENT 3

If location is in fact encoded in the mental representations underlying visual images, then we can ask what sort of coordinate frame is used. The

visual input is necessarily organized according to retinal location. In order to navigate through space and manipulate physical objects, some stage of visual processing must include an integration of information from different fixations using a single (spatiotopic) frame of reference. Therefore, we can ask whether location is encoded retinotopically or spatiotopically in these image representations. Determining the coordinate frame used in these mental representations will help to ascertain the level at which these representations are used in regular visual processing. To do this, we must have some idea of where in the stream of visual processing the retinal coordinates are converted to a more useful reference frame.

One possibility is that the coordinate shift occurs early in visual processing. Feldman (1985a) uses this strategy in his Four Frames Model, in which visual information in the "retinotopic frame" is transferred to a part of the "stable feature frame." The location within the stable feature frame for each fixation is determined by eye position. Thus multiple fixations can be integrated into a single, complete representation. The stable feature frame is still spatially organized: shape properties are still an integral part of the representation of each shape. Instead of coding location in terms of retinal position, however, the stable feature frame codes location in relation to head position. Virtually all high-level visual processing is then based on the stable feature frame and not on the retinotopic frame.

Feldman (1985a, 1985b) cites a number of reasons for implementing an early coordinate shift in this model. First, the early shift makes it easy to integrate information across fixations. Even when the viewer is too close to an object to see it all in a single fixation, it is possible to assemble all the parts together into a single coherent representation. Furthermore, the stable feature frame can serve as a necessary "substrate" for our subjective experience of a visual world that is unified across fixations. Feldman also states that it can be used for imagery, although he does not elaborate on the disadvantages of implementing imagery in the retinotopic rather than the spatiotopic frame. Finally, he lists perceptual experiments that are consistent with the presence of the stable feature frame. For example, Davidson, Fox, and Dick (1973) presented subjects with a letter array, had the subjects move their eyes, and then presented a mask at one of the letter locations. When the subjects were asked to report the mask location, they usually reported its correct spatiotopic location.

If the coordinate shift occurs as early in visual processing as Feldman claims, then image representations are almost certainly not coded retinotopically. In spite of his evidence, however, there are other reasons to believe that the coordinate shift occurs later in processing; perhaps *after* stimuli are compared with these image representations. One reason arises from Davidson et al.'s data. Although subjects reported the mask at its

spatiotopic location, the mask interfered with the letter that was at the same *retinotopic* location. Using a similar methodology, Irwin, Brown, and Sun (1988) also found that mask interference depended on retinotopic location. They then tested the integration of information across saccades more carefully by replacing the mask with a small bar and asking subjects to report the letter that occurred at the same location as the bar. As long as the delay between the letter array and the bar was short, subjects reported the letter at the same retinotopic location more often than the letter at the same spatiotopic location, even though in many trials they could also correctly report the bar's spatiotopic location. With a longer delay, they more often reported the letter at the same spatiotopic location, but they also reported a different visual experience, with the stimuli before and after fixation no longer fused together as they were with the shorter delay. In their final experiment, Irwin et al. used a task that required the fusion of two dot patterns into a single form. Performance on this task was better when the two patterns occupied the same retinotopic location than when they occupied the same spatiotopic location.

Irwin et al.'s results indicate that, at least over short time intervals, information across saccades is superimposed within a retinotopic coordinate frame, rather than within a spatiotopic or head-centered system as it would be in the stable feature frame. They concluded that when the delay is longer and the information from different fixations is integrated spatiotopically, this integration is not done within the early stages of processing that are the basis for visual persistence. More recently, Irwin, Zacks, and Brown (1990) have collected additional evidence by testing for a spatial-frequency-specific priming effect. Normally, subjects are less accurate at detecting a grating stimulus when another grating of the same spatial frequency has just appeared at the same location. Irwin et al.'s subjects moved their eyes after the presentation of one grating and before another. They showed no decrease in performance, even though the two gratings occupied the same spatiotopic location.

More serious doubts about the early coordinate shift arise from neuroanatomical and neurophysiological studies of the visual system. Although these studies have discovered numerous brain regions devoted to different aspects of visual processing, the receptive fields of the cells in almost all of these regions appear to be retinotopically organized. One partial exception is in posterior parietal cortex, where Andersen, Essick, and Siegel (1985) found individual units that responded to stimuli at a particular retinotopic location, but the strength of the response was gated by eye position. These units could be part of a distributed representation of location in head-centered coordinates (Zipser & Andersen, 1988). This area of the brain, however, appears to be devoted to the processing of location, while working in conjunction with another system in the tem-

poral lobes that processes identity (Ungerleider & Mishkin, 1982). Lesions in this area interfere with monkeys' ability to detect stimulus location, but not to identify visual patterns. Thus, the role of posterior parietal cortex in visual processing is much later than Feldman's stable feature frame. When Kosslyn, Flynn, Amsterdam, and Wang (1990) constructed a model of higher-level visual processing, the accumulated evidence from neuroanatomy and neurophysiology led them to omit an early spatiotopic frame. In their model the spatiotopic transform occurs in the location subsystem, after location information has been factored apart from shape information. The shape information that is matched against memory representations and categorized by the identity subsystem has not been transformed into spatiotopic coordinates.

Experiment 3 is designed to test whether the representations used in mental rotation are coded retinotopically or spatiotopically. The answer to this question is important in determining how these representations are used in visual processing. It could also have general implications for the coding of location in all types of visual processing, and for the way that information is integrated across fixations. If retinotopic representations are used in mental rotation, then a coordinate shift is unlikely to occur early in visual processing. Conversely, mental rotation itself could not be so late in the processing stream that it occurs after spatiotopic recoding.

The general strategy in Experiment 3 is to measure the response time advantage that occurs when the image occupies the stimulus location, as was demonstrated in Experiment 2. By introducing a saccade in between the image cue and the test stimulus, we can measure whether this advantage is associated with the retinotopic or the spatiotopic location.

Method

Subjects. As in Experiment 2, we used a pupil-tracker to monitor eye movements. Most subjects required practice before they could respond to a stimulus in the periphery without moving their eyes, and some were unable to finish the experiment or chose to quit early. Once the data for a subject were collected, all response times that were more than three standard deviations from the mean for that subject were removed. After the outliers were removed, a few subjects did not have data for every combination of conditions and thus their data were not included. In all, a total of 75 subjects completed the experiment with a full set of data, and another 44 subjects did not. All subjects were from the M.I.T. Department of Brain and Cognitive Sciences subject pool, and all were paid for their time, whether or not they completed the experiment. Most were M.I.T. undergraduates, and their vision was normal or corrected to normal. These subjects did not participate in the earlier experiments.

Apparatus. The computer, display, pupil-tracker, and video camera used in Experiment 2 were all used in this experiment.

Stimuli. The test stimuli were the same three characters, either normal or mirror-reversed, displayed at the same four orientations, 45, 135, 225, and 315°. The two cue types were also the same: each shape cue was a normal character oriented appropriately, and each no-shape cue was a rectangle with a small arrow indicating the top, also oriented appropriately.

In this experiment, there were three possible locations at which the cue and test stimulus

could appear, as shown in Fig. 8: one at the center, one at the far left, and one at the far right. The fixation cross could appear at one of two locations: either between the center and left stimulus locations or between the center and right locations. A cue or test stimulus could only occur in one of the two locations next to the current fixation cross. The distance between each fixation cross and each of the two neighboring stimulus locations was 4.8 cm (5.5°). With this arrangement, a test stimulus occurring to the right of the left fixation cross would occupy the same screen location as a stimulus to the left of the right fixation cross.

Suppose that a subject fixates on the left fixation cross and then views a cue to the right of the fixation cross (in the center of the screen). The subject then saccades to the right fixation cross. If a test stimulus now appears to the left of the new fixation cross, it will be in the same spatiotopic location as the cue. If it instead appears to the right of the fixation cross, it will be in the same retinotopic location as the cue. In this experiment, the cue and test stimulus always appeared at one of the two locations next to the current fixation cross. Therefore, the distance of the cue and test stimulus from the fixated location was always constant.

Procedure. As in the previous experiments, the subject was seated in front of the CRT in a dimly lit room. A chin rest and forehead restraint were used to maintain a constant viewing distance and prevent head movements. The subject was instructed to keep one hand on each of the response keys.

Because the eye-tracker was used to monitor saccades in this experiment, a short calibration procedure was necessary for each subject prior to testing. During this procedure, the eye-tracker monitored eye position while the subject performed a simple visual task that required saccades back and forth between two points on the screen. The calibration program calculated the median difference between the pupil locations before and after each saccade, and this value was used in the regular experiment to predict the correct eye position after a saccade was cued. This calibration procedure was repeated at the beginning of each testing session.

Figure 9 illustrates the various steps of the experiment itself. Each trial began with the appearance of a fixation cross at one of the two possible locations. The subject was instructed to fixate on the cross when it appeared. After the cross had been visible for 1500 ms, the eye position was recorded and the cue appeared, either to the left or the right of the fixation cross. The cue was present for 700 ms and then disappeared, leaving only the fixation cross for another 1500 ms. The subject was instructed to remain fixated on the cross during this entire time, and if the computer detected a substantial eye movement the trial

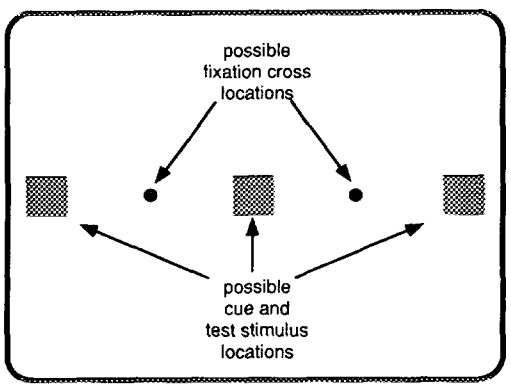

FIG. 8. The two locations at which a fixation cross could appear, and the three locations at which a cue or test stimulus could appear.

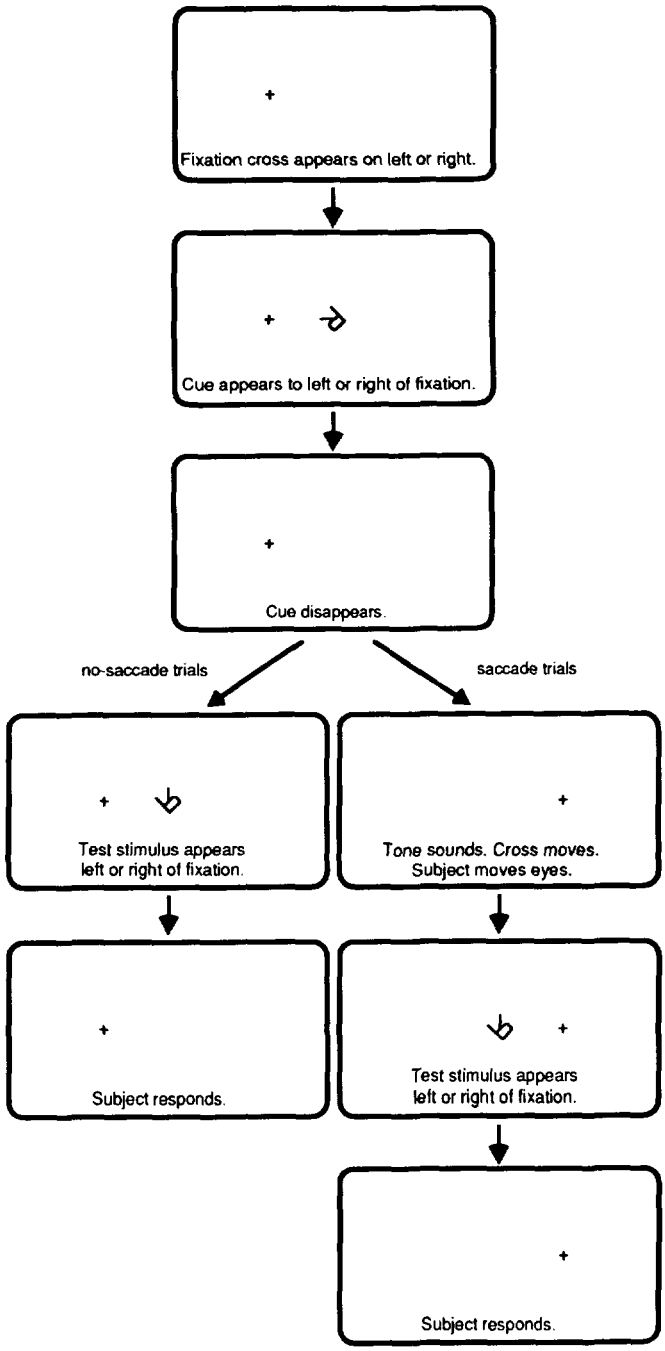

FIG. 9. The sequence of events for no-saccade trials and saccade trials in Experiment 3.

was aborted, as in Experiment 2. Note that once the fixation cross appeared, the cue was equally likely to appear 4.8 cm to the left or to the right of the cross. Therefore the subject's best strategy was to follow the instructions and to fixate on the cross, half-way between the two possible cue locations.

The next step in the procedure varied between the two different trial types. In the no-saccade trials, the test stimulus was presented for 120 ms, either at the cued location on one side of the fixation cross or at the uncued location on the other side. The subject then responded as in the previous experiments. In the saccade trials, however, a tone sounded, and the fixation cross moved to the other fixation location on the other side of the display. This was the subject's signal to shift fixation to a new location. At this point, the computer no longer compared the eye position to the standard eye position measured at the beginning of the trial. Instead, it calculated a new standard position by adjusting the old position according to the values from the calibration procedure. It then began comparing the eye position against the new standard, using the same threshold as before. If the eye position did not come within the threshold range of the new standard within 2000 ms, the trial was aborted. Once the eye position came within the threshold range, it was compared for a two additional pupil-tracker cycles (approximately an additional 33 ms) to ensure that the eye was no longer moving.

After this third eye position comparison, the current eye position was taken as the new standard for the rest of the trial. The test stimulus then appeared on either side of the new fixation cross, and the subject responded appropriately. The test stimulus was equally likely to occur 4.8 cm to the left or right of the new fixation cross. Thus, as before, once the new fixation cross appeared, the subject's best strategy was to follow the instructions and to fixate on the cross, half-way between the two possible stimulus locations.

After an incorrect response, the screen flashed blue and a long tone sounded. In both trial types, eye movements were monitored until the response. If an uncued saccade occurred at any stage of the trial, the screen flashed red and a long tone sounded. Using the same procedure as in Experiment 1, trials with incorrect responses or incorrect eye movements were repeated after each block of 96 trials, and this process continued until all trials in the block had been completed correctly. Subjects were given an opportunity for a break after every 20 trials.

In this experiment, half the trials were saccade trials and half were not. Within each type, half were shape cue trials and half were not. The fixation could be either to the left or right of center, and the cue could be to the left or right of fixation. The test stimulus could be at the cued or uncued location, it could be one of three shapes, it could be at one of four orientations, and it could be either normal or mirror-reversed. All the different combinations produce a total of 768 trials for each subject. A different random order was generated for each. Most of the 75 subjects who completed the entire set of trials required four or five sessions, each lasting about an hour.

Given that subjects can adjust represented location in images quickly, they might be able to produce either retinotopic or spatiotopic results by "moving" their images as they move their eyes. Assuming for the moment that images are coded spatiotopically, if the subject believes that the image should always be in the same retinotopic location as the cue, then after a saccade the subject could shift the image to the physical screen location that now corresponds to the cue's retinotopic location. Likewise, if images are coded retinotopically, subjects could move the image to the spatiotopically correct physical location after the saccade. To investigate this possibility, we randomly split the 75 subjects into three groups of 25 each, and gave each group slightly different instructions. All three groups were instructed to prepare for the test stimulus at the location occupied by the cue. Subjects in the retinotopic group were told that on saccade trials, they should expect the test stimulus at the same location relative to the fixation cross as the cue. Those in the spatiotopic group were told to expect the test stimulus at the same location on the screen as the cue. Those in the

neutral group received no instructions either way. (As in the previous experiments, there were no specific instructions to make a mental image or to put an image at a particular location.) If subjects are able to adjust image location before the stimulus appears, then the location for which subjects respond more quickly should vary with the instructions they receive. For all three subject groups, there were equal numbers of trials at the cued and the uncued locations.

Results and Discussion

To determine whether the image representations used in this task are retinotopic or spatiotopic, we must look at those trials in which the test stimulus could occur at either the same retinotopic location or the same spatiotopic location as the cue. Therefore, we discarded data from all trials in which the cue appeared at the far left or right of the screen, because after a saccade the test stimulus could never occur at the same spatiotopic location in these trials. These trials were included in the experiment to ensure that subjects fixated at the appropriate location.

The trials without saccades serve as a control to test whether the location-specificity found in Experiment 2 can be measured in the current experiment. These trials demonstrated that subjects responded more quickly when the test stimulus occurred at the same location as the cue. The response time data for trials with and without saccades were submitted to separate ANOVAs, with cue, orientation, location, handedness, shape (R, J, or 4), and type of instructions as factors. Subjects made incorrect responses on 2% of the trials. These trials were excluded from the analyses and new trials were added to replace them. Because the results from Experiments 1 and 2 indicated that any retinotopic or spatiotopic advantage would be very subtle, we wanted to ensure that any small effects were not obscured by long response times from trials in which subjects lost their concentration. Therefore we excluded trials with response times more than three standard deviations from the subject's mean. (On the average, only 7 of the 768 correct-response trials for each subject were excluded.) In the results described below, the values F_f and p_f are from the analysis of trials in which the eyes remained fixed, and the values F_s and p_s are from the analysis of trials with a saccade.

In both saccade and no-saccade trials, the results indicated that subjects were using images when possible, as they had in the previous experiments. Responses were faster when subjects knew the shape beforehand, $F_f(1,72) = 170.9$, $p_f < .001$, $F_s(1,72) = 167.4$, $p_s < .001$. Contrasts indicated that they were also faster for the less misoriented stimuli (45 and 315°) than for the more misoriented stimuli (135 and 225°), $F_f(1,216) = 761.2$, $p_f < .001$, $F_s(1,216) = 712.7$, $p_s < .001$, and that the disadvantage for the larger rotations was less when subjects knew the shape, $F_f(1,216) = 48.0$, $p_f < .001$, $F_s(1,216) = 30.4$, $p_s < .001$, suggesting once again that shape knowledge allowed subjects to form images. As before, responses

were slower for mirror-reversed stimuli, $F_f(1,72) = 128.9$, $p_f < .001$, $F_s(1,72) = 92.5$, $p_s < .001$. The three different instruction sets had no significant overall effect on response times, $F_f < 1$, $F_s < 1$.

When the eyes remained fixed, subjects responded more quickly when the test stimulus occurred at the cue location, $F_f(1,72) = 74.4$, $p_f < .001$. This location advantage was stronger in the Image (Shape Cue) than in the No-Image (No-Shape Cue) condition, $F_f(1,72) = 33.7$, $p_f < .001$, as Fig. 10 shows. This pattern confirms the finding from Experiment 2 that the faster responses for the cued location are not just a general attentional affect, but an indication that image representations include location information.

The results from the no saccade trials confirm that this particular task can be used to measure the advantage that comes from having an image in the right location. Thus the data from the saccade trials should indicate whether location information in images is encoded retinotopically or spatiotopically. Overall, subjects responded faster to stimuli in the same retinotopic location as the cue than to those in the same spatiotopic location as the cue, $F_s(1,72) = 7.7$, $p_s < .01$. More importantly, the retinotopic advantage was stronger in the Image condition than in the No-Image condition, $F_s(1,72) = 5.9$, $p_s < .02$, as is shown in Fig. 11. From this difference we conclude that image representations are encoded in a retinotopic coordinate system.

Responses to stimuli in the cued and uncued locations in the trials without saccades provided more evidence that subjects occasionally rotated stimuli even in the Image condition. When the shape was cued, there was a greater response time difference between larger and smaller rotations for stimuli that appeared at the uncued location than for those that appeared at the cued location; without the shape cue, the orientation

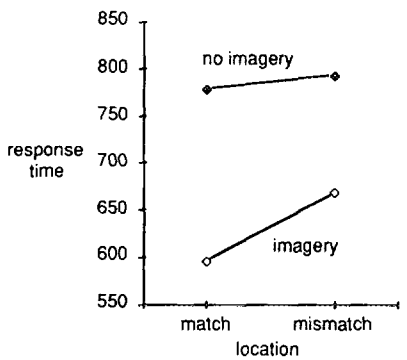

FIG. 10. Mean response time for no-saccade trials when test stimulus location matched the cued location and when it did not.

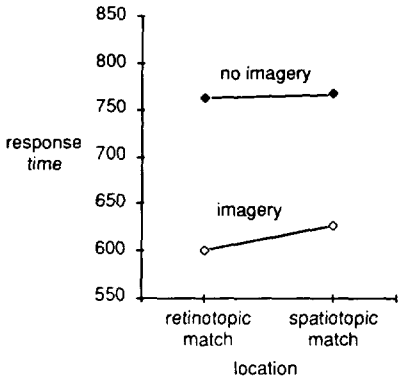

FIG. 11. Mean response time for saccade trials when test stimulus location matched the retinotopic location and when it matched the spatiotopic location.

difference was not greater for the uncued location and was actually somewhat less, $F_f(3,216) = 4.3$, $p_f < .01$. This pattern suggests that when subjects know the shape, they are more likely to rotate the stimulus if it occurs at the uncued location. This is the same sort of pattern seen in Fig. 7 in Experiment 2. The same general pattern also appeared in the saccade trials for Experiment 3 (a larger orientation difference with spatiotopic stimuli than with retinotopic in the image condition), but the effect was not significant, $F_s(3,216) = 1.0$. Presumably the same explanation accounts for all three cases. If the stimulus is at the uncued location, then the image that was prepared beforehand must be moved to the new location before it can be used, and thus subjects are more likely to abandon the image and rotate the stimulus.

Both the saccade and no-saccade analyses produced other significant effects that we did not plan to test a priori and that will not be considered here. The error rates in this experiment generally covaried with response times. One small exception appeared in the error rates for the no-saccade trials, in which there was an advantage for the cued location in the Image condition that was balanced by an advantage for the *uncued* location in the No-Image condition, $F_f(1,72) = 6.4$, $p_f < .02$. This difference raises the possibility that the faster response times at the cued location might be due in part to a speed/accuracy trade-off rather than to the allocation of visual attention, as suggested earlier. Whatever the explanation for the no-image data, there is no hint of a speed/accuracy trade-off in the Image condition, providing no reason to doubt that performance improves when the stimulus appears at the same location as the visual image. Furthermore, the error rates in the saccade trials give no indication that the response-time retinotopic advantage might be due to a speed/accuracy trade-off.

Varying the instructions. Both the saccade and no-saccade trials produced little overall difference in response time among the three instruction groups, $F_f < 1$, $F_s < 1$. There was also no significant difference between the three groups in the advantage for retinotopic over spatiotopic location in the saccade trials, $F_s(2,72) = 1.5$, $p_s < .2$, or in the interaction between retinotopic/spatiotopic location and the Image vs No-Image conditions, $F_s < 1$. Figure 12 confirms that the interaction between imagery and retinotopic/spatiotopic location is found in all three instruction conditions, with only slight modulations imposed by the instructions.

Clearly, the retinotopic location enjoys an advantage in the Image condition regardless of the instructions. Although the instructions had little effect on the use of images, there are nonsignificant trends suggesting that the instructions affected the allocation of attention. The data from the

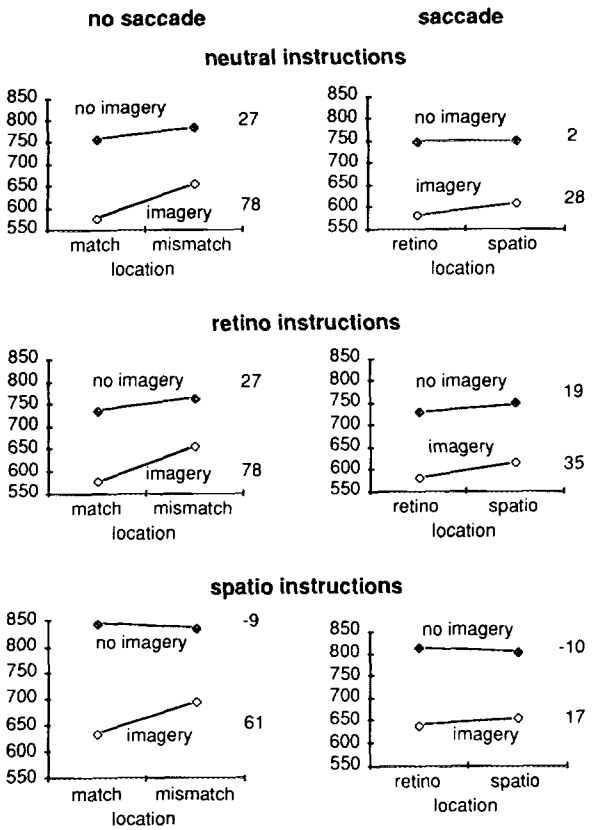

FIG. 12. Response times presented separately for the three instruction groups. The numbers to the right of each graph are the differences between the response times for the two location conditions.

saccade trials are the most relevant. The results in the No-Image condition seem to reflect the instructions: responses to the retinotopic location were faster with retinotopic instructions, responses to the spatiotopic location were faster with spatiotopic instructions, and response times were almost equal for the two locations with neutral instructions. This pattern suggests that subjects took note of the instructions and followed them when allocating their attention, even though there were always equal numbers of stimuli at each location. At first this flexibility of attention seems to run counter to the claims of Posner and Cohen (1984) that attention is allocated in retinotopic coordinates. However, Posner and Cohen also infer that subjects have "considerable voluntary control" over attention allocation, because in their experiments they can manipulate attentional facilitation by manipulating the probable target locations. Also, they conclude from another experiment that "inhibition of return" operates in spatiotopic coordinates. Therefore in our saccade trials, inhibition of return could be negating the effects of retinotopic attentional facilitation.

The same attentional effects from the instructions appear in the Image condition. In each instructional group, however, a retinotopic advantage appears in the Imagery condition, regardless of attentional fluctuations. These results make a strong case for the retinotopic coding of location in the image representations used in mental rotation, regardless of instructions and attention allocation.

GENERAL DISCUSSION

From Experiment 3 we conclude that the visual image representations used in mental rotation interact with stimulus representations at a level before any spatiotopic coordinate transform. There is an alternative explanation, however, that must be considered. Perhaps some part of the visual system uses a "scene-based" reference frame, in which locations are measured not by retinotopic location, nor by location within the whole environment, but by relative location to a particular subset of the objects in the visual field. In this case, the "scene" would consist of the stimulus and the fixation cross. If an image were placed according to scene-based coordinates, then it would always be placed at the same location with respect to the fixation point and we would expect the same pattern of results we find in Experiment 3. One way to test the scene-based account against the retinotopic account would be to keep both fixation points visible at all times and to surround each potential stimulus location with a different geometric shape. The shapes and the fixation points could each be of a different color.

Without the benefit of such a test, we believe that the retinotopic account is more plausible for two reasons. First, it is difficult to believe that

the visual system would choose to define the scene only by the small fixation cross and ignore the edge of the display monitor and other environmental stimuli surrounding it. Second, if the visual system goes to the trouble of transforming the coordinate system, then why not take advantage of information about saccades and create an environmental reference frame, rather than just a scene-based frame?

If mental rotation does operate before a coordinate transform, then an ordering is imposed on visual information processing: The representations used in mental rotation cannot be at so late a stage that they follow a spatiotopic transform; the spatiotopic transform cannot be at so early a stage that it precedes these image representations. In particular, results from Experiment 3, along with those of Irwin, Brown, and Sun and those from visual neurophysiology suggest that the spatiotopic transform does not occur at a very early processing stage, despite Feldman's arguments. If not, then at what processing stage could we expect the shift to occur?

It might be useful to try to place the coordinate shift in relation to another landmark on the visual processing pathway: The point at which shapes in the visual input are matched against memory representations and categorized. Assuming that there is a single such stage in visual processing, it is the stage at which spatial properties such as location, size, and orientation are factored out and represented separately from shape. In other words, this stage marks the point at which spatially organized representations are converted to abstract representations. Thus we can ask whether the spatiotopic transform occurs before or after the shift to abstract representations.

As mentioned earlier, the results from Cooper and Shepard's experiment and from numerous other imagery experiments indicate that the representations used in mental rotation are spatially organized (though see Pylyshyn, 1973, 1981, 1984). Likewise, the dot pattern task used by Irwin et al. (1988) also seems to be an "imagery task." In other words, it relies on a spatially organized representation as well. In this task, subjects must determine the single position in a 3×3 grid that is not occupied by a dot in either of two patterns. When the two patterns are superimposed, this task is trivially easy. If the mirror-reversal task and the dots task are both done with retinotopic representations, then are all spatially organized representations coded retinotopically? The alternative is that a representation that is spatially organized and spatiotopic exists somewhere in the visual processing stream after the representations used in mental rotation and before the abstract coding. In Feldman's terms, there would be a retinotopic frame and a stable feature frame, but for some reason the stable feature frame would not be the basis for visual imagery (or at least for mental rotation), even though it would be spatially organized. If the spatiotopic transform is postponed until after the level of image represen-

tations, then is there much to be gained by doing it before object identification, or might it be postponed until the level of abstract coding?

This question is best considered in relation to Ungerleider and Mishkin's (1982) extensive evidence that there is one subsystem in the temporal lobe to identify shapes and a separate subsystem in the posterior parietal lobe to process locations. If a dedicated subsystem is handling location information, then this subsystem probably performs the spatiotopic transform. One advantage would be that its representations would be specialized for recording location and would not be complicated by the presence of complex shape information, making the spatiotopic transform that much simpler. Therefore, the representations that contained spatiotopic location coordinates would not include shape information, and the representations in the identity subsystem that included shape information would not be spatiotopically coded. There would be no spatially organized representation that included shape information and was encoded spatiotopically.

In light of the division of labor between the identity and location subsystems, the findings from Experiments 2 and 3 have important implications for the role of image representations in visual processing. If the identity subsystem specializes in recognizing shapes regardless of location, then it probably utilizes representations that are location-independent. If the representations used in mental rotation are location-specific, as Experiment 2 demonstrated, and if these representations are coded in a retinotopic reference frame, as Experiment 3 demonstrated, then these representations are probably situated at a processing level before the identity-processing stream and the location-processing stream diverge.

One way in which visual imagery might be implemented at this processing level is illustrated in a model of high-level visual processing by Kosslyn, Flynn, Amsterdam, & Wang (1990). This model also demonstrates how the spatiotopic shift might be implemented in a system with separate subsystems for identity and location. The first stage of their model uses retinotopically coded representations of the input at a number of different scales. An "attention window" acts to select a particular area within one of these scales. The selected area is normalized to produce a location-independent and size-independent representation that is then passed on to the identity subsystem. In this process, the retinotopic location of the selected area is passed on to the location subsystem, which also uses head and body position information to transform retinotopic coordinates to spatiotopic coordinates. The abstract representation produced by the identity subsystem is recombined with the spatiotopic location information in an associative memory to make a complete, abstract representation of the visual input. Information can be fed backwards from

the associative memory through both the identity and location subsystems to form visual images within the original retinotopic representations.

If there is no spatially organized representation that is coded spatiotopically, then we must conclude that information from different fixations is not integrated at the level of spatially organized representations, as Feldman claimed, but instead at the level of abstract representations (see Irwin, 1991; Pollatsek, Rayner, & Henderson, 1990). This point is also well-illustrated in the model by Kosslyn, Flynn, Amsterdam, & Wang. As the eyes move from position to position, and also as the attention window moves within each fixation, different abstract shape representations produced by the identity subsystem accumulate in the associative memory, each coupled with its spatiotopic coordinates provided by the location subsystem. Together they form an abstract representation of the environment, coded in spatiotopic coordinates.

In conclusion, these results provide evidence that subjects engaged in a shape discrimination task utilize image representations in which location information is coupled with shape information. The represented location can be adjusted, but these adjustments are performed gradually, either smoothly or in small steps. Thus location in these representations appears to be adjusted in a manner similar to that used for size and orientation adjustments. These representations are coded retinotopically, but it makes little difference in everyday processing, because represented location can be adjusted quickly and easily whenever necessary.

REFERENCES

Andersen, R. A., Essick, G. K., & Siegel, R. M. (1985). Encoding of spatial location by posterior parietal neurons. *Science*, **230**, 456–458.

Bundesen, C., & Larsen, A. (1975). Visual transformation of size. *Journal of Experimental Psychology: Human Perception and Performance*, **1**, 214–220.

Bundesen, C., Larsen, A., & Farrell, J. E. (1981). Mental transformations of size and orientation. In J. Long & A. Baddeley (Eds.), *Attention and performance IX* (pp. 279–294). Hillsdale, NJ: Erlbaum.

Cave, K. R., & Kosslyn, S. M. (1989). Varieties of size-specific visual selection. *Journal of Experimental Psychology: General*, **118**, 148–164.

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

Davidson, M. L., Fox, M. J., & Dick, A. O. (1973). Effect of eye movements on backward masking and perceived location. *Perception & Psychophysics*, **14**, 110–116.

Downing, C. J., & Pinker, S. (1985). The spatial structure of visual attention. In M. I. Posner & O. S. M. Marin (Eds.), *Attention and performance XI*. Hillsdale, NJ: Erlbaum.

Farah, M. J. (1984). The neurological basis of mental imagery. A componential analysis. *Cognition*, **18**, 245–272.

Farah, M. J. (1985). Psychophysical evidence for a shared representational medium for mental images and percepts. *Journal of Experimental Psychology: General*, **114**, 91–103.

Farah, M. J. (1986). The laterality of mental image generation: A test with normal subjects. *Neuropsychologia*, **24**, 541–551.

Feldman, J. A. (1985a). Four frames suffice: A provisional model of vision and space. *The Behavioral and Brain Sciences*, **8**, 265–289.

Feldman, J. A. (1985b). Tunnel vision will not suffice. *The Behavioral and Brain Sciences*, **8**, 302–309.

Finke, R. A., & Pinker, S. (1982). Spontaneous imagery scanning in mental extrapolation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **8**, 142–147.

Finke, R. A., & Pinker, S. (1983). Directional scanning of remembered visual patterns. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **9**, 398–410.

Finke, R., & Shepard, R. N. (1986). Visual functions of mental imagery. In Kaufman, L. & Thomas, J. (Eds.), *Handbook of perception and human performance*. New York: Wiley.

Hinton, G. E., & Parsons, L. M. (1981). Frames of reference and mental imagery. In Long, J. & Baddeley, A. (Eds.), *Attention and Performance IX*. Hillsdale, NJ: Erlbaum.

Irwin, D. E. (1991). Information integration across saccadic eye movements. *Cognitive Psychology*, **23**, 420–456.

Irwin, D. E., Brown, J. S., & Sun, J. S. (1988). Visual masking and visual integration across saccadic eye movements. *Journal of Experimental Psychology: General*, **117**, 276–287.

Irwin, D. E., Zacks, J. L., & Brown, J. S. (1990). Visual memory and the perception of a stable visual environment. *Perception & Psychophysics*, **47**, 35–46.

Kosslyn, S. M. (1973). Scanning visual images: Some structural implications. *Perception & Psychophysics*, **14**, 90–94.

Kosslyn, S. M. (1978). Measuring the visual angle of the mind's eye. *Cognitive Psychology*, **10**, 356–389.

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

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

Kosslyn, S. M., Ball, T. M., & Reiser, B. J. (1978). Visual images preserve metric spatial information: Evidence from studies of image scanning. *Journal of Experimental Psychology: Human Perception and Performance*, **4**, 47–60.

Kosslyn, S. M., Flynn, R. A., Amsterdam, J. B., & Wang, G. (1990). Components of high-level vision: A cognitive neuroscience analysis and accounts of neurological syndromes. *Cognition*, **34**, 203–277.

Kubovy, M., & Podgorny, P. (1981). Does pattern matching require the normalization of size and orientation? *Perception & Psychophysics*, **30**, 24–28.

Larsen, A. (1985). Pattern matching: Effects of size ratio, angular difference in orientation, and familiarity. *Perception & Psychophysics*, **38**, 63–68.

Larsen, A., & Bundesen, C. (1978). Size scaling in visual pattern recognition. *Journal of Experimental Psychology: Human Perception and Performance*, **4**, 1–20.

Pinker, S. (1984). Visual cognition: An introduction. *Cognition*, **18**, 1–63.

Pinker, S., Choate, P. A., & Finke, R. A. (1984). Mental extrapolation in patterns constructed from memory. *Memory & Cognition*, **12**, 207–218.

Pollatsek, A., Rayner, K., & Henderson, J. M. (1990). Role of spatial location in integration of pictorial information across saccades. *Journal of Experimental Psychology: Human Perception and Performance*, **16**, 199–210.

Posner, M. I., & Cohen, Y. (1984). Components of visual orienting. In H. Bouma & D. G.

Bouwhuis (Eds.), *Attention and performance X: Control of language processes*. Hillsdale, NJ: Erlbaum.

Posner, M. I., Nissen, M. J., & Ogden, W. C. (1978). Attended and unattended processing modes: The role of set for spatial location. In H. L. Pick & I. J. Saltzman (Eds.), *Modes of perceiving and processing information*. Hillsdale, NJ: Erlbaum.

Posner, M. I., Snyder, C. R. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General*, **109**, 160–174.

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. (1981). The imagery debate: Analogue media versus tacit knowledge. *Psychological Review*, **88**, 16–45.

Pylyshyn, Z. W. (1984). *Computation and cognition: Toward a foundation for cognitive science*. Cambridge, MA: MIT Press.

Rizzolatti, G., Riggio, L., Dascola, I., & Umiltá, C. (1987). Reorienting attention across the horizontal and vertical meridians: Evidence in favor of a premotor theory of attention. *Neuropsychologia*, **25**, 31–40.

Rock, I. (1983). *The logic of perception*. Cambridge, MA: MIT Press.

Rosenthal, R., & Rosnow, R. L. (1985). *Focused comparisons in the analysis of variance*. Cambridge, England: Cambridge Univ. Press.

Sekuler, R., & Nash, D. (1972). Speed of size scaling in human vision. *Psychonomic Science*, **27**, 93–94.

Shepard, R. N., & Cooper, L. A. (1982). *Mental images and their transformations*. Cambridge, MA: MIT Press.

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

Skelton, J. M., & Eriksen, C. W. (1976). Spatial characteristics of selective attention in letter matching. *Bulletin of the Psychonomic Society*, **7**, 136–138.

Tarr, M. J., & Pinker, S. (1989). Mental rotation and orientation-dependence in shape recognition. *Cognitive Psychology*, **21**, 233–282.

Tarr, M. J., & Pinker, S. (1990). When does human object recognition use a viewer-centered frame of reference? *Psychological Science*, **1**, 253–256.

Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale, & R. J. W. Mansfield (Eds.), *The analysis of visual behavior*. Cambridge, MA: MIT Press.

Zipser, D., & Andersen, R. A. A back-propagation programmed network that simulates response properties of a subset of posterior parietal neurons. *Nature*, **331**, 679–684.

Accepted February 25, 1993