An eye movement analysis of “mental rotation” of simple scenes

CHIE NAKATANI

RIKEN Brain Science Institute, Hirosawa, Japan

and

ALEXANDER POLLATSEK

University of Massachusetts, Amherst, Massachusetts

Participants saw a standard scene of three objects on a desktop and then judged whether a comparison scene was either the *same*, except for the viewpoint of the scene, or *different*, when one or more of the objects either exchanged places or were rotated around their center. As in Nakatani, Pollatsek, and Johnson (2002), judgment times were longer when the rotation angles of the comparison scene increased, and the size of the rotation effect varied for different axes and was larger for *same* judgments than for *different* judgments. A second experiment, which included trials without the desktop, indicated that removing the desktop frame of reference mainly affected the *y*-axis rotation conditions (the axis going vertically through the desktop plane). In addition, eye movement analyses indicated that the process was far more than a simple analogue rotation of the standard scene. The total response latency was divided into three components: the *initial eye movement latency*, the *first-pass time*, and the *second-pass time*. The only indication of a *rotation effect* in the time to execute the first two components was for *z*-axis (plane of sight) rotations. Thus, for *x*- and *y*-axis rotations, rotation effects occurred only in the probability of there being a second pass and the time to execute it. The data are inconsistent either with an initial rotation of the memory representation of the standard scene to the orientation of the comparison scene or with a holistic alignment of the comparison scene prior to comparing it with the memory representation of the standard scene. Indeed, the eye movement analysis suggests that little of the increased response time for rotated comparison scenes is due to something like a time-consuming analogue process but is, instead, due to more comparisons on individual objects being made (possibly more *double checking*).

In the 1970s and 1980s, Shepard and his collaborators published a series of studies on *mental rotation* (see Shepard & Cooper, 1982, for an overview), which are among the best-known studies in experimental psychology. For example, in the original study, Shepard and Metzler (1971) presented two 2-D images that were projections of abstract 3-D figures that were assembled from several cubes as “building blocks.” In a typical task, one of the 3-D figures was rotated in 3-D space with respect to the orientation of the other, and participants were asked to judge whether the 3-D objects represented by the two images were the same or different. (When the figures were different, they were mirror images of each other.) Shepard and Metzler’s main finding was that the response time (RT) to process

the pair of objects increased linearly with the angle that the second object was rotated from the orientation of the first object. They called this linear function the *mental rotation function* and claimed that the increase in RT was proportional to the angle of rotation, because a 3-D representation of one of the objects was mentally rotated at a constant angular velocity before it was compared with the mental representation of the other object. This claim is often called the *mental rotation hypothesis*.

The mental rotation paradigm has recently been extended to the recognition of the layout of objects in a scene (Diwadkar & McNamara, 1997; Nakatani, Pollatsek, & Johnson, 2002). In these *scene rotation* studies, a stimulus consisted of several familiar objects that were placed on a background to compose a scene. The scene stimuli employed thus had a more complex global–local structure than did a single Shepard–Metzler object. A major question of interest has been whether a linear mental rotation function would be observed when two of these multiobject scene images were seen from different viewpoints and people were asked to judge whether the 3-D realities of the two scenes were the same or different. Diwadkar and McNamara employed grayscale photographic images of an array of six common objects (e.g.,

This research was supported by Grant HD26765 from the National Institutes of Health, by KDI Grant 9980045 from the National Science Foundation, and by a grant from the General Electric Fund. The present experiments were conducted in partial fulfillment of a PhD for C.N. We thank the other members of the committee, Neil Berthier, Donald Fisher, and Keith Rayner, for their helpful suggestions on the experiments. Correspondence concerning this article should be sent to C. Nakatani, Laboratory for Perceptual Dynamics Laboratory, RIKEN Brain Science Institute, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan (e-mail: cnakatani@brain.riken.jp).

a light bulb or a mug) on a round table. In the comparison scene, the array was rotated around the vertical axis (i.e., the axis perpendicular to the desktop) with respect to the standard scene, and participants were asked to report whether the relative locations of the objects were the same in the two scenes. The results were consistent with a mental rotation function: RTs increased more or less linearly between the 0° and the 135° rotations.

Nakatani et al. (2002) reported a similar result with a three-object array. The experiments, however, differed from Diwadkar and McNamara's (1997) experiments in two ways. First, Nakatani et al. examined rotations around four axes; in addition to rotations around the vertical (*y*) axis, there were rotations around the left-right (*x*) axis, the line-of-sight (*z*) axis, and an oblique axis. Second, two different types of changes, *location change* and *orientation change*, were employed. In the location change conditions, which were similar to all of Diwadkar and McNamara's *different* trials, either two or three objects switched their locations on the desktop (e.g., from mug-right and pen-left to mug-left and pen-right). In the orientation change conditions, either one or all three of the individual objects were rotated 90° around their own vertical axes (e.g., from the lamp facing the front of the desk to the lamp facing the side of the desk) with their locations unchanged.

Nakatani et al. (2002) observed mental rotation functions around all three major axes and the oblique axis, but the slopes were different for the axes of rotation, with the *y*-axis rotation having the largest slope. Furthermore, the slope varied with the type of change made in the scene: The slope of the mental rotation function was steeper in the orientation change condition than in the location change condition. The former result is inconsistent with a single rotation operation independent of axis of rotation, and the latter result is inconsistent with a two-stage *rotate and compare* model implied by the mental rotation hypothesis. That is, the mental rotation hypothesis tacitly assumes that the alignment/rotation of the entire stimulus is completed before the comparison process begins. If such a two-stage process were responsible for mental rotation effects, the slope (i.e., the rate of alignment) should be the same for the location change and the orientation change conditions (although the absolute times might differ, because the comparison stage might be easier for location changes). Instead, the results suggest either that the alignment process was applied to *pieces* of the scene stimulus or that the rotation and the comparison processes are intertwined.

Some piecemeal mental rotation models have been proposed (Bethell-Fox & Shepard, 1988; Folk & Luce, 1987) that were designed to explain the mental rotation of complex stimuli, such as irregular checkerboard patterns. These models assume that pieces of the complex stimulus are aligned individually and, as a result, the rate of rotation becomes slower as the number of pieces increases. However, it is not straightforward to apply these models to scene rotation unless one assumes that the in-

dividual objects are the pieces. However, this seems a bit simplistic, because a scene has a hierarchical structure—an object is “a piece” of a scene, but a part of the object could also be “a piece.” In addition, the relationships between the objects and their relation to the scene framework seem hard to explain merely by counting “pieces.” As a result, Nakatani et al. (2002) proposed a model that takes into account this more complex nature of a scene. In an earlier stage, a representation of the scene is rotated that merely contains crude *tokens* of the objects, which would be good enough for deciding whether two or more of the objects were in different locations, and in a second stage, more detailed representations of the objects are rotated (perhaps one at a time) in order to decide whether objects have been rotated around their individual axes. (We will return to this issue in more detail later.)

RTs, however, seem like a limited tool for examining a complex task that takes well over a second. Thus, in order to get a better understanding of the dynamics in such a task, we decided to examine the pattern of eye movements when people were doing the task. There is now a large literature concerning eye movements as an *on-line* measure of cognitive processes in general (see Rayner, 1998, for a recent review) and, more specifically, in scene recognition studies (Antes, 1974; Antes & Penland, 1981; Blackmore, Brelstaff, Nelson, & Troscianko, 1995; Boersema, Zwaga, & Adams, 1989; Boyce & Pollatsek, 1992; De Graef, De Troy, & d'Ydewalle, 1992; Friedman, 1979; Henderson & Hollingworth, 1999; Henderson, Weeks, & Hollingworth, 1999; Loftus & Mackworth, 1978; Mackworth & Morandi, 1967; Nelson & Loftus, 1980; Rayner & Pollatsek, 1992). In these studies, various eye movement indices were employed, although the sum of the durations of the fixations on an object was widely used as an index of the processing time of the object in a scene. However, there are few studies in which mental rotation effects have been investigated using fixation durations, and those that have (Just & Carpenter, 1985) measured eye movements in an object rotation task, but not in a scene rotation task.

The way we used eye movement records to understand the scene rotation task will probably become clear only when we have explicated the details of our data analysis method, but perhaps the following example will serve to illustrate the general idea of how an eye movement record can be of service in understanding the processing that occurs during the task. Suppose, for example, that a coarse computation of the relative viewpoints of the standard and the comparison scenes is done quickly, and then details/parts of the comparison scene are encoded, aligned further, and compared with corresponding parts of the standard scene (cf. Nakatani et al., 2002). To avoid an incorrect piecemeal process, the coarse computation should be completed before the eyes move to process the parts. If so, one should start to see *rotation effects* in the eye movement record (e.g., latency of the first saccade and fixation duration) fairly early in the trial—certainly

in the first half of a trial, which lasts over a second. In contrast, if mental rotation occurs only after lengthy encoding of the comparison scene or if early comparison processes do not require an alignment process, one may not observe mental rotation effects until the later portions of the eye movement record. As we will see, the latter pattern is quite close to what we observed.

EXPERIMENT 1

The task was the scene rotation task of Nakatani et al. (2002), discussed earlier. The participants saw the same standard scene (which always appeared from the same viewpoint) on each trial of a block of trials, followed by a varying comparison scene. The comparison scene always had the same objects as the standard scene. On half of the trials, the comparison scene was identical to the standard scene, except for a possible change in viewpoint; on the other half of the trials, either the locations of the individual objects were switched, or individual objects were oriented differently with respect to the desktop (in addition to a possible viewpoint change). The task was to judge whether the two scenes were identical, except for a possible change in the viewpoint of the observer.

Method

Participants. Twenty-one members of the University of Massachusetts community participated in the experiment. All had normal

or corrected-to-normal vision. They received either \$8 or experimental credits in psychology courses for their participation.

Stimuli and Design. The stimuli were computer-generated images of office objects on a desktop. There were four different standard scenes. Each had three office objects on a square desktop: Scene 1, briefcase, mug, and calculator; Scene 2, stapler, keyboard, and monitor; Scene 3, pen, telephone, and tape dispenser; Scene 4, desk lamp, document box, and index card holder. The desktop and the objects were made by 3-D graphic models with the Infini-D 2.5 software package. Each scene was rendered as an 800×600 pixel image with naturalistic colors and shading. The objects were carefully placed on the desktop to prevent occlusion of any of the objects in any view.

The standard viewpoint was from the front of the desktop, 5° above the gravitational horizontal plane of the desktop (see Figure 1). There were seven alternative viewpoints created by rotating a hypothetical camera around the desktop, starting from the standard viewpoint. Three orthogonal axes and an oblique axis of rotation were used. The x -axis went from left to right, so that x -axis rotations were equivalent to bringing the viewer over the desktop, with close to a bird's-eye view of the desktop at the 70° rotation. The y -axis rotations were clockwise in the horizontal plane around the vertical axis, so that the views were as if one were walking around the desk to the left. The z -axis went straight out to the viewer, so that a z -axis rotation was identical to a counterclockwise rotation of the camera in the picture plane. For each axis of rotation, there were two levels of rotation, 35° and 70° .¹ The seven viewpoint conditions are shown in Figure 1, the no-rotation condition (000) and the X35, X70, Y35, Y70, Z35, and Z70 conditions, together with a Y70–X70 (double-axis) condition (see note 1); these were the rotation conditions employed in Experiments 2 and 3 of Nakatani et al. (2002).

Figure 1. Scene stimuli.

Figure 2. Same-scene and different-scene conditions.

There were three types of comparison conditions for each of the seven comparison scene viewpoints (see Figure 2). For *same* comparison scenes, neither the location nor the orientation of any of the three objects (relative to the desktop) was changed from the standard scene. There were two types of *different* conditions. First, in the *location change* conditions, either the locations of two objects were switched (*Location 2 condition*), or the locations of all three objects were switched (*Location 3 condition*). Second, in the *orientation change* conditions, either one object was rotated 90° around its own vertical axis (*Orientation 1 condition*), or each of the three objects was independently rotated 90° around its vertical axis (*Orientation 3 condition*). (There were no location changes in the orientation change conditions.) All of these changes were approximately counter-balanced across viewpoints and particular objects.

Procedure. The task was to judge whether the object arrangements in the standard and the comparison scenes represented the same 3-D scene. The standard scene, which was always in the standard viewpoint, was presented before each trial began, and the participants could view it as long as they needed. When they pressed the “ready” key (the space bar), a pattern mask that covered the entire screen was presented for 300 msec, followed by a fixation point in the center of the screen. As soon as the participant fixated on the fixation point, the comparison scene was presented and remained on until the participant responded *same* or *different*. The “/” key was assigned to *same* responses, and the “z” key was assigned to

different responses. The participants were asked to respond as quickly and accurately as possible.

The session started with a block of 32 practice trials, in which feedback was provided about whether the response was correct after each trial, and then continued with the 256 trials of the experimental session, during which no feedback was provided. The practice trials employed a different scene from the four scenes used in the experimental session, but the trials were otherwise similar. The 256 experimental trials were divided into four blocks of 64 trials. In each block, all the trials were with the same standard scene, and half were *same* trials and half were *different* trials. For the 32 *different* trials in each block, two location change trials and two orientation change trials were presented for each of the eight viewpoints. To make the number of *same* and *different* trials equal, each *same* comparison stimulus at each of the eight viewpoints was repeated four times. The order of the 64 trials within a block was randomized separately for each participant, and the order of the standard scenes was counterbalanced over participants.

The eye movements during the scene rotation task were measured by an SMI EyeLink system. The system consists of a lightweight helmet that has an infrared sensor for head movements and two CCD cameras for eye tracking, two Compaq Deskpro PC compatible computers, and a 17-in. Viewsonic 17PS monitor. After the practice session, the participants donned the helmet and were seated in front of the display. Head movements were monitored and com-

pensated for by the infrared sensor system; thus, no head support was used. The distance between the display and the eyes was approximately 80 cm. Eye position was recorded only from the right eye.

At the beginning of the session, the camera position and image level were adjusted prior to an initial calibration procedure, using a nine-point grid pattern. The participant was recalibrated at the beginning of each trial block and whenever else it was needed. (An automatic drift correction was made around the center point before each trial, and a recalibration was inserted when the current measurement fell out of the margin of the drift correction.) The refresh rate of the display was 75 Hz, and the sampling rate of the eye position was 250 Hz. The sample was filtered, and fixation durations and fixation locations were extracted, along with other saccade-related indices. The filter was a part of the EyeLink system, which computes the saccade-related indices on the basis of acceleration.

Response measures. The mean RT and error rate were computed in each condition. There were also three levels of eye movement indices. At the finest level, we recorded both the location of each individual fixation and the duration of that fixation. Each fixation was counted as being on one of the three objects, and the object whose center was the smallest Euclidean distance from the fixation point was counted as the object being fixated. At the next level, analogous to the procedure in the reading literature (Just & Carpenter, 1980; Rayner, 1998), we grouped fixations into *gazes*. A gaze is a sequence of fixations that are all on the same object, and the sum of these successive fixation durations is a *gaze duration*.

At the most global level, we divided the trial into three components.² The first was the *initial latency*, the time between the onset of the comparison stimulus and the initial saccade off of the central fixation point. (This time did not count as a fixation time on any object.) The remainder of the trial was divided into a *first pass* and a *second pass*. The first pass was defined as ending either when one of the objects was *revisited* or when a response was made before any revisit. An object was defined as revisited if it had been previously fixated but another object had been fixated between the current fixation and a prior fixation on the object. Thus, the *first-pass time* is the sum of the gaze durations before the first revisit to a previously fixated object, and the *second-pass time* is the sum of the fixation durations beginning with the first revisit to a previously fixated object. In the principal analyses below, if there was no regression to a previously fixated object, the first-pass time was the sum of the gaze durations prior to the button push, and the second-pass time for these trials would be counted as zero. In other analyses, we excluded the trials on which there was no second pass; thus, the mean second-pass duration in those analyses is the mean duration of a second pass, given that there was a second pass. We also examined the number of gazes and the mean gaze duration for each of the two passes. Perhaps the coding scheme is best clarified by an example. Suppose the sequence of fixations on a trial was the following: display center, Object 1, Object 1, Object 2, Object 3, Object 1, Object 3, Object 3, and Object 1.³ The initial latency, as indicated earlier, would be the time between when the trial started and the initial saccade to Object 1. The first-pass time would be the sum of the durations of the first four fixations (i.e., before Object 1 is revisited), and the second-pass time would be the sum of the last four fixations. The gazes are best captured by bracketing the fixations as follows: [Object 1, Object 1], [Object 2], [Object 3], [Object 1], [Object 3, Object 3], and [Object 1].

This division of the trial into a first and a second pass is analogous to what is often employed when people read text while their eye movements are monitored (Rayner, 1998). In those studies, there are two types of first-pass times employed. First of all, the initial gaze duration on a word (not including regressive fixations back to the word) is often used as a first-pass measure, presumably indexing the initial encoding of the word. Second, first-pass time is also used on a region of text (e.g., a phrase or clause) in an analogous

manner; that is, this time is used as an index of initial encoding of the phrase, which excludes later processing, such as attempting to repair misconstruals of the meaning in the initial encoding (see Rayner, 1998, and Rayner & Pollatsek, 1992, for detailed explanations of the use of eye movement measures in reading and scene perception). Although the analogy with reading is not perfect, since there is no a priori forward direction through the scene, we thought it was a reasonable working hypothesis that the first-pass time would capture the initial encoding of the comparison scene and the second-pass time would capture processing subsequent to that.⁴

One motivation for examining gaze durations was to understand how *object based* the rotation or alignment operation is. For example, if rotation of the scene occurs either prior to comparison of the component objects or in parallel with the comparison, it should be largely irrelevant where fixations are during the process. Thus, one might expect that a pass time would increase with increasing difficulty of alignment but that the increase could be reflected in some mixture of increased numbers of gazes and in individual gaze durations (i.e., gazes would have no particular importance as measures). In contrast, if individual objects are being rotated or aligned, one might expect that alignment would be reflected mainly in increased gaze durations (as the participant attempts to align the object being gazed on with its counterpart in the standard array) and that the number of gazes might be relatively unaffected by increasing difficulty of alignment. (There might be some increase in number of gazes for more difficult alignments because of increased double-checking.)

Results

General modes of analysis. The design was not completely factorial, because there was only a single 0° rotation condition. In our primary analyses, rotation effects were assessed by computing the slope of the rotation function for each axis of rotation: the difference between the 70° rotation condition for that axis and the no-rotation (000) condition, divided by 70°. (For three equally spaced values, the linear trend test is a contrast between the extreme values.) Because the no-rotation condition was a common baseline, differences in the slope of rotation among the axes are identical to tests comparing the 70° rotation conditions. We also assessed rotation effects by comparing the 35° and the 70° rotation conditions; however, these are commented on only when they presented a different picture than the primary slope analysis.

All the eye movement indices were computed from correctly answered trials. Thus, 132 error trials (2.47% of all the trials) were excluded from the main analyses, as were 10 trials in which the calibration of the eye movements was off. On 0.4% of the trials, the buttonpress was prior to any eye movement, and these *no eye movement* trials were also excluded from the main analyses because they seemed qualitatively different; these will be discussed separately. (Interestingly, all the no eye movement trials were answered correctly.) Moreover, fixations that fell outside the borders of the screen image (600 × 800 pixels) were excluded from the eye movement analyses. The total number of excluded fixations was 700, which was 2.26% of the total number of fixations.

Since many of our analyses were object based, we did a preliminary analysis to check whether most of the fixations were, indeed, on the objects. The Euclidean dis-

tance between the center of an object⁵ and each fixation location was computed, and the average distance between a fixation location and the center of the closest object was 45 pixels. Since the objects were approximately 100×100 pixels on average, it is clear that most of the fixations were indeed on or quite close to the objects, especially since the objects occupied a relatively small minority of the total image area (see Figures 1 and 2). Thus, we are reasonably confident that our rule of assigning all fixations to one of the three objects indexed which object was being processed on a large majority of fixations. However, there was a *drift* over the course of a trial (caused mostly by small slippage of the headband) that did add error to the measurement of object location. To check whether such drift was likely to cause a significant number of misclassifications of which object was fixated, we conducted the following subsidiary analysis. At the end of each trial, the drift was computed as the amount that the fixation point had to be adjusted on the next trial. (This overestimated the amount of drift during the prior trial.) We then took each fixation on the trial and added the drift to the *x*- and *y*-coordinates of each fixation during the trial and computed which object these *adjusted* fixations were closest to. If the computation was different from the original, we counted this a *doubtful* fixation. Overall, fewer than 2% of the fixations were doubtful. In addition, we recomputed first- and second-pass times, excluding trials on which there were any doubtful fixations, and the values were virtually identical to the ones reported below. Thus, we feel quite confident that our division of the trial into three components is not a function of any eye movement recording artifact.

Response times. As can be seen in Table 1, the RTs in the *same* conditions increased significantly (as assessed by the slopes of their rotation effects) for all three major axes of rotation [$ts(20) > 3.1$, $ps < .05$].⁶ The pattern for *different* responses was somewhat different, in that there was little rotation effect in the *x*-axis rotation condition and a somewhat larger rotation effect in the *z*-axis condition [see Table 1; $t < 1$, $t(20) = 4.19$, $p < .001$, and $t(20) = 2.92$, $p < .01$, for the *x*-, *y*-, and *z*-axis conditions, respectively]. However, as was indicated above, the major interest was in partitioning the total time in the trial into three major components: (1) the ini-

tial latency, the time from the onset of the comparison scene until the first eye movement; (2) the first-pass time, the sum of the fixation times before the first regression back to a previously fixated object; and (3) the second-pass time, the sum of the fixation durations after that regression until the response was made. In our primary analyses, if there were no regressions back to a prior object, second-pass time was counted as a zero.

Components of processing times. Since the pattern for the initial latency was the same for *same* and *different* trials, we will discuss them both here. The pattern was not related in any simple way to rotation angle. Instead, there was a difference between the conditions in which there was no rotation in depth (the 000, Z35, and Z70 conditions), which had a mean latency of 276 msec, and the conditions in which there was a rotation in depth (the X35, X70, Y35, and Y70 conditions), which had a mean latency of 216 msec (see Table 2). A one-way *F* test for the *z*-axis rotation conditions showed no significant difference between the Z35 and Z70 (picture plane rotation) conditions and the 000 (no-rotation) condition ($F < 1$), whereas the latency in the 000 condition was longer than those in the *x*- and *y*-axis rotation conditions [$F(2,40) = 19.8$, $p < .001$, and $F(2,40) = 39.5$, $p < .001$, respectively]. Post hoc pairwise contrasts showed that the latency in the 000 condition was longer than those in the X35, X70, Y35, and Y70 conditions [$ts(20) > 4.19$, $ps < .05$, adjusted using the Bonferroni method]. For both of the *same* and *different* conditions, the initial latencies in the 000, Z35, and Z70 conditions were longer than those in the rest of the conditions, and there was no significant difference among the 000, Z35, and Z70 conditions [$F(2,40) = 1.46$, $p > .1$, and $F < 1$, for *same* and *different* conditions, respectively]. Moreover, when the initial latency in the shortest condition of the three above was compared with that in the X35, X70, Y35, and Y70 conditions, the difference was significant for all comparisons [$ts(20) > 2.95$, $ps < .02$, and $ts(20) > 3.29$, $ps < .02$ for *same* and *different* conditions, respectively, each adjusted using the Bonferroni method]. The data thus indicated that the eyes stayed longer at the center of the display when the comparison stimulus did not involve a rotation in depth of the standard scene and was not a function of rotation angle.

Table 1
Response Times (in Milliseconds) and Error Rates (in Percentages)
in the Same-Scene and Different-Scene Conditions in Experiment 1

Response	Axis of Rotation	Response Time			Slope (msec/deg)	Error Rate			Slope per Degree
		0° Rot.	35° Rot.	70° Rot.		0° Rot.	35° Rot.	70° Rot.	
Same	<i>x</i>	1,913	2,180	2,219	4.37**	0.89	0.89	1.79	0.01
	<i>y</i>	1,913	2,325	2,285	5.31***	0.89	3.27	2.08	0.02
	<i>z</i>	1,913	1,998	2,188	3.92**	0.89	1.49	0.60	0.00
Different	<i>x</i>	1,961	1,855	2,025	0.90	2.68	0.89	3.87	0.02
	<i>y</i>	1,961	1,931	2,276	4.49***	2.68	3.27	5.36	0.04*
	<i>z</i>	1,961	2,034	2,375	5.91***	2.68	2.38	3.27	0.01

Note—The 0° rotation conditions were the same for all axes of rotations. Paired *t* tests were performed between the 0° and 70° conditions. * $p < .1$. ** $p < .05$. *** $p < .01$. Probabilities were adjusted using the Bonferroni method.

Table 2
Initial Latency, First-Pass Time, and Second-Pass Time (in Milliseconds) for the Same-Scene and Different-Scene Conditions in Experiment 1

Phase of Processing	Response	Axis of Rotation	Response Time			Slope (msec/deg)
			0° Rot.	35° Rot.	70° Rot.	
Initial latency	Same	<i>x</i>	272	220	212	—
		<i>y</i>	272	227	207	—
		<i>z</i>	272	284	269	—
	Different	<i>x</i>	280	222	217	—
		<i>y</i>	280	228	200	—
		<i>z</i>	280	275	277	—
First pass	Same	<i>x</i>	906	947	915	0.13
		<i>y</i>	906	887	883	−0.33
		<i>z</i>	906	914	945	0.55
	Different	<i>x</i>	932	935	937	0.07
		<i>y</i>	932	893	912	0.29
		<i>z</i>	932	973	1,029	1.39*
Second pass	Same	<i>x</i>	486	673	695	2.99**
		<i>y</i>	486	856	825	4.84***
		<i>z</i>	486	532	690	2.91**
	Different	<i>x</i>	548	407	502	−0.66
		<i>y</i>	548	516	832	4.06***
		<i>z</i>	548	534	783	3.36***

* $p < .1$. ** $p < .05$. *** $p < .01$. Probabilities were adjusted using the Bonferroni method.

Perhaps the most striking and unexpected result was that the first-pass time did not appear to be a function of rotation angle, except for *z*-axis rotations (see Table 2), despite the fact that the first-pass time was almost a second, so that the time when the second pass began was well over a second after the comparison stimulus had been presented. In contrast, there was a clear effect of rotation angle on second-pass time for all three axes in the *same* condition, but only for the *y*- and *z*-axis rotation conditions for the *different* conditions. For the *same* responses, the intercept for first-pass time was 906 msec, and the slopes were not significantly different from zero for any axis; for the *different* responses, the intercept for first-pass time was 932 msec, and only the slope for the *z*-axis rotation was significant (see Table 2). As was indicated above, second-pass time can be computed several different ways. The one most consistent with “book-keeping” for the total RT is to count second-pass time as zero when there is no second pass (then the total trial time equals the sum of initial latency, first-pass time, and

second-pass time). Using this measure of second-pass time, the intercept for *same* trials was 486 msec (with significant slopes for all three axes), the intercept for *different* trials was 548 msec, and the slope was significant for *y*- and *z*-axis rotations (see Table 2).

The analyses above thus indicate that, with the exception of *z*-axis rotations, there is little or no rotation effect until the end of the first pass. In contrast, the second-pass times showed significant rotation effects, with the exception of the *x*-axis *different* condition. We next examined the portion of the trial after the end of the first pass, to determine whether the differences above in mean second-pass duration were due to (1) there being differences in the probability of a second pass (i.e., the probability of a *revisit* to a previously fixated object), (2) longer second-pass times, given that a revisit to a previously fixated object was made, or (3) both. For *same* trials, the probability of a revisit in the no-rotation condition was .62, and for *different* trials, the probability of a revisit in the no-rotation condition was .56. This probability increased signifi-

Table 3
Probability of a Second Pass and Second-Pass Time Conditional on There Being a Second Pass in Experiment 1

Response	Axis of Rotation	Probability of a Second Pass				Second-Pass Time Conditional on a Second Pass			
		Probability			Slope per Degree	Second-Pass Time (msec)			Slope (msec/deg)
		0° Rot.	35° Rot.	70° Rot.		0° Rot.	35° Rot.	70° Rot.	
Same	<i>x</i>	.62	.74	.73	0.002	713	871	896	2.62*
	<i>y</i>	.62	.84	.86	0.003***	713	971	942	3.27**
	<i>z</i>	.62	.69	.71	0.001*	713	711	920	2.97**
Different	<i>x</i>	.56	.49	.53	−0.000	861	661	953	1.31
	<i>y</i>	.56	.61	.72	0.002***	861	827	1,157	4.23**
	<i>z</i>	.56	.53	.58	0.000	861	868	1,387	7.51***

* $p < .1$. ** $p < .05$. *** $p < .01$. Probabilities were adjusted using the Bonferroni method.

cantly for *y*-axis rotation conditions (see Table 3). It thus appears that, even though the time of first-pass processing is largely unaffected by rotation angle (except for *z*-axis rotations), the probability that a judgment could be made on the basis of the first pass was affected by it. However, the second-pass time effects reported in Table 2 were not merely due to the probability of there being a second pass. For the *same* trials and the *different* trials on which there was a revisit, the second-pass times in the no-rotation conditions were 713 and 861 msec, and there were significant rotation effects for the *y*- and *z*-axis rotation conditions (see Table 3). Thus, the overall rotation effects in second-pass time reported in the paragraph above were due both to a greater probability of there being a second pass when the comparison scene was rotated and to a longer second-pass processing time, given that a second pass was embarked on. (The *x*-axis rotation effects appeared to be due mainly to increases in the conditional second-pass times, but the effect was only marginally significant.)

Gaze measures. A finer analysis of the eye movement record can be carried out at the object level: Consecutive fixations are grouped as *gazes* when they are on the same object. These analyses allow one to determine whether increases in time during a component are due to a greater number of gazes (i.e., objects visited more often), to individual gaze durations being lengthened, or to both. As with the overall first-pass analyses, there was little rotation effect either in number of gazes or in mean gaze duration for the first pass, except for *z*-axis rotations. For both *same* and *different* trials, there was only a sizeable increase in the number of gazes for *x*-axis rotations, and a rotation effect on the mean gaze duration only for *z*-axis rotations in the *different* trials (see Table 4). (In fact, the mean gaze duration decreased with increasing rotation for the *x*-axis rotations, which appeared to trade off with the number of gazes.) For the second pass, conditional on there being a second pass, there was a significant rotation effect on the number of gazes only for *y*-axis rotation *same* trials and *z*-axis *different* trials. For

mean second-pass gaze duration, conditional on a second pass, there were significant rotation effects only for *x*- and *z*-axis rotations (see Table 4). Thus, there was a suggestion that the second-pass rotation effect may have been different for *y*-axis rotations than for *x*- and *z*-axis rotations.

To summarize, the initial latency was not a function of the angle of rotation; instead, there was a two-group pattern—the 000, Z35, and Z70 conditions versus the other conditions. First-pass time was affected by the angle of rotation only for the *z*-axis (picture plane) rotation condition, whereas second-pass time was affected by the angle of rotation for all rotation conditions except the *x*-axis different condition.⁷ This indicates that the time course of the alignment process for rotations in depth is different from that in the picture plane. Furthermore, the data indicated that second-pass times increased both because the probability of a second pass increased and because the time spent in the second pass (conditional on their being a second pass) increased as well—generally, both because there were more gazes and because mean gaze durations increased.

Location change and orientation change trials compared. For ease of exposition, the analyses above were collapsed over the various types of *different* trials. There were some differences of interest among these conditions worth noting. First, as might be expected, there were more errors in the Orientation 1 condition (10.20%, averaged over the seven viewpoints) than in the Location 2, Location 3, and Orientation 3 conditions (1.36%, 0.34%, and 0.51%, respectively), as indicated by a significant interaction between location versus orientation change and the size of change [i.e., Location 2 and Orientation 1 vs. Location 3 and Orientation 3; $F(1,20) = 27.2, p < .001$]. Second, the rotation effects in the RTs were more pronounced for the orientation change conditions. For the location change conditions, the only significant positive slope was for *z*-axis rotations in the Location 3 condition [$t(20) = 4.77, p < .01$], whereas there were significant rotation effects for all of the orientation change

Table 4
Measures of Individual Gazes on Objects in Experiment 1

		Axis of Rotation	Mean Number of Gazes			Slope per Degree	Mean Gaze Duration (msec)			Slope (msec/deg)
Response			0° Rot.	35° Rot.	70° Rot.		0° Rot.	35° Rot.	70° Rot.	
First pass	Same	<i>x</i>	2.55	2.63	2.69	0.002	365	359	355	−0.15
		<i>y</i>	2.55	2.51	2.46	−0.001	365	360	358	−0.10
		<i>z</i>	2.55	2.62	2.60	0.001	365	347	364	−0.01
	Different	<i>x</i>	2.26	2.43	2.41	0.002*	414	416	414	0.01
		<i>y</i>	2.26	2.35	2.28	0.000	414	401	408	−0.09
		<i>z</i>	2.26	2.38	2.23	−0.000	414	406	479	0.93***
Second pass	Same	<i>x</i>	2.16	2.39	2.23	0.001	298	329	443	2.07***
		<i>y</i>	2.16	2.64	2.60	0.006***	298	335	329	0.45
		<i>z</i>	2.16	2.13	2.43	0.004	298	319	340	0.67**
	Different	<i>x</i>	2.25	1.91	2.24	−0.000	361	311	389	0.40
		<i>y</i>	2.25	2.26	2.50	0.004	361	356	451	1.29
		<i>z</i>	2.25	2.25	2.85	0.009**	361	347	442	1.16*

* $p < .1$. ** $p < .05$. *** $p < .01$. Probabilities were adjusted using the Bonferroni method.

conditions [$t(20) > 2.52$, $ps < .06$],⁸ except for the y -axis rotations in the Orientation 1 condition⁹ and x -axis rotations in the Orientation 3 condition ($ts < 1.1$). There were also differences between the two conditions for the first- and second-pass times. The first-pass times for the Location Change 2 and 3 conditions (926 and 916 msec) were about the same as that for the *same* condition (913 msec; $F < 1$), but the first-pass times for the orientation change conditions for the Orientation 1 and Orientation 3 conditions (991 and 945 msec) were longer than those in the location change conditions [$F(1,20) = 5.40$, $p < .05$, for the main effect of the orientation vs. location change conditions; $F(1,20) = 1.54$, $p > .10$, for the interaction of orientation location and size of the changes]. This suggests that at least a tentative identification of orientation changes occurred during the first pass of processing and that this process was time consuming.

We also compared the probability of a revisit (i.e., that there was a second pass) for the *same* trials and the various types of *different* trials. Averaged over the seven viewpoints, the revisit probability for *same* trials (.74) was significantly higher than that for any of the *different* conditions (.57, .57, .57, and .57 for the Location 1, Location 3, Orientation 1, and Orientation 3 conditions, respectively; $ts > 3.15$, $ps < .05$, for comparisons between the *same* condition and the respective *different* conditions]. As is shown in Table 3, the rotation effect on the probability of a revisit was seen mostly in the y -axis rotation condition. When *different* trials were analyzed separately for location change and orientation change conditions, the rotation effect was significant only in the Location 3 condition [$t(20) = 2.62$, $p < .05$] and in the Orientation 3 condition [$t(20) = 5.89$, $p < .01$, for the y rotations; probabilities were adjusted using the Bonferroni method].

It is thus clear that detection of a difference in the first pass can be certain enough to produce a response. However, it is equally clear that interpretation of the components is not nearly as simple as the following: Processing in the first component is looking for a location change and processing in the second component is looking for an orientation change. If this had been the case, (1) the revisit probabilities should have been markedly lower for location changes than for orientation changes, and (2) the probability of a revisit should have been close to 1 in the *same* and the orientation change conditions (since there would have had to have been a second component to re-examine the array for an orientation change). Instead, it appears that, on a significant fraction of the trials, a definite determination could be made during the first pass of the existence or nonexistence of a change for both types of changes. Thus, a significant part of the comparison process occurred in the first pass, even though there was no evidence of a rotation effect on the time to complete the first pass for the rotations in depth.

Use of information from parafoveal vision. Object identification studies (Henderson, 1997; Henderson & Anes, 1994; Henderson, Pollatsek, & Rayner, 1987; Meyer & Dobel, 2003; Pollatsek, Rayner, & Collins, 1984) have

indicated that parafoveal vision facilitates foveal processing of objects. In the scene rotation task, parafoveal vision might also facilitate foveal processing by allowing a preliminary judgment about where there is a change and then guiding the eyes to an object that has possibly changed. If so, a changed object should be fixated earlier than an unchanged object. To test this speculation, the number of gazes before a changed object was fixated was recorded (1) in the Orientation 1 condition and (2) in the Location 2 condition. The number of gazes in (1) was compared with the number of gazes in the *same* condition (for the same scene) until the object that had been changed in the Orientation 1 was fixated, and the number of gazes in (2) was compared with the number of gazes in the *same* condition (for the same scene) until either of the objects changed in the Location 2 condition was fixated. For example, in the Orientation 1 condition, if the mug was changed in the X35 viewpoint, the number of gazes until the mug in the X35 *same* condition was fixated was used as the baseline.

Averaged over all viewpoints, it took 1.31 gazes to reach one of the changed objects in the Location 2 condition, which was smaller than its baseline of 1.48 gazes [$t(20) = 5.10$, $p < .001$]. In contrast, it took 2.28 gazes to reach the changed object in the Orientation 1 condition, which was actually slightly more than the baseline value of 2.18 gazes [$t(20) = -1.43$, $p < .20$]. We suspect that the latter effect was merely a Type I error (it was not replicated in Experiment 2); the former effect indicates that parafoveal vision helped to guide the eyes to the changed object in the location change conditions.

No eye movement trials and error trials. There were 25 trials in which the participant did not make an eye movement. Seventeen were made by 2 participants, so that the no eye movement trials might be a result of the strategies of only a few individuals. Of greatest interest was that the responses for all these trials were correct. The overall mean RT for these trials was 1,191 msec. Twenty-two out of the 25 occurred in the 000, Z35, and Z70 viewpoint conditions, and 21 out of the 25 were *different* trials, mainly Location 3 and Orientation 3 trials. These results seem to indicate that the scene rotation task can be solved without eye movements, especially if (1) the comparison stimulus is identical or a rotation in the picture plane and (2) the change itself was large (i.e., all three objects were changed).

Out of 132 error trials, 41 were made in the *same* condition, 9 in the Location 2 condition, 3 in the Location 3 condition, 76 in the Orientation 1 condition, and 3 in the Orientation 3 condition. To examine the difference in eye movements between the error trials and the correct trials, the error trials in the Orientation 1 condition were analyzed. Although the Orientation 1 condition had the largest number of error trials, since the number of samples in each condition varied considerably, indices were averaged over all viewpoints, and no statistical tests were performed.

The largest difference was in the first-pass time. The first-pass time in the error trials (907 msec) was 116 msec shorter than that in the correct trials (1,023 msec), whereas

the second-pass time in the error trials (593 msec) was only 44 msec shorter than that in the correct trials (627 msec). The number of gazes in the error trials (4.06) was actually slightly larger than that in the correct trials (3.78). Also, all the objects were more likely to be visited in the error trials than in the correct trials. The probability that all the objects were visited was .74 in the error trials and .66 in the correct trials, and the changed object was only slightly less likely to be visited in the error trials than in the correct trials (.93 vs. .98). The results suggested that errors were caused primarily by the eyes not spending enough time during the first pass, rather than by a premature termination during the second pass.

Summary. Our RT measures indicated that there were rotation effects for all three axes. When this time was partitioned into the three components (initial latency, first-pass time, and second-pass time), using the eye movement data, however, there appeared to be differences in the rotation effects depending on the axis of rotation. The first component prior to an eye movement (which produced a response only on 25 trials) took less time for rotations in depth than rotations in the plane or when there was no rotation. We do not have a good explanation for this. Most strikingly, the next component, the first-pass time, was significantly affected by rotation angle only for *z*-axis rotations (i.e., rotations in the plane). Moreover, this apparent rotation effect may not have been due to aligning the comparison stimulus to the standard, but merely to the additional time needed to encode the comparison stimulus because it was in an unusual orientation. Nonetheless, roughly one quarter of the *same* judgments and one third of the *different* judgments were completed after the first component. The data thus indicated (with the possible exception of *z*-axis rotations) that the time to make a judgment in this component is unaffected by rotation angle, although the probability of successful completion was significantly affected by rotation angle for *same* and orientation change conditions. Instead, the increase in RT to make a judgment for rotations in depth was due both to an increased probability that there was a second component (especially for *same* trials) and to increased times in the second component. As a result, the rotation effect for *x*- and *y*-axis rotations could be interpreted as due both to increasing the probability of double-checking and to increasing the time that this component takes. If so, this is quite a different picture of mental rotation than that inferred from the data with single objects, where it is assumed that, after some brief initial encoding, an analogue rotation process occurs on virtually all trials.

EXPERIMENT 2

The results of Experiment 1 suggested that a time-consuming mental rotation process does not begin for the *x*- and *y*-axis rotations until the onset of the second pass—over a second after the presentation of the comparison stimulus. It is possible that a reasonably complete 3-D

interpretation of a 2-D image may take this long to emerge (e.g., Sekuler & Palmer, 1992), especially for the detailed representation of objects and how they relate to the scene as a whole. If such high-level information becomes fully available only by the time the second-pass process starts, altering certain aspects of the scene that affect the ability to construct a 3-D representation of it may have an effect only on second-component processing. One such aspect is the presence of the background (in this case, the table top). For example, Aks and Enns (1996) reported that removing background information, such as the texture gradient and local frames, affected the 3-D interpretation of a 2-D multiobject array. In the scene rotation task, Nakatani and Pollatsek (2001) reported that the presence of the desktop was critical for *same-different* judgments for rotation in depth. In the study, the components of a square desktop (the edges and the hard surface of the desk) were removed systematically, and the results showed that the desktop, especially the four edges of the desk, were critical for a correct judgment; without the edges, the same scene was wrongly judged *different*, particularly in the *y*-axis rotation condition.

In Experiment 2, the nature of the processing during the first pass and second pass was probed. On the basis of the results of Nakatani and Pollatsek (2001), the presence/absence of the desktop was employed to manipulate the availability of the 3-D structure of the scene reconstructed from a 2-D image. Experiment 2 was essentially a replication of Experiment 1, with one new manipulation: On half the trials, the three objects were presented as in Experiment 1 (the *with-desk* condition), whereas on the other half of the trials, there was no desktop, and the three objects were presented against a uniform black background (the *no-desk* condition). If the processing during the second pass is merely a reworking of the processing during the first pass, removal of the desktop should affect first-pass and second-pass processing similarly. On the other hand, if only the processing during the second pass is crucially dependent on utilizing this high-level 3-D information, only the second pass should be affected by the absence of the desktop, and perhaps only for rotations in depth.

Method

Participants. Eighteen members of the University of Massachusetts community participated in the experiment. All had normal or corrected-to-normal vision. They received either \$8 or experimental credits in psychology courses for participating.

Stimuli and Design. Two versions of computer-generated images, scenes with the desktop and scenes without the desktop, were used in Experiment 2. The images in the *with-desk* condition were identical to those in Experiment 1, and the images in the *no-desk* condition were identical to those in Experiment 1 except that the desktop was eliminated and the objects were placed against a uniform black background (for both standard and comparison stimuli). The five stimulus types (*same*, Location 2, Location 3, Orientation 1, and Orientation 3) and the seven viewpoints (000, X35, X70, Y35, Y70, Z35, and Z70) were the same as those in Experiment 1.¹⁰

Procedure. The procedure in Experiment 2 was the same as that in Experiment 1, except that the experimental session was split into 2 days. On the 1st day, the participants were given 32 practice trials with feedback, followed by four blocks of 64 trials without feedback.

On the 2nd day, there were four more blocks of 64 trials without feedback. As in Experiment 1, there were four standard scenes, each with three office objects, and one standard scene was used for each block. On each day, two of the blocks were the no-desk condition, and the other two were the with-desk condition. The order of the no-desk and with-desk conditions was counterbalanced (e.g., if a participant was given Scene 1 and Scene 2 in the no-desk condition and Scene 3 and Scene 4 in the with-desk condition on the 1st day, he or she had Scene 1 and Scene 2 in the with-desk condition and Scene 3 and Scene 4 in the no-desk condition on the 2nd day). Half of the participants started with the no-desk condition and then moved on to the with-desk condition on Day 1, with the order of the no-desk and with-desk conditions reversed on Day 2, and the order was reversed for the other half of the participants. The mean interval between the two sessions was 5.22 days. The eye movements during the trials were recorded by the EyeLink system as in Experiment 1.

Results

Excluded data. One of the participants reported that she deliberately tried not to move her eyes; her data were excluded from the analyses. As in Experiment 1, error and no eye movement trials were excluded from computations of the RTs and the eye movement indices. There were 155 error trials (3.56%) and 65 no eye movement trials (1.49%) in the with-desk condition and 273 error trials (6.30%) and 34 no eye movement trials (0.78%) in the no-desk condition. Fixations outside of the display—320 (1.24%) in the with-desk condition and 306 (1.10%) in the no-desk condition—were also excluded from the computation of eye movement indices.

Response times and errors. As can be seen in Table 5, RTs and errors in the with-desk condition were quite similar to those in Experiment 1. Moreover, perhaps surprisingly, the RTs averaged over *same* and *different* responses were only 86 msec greater in the no-desk condition than in the with-desk condition [$F(1,16) = 1.98$, $p > .10$]. Furthermore, virtually the whole difference was in the y -axis rotation conditions. Both the interaction of presence/absence of the desktop and axis of rotation

[$F(2,32) = 4.61$, $p < .002$] and the 300-msec difference between the no-desk and the with-desk conditions in the y -axis rotation conditions [$F(1,16) = 8.07$, $p < .02$] were significant. The error data showed a similar pattern: The error rate in the no-desk condition was significantly greater than that in the with-desk condition for the y -axis rotations [$F(1,16) = 10.24$, $p < .001$], but not for either the x - or the z -axis rotations [$F(1,16) = 2.62$, $p > .10$, and $F < 1$, respectively]. Thus, both RTs and errors indicated that processing in the y -axis rotation condition was more difficult when the desktop was removed, whereas the x - and z -axis rotation conditions were hardly affected by the absence of the desktop. Since the absence of the desktop had so little effect (except for y -axis rotations), we will not comment on desktop effects below unless they were significant. (In fact, most of the tests of either the main effect of desktop or an interaction of a rotation effect with desktop had an F less than 1.)

For *same* trials, the slopes for the overall RTs were fairly similar to those in Experiment 1, and there was an appreciable difference in slope between the with-desk and the no-desk conditions only for the y -axis (see Table 5). However, even the y -axis slope was not significantly affected by the absence of the desktop [$t(16) = 1.39$, $p > .10$]. There were also significant rotation effects on the error rates for x - and y -axis rotations, and the interaction of background with the y -axis rotation effect was obviously large [$t(16) = 3.31$, $p < .01$]. These high error rates for the *same* trials in the y -axis conditions are anomalous, since the error rates were higher for the *different* trials in virtually all the other conditions. For *different* trials, the RT slopes were somewhat smaller than those for the *same* trials. They were quite similar for the with-desk and the no-desk conditions and were not that different from those in Experiment 1 (see Table 5). The rotation effects on errors for *different* trials were all quite small, and there were no significant interactions of background

Table 5
Response Times (in Milliseconds) and Error Rates in the
Same-Scene and Different-Scene Conditions in Experiment 2

Same Scene and Different Scene Conditions in Experiment 2									
Response	Axis of Rotation	Response Time			Slope (msec/deg)	Error Rate			Slope per Degree
		0° Rot.	35° Rot.	70° Rot.		0° Rot.	35° Rot.	70° Rot.	
With-Desk Condition									
Same	<i>x</i>	1,860	1,983	2,308	6.41**	0.37	1.84	3.68	0.05
	<i>y</i>	1,860	2,222	2,352	7.04**	0.37	1.84	2.21	0.03
	<i>z</i>	1,860	2,038	2,152	4.17**	0.37	0.74	1.47	0.02
Different	<i>x</i>	1,914	1,780	2,097	2.61	4.41	4.41	5.15	0.01
	<i>y</i>	1,914	1,954	2,245	4.73**	4.41	3.68	5.15	0.01
	<i>z</i>	1,914	2,096	2,177	3.76*	4.41	6.25	6.62	0.03
No-Desk Condition									
Same	<i>x</i>	1,853	1,907	2,331	6.82*	1.10	2.57	6.25	0.07*
	<i>y</i>	1,853	2,552	2,559	10.09***	1.10	23.90	14.71	0.19***
	<i>z</i>	1,853	1,893	2,199	4.94**	1.10	1.10	2.57	0.02
Different	<i>x</i>	2,051	1,902	2,168	1.67	5.15	4.04	8.46	0.05*
	<i>y</i>	2,051	2,163	2,501	6.42***	5.15	2.57	5.51	0.01
	<i>z</i>	2,051	2,028	2,273	3.16**	5.15	6.25	5.58	0.01

* $p < .1$. ** $p < .05$. *** $p < .01$. Probabilities were adjusted using the Bonferroni method.

with these effects ($ps > .10$). Thus, the results in both experiments were quite consistent, with strong rotation effects for the y - and z -axes for both *same* and *different* trials and a significant rotation effect for x -axis *same* trials and a nonsignificant rotation effect for x -axis *different* trials.

Components of processing times. As in Experiment 1, the initial latency was not consistent with a mental rotation effect, since it was longer in the 000 (no-rotation) condition and the z -axis rotation conditions than in the x - and y -axis rotation conditions for both the with-desk condition [$ts(16) > 3.07, ps < .05$] and the no-desk condition [$ts(16) > 2.66, ps < .05$]. However, averaged over the seven viewpoints, the average initial latency in the no-desk condition (240 msec) was actually 16 msec less than that in the with-desk condition [256 msec; $t(16) = 2.55, p < .025$], suggesting that it was advantageous to start processing object-specific information earlier in the no-desk condition, perhaps because global features were absent (see Table 6).

As in Experiment 1, the first-pass times showed virtually no rotation effect for either the no-desk or the with-desk conditions, except for the z -axis rotations (see Table 6). Moreover, the presence/absence of the desktop had little effect on first-pass times. The only significant effect was that the rotation effect in first-pass time for z -axis rotations was significant averaged over *same* and *different* trials [$t(16) = 2.44, p < .05$]. The absence of the desktop had little effect on first-pass times [the average first-pass time over all viewpoints was actually 37 msec less in the no-desk condition than in the with-desk condition; $t(16) = 1.50, p > .1$], and all interactions between desktop and slopes of rotation had $ts < 1$.

In contrast, most of the rotation effects were significant in second-pass times for the analysis counting second-pass time as zero when there was no second pass (see Table 6). The z -axis rotation for *different* trials was also significant when averaged over the with-desk and the no-desk conditions [$F(1,16) = 18.25, p < .01$]. Moreover, when the second-pass times for the seven viewpoints were averaged over *same* and *different* trials, the average second-pass time in the no-desk condition (793 msec) was 120 msec longer than that in the with-desk condition [673 msec; $F(1,16) = 6.10, p < .05$], but the interaction of that effect with *same* and *different* was not close to significant ($F < 1$). The absence of the desktop appeared to influence the rotation effect only for the y -axis rotations [$F(1,16) = 6.55, p < .05$, for the main effect of desktop on slopes averaged over *same* and *different* trials], since there was virtually no effect of the desktop on x - and z -axis rotations ($Fs < 1$, for the main effect in x - and z -axis rotations).

As in Experiment 1, we analyzed processing after the first pass to determine how much of the effects on second-pass times were due to differing probabilities of a second pass and how much were due to longer second-pass times, given that there was a second pass. There were large effects of rotation angle on the probability of a second pass (i.e., a revisit to a previously fixated object) for *same* trials for all three axes, but for *different* trials the increases were significant only for y -axis rotations (see Table 7). The probability of a revisit increased when the desktop was removed. For *same* trials, the average probability over the seven viewpoints was .69 in the with-desk condition and .72 in the no-desk condition [$t(16) = 2.09, p < .10$], and for *different* trials, the probabilities

Table 6
Initial Latency, First-Pass Time, and Second-Pass Time (in Milliseconds)
for the Same-Scene and Different-Scene Conditions in Experiment 2

Phase of Processing	Response	Axis of Rotation	With-Desk Condition				No-Desk Condition			
			Response Time			Slope (msec/deg)	Response Time			Slope (msec/deg)
			0° Rot.	35° Rot.	70° Rot.		0° Rot.	35° Rot.	70° Rot.	
Initial latency	Same	x	279	240	239	—	277	234	236	—
		y	279	255	218	—	277	188	183	—
		z	279	297	267	—	277	302	274	—
	Different	x	277	238	251	—	279	240	226	—
		y	277	246	212	—	279	234	189	—
		z	277	289	277	—	279	251	248	—
First pass	Same	x	857	805	859	0.02	842	790	791	−0.73
		y	857	803	858	0.02	842	774	858	0.22
		z	857	883	906	0.71	842	826	867	0.36
	Different	x	875	847	891	0.22	864	842	852	−0.17
		y	875	878	876	0.01	864	797	860	−0.05
		z	875	909	938	0.89	864	834	921	0.83
Second pass	Same	x	490	631	839	4.98**	500	594	917	5.95**
		y	490	845	899	5.83**	500	1,224	1,121	8.87***
		z	490	612	714	3.20**	500	537	804	4.35**
	Different	x	549	436	627	1.12	667	549	751	1.20
		y	549	580	824	3.92*	667	831	1,075	5.83***
		z	549	655	719	2.43	667	688	839	2.45

* $p < .1$. ** $p < .05$. *** $p < .01$. Probabilities were adjusted using the Bonferroni method.

Table 7
Probability of a Second Pass and Second-Pass Time Conditional on There Being a Second Pass in Experiment 2

Response	Axis of Rotation	Probability			Slope per Degree	Second-Pass Time (msec)			Slope (msec/deg)
		0° Rot.	35° Rot.	70° Rot.		0° Rot.	35° Rot.	70° Rot.	
With-Desk Condition									
Same	<i>x</i>	.57	.67	.71	0.002***	832	835	1,060	3.26
	<i>y</i>	.57	.75	.81	0.003***	832	1,007	1,037	2.91
	<i>z</i>	.57	.62	.72	0.002***	832	815	903	1.01
Different	<i>x</i>	.50	.57	.57	0.001	939	801	951	0.17
	<i>y</i>	.50	.65	.66	0.002**	939	920	1,141	2.87
	<i>z</i>	.50	.57	.59	0.001	939	1,096	1,317	5.39*
No-Desk Condition									
Same	<i>x</i>	.58	.69	.73	0.002***	817	766	1,144	4.69
	<i>y</i>	.58	.88	.87	0.004***	817	1,479	1,337	7.43†
	<i>z</i>	.58	.64	.72	0.002***	817	745	966	2.13
Different	<i>x</i>	.64	.62	.69	0.001	959	775	955	0.05
	<i>y</i>	.64	.71	.79	0.002***	959	1,136	1,291	4.74***
	<i>z</i>	.64	.60	.69	0.001	959	1,064	1,089	1.86

* $p < .1$. ** $p < .05$. *** $p < .01$. Probabilities were adjusted using the Bonferroni method. †The large slope was due to an outlier in the Y70 condition.

were .60 in the with-desk condition and .70 in the no-desk condition [$t(16) = 3.65$, $p < .01$]. There were also increases in the second-pass time, conditional on there being a second pass, although some of the effects, although large, were not significant. There was virtually no effect of the presence/absence of the desktop on these conditional second-pass times (the p values in all tests of main effects and interactions were greater than .10).

To summarize, the data on pass times replicated those in Experiment 1 quite well, since there were virtually no rotation effects on first-pass times, except for z -axis rotations, and fairly healthy rotation effects on second-pass times, with the exception of the x -axis *different* condition. Moreover, the desktop had surprisingly little effect on rotation effects on either first- or second-pass times, with the exception of y -axis rotations. Furthermore, as in Experiment 1, when these second-pass times were broken down into the probability of a revisit's occurring and second-pass time, given that a revisit had occurred, there generally appeared to be effects on both measures. Moreover, the effect of absence of the desktop on second-pass time for y -axis rotations appeared to be almost completely due to increasing the probability of a second pass; the absence of the desktop also increased the second-pass time, conditional on the second pass, in *different* trials as well.

Gaze measures. In the first pass, there was the same strange pattern for x -axis rotations as in Experiment 1: With increasing rotation, there was an increase in the number of gazes but a decrease in the mean gaze duration. There was also a significant increase in the number of gazes for z -axis rotations in the with-desk condition (see Table 8). The increase in first-pass mean gaze duration for z -axis rotations observed in Experiment 1 was virtually completely absent in the with-desk condition. As was indicated above, the effects on the second-pass time, conditional on there being a second pass, were not

particularly large. As a result, it is not that surprising that when this effect is broken down into number of gazes and mean gaze duration, the effects are not particularly reliable. The mean number of gazes in the second pass (conditional on there being a second pass) increased significantly only for y -axis rotations in the no-desktop condition; however, there was also a sizable increase in the y -axis desktop condition, which was actually slightly larger than that in Experiment 1. There was a significant effect of desktop on the y -axis rotation effect [$F(1,16) = 6.59$, $p < .05$, for the main effect of desktop in slopes averaged over *same* and *different* trials; see Table 8]. For mean gaze durations, conditional on there being a second pass, no rotation effects were close to significant (all $ps > .10$), although the pattern mirrored that in Experiment 1, with generally larger rotation effects for x - and z -axis rotations than for y -axis rotations. There were also no significant interactions with the desktop manipulation (all $ps > .10$).

Location change and orientation change trials compared. As in Experiment 1, the mental rotation effect was larger in the orientation change condition than in the location change condition. The RT slopes in the location change condition ranged from 1 to 3 msec/deg, whereas those in the orientation change condition ranged from 5 to 10 msec/deg [$F(1,16) = 5.64$, $p = .03$]. This was not due to a speed-accuracy tradeoff, since the error slopes in the location change and the orientation change conditions were about the same ($F < 1$).

The effect of the removal of the desktop was larger in the location change conditions than in the orientation change conditions, but the difference was clear only in the error rates. For the location change conditions, the mean RT in the no-desk condition was 166 msec more than that in the with-desk condition [averaged over the seven viewpoints of the Location 2 and Location 3 conditions; $t(16) = 1.35$, $p > .1$], and in the orientation

Table 8
Measures of Individual Gazes on Objects in Experiment 2

		Axis of Rotation	Mean Number of Gazes			Slope per Degree	Gaze Duration			Slope (msec/deg)
Response			0° Rot.	35° Rot.	70° Rot.		0° Rot.	35° Rot.	70° Rot.	
With-Desk Condition										
First pass	Same	<i>x</i>	2.38	2.54	2.53	0.002*	374	328	341	−0.48
		<i>y</i>	2.38	2.34	2.42	0.001	374	356	358	−0.23
		<i>z</i>	2.38	2.47	2.53	0.002**	374	371	367	−0.06
	Different	<i>x</i>	2.15	2.31	2.35	0.003***	424	374	387	−0.53*
		<i>y</i>	2.15	2.24	2.23	0.001	424	398	406	−0.25
		<i>z</i>	2.15	2.23	2.20	0.001	424	411	431	0.10
Second pass	Same	<i>x</i>	2.22	2.51	2.65	0.006	325	293	368	0.62
		<i>y</i>	2.22	2.79	2.85	0.009	325	353	327	0.03
		<i>z</i>	2.22	2.39	2.40	0.003	325	295	346	0.29
	Different	<i>x</i>	2.50	2.24	2.35	−0.002	366	339	362	−0.06
		<i>y</i>	2.50	2.38	2.80	0.004	366	334	383	0.25
		<i>z</i>	2.50	2.86	2.68	0.003	366	378	404	0.55
No-Desk Condition										
First pass	Same	<i>x</i>	2.35	2.52	2.46	0.002	379	320	323	−0.80
		<i>y</i>	2.35	2.06	2.20	−0.002	379	346	371	−0.11
		<i>z</i>	2.35	2.39	2.46	0.002	379	356	363	−0.23
	Different	<i>x</i>	2.26	2.40	2.34	0.001	386	367	368	−0.25
		<i>y</i>	2.26	2.16	2.25	−0.000	386	391	391	0.08
		<i>z</i>	2.26	2.20	2.25	−0.000	386	393	420	0.49
Second pass	Same	<i>x</i>	2.43	2.44	2.92	0.007	308	291	335	0.38
		<i>y</i>	2.43	3.66	3.37	0.013**	308	353	344	0.52
		<i>z</i>	2.43	2.42	2.64	0.003	308	268	318	0.14
	Different	<i>x</i>	2.41	2.34	2.61	0.003	404	330	343	−0.88
		<i>y</i>	2.41	2.92	3.18	0.011***	404	346	369	−0.51
		<i>z</i>	2.41	2.77	2.48	0.001	404	362	416	0.17

* $p < .1$. ** $p < .05$. *** $p < .01$. Probabilities were adjusted using the Bonferroni method.

change conditions, the average RT was 109 msec longer in the no-desk condition than in the with-desk condition [averaged over the seven viewpoints in the Orientation 1 and Orientation 3 conditions; $t(16) = 1.40, p > .1$]. For the location change conditions, the 1.47% increase in error rates in the no-desk condition was significant [$t(16) = 2.14, p < .05$], whereas there was actually a 0.84% decrease in the error rates in the orientation change conditions ($t < 1$).

As in Experiment 1, the first-pass time in the *same* and location change conditions were about the same: averaged over the viewpoints, 821, 825, and 830 msec for the *same*, Location 2, and Location 3 conditions in the no-desk condition and 853, 851, and 854 msec for the *same*, Location 2, and Location 3 conditions in the with-desk condition, respectively ($F < 1$ for the main effect of condition). All were slightly longer than those in the orientation change conditions (905 and 850 msec for the Orientation 1 and Orientation 3 conditions in the no-desk condition and 951 and 894 msec for the Orientation 1 and Orientation 3 conditions in the with-desk condition, respectively). Although the main effect of location change versus orientation change conditions was significant [$F(1,16) = 6.58, p < .025$], the main effect and interactions of desktop were not [$F_s(1, 16) < 2.1, p_s > .1$]. Thus, the absence of the desktop did not affect the first-pass process in any of the four *different* conditions.

As can be seen in Table 7, the probability of a revisit in *different* trials was higher in the no-desk condition (averaged over the viewpoints, .70, .65, .69, and .67 for the Location 2, Location 3, Orientation 1, and Orientation 3 conditions, respectively) than in the with-desk condition [.60, .58, .59, and .55, respectively; $F(1,16) = 15.42, p < .001$, for the main effect of desktop], but the difference between location change and orientation change was not significant ($F < 1$). The second-pass times, conditional on there being a second pass, in the no-desk condition (1,114, 1,100, 1,232, and 1,248 msec for Location 2, Location 3, Orientation 1, and Orientation 3 conditions, respectively) were somewhat longer than those in the with-desk condition (996, 1,176, 1,153, and 1,266 msec); however, neither the main effect of desktop nor the interaction between desktop and location-orientation conditions were significant ($F_s < 1$). Prior analyses indicated that the effect of desktop was limited in the *y*-axis rotation conditions. Thus, we ran separate analyses for the *y*-axis condition; however, the interaction between desktop and location-orientation was not significant ($F < 1$). The results thus indicated that the effect of the desktop was basically the same for the four types of *different* conditions when the probability of a second pass was taken into account.

The number of gazes to reach a location-changed object. As was mentioned earlier, one of the purposes of Experiment 2 was to discover the role of the desktop in

helping to guide the eyes to a changed object. As in Experiment 1, the number of gazes to a changed object in the Location 2 and Orientation 1 conditions was compared with the number of gazes to the same object in the same-scene condition. In the with-desk condition, the number of gazes to a location-changed object (1.38 gazes, averaged over seven viewpoints) was smaller than that to the same object in the same-scene condition [1.61 gazes; $t(16) = 4.03, p < .002$], whereas in the no-desk condition, the number of gazes to a location-changed object (1.40 gazes) was about the same as that in the same-scene condition [1.46 gazes; $t(16) = 1.15, p > .20$]. The presence/absence of the desktop $\times$ Location 2 versus same-scene interaction was also significant [$F(1,16) = 5.68, p < .05$]. It thus appears that the parafoveal information that helped to guide the eyes to a location-changed object was the relationship of the object with the desktop. In the Orientation 1 condition, the number of gazes until reaching the changed object was just about the same as that for the baseline condition (2.25 vs. 2.36 gazes in the no-desk condition and 2.36 vs. 2.33 gazes in the with-desk condition; $t_s < 1$). Thus, it appears that the large difference of a changed object can direct the eyes to the changed object, but the subtler difference of a rotated object does not.

No eye movement trials. In Experiment 2, there were 99 trials in which the participants did not move their eyes (all but 1 correctly responded to). Besides the small number of trials, as in Experiment 1, there was large participant variability—some participants had many no eye movement trials, but others had none—so no statistical tests were performed on the data. Two thirds (65) of the no eye movement trials occurred in the with-desk condition. The major difference from the no eye movement trials in Experiment 1 was that only about half of them in the with-desk condition (32 trials) and less than half of them in the no-desk condition (14 trials) were in the 000, Z35, and Z70 conditions. (In Experiment 1, the large majority of the no eye movement trials came from these three conditions.) This suggests that the processing in these trials may be more than matching 2-D representations.

Summary. The data in the with-desk condition in Experiment 2 replicated almost all the significant findings in Experiment 1. The removal of the desktop made processing somewhat more difficult, although not qualitatively so, with the possible exception of the y -axis rotation condition. Removal of the desktop increased the mean RTs by about 50–100 msec, increased the slopes of the rotation functions only slightly, except for y -axis rotations, but did increase the probability of a revisit for all three axes for *different* trials. The y -axis rotations became quite a bit more difficult (longer RTs and many more errors), and furthermore, the pattern of errors changed markedly, since *same* trials were responded to as “different” an order of magnitude more in the no-desk y -axis rotation conditions than in any other conditions. In addition, the removal of the desktop largely eliminated the parafoveal detection of a location change observed in the with-

desk conditions in Experiments 1 and 2. In addition, the presence/absence of the desktop affected only the second-pass time. However, since this effect was due chiefly to an increase in the probability of a revisit, it is a bit hard to interpret. It does suggest, however, that removing the desktop does not affect the time to perform any operation early in processing, although it influences the probability that those early operations will lead to a response.

GENERAL DISCUSSION

The major goal of the present experiments was to extend the work of Nakatani et al. (2002) on scene rotation through a careful examination of the pattern of eye movements while people were engaged in the scene rotation task. The RT and error data in the with-desk conditions replicated the earlier findings—specifically, in finding a different rotation effect for location and orientation changes—indicating that Shepard and Metzler’s (1971) model of a holistic analogue rotation was not valid for simple scenes of the type we studied. That is, if a mental rotation of either the standard or the comparison scene preceded any comparison process, one would expect the same rotation effect for location and orientation changes. Of course, one might get a main effect of comparison type, indicating that location changes are easier to detect than orientation changes once the rotation has been accomplished. Moreover, as we will argue below, the data also go against our earlier conjecture (Nakatani et al., 2002) that the rotation effect observed in RT was due to an accumulation of time-consuming analogue *piecemeal* (largely object by object) rotations.

Our eye movement analyses revealed several additional things about how judgments are made in the scene rotation task. Most obviously, it indicated that an analogue time-consuming alignment process was not occurring in the early part of the trial (about the first two thirds of the trial), with the possible exception of the z -axis rotations. That is, except for z -axis rotations, neither the duration of the initial component of processing (terminated by the initial saccade) nor the duration of the first pass (terminated by the first revisit) was affected by angle of rotation. However, the probability that a response could be made on the basis of completing these first two components did decrease with increasing rotation angle. This implies that a comparison process occurred (on at least some of the trials) whose success, rather than the time consumed, was affected by rotation angle. Moreover, elimination of the desktop in Experiment 2 had little effect on the duration of either of these first two components, but it did affect the probability that a response could be made without a second pass, especially for y -axis rotations. In contrast to what was observed for the first two components, the time spent in the final component of processing (i.e., the second-pass time) was decidedly influenced by rotation angle for all axes of rotation. In this last component of processing, both individual gaze durations on objects and the num-

ber of gazes were affected by rotation angle. However, the duration of this last stage was little affected by the absence of the desktop frame. An additional finding with respect to the role of the desktop frame was that there were many more errors in the *y*-axis *same* condition than in the other conditions.

How can one best explain our observed pattern of results? Perhaps the place to start is to try to explain what is occurring in the first pass of processing. In this analysis, we will take as a working assumption (and will justify below) that if there is an analogue mental rotation process, it should be evident in increased gaze durations on objects in the comparison stimulus that is being viewed.¹¹ That is, we will assume that operations on an object are reflected by the time that is taken to view an object.

Given that there were no rotation effects on first-pass processing time (or on individual gaze durations during the first pass), we think it is parsimonious to assume that there was no analogue rotation process occurring during the first pass, with the possible exception of *z*-axis rotations. Alternatively, the object array might be encoded in some reference frame that allows a participant to make a moderately adequate *same*–*different* judgment without an alignment. We think that the three coordinate frames that people would be likely to use for encoding the three objects are (1) viewer centered, (2) desk centered, and (3) object centered. These codes could be thought of as verbal labels, although they would not necessarily be so.

The first two frames seem quite straightforward. In the viewer-centered frame, the objects are coded in terms of both their locations with respect to the viewer and their orientations with respect to the viewer. These could be in terms of either the 2-D representation on the screen or a 3-D representation. However, for these particular stimuli, it would not make much difference which was used. For example, “left” and “right” for location would be the same in both 2-D and 3-D representations, “up” and “down” in 2-D would be replaced by “near” and “far” in a 3-D representation, and the orientations of the objects with respect to the viewer would be roughly the same for both 2-D and 3-D representations. There could be several desk-centered ways to code objects, but we think the most natural would be to code locations in terms of the side(s) that the object was near (remember that the sides were color coded) and orientations in terms of how the objects were pointing—that is, most naturally coded for their orientation with respect to side of the desk nearest the object.¹² (We will return to object-centered frames later.)

For *x*-axis rotations, either of these frameworks would suffice to get correct answers, and thus, the desktop frame is not needed to make judgments (although it might aid them). For *z*-axis rotations, the situation is a bit less clear. However, for both the *z*-axis rotations we employed, the left-to-right order of the objects is preserved, and their orientations are also preserved (e.g., the laptop computer is facing the viewer in both *x*-axis rotations; see Figure 1). Thus, again, a viewer-centered represen-

tation is adequate (although it might take longer to compute the code for greater *z*-axis rotation angles). Obviously, the situation is different for the *y*-axis rotations, since the left–right order is changed as well as near–far and top–bottom codings (see Figure 1). Thus, something is definitely needed for these rotations besides a viewer-centered frame.

This analysis thus makes it reasonable that removing the desktop frame had a selective effect on *y*-axis rotations. That is, one explanation for the large increase in errors on *same* trials when the desktop was absent is that people were still using a viewer-centered representation on at least some trials and, as a result, were classifying *same* stimuli as “different” when the viewer-centered order changed. This would not necessarily have led to many additional errors on *different* trials, since the order would have still been different from the standard on most trials (i.e., they would make the correct response for the wrong reason). However, if people had relied exclusively on a viewer-centered representation for their judgments, they would have had close to 100% errors on *same* trials. Thus, on a majority of trials, they must have either (1) initially used another representation or (2) double-checked an initial judgment that relied on a viewer-centered representation, using a different representation. What remains unclear is why people would ever have relied on a viewer-centered representation in the desktop-absent conditions (rather than at least double-checking with another representation).

Our analysis thus far has the following implications. First, a desk-centered frame of reference appears to be used when it can be used, because the error rates were so much lower when it was present in the *y*-axis rotation conditions. Second, when the desktop frame is absent, a viewer-centered frame may suffice for judgments for *x*- and *z*-axis rotations, and thus, the absence of the desktop has no effect on performance in these two conditions. This analysis is useful as a start but does not explain (1) why there is any second-pass processing for *x*- or *z*-axis rotations, (2) why reliance on second-pass processing is markedly greater in the *y*-axis rotation conditions when the desktop is absent, or (3) why first-pass processing times are the same in the desk and the no-desk conditions.

We think a part of the answer is that some sort of object-centered frame must be employed at least some of the time but the object-centered coding system is less reliable than the other two systems and that, furthermore, the viewer is aware of this and, thus, aware of the need for double-checking. This system must be used at least some of the time in the *y*-axis no-desk condition; otherwise, people would never be correct in this condition by the end of the first pass. That is, with the desktop removed, they could not use the desk frame, and if they used a viewer-centered frame, they would make errors in the *y*-axis rotation *same* conditions most of the time. In an object-centered frame, the codes would relate the objects to each other, but potentially there would be many

such frames. One distinction is that between coding the relationship between the objects in two dimensions and coding the relation in three dimensions. In a 2-D frame, one might code the three object locations in an imaginary circle, so that, for example, the mug is clockwise from the computer (see Figure 1). Orientations could be coded similarly, so that the computer is pointing a bit to the clockwise side of the mug. Such a coding actually has the potential to work in the task, but it might be hard to compute the orientations in some of the views. A second possibility, of course, would be to code the relationships of the objects with respect to each other in three dimensions. The coding would be similar to the 2-D version, but presumably in depth. The latter representation should, if computed correctly, be perfectly adequate even to do the orientation change task, in general, although the relationships might be quite a bit harder to compute than relationships with respect to a frame, such as the desktop. It seems intuitively clear that coding the relations of the objects to each other is more difficult and less reliable than coding the relations of the objects to the desktop.

To recap, we think that the most parsimonious explanation of the processing in the first pass is that people code stimuli in terms of some combination of these three frames and simply judge whether the relational codings of the objects in the standard and the comparison scenes are the same. That is, no analogue mental rotation process is involved. Why would anything more be needed (especially if the desktop is present)? It seems likely that such coding, either of the standard or the comparison scene, is likely often to be either fallible or incomplete. Thus, further double-checking processing may often be needed. This presumably is the second pass.

Is this second pass qualitatively different from the first pass, or just a repeat of it? We think that, although some of this processing may be just a repeat of first-pass processing, there is evidence that something different is occurring in the second pass. Most notable, of course, is the fact that second-pass times were affected by rotation angle. If the second pass were just a partial repeat of the first pass, one would expect the same pattern as the first pass, with a rotation effect only for z-axis rotations. In addition, whatever process was going on in this stage was little affected by the absence of the desktop even though the probability of making a second pass was affected by the absence of the desktop. This again seems to rule out the hypothesis that the second pass is just a re-tracing of part of the first pass. Furthermore, although the effect was not huge, there was an increase in individual gaze durations in the second pass, most notably in Experiment 1. This suggests that some process was occurring with the processing of each object that took more time, the greater the rotation angle. An analogue *mental rotation* process seems like a good candidate.

As was indicated above, a key assumption in the argument above is that where people are looking in this task is an index of what is being processed. Otherwise,

one could argue that individual gaze durations and the first- and second-pass times are irrelevant indicators of the processing underway. That is, it is possible (and the no eye movement trials indicate that it happens occasionally) that people do not need to move their eyes in order to accomplish this task—unlike when one reads text). Nonetheless, there is a considerable amount of data indicating that the place where people are looking is a good indicator of what information is being processed and/or attended to. First, we should point out that our measure of which object was fixated was not merely a convenient fiction; most fixations were either on or quite close to one of the three objects. This makes sense, since detecting the way an object was facing (relevant to detecting orientation changes) would have generally been difficult (although not impossible) without fixating the object. Certainly the work of Henderson and Hollingworth (1999) and others shows that detection of changes in a display is much more difficult when the object is not fixated either just before the change or just after it.

Moreover, there are now a large number of experiments indicating that fixation of an object also indexes attending to that object even when the visual information has certainly already been extracted. For example, Meyer and Dobel (2003) have shown that speech production is tied to where people are looking. For example, line drawings whose names take longer to say but are not more visually complex are fixated longer in an object-naming task. That is, the eyes are apparently not allowed to move until the speech production system has achieved some sort of resolution. This relates to work (e.g., Griffin & Bock, 2000) indicating that people's eye movements when they are giving verbal descriptions of simple actions (represented in line drawings) mirrors the sequence of actions being described. In addition, in more complex real-world events, such as making a cup of tea (e.g., Ballard, Hayhoe, & Pelz, 1995; Land, Mennie, & Rusted, 1999), people fixate an object well beyond the time needed to encode the object (presumably, to keep the object in the focus of attention for the action being performed). Thus, although it may be logically possible for people to rotate the objects in the scene (or the whole scene) in a way that is irrelevant to where they are fixating, we think it quite unlikely. Thus, the fact that individual gaze durations on objects do not increase much as rotation angle is increased suggests that little or none of the alignment process involves a time-consuming analogue mental rotation process.

Another interesting phenomenon that the eye movement record revealed was that the eyes were drawn to an object in a changed location, but not to an object that was in the wrong orientation. This makes sense if eye movements are drawn to object locations and the choice of a location is determined not only by some predetermined order of scanning, but also by the identities of objects seen in parafoveal vision (Henderson & Hollingworth, 1999). The fact that this phenomenon was eliminated by the absence of the desktop is also interesting. It suggests

that the change of location is easily detected (and hence, draws the viewer's eye to it) only when the object is seen against the frame of the desktop. This makes sense, since the desktop edges were colored, so that the object's being seen against the "wrong" color (or luminance) may have been the signal. It is clearly an interesting question to consider what other *frame* cues would also allow the changed object to be salient and, hence, draw attention.

In sum, the process of judging the identity of a simple scene given the possibility of a change of viewpoint is a complex one. The increased times when scenes are rotated appear to be due mainly to processes other than analogue mental rotation. Instead, it appears that the early components do not involve any such time-consuming analogue rotation process but, instead, involve processes that involve comparing the objects of the comparison scene with some sort of mental representation of the standard scene in which there is some sort of coding of the locations and orientations of the objects—probably in some mixture of three different representations: object centered, desktop centered, and viewer centered. However, when there is a rotation, especially around the *y*-axis, the coding of these representations becomes less adequate for successfully doing a comparison during the first pass, and the probability that a second pass is needed increases. This change in probability accounts for quite a bit of the rotation effect observed in RTs. Increased second-pass times also account for a considerable part of the rotation effect. They appear to be a bit more analogue in nature, since individual gaze durations are somewhat affected by angle of rotation. However, since most of the increase in second-pass time is due to there being more gazes on the objects (especially for *y*-axis rotations), the increase in second-pass time appears to be due mainly to the making of more comparisons—perhaps more double-checking.

REFERENCES

- AKS, D. J., & ENNS, J. T. (1996). Visual search for size is influenced by a background texture gradient. *Journal of Experimental Psychology: Human Perception & Performance*, **22**, 1467-1481.
- ANTES, J. R. (1974). The time course of picture viewing. *Journal of Experimental Psychology*, **103**, 62-70.
- ANTES, J. R., & PENLAND, J. G. (1981). Picture context effects on eye movement patterns. In D. F. Fisher, R. A. Monty, & J. W. Senders (Eds.), *Eye movements: Cognition and visual perception* (pp. 157-170). Hillsdale, NJ: Erlbaum.
- BALLARD, D. H., HAYHOE, M. M., & PELZ, J. B. (1995). Memory representations in natural tasks. *Journal of Cognitive Neuroscience*, **7**, 66-80.
- BETHELL-FOX, C. E., & SHEPARD, R. N. (1988). Mental rotation: Effects of stimulus complexity and familiarity. *Journal of Experimental Psychology: Human Perception & Performance*, **14**, 12-23.
- BLACKMORE, S. J., BRELSTAFF, G., NELSON, K., & TROSCIANKO, T. (1995). Is richness of our visual world an illusion? Transsaccadic memory for complex scenes. *Perception*, **24**, 1075-1081.
- BOERSEMA, T., ZWAGA, H. J. G., & ADAMS, A. S. (1989). Conspicuity in realistic scenes: An eye-movement measure. *Applied Ergonomics*, **20**, 267-273.
- BOYCE, S. J., & POLLATSEK, A. (1992). Identification of objects in scenes: The role of scene background in object naming. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **18**, 531-543.
- DE GRAEF, P., DE TROY, A., & D'YDEWALLE, G. (1992). Local and global contextual constraints on the identification of objects in scenes. *Canadian Journal of Psychology*, **46**, 489-508.
- DIWADKAR, V. A., & MCNAMARA, T. P. (1997). Viewpoint dependence in scene recognition. *Psychological Science*, **8**, 302-307.
- FOLK, M. D., & LUCE, R. D. (1987). Effects of stimulus complexity on mental rotation rate of polygons. *Journal of Experimental Psychology: Human Perception & Performance*, **13**, 395-404.
- FRIEDMAN, A. (1979). Framing pictures: The role of knowledge in automatized encoding and memory for gist. *Journal of Experimental Psychology: General*, **108**, 316-355.
- GRIFFIN, Z. M., & BOCK, K. (2000). What the eyes say about speaking. *Psychological Science*, **11**, 274-279.
- HENDERSON, J. M. (1993). Eye movement control during visual object processing: Effects of initial fixation position and semantic constraint. *Canadian Journal of Experimental Psychology*, **47**, 79-98.
- HENDERSON, J. M. (1997). Transsaccadic memory and integration during real-world object perception. *Psychological Science*, **8**, 51-55.
- HENDERSON, J. M., & ANES, M. D. (1994). Roles of object-file review and type priming in visual identification within and across eye fixations. *Journal of Experimental Psychology: Human Perception & Performance*, **20**, 826-839.
- HENDERSON, J. M., & HOLLINGWORTH, A. (1999). The role of fixation position in detecting scene changes across saccades. *Psychological Science*, **10**, 438-443.
- HENDERSON, J. M., POLLATSEK, A., & RAYNER, K. (1987). Effect of foveal priming and extrafoveal preview on object identification. *Journal of Experimental Psychology: Human Perception & Performance*, **13**, 449-463.
- HENDERSON, J. M., WEEKS, P. A., & HOLLINGWORTH, A. (1999). The effect of semantic consistency on eye movements during complex scene viewing. *Journal of Experimental Psychology: Human Perception & Performance*, **25**, 210-228.
- JUST, M. A., & CARPENTER, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, **87**, 329-354.
- JUST, M. A., & CARPENTER, P. A. (1985). Cognitive coordinate systems: Accounts of mental rotation and individual difference in spatial ability. *Psychological Review*, **92**, 137-172.
- LAND, M., MENNIE, N., & RUSTED, J. (1999). The roles of vision and eye movements in the control of activities of daily living. *Perception*, **28**, 1311-1328.
- LOFTUS, G. R., & MACKWORTH, N. H. (1978). Cognitive determinants of fixation location during picture viewing. *Journal of Experimental Psychology: Human Perception & Performance*, **4**, 565-572.
- MACKWORTH, N. H., & MORANDI, A. J. (1967). The gaze selects informative details within pictures. *Perception & Psychophysics*, **2**, 547-552.
- MEYER, A. S., & DOBEL, C. (2003). Application of eye tracking in speech production research. In J. Hyönä, J. R. Radach, & H. Deubel (Eds.), *The mind's eye: Cognitive and applied aspects of eye movement research* (pp. 253-272). Oxford: Elsevier.
- NAKATANI, C. (2001). *Eye movements during recognition of a rotated scene*. Unpublished doctoral dissertation, University of Massachusetts, Department of Psychology.
- NAKATANI, C., & POLLATSEK, A. (2001). The role of the frame of reference in the scene rotation task. *Abstracts of the Psychonomic Society*, **6**, 100.
- NAKATANI, C., POLLATSEK, A., & JOHNSON, S. (2002). Viewpoint-dependent recognition of scenes. *Quarterly Journal of Experimental Psychology*, **55A**, 115-139.
- NELSON, W. W., & LOFTUS, G. L. (1980). The functional visual field during picture viewing. *Journal of Experimental Psychology: Human Learning & Memory*, **6**, 391-399.
- POLLATSEK, A., RAYNER, K., & COLLINS, W. E. (1984). Integrating pictorial information across eye movements. *Journal of Experimental Psychology: General*, **112**, 426-442.
- RAYNER, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, **124**, 372-422.
- RAYNER, K., & POLLATSEK, A. (1992). Eye movements and scene perception. *Canadian Journal of Experimental Psychology*, **46**, 342-376.
- SEKULER, A. B., & PALMER, S. E. (1992). Perception of partly occluded

- objects: A microgenetic analysis. *Journal of Experimental Psychology: General*, **121**, 95-111.
- 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*, **191**, 952-954.
- STERNBERG, S. (1969). Memory scanning: Mental processes revealed by reaction-time experiments. *American Scientist*, **57**, 421-457.

NOTES

1. There was also a double axis rotation, which could be viewed as a composite rotation around two axes: a 70° rotation around the y-axis, followed by a 70° rotation around the x-axis. To simplify exposition, we will not report on it here. Essentially, the data from that condition were similar to those in the X70 condition. For a complete analysis of that condition, see Nakatani (2001).
2. The word *component* is intended to be a relatively neutral term. That is, our analysis intends to “carve up” the total trial time into the three component times that are divided by signal events in the trial: the first saccade to an object and the first saccade back to a previously fixated object. Of course, any such division presupposes as a working hypothesis that somewhat different processing is occurring in the three components, and we believe that our analyses below indicate that this working hypothesis is justified. However, we do not see them as formal processing *stages* (Sternberg, 1969), in the sense that they are defined by qualitatively different processing. (Our three components do not exactly add up to the total RT, since saccade times are not included in the component analysis.)
3. Note that the three objects were relatively far from each other and that the initial fixation point in the center of the display was not close to any of the three objects.
4. We should also note that Henderson (1993) has used initial gaze duration on an object to capture its initial encoding during scene perception. An alternative measure of first-pass processing that is perhaps closer to his is to sum the initial gaze durations on all three objects regardless of whether revisits to other objects occur before the third object is fixated. Second-pass times would then be the sum of the fixation times when objects are revisited. If fewer than three objects are fixated

before a response is made, this measure—like our first-pass measure above—would be the sum of the gaze durations on whichever objects were fixated (and second-pass time would be zero). We will occasionally report the results of this partition of a trial. However, we will focus on the partition described initially, because we think it captures the initial encoding of the scene better.

5. The x- and y-coordinates of the centers of the objects were estimated by experimenter C.N.
6. We will directly compare the *same* and the two types of *different* responses later. Also, as we will indicate later, the pattern in the location change condition in this study was slightly different from that in Nakatani et al. (2002).
7. It is worth noting that the increase in gaze durations during the second pass was not due to individual fixation durations increasing with increasing angle of rotation ($t_s < 1$ for all three axes). Thus, it appears that the objects, rather than individual fixations, are the relevant units of analysis. Since there were no significant effects in individual fixation durations in either experiment, we will not report on them further.
8. The p values were adjusted using the Bonferroni method.
9. This nonsignificant RT difference might be due to a speed-accuracy tradeoff. The mean error rate was the highest in the Y70 condition (16.67%), as compared with those in X70 (13.10%) and Z70 (11.91%) conditions.
10. Again, X70-Z70 was also included, but to simplify exposition, we will not discuss it. See Nakatani (2001) for a presentation and discussion of these data.
11. We are not assuming that such increases are necessarily due to analogue rotation processes, but merely that any time-consuming process such as analogue rotation would have to be reflected in increased gaze duration on an object.
12. In an unpublished study (in which eye movements were not monitored), we compared the no-desk condition with one in which there was an “infinite desk top” covering the screen. There was little difference between these conditions, indicating that the major function of the desk-top was to provide well-defined edges that could be used as landmarks.

(Manuscript received May 27, 2003;
revision accepted for publication December 5, 2003.)