Research Article

VIEWPOINT DEPENDENCE IN SCENE RECOGNITION

Vaibhav A. Diwadkar and Timothy P. McNamara

Vanderbilt University

Abstract—*Two experiments investigated the viewpoint dependence of spatial memories. In Experiment 1, participants learned the locations of objects on a desktop from a single perspective and then took part in a recognition test; test scenes included familiar and novel views of the layout. Recognition latency was a linear function of the angular distance between a test view and the study view. In Experiment 2, participants studied a layout from a single view and then learned to recognize the layout from three additional training views. A final recognition test showed that the study view and the training views were represented in memory, and that latency was a linear function of the angular distance to the nearest study or training view. These results indicate that interobject spatial relations are encoded in a viewpoint-dependent manner, and that recognition of novel views requires normalization to the most similar representation in memory. These findings parallel recent results in visual object recognition.*

Human actions often rely on memories of where objects are located in the environment. These memories are produced by perceptual and motor systems that have a point of view: seeing, reaching and grasping, and locomoting. Because humans are by nature mobile organisms and able to approach an environment from many directions, they often need to recognize familiar spaces from unfamiliar views.

The difficulty of recognizing novel views of a space will depend on how spatial relations are encoded. If spatial relations are encoded independently of viewpoint, then familiar and novel views of the space will be equally accessible (e.g., Presson, DeLange, & Hazelrigg, 1989). However, if spatial relations are encoded in a view-specific manner, then familiar views will be more accessible than novel views; indeed, the latter may not be recognized at all (e.g., Rock & DiVita, 1987). It is widely acknowledged that mental representations of spatial structure may be viewpoint independent even though most perceptual systems possess aspect (e.g., Biederman, 1987; Marr, 1982).[1]

There is a growing body of evidence that spatial memories are viewpoint dependent (e.g., Levine, Jankovic, & Palij, 1982; Rieser, 1989). For instance, Shelton and McNamara (1997) had subjects learn the locations of objects in a large room. The participants then made judgments of relative direction using memory (e.g., "Imagine you are standing at the book facing the clock. Point to the shoe."). These judgments were faster and more accurate when the imagined heading (e.g., "at the book facing the clock") was aligned with a view that the participants actually experienced at the time of learning than when the imagined heading was misaligned with an experienced view of the space.

Viewpoint dependence has also been of great interest in the domain of visual object recognition. Recent investigations indicate that intraobject spatial relations are represented in a viewpoint-dependent manner, and that recognition of a novel view of an object is achieved by transforming the novel view into a represented one. Multiple views of an object seem to produce multiple viewpoint-dependent representations, not a single viewpoint-independent representation (e.g., Tarr & Bülthoff, 1995; but see Biederman & Gerhardstein, 1995).

The question we asked in the present research was whether the findings in these two domains reflect the same set of constraints on spatial representations and processing. Direct comparisons are rendered tenuous by vast differences in methods. Studies of interobject spatial representations have typically used computationally intensive tasks that require explicit mental reorientation and the computation of heading from memory, whereas studies of intraobject spatial representations have typically used relatively automatic tasks, such as recognition or naming. Heading and direction between objects may be difficult to compute from novel viewpoints, but this does not imply that interobject relations will be difficult to recognize from novel viewpoints.

In our experiments, observers studied a layout of objects under normal viewing conditions, and then their ability to recognize the layout was tested. The test scenes included familiar views of the layout, novel views of the layout, and views of other layouts (which contained the same objects but in different spatial configurations). The task was to discriminate familiar and novel views of the studied layout from views of other layouts. The viewpoint dependence of spatial memories had never been tested in this way.

Even in situations in which object recognition can be characterized as viewpoint dependent, there is evidence of viewpoint-independent recognition of subsets of test views, in particular, those between closely spaced study views. For example, if the study views are arbitrarily labeled 0° and 45°, then novel test views between 0° and 45°, such as 15° and 30°, may be recognized as well as the study views and better than novel views of 345° and 60°. The processing of these two sets of test views has been referred to as interpolation and extrapolation, respectively (Bülthoff & Edelman, 1992). The mechanisms responsible for this effect are not well understood. However, viewpoint-independent recognition in the interpolation range is problematic for alignment models (e.g., Ullman, 1989) but is predicted by a class of models referred to as view-interpolation models (e.g., Bülthoff, Edelman, & Tarr, 1995). The use of a visual recognition task allowed us to test whether differences between interpolation and extrapolation are evident in scene recognition, an issue that had never before been investigated.

Several aspects of our methods are important and deserve comment: We used an old-new recognition task with full knowledge that in the domain of object recognition, implicit tasks, such as naming, have been advocated as being superior to explicit tasks, such as old-new recognition, on the grounds that they permit purer access to the underlying viewpoint-independent representations of objects (e.g., Biederman & Gerhardstein, 1993). We used old-new recognition for several reasons: First, there is no reason to assume that implicit tasks

Address correspondence to Timothy P. McNamara, Department of Psychology, Vanderbilt University, 111 21st Ave. South, Nashville, TN 37240.

1. Viewpoint dependence can be distinguished from orientation dependence. In the latter, all views parallel to an experienced view should be equally accessible because the orientation is constant, even though the viewing location has changed. This distinction is not tested in the present experiments.

re more representative than explicit tasks of everyday object or scene recognition (Tarr & Bülthoff, 1995). Second, even in object recognition, viewpoint dependence has been evinced in implicit tasks (e.g., Srinivas, 1995; Tarr & Pinker, 1989). Third, the paradigmatic implicit task, naming, would be a poor choice for studying spatial memories because names are not intrinsic properties of most scenes. And finally, several theoretically important results, such as the difference between interpolation and extrapolation, have been documented primarily in explicit recognition.

An important aspect of the recognition task was that the target and the distractor scenes contained the same objects; they differed only in the spatial relations among the objects. These six familiar and distinctive objects were presented on every trial. The task was therefore spatial, and if object recognition contributed to performance, it should have led to viewpoint-independent recognition. The scenes also had the property that their spatial structure was visible from all perspectives. Viewpoint-dependent recognition of objects has been attributed to the occlusion of features as objects are rotated in depth (e.g., Biederman & Gerhardstein, 1993). If this principle applies to our experiments, it predicts viewpoint-independent performance because the spatial structure of the scenes was visible from all views.

Finally, a strength of our methods is that novel stimuli can be constructed from common objects simply by rearranging their positions. Our stimuli contrast sharply with the stimuli used in many studies of object recognition. To control for prior experience, researchers have been forced to use artificial, and sometimes confusing, nonsense objects as stimuli. The stimuli in our experiments were configurations of objects that one might find on the living room floor of a household with young children or even on the desk of a college professor.

In summary, our experiments were designed to answer two questions: Is recognition of multiobject scenes viewpoint dependent? If so, do important results evidenced in object recognition, such as the functional role of multiple views and the differences between interpolation and extrapolation, generalize to scene recognition?

EXPERIMENT 1

Method

Subjects

Twenty-four students participated in the experiment to satisfy a course requirement. Data from 1 subject were lost because of equipment failure.

Materials

Six spatial layouts were contructed by placing objects on a circular desktop (diameter = 106.7 cm), which rested on the floor of a room (see Fig. 1). Each layout contained the same six familiar objects (coffee mug, lightbulb, scissors, screwdriver, stapler, strainer), but in a unique configuration. Layouts were constructed by randomly placing objects on a virtual grid imposed on the desktop, with the constraint that objects were never placed in adjacent grid locations. Each layout was photographed from 24 viewpoints in increments of 15° around the viewing circle. The camera was placed 259 cm from the layout and rested at a height of 132 cm. Pictures were digitized as 8-bit gray-scale images. The materials used in the recognition test consisted of these 144 digitized pictures.

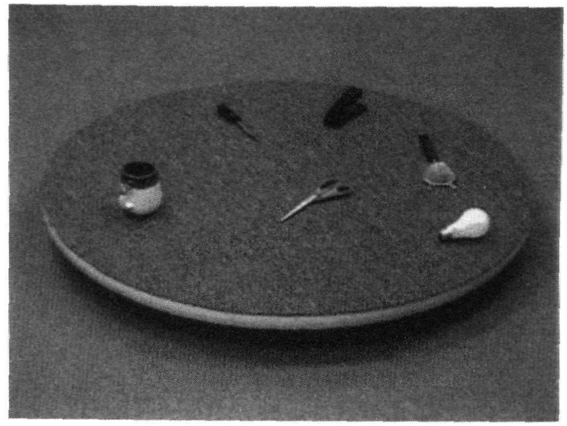

Fig. 1. One of the layouts used in the experiments.

Procedure

Each subject learned one of the six layouts from a single perspective under normal viewing conditions. Subjects were stationed 259 cm from the layout and studied it for 30 s. Subjects then were asked to close their eyes and point to and name all of the objects in the layout. Study and naming were repeated three times. All subjects were accurate on each of the three trials.

Subjects were taken to a different room for the recognition test. Digitized pictures of the layouts were displayed one at a time in a random order on a Macintosh IIci. Subjects had to decide whether the layout depicted in a picture was the layout they had studied or a different one. They responded by pressing one of two keys on the keyboard; errors were signaled with a tone. Subjects were allowed to take periodic breaks. The recognition test included 240 pictures of the target layout (10 repetitions of each of the 24 views) and 240 distractors. The distractors were randomly and equally sampled from the set of 120 pictures of the five layouts that a subject had not studied. Across subjects, all pictures served as both targets and distractors. The entire recognition session was conducted using PsyScope (J.D. Cohen, MacWhinney, Flatt, & Provost, 1993).

Results and Discussion

Analyses were based on mean latencies computed for each subject and each condition. One outlying latency greater than 10 s was excluded from the computation of means. The overall miss and false alarm rates were both 2%. Analyses of variance (ANOVAs) included layout as a between-subjects variable. An alpha level of .05 was used.

Mean latency for each test view is plotted in Figure 2. The test view of 0° corresponds to the studied view; it is replotted at 360° to emphasize symmetry. The data in Figure 2 reveal three important results. First, responses to the study view were faster than responses to all other views. Second, there is a roughly linear increase in response latency with angular distance from the study view. Third, the function is nonmonotonic between 135° and 225°, suggesting a savings in recognition in this angular range. This savings may occur

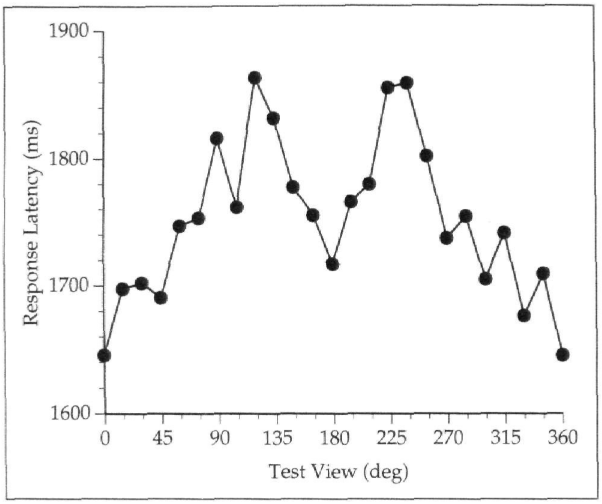

Fig. 2. Response latency as a function of test view in Experiment 1.

because the spatial structure of multiobject layouts is highly self-similar under rotations of 180°. Studies of object recognition in humans and monkeys have shown that such regularities in the geometric structure of objects can result in the creation of virtual views that are easy to recognize even though they are unfamiliar (e.g., Logothetis & Pauls, 1995; Vetter, Poggio, & Bülthoff, 1994). A similar effect may have occurred in this experiment. An ANOVA with variables corresponding to layout and to test view revealed that the effect of test view alone was reliable, $F(23, 391) = 1.72$, $MSE = 47{,}952$.

To investigate further the linear portions of the function in Figure 2, we redefined test views in terms of their angular distance from the study view. Test views between 135° and 225° were excluded from this analysis. This transformation amounts to averaging the curve from 0° to 135° with the curve from 225° to 345°. These data are plotted in Figure 3 (±1 standard error of the mean, as estimated from the ANOVA). An ANOVA including layout and angular distance showed that the effect of distance alone was reliable, $F(9, 153) = 4.42$, $MSE = 24{,}879$. The linear relation in Figure 3 was reliable, $F(1, 153) = 35.12$, and the slope of the regression line was 1.43 ms/°, or expressed as a rate, 699°/s.

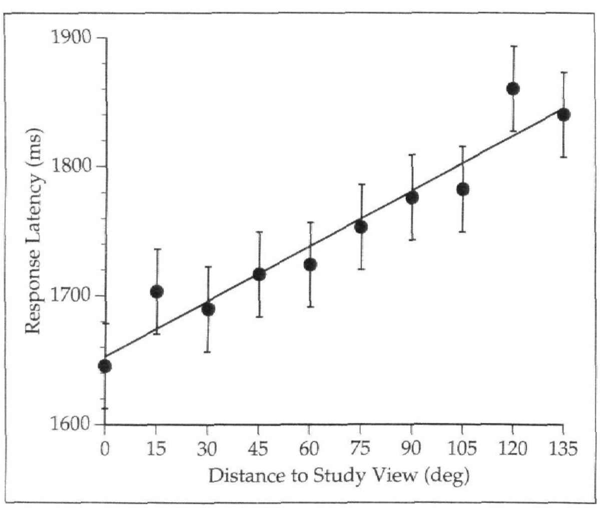

Fig. 3. Response latency as a function of the angular distance to the study view in Experiment 1.

These results indicate that a novel view of a familiar scene is recognized by effecting a transformation between the novel view and the view represented in memory. This transformation consumes more time as the angular distance over which it must be carried out increases. The bilateral symmetry of the function in Figure 2 suggests that the transformation is conducted along the shortest path in angular distance (e.g., Shepard & Cooper, 1982).

Our second experiment was designed to determine whether the multiple-views-plus-transformation model outlined by Tarr (1995) would hold in scene recognition. As in Experiment 1, observers studied a collection of objects from a single perspective. In two training blocks, observers then learned to recognize the layout from the study view and three additional training views. In a surprise test block, all 24 views were tested. If multiple views are represented and used in recognition, the best predictor of performance in recognizing novel views should be their distance to the nearest view in memory.

EXPERIMENT 2

Method

Subjects

Twenty-four students participated in the experiment to satisfy a course requirement.

Materials

The stimuli were identical to those used in Experiment 1.

Procedure

Subjects learned a layout from a single perspective, as in Experiment 1. After learning a layout, subjects participated in two training blocks in which they learned to recognize the layout from the study view (0°) and three additional training views (45°, 90°, and 270°). A schematic of the training perspectives is shown in Figure 4. The training perspectives were chosen so as to allow us to assess potential differences between interpolation and extrapolation.

In the training blocks, subjects had to discriminate four views of the studied layout (0°, 45°, 90°, and 270°) from views of unstudied layouts. In one version of the experiment, 12 subjects saw, across the two blocks, the study view (0°) 24 times and the three training views 12 times each; a separate group of 12 subjects saw each of the four views 20 times each. Following the two blocks of training, all subjects received a surprise test block in which all 24 views of the layout were tested 10 times each. The surprise test block was identical to the recognition test used in Experiment 1. Errors were signaled with a tone in all three blocks.

Results and Discussion

Analyses were based on mean latencies computed for each subject and each condition. Eight outlying latencies greater than 10 s were excluded from the computation of means. The overall miss rate was 3%, and the overall false alarm rate was 2%. ANOVAs included

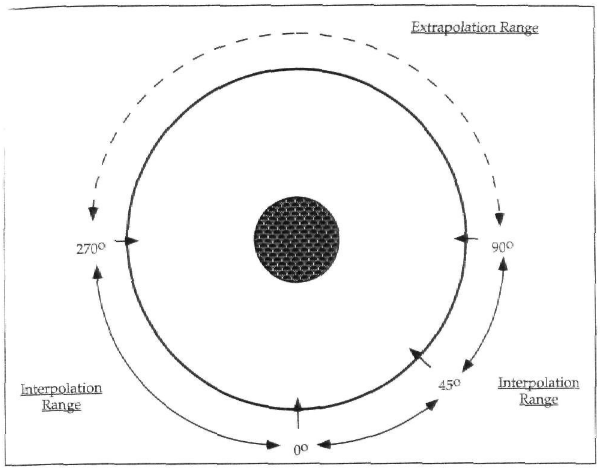

Fig. 4. Schematic of the study and the training views used in Experiment 2.

version of training and layout as between-subjects variables, with subjects nested in the combination. An alpha level of .05 was used, except where noted.

The major result of the training blocks was that speed of responding increased with practice. Mean latencies were slower in Block 1 (M = 2,473 ms) than in Block 2 (M = 1,729 ms), $F(1, 12) = 74.43$. There was also an effect of test view, $F(3, 36) = 3.70$, but this effect interacted with the two versions of training, $F(3, 36) = 3.74$. The interaction resulted because the test view of 0° was faster than the other views for subjects who received more training on 0° (in milliseconds, Ms = 1,885 for 0°, 2,238 for 45°, 2,220 for 90°, and 2,331 for 270°), but not for subjects who received the same amount of training on the four test views (in milliseconds, Ms = 2,103, 2,054, 2,064, and 2,138, respectively). The four-way interaction between version of training, layout, block, and test view was reliable but did not compromise any of the effects mentioned. No other effects reached significance.

The most important data were those for the surprise block in which all 24 views of the layout were tested. These data are displayed in Figure 5. The three training views of 45°, 90°, and 270° are marked with arrows. An ANOVA with variables corresponding to training version, layout, and test view revealed that test view alone was reliable, $F(23, 276) = 3.84$, $MSE = 47{,}225$. Mean latencies at 0° were faster than the average of the training views (45°, 90°, and 270°), $t(276) = 1.93$, $p < .05$, one-tailed.

The primary goal of this experiment was to determine whether novel views in the surprise test block would be normalized to the training views. The function in Figure 5 suggests that they were: The training views seem to be local minima in the response time function. To investigate this relation more quantitatively, we redefined test views in the surprise block in terms of their angular distance from the nearest study or training view. As in Experiment 1, the range from 135° to 225° evinced nonlinearities, although the pattern is not as symmetric in Experiment 2 as in Experiment 1. The irregularity in Experiment 2 may have arisen because multiple views and geometric similarities both contribute to recognition of views near 180°. Because a multiplicity of mechanisms may contribute to recognition in this range, test views between 135° and 225° were excluded from this analysis.

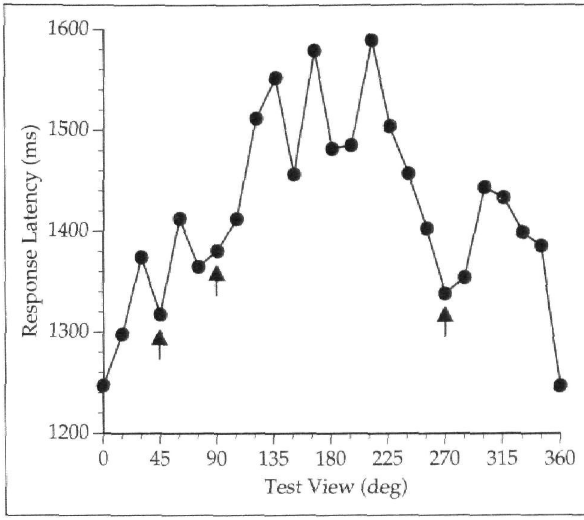

Fig. 5. Response latency as a function of test view in Experiment 2. The training views are marked with arrows.

Mean recognition latencies as a function of angular distance are plotted in Figure 6. These results suggest strongly that novel views of a studied scene were recognized by normalization to the most similar representation in memory. An ANOVA with variables corresponding to training version, layout, and angular distance produced only one reliable effect, for angular distance, $F(3, 36) = 14.34$, MSE = 10,122. The linear component was reliable, $F(1, 36) = 42.60$. Expressed as a rate, the slope of the regression line in Figure 6 is 250°/s. The slopes were very similar for interpolation (289°/s) and for extrapolation (263°/s). This finding differs from some findings in object recognition (e.g., Bülthoff & Edelman, 1992), and indicates that scene recognition mechanisms were similar in the interpolation and the extrapolation ranges.

In a final analysis, we examined whether repeated processing of novel views in the surprise block would lead to those views being represented in memory, thus mitigating the need for normalization (e.g., Jolicoeur, 1985). The data summarized in Figure 6 were divided into *early* and *late* trials. The first five instances of a novel view's occurrence were categorized as early, and the last five instances were categorized as late. The relationship between latency and angular distance for early versus late trials is graphed in Figure 7. The interaction between early versus late and angular distance was reliable, $F(3, 36) = 4.64$, $MSE = 18{,}538$. The slopes of the regression lines, expressed as rates, are 166°/s for early trials and 498°/s for late trials. This pattern is predicted by multiple-view models of scene (and object) recognition.[2]

GENERAL DISCUSSION

In two experiments, observers learned the locations of objects in a spatial layout and then received a recognition test in which they dis-

2. A similar analysis on the relevant data from Experiment 1 revealed a similar pattern of differences in the values of the slopes (571 vs. 926°/s).

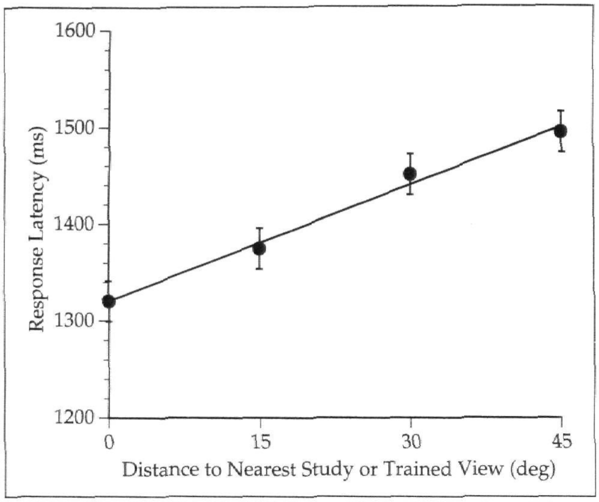

Fig. 6. Response latency as a function of distance to the nearest study or training view in Experiment 2.

criminated familiar and novel views of the layout from views of unstudied layouts. Recognition performance indicated that (a) spatial memories were encoded in a viewpoint-dependent manner; (b) novel views of a familiar scene were recognized by normalization to the most similar representation in memory; (c) observers could be trained to recognize a layout from unfamiliar views, and these training views were functional in the subsequent recognition of novel views; and (d) with sufficient practice, novel views were also represented in memory, mitigating the need for normalization.

These experiments provide a unique supplement to prior research on the view specificity of spatial memories (e.g., Shelton & McNamara, 1997). Moreover, they show that recognition of spatial structure is viewpoint dependent. This finding should put to rest concerns that viewpoint-dependent recognition is limited to nonsense objects or is caused by the occlusion of features with rotation in depth.

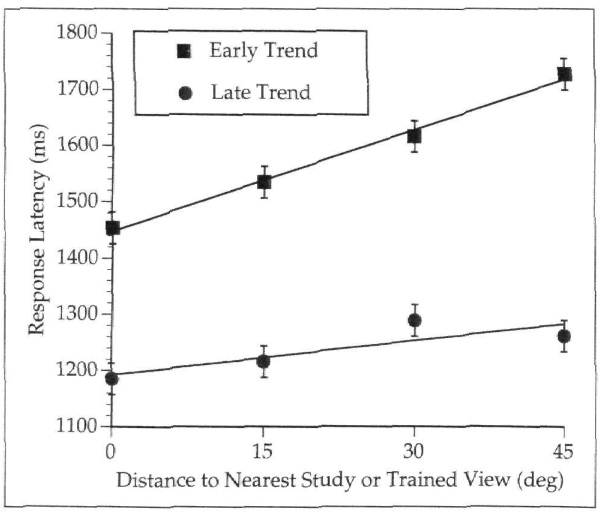

Fig. 7. Response latency as a function of distance to the nearest study or training view, for early versus late trials, in Experiment 2.

Our results invite strong comparisons to research on visual object recognition: Viewpoint dependency, recognition by normalization, and the functional role of multiple views have been documented in investigations of object recognition (e.g., Tarr, 1995). Furthermore, the rates of normalization in our experiments are well within bounds established by studies demonstrating the viewpoint dependence of object recognition and identification (approximately 1,000°/s; e.g., D. Cohen & Kubovy, 1993). Collectively, these data suggest that interobject and intraobject spatial relations may be mentally represented and processed in similar ways.

Two of our findings, however, are inconsistent with at least some findings on object recognition. Error rates in both experiments were very low, and the linear relation between latency and angular distance in Experiment 2 was identical for views in the extrapolation and the interpolation ranges. In contrast, Bülthoff and Edelman (1992) found that error rates were higher in the extrapolation than in the interpolation range, and Tarr (1995) found that latency increased linearly with angular distance in the extrapolation but not the interpolation range. These two patterns of results support different classes of models of object recognition.

Our results are consistent with alignment models of object recognition (e.g., Ullman, 1989); indeed, an effect of viewpoint on latency but not on error rate has been cited as the signature of models that use view-specific three-dimensional (3-D) representations combined with robust normalization (Bülthoff et al., 1995). According to one version of this class of models, views of an object (and by extension, a scene) are mentally represented as view-specific 3-D models. Recognition is achieved by first aligning the two-dimensional (2-D) input image with the best fitting 3-D model using simple low-level features, and then matching the image with a 2-D projection of the model. The outcome of the recognition process is the model whose projection most closely matches the input image after the two have been aligned. The linear relation between latency and angular distance indicates that the complexity of the alignment process scales with the magnitude of the transformation. The low error rates show that the alignment process is robust.

The data reported by Bülthoff and Edelman (1992) and by Tarr (1995) are more consistent with view-interpolation models. In this class of models, views of an object are represented by 2-D view-specific representations. Recognition mechanisms differ across models, and include the use of generalized radial basis functions (e.g., Poggio & Girosi, 1990) and fuzzy template matching (e.g., Edelman & Weinshall, 1991). Predictions of this class of models depend on the recognition mechanism and on details of the implementation, but these models usually predict higher accuracy on views in the interpolation than in the extrapolation range. These models may be able to predict an increase in latency with distance from a familiar view if the interpolation process is incremental.

We are reluctant to endorse these apparent differences between scene and object recognition, for two reasons. First, in another line of research, we have evidence of viewpoint-independent recognition in the interpolation range. This research has examined memory for arrays of dots displayed in depth (using linear perspective). On each trial, two views of an array, separated by 75°, are displayed in immediate succession; after a brief interval of time, a test scene is presented. The task is old-new recognition. In these experiments, response latency is flat for test views in the interpolation range but increases linearly with angular distance in the extrapolation range.

An important feature of the dot-array experiments (as well as some nvestigations of object recognition; e.g., Bülthoff & Edelman, 1992) s that the studied views were experienced close together in time and n space (this was not true of Tarr, 1995); in contrast, in Experiment 2, the training views were presented in a random order and separated in time by several seconds. The successive presentation of neighboring views in the dot-array experiments may allow access to the intervening geometry of a scene or of an object. A more restricted possibility is that apparent motion between study views (which was attained by subjects on a majority of trials) may have created views in memory in the interpolation range.

The other reason we want to withhold judgment on differences between scene and object recognition is that the stimuli in the two types of studies differ in an important way, namely, the presence of occlusion. In our stimuli, occlusion was virtually nonexistent, and the spatial structure was visible from all views. The "amoebas" and, to a lesser extent, the "wires" used by Bülthoff and Edelman (1992) and the block figures used by Tarr (1995) share the property that local features and parts can be occluded with rotation. If successful recognition depends on establishing correspondence between features in the study and in the test views, and occlusion interferes with this process, then an effect of test view on accuracy may be expected for self-occluding stimuli. Occlusion cannot account for the interpolation-extrapolation difference in latency, however, because this difference was found in our dot-array experiments, in which occlusion was absent.

In general, the limiting conditions on the interpolation-extrapolation difference are not known, even in object recognition (e.g., Logothetis & Pauls, 1995). We are testing alternative explanations in current experiments, using multiobject scenes and arrays of dots as stimuli.

For now, we believe that our results justify several conclusions. Interobject spatial relations are mentally represented in a viewpoint-dependent manner, and this viewpoint dependency is manifested in scene recognition. Multiple views of a space produce multiple representations of that space, and unfamiliar views are recognized by normalization to the most similar representation in memory. This constellation of findings is very similar to a pattern that is emerging in investigations of visual object recognition. These affinities between scene and object recognition suggest that the human brain may not honor the distinction between configurations of objects and configurations of object parts in its attempt to preserve the spatial structure of the world.

Acknowledgments—This research was supported by National Science Foundation Grant SBR 9222002 to the second author and by Office of Naval Research Grant N00014-93-1-0578 to Carolyn Cave. We are grateful to Carolyn Cave for allowing us to use video equipment in her laboratory, and we thank Carolyn Cave, Kyle Cave, and two anonymous reviewers for their comments on the article. Portions of this work were presented in November 1996 at the annual meeting of the Psychonomic Society, Chicago.

REFERENCES

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review, 94,* 115–147.

Biederman, I., & Gerhardstein, P.C. (1993). Recognizing depth-rotated objects: Evidence and conditions for three-dimensional viewpoint invariance. *Journal of Experimental Psychology: Human Perception and Performance, 19,* 1162–1182.

Biederman, I., & Gerhardstein, P.C. (1995). Viewpoint-dependent mechanisms in visual object recognition: Reply to Tarr and Bülthoff (1995). *Journal of Experimental Psychology: Human Perception and Performance, 21,* 1506–1514.

Bülthoff, H.H., & Edelman, S. (1992). Psychophysical support for a two-dimensional view interpolation theory of object recognition. *Proceedings of the National Academy of Sciences, USA, 89,* 60–64.

Bülthoff, H.H., Edelman, S.Y., & Tarr, M.J. (1995). How are three-dimensional objects represented in the brain? *Cerebral Cortex, 5,* 247–260.

Cohen, D., & Kubovy, M. (1993). Mental rotation, mental representation, and flat slopes. *Cognitive Psychology, 25,* 351–382.

Cohen, J.D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavior Research Methods, Instruments, & Computers, 25,* 257–271.

Edelman, S., & Weinshall, D. (1991). A self-organizing multiple-view representation of 3D objects. *Biological Cybernetics, 64,* 209–219.

Jolicoeur, P. (1985). The time to name disoriented natural objects. *Memory & Cognition, 13,* 289–303.

Levine, M., Jankovic, I.N., & Palij, M. (1982). Principles of spatial problem solving. *Journal of Experimental Psychology: General, 111,* 157–175.

Logothetis, N.K., & Pauls, J. (1995). Psychophysical and physiological evidence for viewer-centered object representations in the primate. *Cerebral Cortex, 5,* 270–288.

Marr, D. (1982). *Vision.* San Francisco: Freeman.

Poggio, T., & Girosi, F. (1990). Regularization algorithms for learning that are equivalent to multilayer networks. *Science, 247,* 978–982.

Presson, C.C., DeLange, N., & Hazelrigg, M.D. (1989). Orientation specificity in spatial memory: What makes a path different from a map of the path? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 887–897.

Rieser, J.J. (1989). Access to knowledge of spatial structure at novel points of observation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 1157–1165.

Rock, I., & DiVita, J. (1987). A case of viewer-centered object perception. *Cognitive Psychology, 19,* 280–293.

Shelton, A.L., & McNamara, T.P. (1997). Multiple views of spatial memory. *Psychonomic Bulletin & Review, 4,* 102–106.

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

Srinivas, K. (1995). Representation of rotated objects in explicit and implicit memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 1019–1036.

Tarr, M.J. (1995). Rotating objects to recognize them: A case study on the role of viewpoint dependency in the recognition of three-dimensional objects. *Psychonomic Bulletin & Review, 2,* 55–82.

Tarr, M.J., & Bülthoff, H.H. (1995). Is human object recognition better described by geon structural descriptions or by multiple views? Comment on Biederman and Gerhardstein (1993). *Journal of Experimental Psychology: Human Perception and Performance, 21,* 1494–1505.

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

Ullman, S. (1989). Aligning pictorial descriptions: An approach to object recognition. *Cognition, 32,* 193–254.

Vetter, T., Poggio, T., & Bülthoff, H.H. (1994). The importance of symmetry and virtual views in three-dimensional object recognition. *Current Biology, 4,* 18–23.

(RECEIVED 3/12/96; REVISION ACCEPTED 11/13/96)