Available online at www.sciencedirect.com
SCIENCE DIRECT®

Cognitive Psychology 47 (2003) 43–86

Cognitive Psychology

www.elsevier.com/locate/cogpsych

Representation and perception of scenic layout

Thomas Sanocki*

Department of Psychology, PCD 4118, University of South Florida, Tampa, FL 33620-8200, USA

Accepted 4 November 2002

Abstract

This paper presents a cognitive approach to on-line spatial perception within scenes. A theoretical framework is developed, based on the idea that experience with a scene can activate a complex representation of layout that facilitates subsequent processing of spatial relations within the scene. The representations integrate significant, relevant scenic information and are substantial in amount or extent. The representations are active across short periods of time and across changes in the retinal position of the image. These claims were supported in a series of experiments in which pictures of scenes (primes) facilitated subsequent spatial relations processing within the scenes. The prime-induced representations integrated object identity and layout, were broad in scope, involved both foreground and background information, and were effective across changes in image position.

---

1. Introduction

The ability to perceive the layout of a scene is crucial for survival and involves a relatively large portion of the brain (e.g., Gibson, 1979; Goodale, 1995; Kosslyn, 1987; Stiles-Davis, Kritchevsky, & Ursula, 1988). Consequently, the perception and representation of scenic layout is an important topic for cognitive psychology. A fundamental issue is whether basic cognitive mechanisms such as scene representations influence the immediate perception of a scene's layout. However, there has been little research relevant to this issue.

---

* Fax: +813-974-4617.
E-mail address: sanocki@chuma.cas.usf.edu.

0010-0285/03/$ - see front matter
doi:10.1016/S0010-0285(03)00002-1

There has been considerable recent research on psychophysical and ecological aspects of perceiving layout (e.g., Epstein & Rogers, 1995; Gibson, 1979; Sedgwick, 1986), and this research has been successful without invoking cognitive mechanisms. Perhaps spatial perception is not mediated by representations, in contrast to an identification system for which there is extensive evidence of mediation (Neisser, 1994). On the other hand, cognitive approaches have been fruitful in topic-areas related to layout perception, including the processing of spatial relations (e.g., Kosslyn, 1987; Logan, 1995; Pani, 1997), the control of saccades (e.g., Henderson & Hollingworth, 1997), spatial memory (Cooper & Lang, 1996; Shelton & McNamara, 1997, 2001; Tversky, Kim, & Cohen, 1999), and the semantics of scene processing (e.g., Biederman, 1981; Boyce, Pollatsek, & Rayner, 1989; Oliva & Schyns, 1997). Scene representations may facilitate the perception of layout by providing ordered surfaces and landmarks that help define distance relations during subsequent processing. This could speed spatial processing in general, and contribute to navigation through scenes (e.g., Andersen, Hahn, & Saidpour, 2001; Vishton & Cutting, 1995).

To begin exploring influences of scene representations on the perception of spatial layout, Sanocki and Epstein (1997) developed a scenic priming manipulation and measured the rapid perception of distance relations within scenes. A picture of a scene (a scenic prime) was presented for 1 s, or a control prime was presented. The prime display was followed by a brief blank display and then a target picture of a scene. The target had two critical probes not present in the primes (two chairs, or two balls), and the task was to indicate which of the probes (left or right) was closer to viewpoint in the pictorial space. The main result was that responses were 30–50 ms faster when the prime was of the same scene as the target than with control primes or primes from different scenes. The interpretation was that the scenic prime activated a representation of the scene's layout and that this representation facilitated the subsequent spatial processing of same-scene targets. These results provide initial evidence that representations influence the immediate processing of spatial layout.

The purpose of this paper is to further develop the idea that mental representations influence the perception of the spatial layout of scenes. The core idea is that experience with a scene activates a meaningful and substantial representation of the scene, and that this representation facilitates subsequent spatial processing within the scene. The representation is meaningful because it binds together information about objects and their layout. The representation is substantial in amount or extent, encompassing the breadth of a typical scene or the depth of a complex region. The theory is developed in the next section and contrasted with alternative approaches, which make weaker cognitive assumptions. Then a body of evidence from priming experiments is reported.

1.1. Meaningful and substantial representations theory (MSRT)

This approach is a moderately strong cognitive theory about representations that influence the immediate processing of spatial layout. This section begins by outlining some general assumptions. Then MSRT is described and distinguished from some alternative approaches.

1.1.1. Background assumptions

Most approaches to scene perception share the assumption that representations of information in scenes are rapidly constructed in a hierarchical sequence of processing levels. As a result, information sufficient for scene categorization is extracted from the world very quickly (e.g., Biederman, 1981; Oliva & Schyns, 1997; Potter, 1976). Processing begins with features such as edges, colors, textures, and some basic relations (e.g., Enns & Rensink, 1991; Treisman & Gormican, 1988). Processing continues as feature information is bound together into meaningful composite entities such as objects and surfaces. The composite entities might in turn be bound into larger scenic entities such as object groups and relations between objects and surfaces. Some theorists deny that this later stage of processing normally occurs (e.g., O'Regan, 1992; Pylyshyn, 1999). However, most theorists would agree that if this later stage did occur, the resulting representation would be a complex composition of more local mental representations. MSRT assumes that such complex compositions are constructed. Evidence for complex composite representations comes from the finding that under certain conditions, component representations can be wrongly combined, creating illusory conjunctions of features, objects and even scenes (e.g., Intraub, 1989; Treisman & Gelade, 1980).

MSRT also assumes that processing at different levels is cascaded and concurrent (e.g., McClelland, 1979), and involves some feedback between levels (e.g., McClelland & Rumelhart, 1981; Zeki, 1993). Interaction is assumed to be limited to related levels of processing, as suggested by over 30 years of research on contextual influences on letter and word identification during reading. In this domain, moderately interactive models (e.g., Rueckl, Mikolinski, Raveh, & Miner, 1997; Samuel, 1997; Stone, Vanhoy, & Van Orden, 1997) have been supported over more extreme alteratives. There has been little support for strong interactive hypotheses (e.g., Henderson, 1977; Neisser, 1967; Rumelhart, 1977), in which a relatively high level directly influences distant lower levels of processing (e.g., semantic levels modify feature or letter processing). The research is also inconsistent with strong non-interactive approaches because extensive evidence indicates that component levels of processing are influenced by immediately higher levels (e.g., letter processing is affected by word level units; e.g., McClelland & Rumelhart, 1981; Paap, Newsome, McDonald, & Schvaneveldt, 1982). Thus, a moderately interactive approach is assumed to be fruitful in scene perception.

1.1.2. The theory

In this paper the focus is on the nature of the representations induced by a scene. In MSRT, the representations are meaningful compositions. Their content is determined in part by task requirements and in part by learning and general statistical factors (e.g., large objects and certain functional properties are often important and likely to be included in the representation). Thus, scene representations integrate relevant and significant scenic information.

In the context of the spatial relations task, the scene representations integrate information about significant objects and landmarks with other information about layout. The representations function as a network that binds object identity and layout information. This claim distinguishes the present theory from alternative ones in which object identity information is represented independent of spatial information (e.g., Neisser, 1994; Ungerleider & Mishkin, 1982), or in which objects are represented as independent tokens loosely bound to locations (e.g., Rensink, 2000a).

According to the present theory, the representations are also substantial in extent. The representations can be fairly broad in scope, encompassing a typical indoor or outdoor scene as seen from a single vantage point. Relatively complex local regions can also be represented but with a reduction in scope. In general, both foreground entities and significant background entities can be represented. Thus, there can be a lot of information in the representations, much of which is likely to be implicit in nature (Chun & Jiang, 1998; Lewicka, Hill, & Czyzewska, 1992). The claim of broad scope is consistent with evidence that memory for a scene often extends beyond a viewed field to include immediately adjacent areas (e.g., Intraub & Bodamer, 1993; Intraub & Richardson, 1989). The claims about the substantial extent of scene representations distinguish the present theory from other ideas. For example, some theorists argue that scene representations are limited to one or several objects, an idea developed in the following section.

MSRT also makes the claim that scene representations are useful across short periods of time, serving to integrate at least some information across eye movements (e.g., Hochberg, 1978). Similarly, scene representations are predicted to be active across space, surviving moderate changes in the retinal position of the scenic image.

In summary, experience with a scene is assumed to induce a meaningful and substantial scene representation. The representation binds identity and layout information across broad portions of the visual field.

Scene representations can be studied experimentally with priming paradigms (e.g., Sanocki & Epstein, 1997). In MSRT, the assumption is that a scenic prime can activate a broad-scope and somewhat detailed representation of the scene. When the target is presented, consistent portions of the prime-based representation are combined with incoming information and, in particular, with information about the critical probes. The response is initiated after spatial relations between the probes have been established, based on depth cues such as occlusion and relations to the ground plane.[1] Relative to control prime conditions, the prime-based representation provides a head start on this spatial processing, with the result that observers can establish the spatial relations between probes more quickly than when no relevant prior representation was available. Thus, scenic primes induce relatively broad-scope representations that influence subsequent, more local processing related to the target probes, a type of interactive effect (for possible modeling approaches, see, e.g., Massaro & Sanocki, 1993; McClelland & Rumelhart, 1981; Sanocki, 1999, 2001).

The main purpose of the present experiments was to begin testing predictions about scene representations that follow from MSRT. The experimental strategy

---

[1] In all experiments one obvious depth cue, height in the visual field, was invalidated. This was done by placing some close probes higher in the visual field than far probes.

was to manipulate information in the primes and targets and then examine the magnitudes of priming effects in the various conditions. When a strong positive priming effect occurs, we can tentatively conclude that relevant prime information was represented mentally and used to facilitate processing in the spatial task.

1.1.3. Evidence for moderate interactivity

MSRT predicts moderate interactive effects of the contextual scenic primes. This prediction is supported by evidence of moderate interactivity in research on object recognition. In the recognition of isolated objects, sensitivity to local object details can be increased by the immediately prior processing of larger and related object structures (Sanocki, 1991, 1993, 1999, 2001; see also McClelland & Miller, 1978; Williams & Weisstein, 1978). Similarly, sensitivity to the presence of objects within scenes can be increased by prior or concurrent processing of the larger scene (e.g., Biederman, 1981; Boyce et al., 1989). The effective scene representation appears to emerge from (Boyce et al., 1989) or be learned from (Biederman, 1972) arrangements of objects (see also Chun & Jiang, 1998). The present research extends the general idea of context-based facilitation, to the perception of scenic layout.

1.2. A contrasting approach: bottom-up processing and attention

The idea of context-based interaction can be contrasted with an approach emphasizing bottom-up processes. Pylyshyn (1999) argued that scene components are processed in a bottom-up manner that is isolated from larger contextual representations. Bottom-up processing occurs within an encapsulated early vision module. When necessary, scene-based representations can be constructed from outputs of the early vision module, by separate high-level processes. However, in this approach scene-based representations are limited in scope and utility. Information about the scene is continuously available from the world itself (e.g., O'Regan, 1992), obviating the need for scenic representations. Pylyshyn (1999) argues that evidence supports (only) one mechanism of top-down influence in visual perception and that is attention. Attention can influence processing within the early vision module and can construct limited scene-based representations (e.g., Rensink, 2000a). Logan and Zbrodoff (1999) proposed that attention interacts with vision by constructing propositional relations binding several arguments that are scene-based components (e.g., Logan, 1995). Thus, in the bottom-up approach, scene representations are limited attentional constructions involving at most a few components.

The bottom-up view is parsimonious, and bottom-up models of object recognition have been successful (e.g., Biederman, 1987; Tarr & Bulthoff, 1998). Moreover, the existence of scene-context effects on the sensitivity of object recognition has been questioned in a recent experiments (Hollingworth & Henderson, 1998). Forced-choice methods that have revealed sensitivity effects within word and object identification (e.g., Reicher, 1969; Sanocki, 1993) have not produced sensitivity effects of scenic contexts on object identification (Hollingworth & Henderson, 1999). The difficulty of detecting changes in scenes is also consistent with the assumption of limited representations (see, e.g., Grimes, 1996; Rensink, O'Regan, & Clark, 1997;

Simons, 2000). Thus, the bottom-up approach to scene perception has considerable support.

1.3. Theoretical implications of representational scope

MSRT can be distinguished from the bottom-up approach on the basis of the scope of the scene representations that influence spatial relations processing. MSRT predicts that a scenic prime could activate a broad representation, which could in turn facilitate spatial processing throughout a broad-scope scene. In contrast, in the bottom-up approach, representations are limited to at most a few components; therefore, effects of broad-scope scene representations are not predicted. The present experiments serve to develop the idea of a broad-scope representation, and scope is manipulated directly in Experiments 3 and 4.

1.4. Abstract versus image-based scene representation

There was one major empirical issue examined in the experiments: What kind of mechanism underlies broad-scope scene representations? That is, how might broad-scope representations be maintained in memory? The traditional computational approach in scene perception is to assume that visual processing quickly becomes symbolic, with retinal inputs being transformed into symbolic tokens such as edges that compose higher level symbols for surfaces and objects (e.g., Hummel & Biederman, 1992; Marr, 1982; Palmer, 1975). Fundamental motivations for this assumption include the immense computational burdens of scene perception (e.g., Tsotsos, 1990) and the problems of variation across instances and across situational factors such as lighting. These difficulties are reduced or factored out in symbolic representations (see, e.g., Hummel & Biederman, 1992; Marr, 1982; Palmer, 1977). Thus, scene representations could involve an abstract representation of surface layout (e.g., Marr, 1982), and information about objects may be represented by object files (Kahneman, Treisman, & Gibbs, 1992), which appear to be generally invariant across image-based details (Gordon & Irwin, 2000; Henderson, 1994, 1997). Consistent with the emphasis on abstraction, there is evidence of abstract spatial memory (e.g., Franklin & Tversky, 1990; Talyor & Tversky, 1992). In this view, broad-scope scene representations would be activated by a flow of information from early image-based areas to more abstract representational areas. Facilitation in spatial priming would be a interactive effect driven by the more abstract representational areas.

Alternative ideas about representation have emerged from research on the nature of thinking, reasoning, and language comprehension. The argument is that abstract entities are important but they must be supplemented by information that is related to image-based perceptual properties of scenes—information processed by neural structures shared by perception, imagination, and cognition (e.g., Barsalou, 1999; Glenberg, 1997; Miller & Johnson-Laird, 1976). Thus, image-based properties such as color, lightness, and texture can be important to representation.

The possibility examined here is that scene representations are supported in part by intermediate-level representations of image information. The information may include color, brightness, and shading information, organized according to meaningful units such as objects and surfaces. Because of this organization, the representation is not a low-level sensory buffer (e.g., McConkie & Rayner, 1976; see Irwin, 1996 for a review of evidence against a sensory buffer). The linkage between image-based information and more abstract units for objects and surfaces may involve a network that interrelates abstract entities and corresponding image-information (e.g., Barsalou, 1999). Facilitation in spatial priming could be an interactive effect involving both intermediate and more abstract representational areas.

To examine the effects of image-based information, the direction of lighting was varied in the experiments. When a scene is held constant but direction of lighting changes between a prime and target, the luminance and color values that define image information can change drastically while layout remains constant (see Fig. 1). Drastic changes should disrupt a representation that involves image-based information. Thus, if image-based information is important to scene representations, then changes in lighting direction between primes and targets should greatly reduce spatial priming effects.

1.5. Summary

MSRT predicts that experience with a scene induces a complex representation that can facilitate subsequent spatial processing throughout a scene. MSRT claims that the representations bind objects and surfaces to locations, that they can be broad in scope, and that they survive changes in image position on the retina.

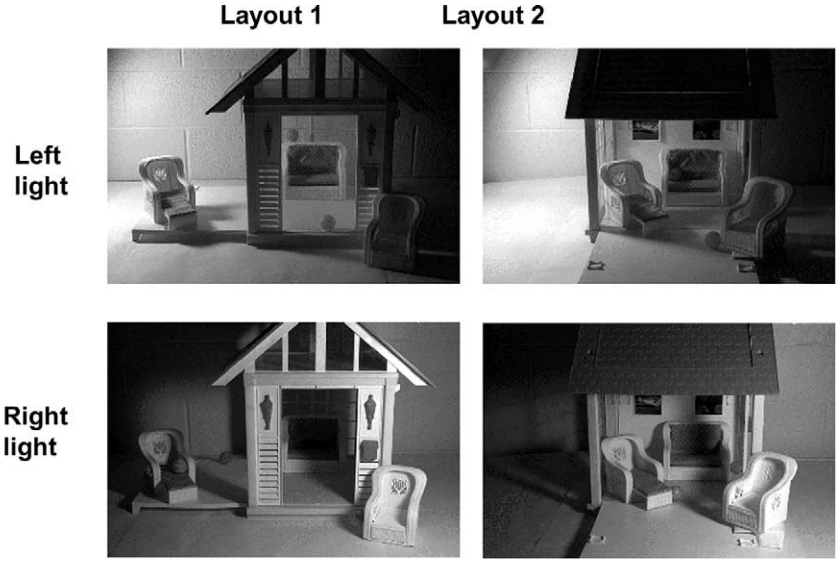


Fig. 1. The scenes used in the experiments. Two layouts were combined with two lighting directions to yield four combinations.

The representations of MSRT may be primarily abstract or they may integrate abstract and image-based information; this is an empirical issue examined in the experiments.

1.6. Previous spatial priming results

The present experiments are an application of information processing methods in general and priming techniques in particular (e.g., Biederman & Cooper, 1991; Meyer & Schvaneveldt, 1971; Sanocki & Epstein, 1997) to spatial processing. As noted earlier, a prime is presented for 1 s, followed by a brief interval (84–150 ms in duration) and then a to-be-processed target. The target contains two probes not present in the prime. The probes are laterally separated, and observers indicate whether the left or right probe is closer in the pictorial space to viewpoint, by pressing a left or right key. Speed and accuracy are stressed. Because location is defined egocentrically, abstract interpretations of "left" and "right" are unnecessary. Note that facilitation effects cannot be attributed to prime-induced response bias because scenic primes facilitate both possible responses (left or right probe closer). High-quality color photographic stimuli are used so that cognitive effects could not be attributed to impoverished stimuli.

In the initial experiments (Sanocki & Epstein, 1997, Experiments 1 and 2), eight everyday scenes were used. On identical scenic prime trials, the prime was identical to the target, except for the critical probes. The probes were pairs of chairs that were salient and appropriate for the scene; they were placed in varying positions throughout the scenes. Reaction times with identical primes were compared to times with a simple control prime (Experiment 1) or primes depicting other scenes from the experiment (different-scene primes, Experiment 2). Responses were 30–50 ms faster with identical primes. Similar results were obtained with two simple, artificial scenes and three different control primes (Experiments 3 and 4). In these later experiments, the critical probes were red balls that should pop out from the blue-green backgrounds. The results are consistent with the idea that prime-induced scene representations facilitate spatial processing of the targets.

The first experiment of that series also provided information about the abstractness of scenic representations. The targets were effectively primed by both the identical photograph primes and by line drawings of those primes. Because the line drawing primes contained no color or lighting information and little surface information, their effectiveness implies that image-based information is not essential to scene representation. However, the line drawings were less effective than photographic primes, suggesting that the photographs provided additional relevant information, which might be image-based in nature.

The results provided evidence against several alternative explanations of the facilitation effects. First, the effects did not depend on sensory persistence from the primes because line drawing primes were effective even though they shared few sensory details with the target pictures. In fact, sensory persistence should be minimal because persistence varies inversely with duration (e.g., DiLollo, 1980; Efron, 1970); consequently, the 1-s prime durations should produce minimal persistence.

Also, primes facilitated responding when shifted laterally relative to targets (Sanocki & Epstein, 1997, Experiment 4), which should not occur if facilitation depended on a sensory integration of prime and target (see also present Experiment 5). The possibility of sensory persistence from the primes was measured directly in the present situation, using the DiLollo (1980) integration method, and there was no evidence of sensory integration (see Method of Experiment 1 and Appendix A).

A second alternative explanation is that scene representations are extremely abstract, such as a general schema for surfaces receding in depth. Different-scene primes provided a general surface receding in depth and could have activated such a general schema; however, they did not facilitate processing (Sanocki & Epstein, 1997, Experiments 2 and 3).

The present experiments were designed to eliminate an additional alternative explanation, in which scenic primes work by providing knowledge of the objects that will appear in the target. Note that blank primes or different-scene primes did not provide such information in the previous experiments. Thus, scenic primes could work by predicting the objects in the target. To counter this explanation, the same objects were used in all of the targets and primes in the present experiments—a doll house and three pieces of furniture (e.g., Fig. 1).

1.7. The present experiments

The present experiments had five goals. The first was to provide additional evidence that scene representations can facilitate spatial relations processing, consistent with the general thrust of MSRT. In each experiment, the same objects appeared in all targets; therefore, scenic primes could provide information about layout but not knowledge of content.

The second goal of the experiments was to test the idea that scene representations bind objects and their locations. The idea was examined in Experiment 1 by comparing facilitation from primes that contain the appropriate objects in the correct locations with that from primes containing the same objects but in somewhat different locations.

The third goal was to test the claim that scenic representation can be broad in scope. The critical probes were distributed throughout the scenes in each experiment, so findings of facilitation imply that much of these scenes was represented. To test the importance of scope directly, a measure of scene-scope was introduced and used in Experiment 3. If the representation is indeed broad in scope, then facilitation effects should occur with small- and large-scope scenes. Also, facilitation should occur for uncued portions of scenes as well as cued portions (Experiment 4).

The fourth goal was to further examine the invariance of the representations across changes in image location on the retina. In Experiment 5, the primes were shifted by varying amounts relative to the targets.

Fifth, the idea that image-based information is important to scene representations was examined in Experiments 2–5, by varying the direction of lighting between primes and targets. If image-based information is important, then we should find substantial facilitation when such information is constant between prime and target, but reduced amounts of facilitation when image-based information varies between prime and target.

2. Experiment 1: Are objects bound to locations?

If objects are bound to locations within scene representations, then scenic primes should facilitate spatial relations processing when they provide the appropriate objects in the appropriate locations, but not when they provide the appropriate objects in different locations. Thus, the layout of objects should be important above and beyond their presence and identity. To test this prediction, a different-layout prime condition was used in the experiment, along with two prime conditions that were standard across all experiments. The standard *similar primes* were identical to the targets except for the absent critical probes—the similar primes contained the same objects in the same locations as the targets. The standard *control prime* was a beige field matched to a background color of the scenes, with red outline rectangles. The advantage in reaction times for similar primes relative to the control prime was the standard measure of facilitation from a scenic prime throughout the present experiments.

The third prime condition in Experiment 1 was used to gauge the importance of layout information. The *different-layout* prime had the other layout (from the target) but the same objects. Thus, it provided the same gist and the same semantic labels (e.g., "dollhouse," "chairs"). In addition, some coarse spatial relations were similar (e.g., central house, similar configurations of chairs and sofa). In Fig. 1, the different-layout prime and target would be the two scenes within either of the two rows. To the extent that facilitation depends on specific spatial layout information, the similar (same layout) primes should provide more facilitation than the different-layout primes. An advantage for similar primes cannot be explained by priming of object identities or tokens loosely bound to location because different-layout primes should prime loosely bound objects equally well.

The critical probes in this experiment were red ovals superimposed on the scenic surfaces. They looked as if they were created by a red flashlight oriented perpendicular to the relevant surface (Figs. 2A and B show examples, with red represented by darkening). This colorizing method leaves the underlying surface texture and color intact and was used in most of the experiments.

2.1. Method

Stimuli and design. The stimuli were derived from four photographs that varied factorially in layout and lighting direction (Fig. 1). Camera position was constant, as was the table-top (covered in white paper) and the background wall. The two layouts were defined by the orientation of the doll house and the arrangement of furniture. The two lighting directions were each produced by one of two desk lamps. One lamp was positioned on the left of the scene (at an angle about 35° above the groundplane) and one on the upper front right (at about 75° above groundplane).

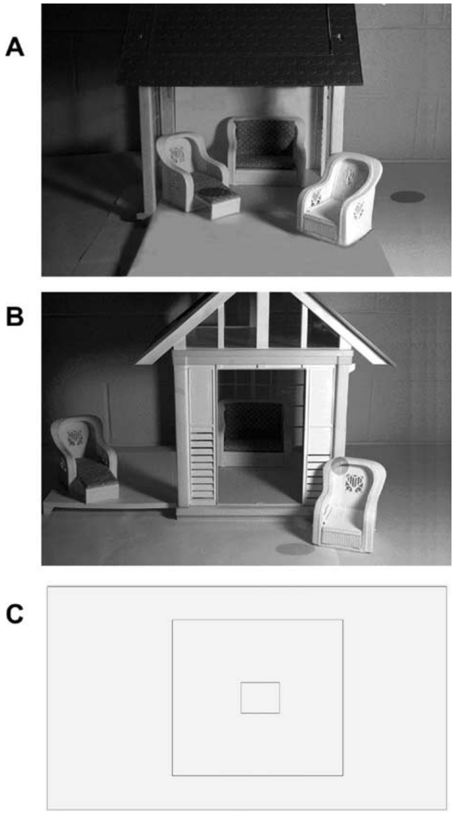


Fig. 2. Examples of stimuli in Experiment 1. (A) and (B) are two targets differing in layout; (C) is the control prime.

Four base-pictures (similar to those in Fig. 1) were used as the scenic primes. Twelve targets were derived from each base-picture by superimposing the critical probes (pairs of red ovals, described below) onto the image (e.g., Figs. 2A and B). The same 12 probe-pair placements were used for the two lighting-versions of a layout. Thus, there were 12 probe-pair placements for each layout and 24 in the total stimulus set. In half of the placements the right probe was closer to viewpoint and in half the left probe was closer.

In the similar prime condition, the targets were preceded by the base-picture they were derived from. In the different-layout condition, targets were preceded by the base-picture depicting the other layout but the same lighting. In the control condition, targets were preceded by the beige control prime. The beige of the control prime was matched to the predominant color of the background walls. The prime had three concentric red outline rectangles of varying sizes (Fig. 2C). During the experiment, each target appeared equally often and was preceded equally often by each of the three prime-types.

The scenic images were the "third edition" versions of the scenes. The original images were photographs of the scenes, exposed at the same exposure settings with a normal focal length and a wide depth of field. The photographs were transferred to CD-ROM by Kodak and saved as 256-color, 640 × 427 pixel images with a good amount of contrast (Fig. 1). Visual angle was 23.5 by 16.1°, filling the horizontal extent of the 13 in display monitor. Intensity values ranged from 0% (black) to 100% (white) within each image. The third edition images (see Figs. 2A and B) had several refinements. First, the exact same objects appeared in each scene. (This differed from the second edition images, which contained several minor objects present in only one view.[2]) In addition, the only reddish colors in the scenes were from the critical probes and the positions of all scenic images were precisely aligned. (These properties differ from the first edition images, as explained in Experiment 2.)

Pairs of critical red oval probes were placed throughout the scene. Each probe-pair was positioned to have a fairly obvious separation in depth. Horizontal separations within pairs varied from large to small. Two precautions prevented inference of the response from a single probe: In some pairs, the furthest probe was higher vertically in the image, and some individual probes participated in two probe-pairs, with different interpretations (close vs. far) in each case. The probes were created to look as if they were part of the relevant scenic surface. Each probe was created by shifting the hue within an oval region toward red while preserving texture, lighting, and color patterns. Elongation of the ovals increased with the departure from normality of the surface (e.g., Fig. 2A). Before the hue change, the intensity was increased in some dark regions and decreased in some light regions, to insure that redness could be perceived easily. The adjustments were done within Adobe Photoshop.

There were 432 test trials, produced by combining 4 scenes of 12 targets each with 3 prime conditions and 3 repetitions. Order of the trials was randomized. In addition, there were 24 practice trials randomly selected from the test conditions.

Procedure. Observers participated individually in sessions lasting approximately 45 min. The 13 in video monitor was enclosed in a large black frame and observers looked through a horizontal cutout. This fixed viewing distance at 60 cm. Each trial began as the prime was presented along with a short auditory "beep." The prime was presented for 1 s, followed by an 84-ms blank (white) interval, and then the target until the response. The images were transferred to the screen within 17 ms by a Macintosh computer, by switching the color table (the screen image) from all white to the image colors. The observers were instructed to indicate which critical object was closer in the scene to viewpoint, by pressing one of two keys (right or left) on a number pad. Speed and accuracy were stressed. Before the trials began the observers were

---

[2] The minor objects were the mailbox and lights in layout 2 and the pictures on the wall within the house in layout 1 (see Fig. 1). For Experiment 1 only, these objects were erased to create the third edition images, which contained exactly the same objects in each layout (e.g., Figs. 2A and B). Also erased were some notches in the floors and one wall, and the small white object supporting the right front chair in layout 1 (Fig. 1, left side). The objects were erased by replacing them with background, in Adobe Photoshop.

shown each prime (including the control prime) and told that it would serve as a ready signal that could help them prepare for the target. After each response, auditory feedback was presented; a nice beep signaled correctness, whereas incorrect responses were followed by redisplay of the target along with a gong sound. The target was redisplayed after incorrect responses.

Sensory persistence study. To determine the effectiveness of the prime-target interval, sensory persistence was measured in a separate study with seven observers, using stimuli and time parameters of the main experiment. The task required integrating an average of 12.5 rectangular regions, from an initial one second display, across the 84-ms blank interval, with a similar number of regions from a subsequent display (e.g., DiLollo, 1980) that was response-limited (as in the main experiments). If the regions of the first display persisted completely, performance should be near 100%, whereas chance performance was 50%. Accuracy averaged 62.7%. Further analyses reported in the Appendix A indicate that the 12.7% above-floor responses can be accounted for by assuming that observers held a few regions in visual short term memory. This is distinct from sensory persistence, which should allow integration of regions across the blank interval and produce performance approaching 100%.

Participants. The observers were students from introductory psychology courses at the University of South Florida, who participated for extra course credit. All reported normal or correct-to-normal vision. A total of 16 students participated (14 female) in the main experiment.

2.2. Results

Responses over 4 s were excluded from the analyses (less than 0.5% of the data in each experiment). The main analyses were planned comparisons of mean reaction times for correct responses in different prime conditions. Facilitation effects (primed RTs minus control RTs) are illustrated in figures, together with the standard error of the differences ($SE_{\text{Difference}}$). Unless stated otherwise, all effects reported were highly reliable ($p < .01$).

Responses averaged 918 ms in the control condition. Facilitation effects are shown in Fig. 3. Relative to the control baseline, the responses were 50 ms faster with similar primes. This advantage indicates that immediately prior experience with the scene can facilitate spatial relations processing, presumably because an appropriate scenic representation has been induced and is helpful. Different-layout primes produced an advantage of 26 ms over control primes that was marginally reliable ($p = .08$). This suggests that information about the content, gist, and general lighting may be somewhat helpful. More important, responses were 23 ms faster with similar primes than with different-layout primes, $F(1, 15) = 7.49$, $p = .02$ ($SE_{\text{Difference}} = 9$ ms), indicating that the specific spatial layout information binding objects to their locations caused facilitation beyond the effects of scene content.

Accuracy was high in each condition (above 97%). It did not differ between conditions ($p > .10$ in Analysis of Variance on Prime Condition).

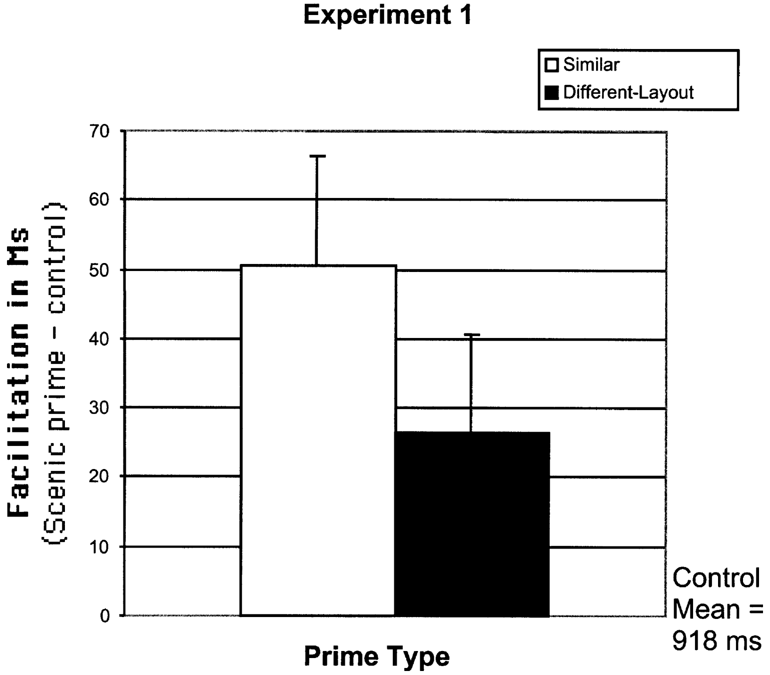


Fig. 3. Facilitation as a function of prime type in Experiment 1 (error bars are standard errors of difference).

2.3. Discussion

There were two main results. First, the 50-ms advantage for similar primes over the control prime is consistent with the idea that immediately prior experience with a scene induces a representation of the scene that facilitates subsequent processing. Because most of the scene was relevant (critical probes appeared throughout the scene), the present effect suggests that the representation was broad in scope. This claim is developed further in Experiment 3.

The second finding was the 23-ms advantage for similar primes over different-layout primes. This effect can be attributed to specific information about spatial layout, because layout was the only difference between the similar and different-layout primes. Semantic factors such as gist or the labels and functions of objects were the same in these primes. In fact, the two layouts were generally similar—the chairs had the same ordinal arrangement and the house had a similar position and orientation. Thus, the effect that was obtained suggests that fairly precise layout information is contained in scene representations. This may include information about relations between the chairs and the house, and perhaps some metric information. This result suggests that objects and surfaces in the scene are bound to their locations. The effect cannot be explained by object representations that are active independent of their locations.

Could these results be explained by assuming a representation of layout that is independent of objects? For example, the critical spatial relations between surfaces might be represented without reference to objects. This idea was examined with a different set of doll furniture scenes in two further experiments. There were three pieces of furniture as shown in Fig. 4; the scenes differed in the identities of the objects (different rows in Fig. 4) or in the layout of objects (different columns in Fig. 4). Experiment 1B was designed to replicate Experiment 1 with the new stimuli and 15 new participants. The mean RT with control primes was 735 ms. Responses were faster with similar primes, a facilitation effect of 23 ms ($SE_{\text{Difference}} = 7\,\text{ms}$, $F[1, 14] = 9.95$). Responses with the different-layout primes were 11 ms *slower* than the control condition ($SE_{\text{Difference}} = 8\,\text{ms}$, $F < 1$). Thus, the main result of Experiment 1, the advantage for same layout and objects over different-layout, was replicated ($F[1, 14] = 25.94$, for similar prime vs. different-layout prime). The different-layout condition fared less well in this experiment than in Experiment 1, perhaps because there was a greater overall change in object position relative to Experiment 1—the objects changed order in Experiment 1B (Fig. 4) but not in Experiment 1 (Fig. 1). Accuracy was uniformly high in this experiment, being 98% or more in each condition ($F < 1$).

Experiment 1C examined the possibility that there could be a representation of layout that is independent of objects. The critical condition was a *same-layout different-object* prime condition, involving the same critical spatial relations but different objects. The objects within each column of Fig. 4 are arranged so that there would be corresponding surfaces occupying the same planes in 3D space—for example, the

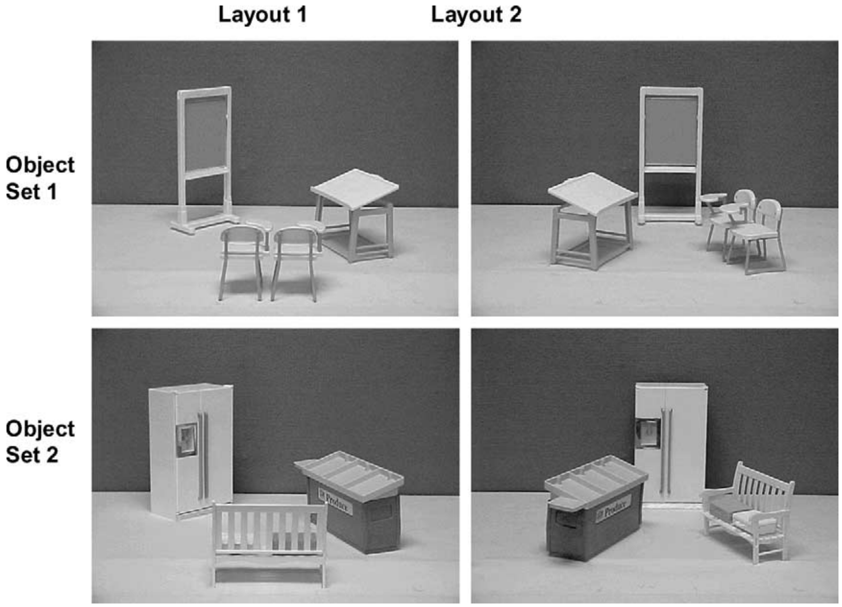


Fig. 4. The scenes in Experiments 1B and 1C. Two layouts were combined with two object-sets to yield four combinations.

fronts of the blackboard and refrigerator, or the tops of the easel and fruit stand. The oval probes were positioned only on these corresponding surfaces, and each probe-pair was used with both corresponding scenes. For example, a pair on the blackboard and easel would also appear on the same locations of the refrigerator and fruit stand. If it is possible for these critical relations to be primed independent of the objects, then there could be substantial facilitation effects for same-layout different-object (SLDO) primes. The standard similar primes and control prime were also used, and there were 15 new participants. The control mean was 705 ms. There was a facilitation effect for the similar primes of 33 ms ($SE_{\text{Difference}} = 10\,\text{ms}$, $F[1,14] = 11.19$), but no reliable effect for the SLDO primes (+8 ms effect, $SE_{\text{Difference}} = 11\,\text{ms}$, $F < 1$). The 25-ms difference between the similar prime and SLDO primes was reliable ($SE_{\text{Difference}} = 4\,\text{ms}$, $F[1,14] = 48.99$). Thus, there was no evidence that the layout could be primed independent of the objects. Accuracy was uniformly high in the experiment, being 98% or more in each condition ($F < 1$). These results are consistent with the claim of MSRT that the layout and the identities of objects are bound together in scene representations.

3. Experiment 2: Effects of lighting direction

The prior results as well as Experiment 1 are consistent with the idea of a broad-scope scene representation that influences spatial relations processing. How might a broad-scope representation be maintained? Does image-based information play a role? In more traditional symbolic approaches, meaningful scenic representations are quickly abstracted away from image-based information. In contrast, mental model approaches are based on the idea that neural representations common to perception and cognition play a critical role in representation. In the present context, image-based information might support scenic representations. The present experiment introduces the manipulation of consistency versus change in the direction of

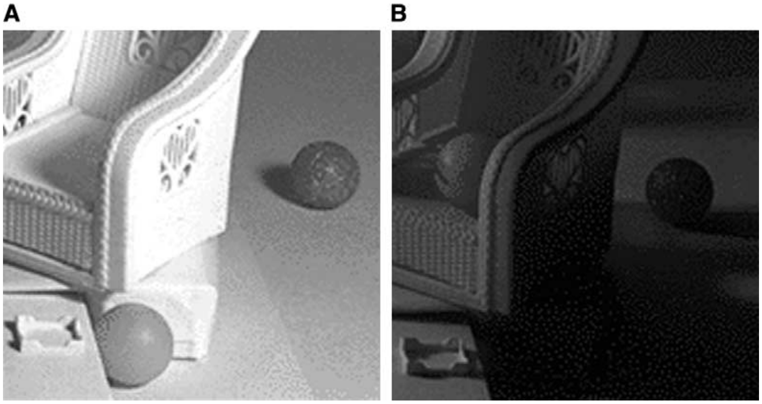


Fig. 5. Closeups of critical probes in Experiment 2 (in the right most example, one probe is in the chair).

lighting. When the scene is held constant but direction of lighting changes between a prime and target, image information can change drastically while layout remains constant. As noted earlier in reference to Fig. 1, the present scenes were lit in two different ways, by one of two desk lamps in different, above-lateral positions. The variation in lighting produced marked differences in local image features (e.g., the sign, strength, and existence of edges, textures and shading variations) and in larger image features such as shadows, objects, and surfaces. For example, the closest objects and surfaces have the lowest intensities in Fig. 1 top row but the highest in Fig. 1 bottom row. Such changes should disrupt a representation that involves image-based information. Thus, if image-based information is important to scene representations, drastic changes in image-based information should greatly reduce spatial priming effects.[3]

Because of the desire to produce strong and highly task-relevant lighting effects, the critical probes in this experiment were small balls that were placed directly in the scene. The lighting variations from location to location produced marked differences in the critical ball's appearance, as illustrated in Fig. 5. In summary, Experiment 2 introduced a *different-lighting* prime condition, in which the lighting direction changed between the prime and target, causing considerable changes in image-based information. The experiment also included the standard similar prime and control prime conditions.

3.1. Method

Design and stimuli. The three prime conditions—similar, different-lighting, control—preceded each target equally often. The stimuli were the first edition version of the doll house images. As noted, the critical probes were two small balls that were moved around in the scene to produce each target. The balls were placed on various surfaces, subject to normal gravitational and lighting constraints, and a photograph was taken of each critical-probe-pair placement. The color of one ball was a dirty yellow and the other a dirty purple, but their luminance and color values varied markedly with local lighting conditions (e.g., Fig. 5). Eight different placements of the probe-pairs were used with each of the four layout/lighting combinations. Thus,

---

[3] The importance of lighting and the effects of lighting-related information have been studied in recent research. Lighting-related information (shading, shadows) is processed efficiently in early vision, but the specific patterns of shading or shadow do not provide strong constraints on visual search and depth perception (e.g., Aks & Enns, 1992; Cavanagh & Leclerc, 1989; Kleffner & Ramachandran, 1992). Lighting-related information may directly constrain face identification (e.g., Hill & Bruce, 1996), and aspects of object recognition (e.g., Srinivas, 1996; Tarr, Kersten, & Bulthoff, 1998). Thus, while lighting information is processed during perception, the computational role of lighting information in scene representation is unknown. Prior research also indicates that changing lighting direction between above and below the scene is more disruptive than changes between lateral directions (Kleffner & Ramachandran, 1992) or between two top-light directions (Hill & Bruce, 1996). Lighting from below is highly unusual and disrupts perceptual processing even when direction is constant (Hill & Bruce, 1996; Kleffner & Ramachandran, 1992). Because of the present interest in normal scene perception, above-ground lighting directions were used in the present experiments.

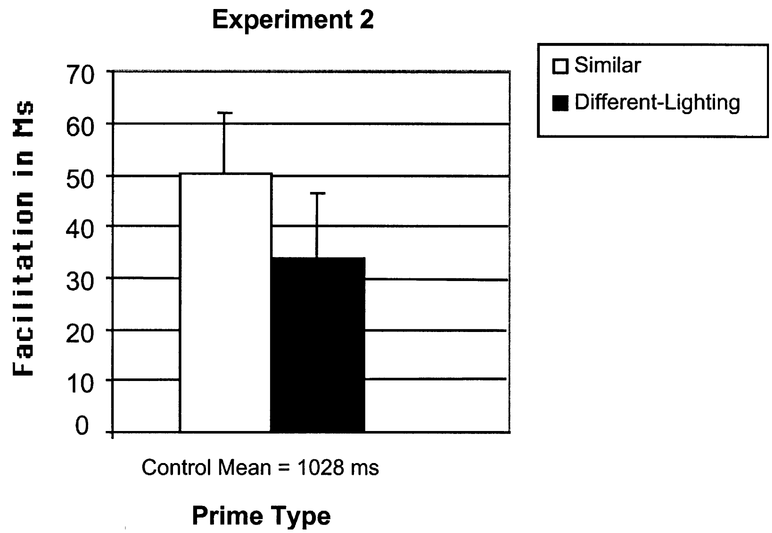


Fig. 6. Facilitation as a function of prime type in Experiment 2.

there was a total of 32 targets or probe-pair placements in the experiment. This contrasts with the other experiments, in which the two lighting-versions of the same layout had the exact same set of probe-pair placements. Also unlike the other experiments, the horizontal positions of the images on the screen varied somewhat, in an unsystematic manner. The differences ranged as high as 1.2° and can be attributed to variation in how the images were placed on the CD.[4]

There were 480 test trials in the experiment, produced by combining 32 targets, 3 prime conditions, and 5 repetitions. Test trials were preceded by 24 practice trials. Twenty two students participated (20 female).

3.2. Results

Responses averaged 1028 ms in the control condition. This relatively long baseline mean is probably due to the increased difficulty of finding some of the critical probes because of the strong lighting effects.

The facilitation effects are shown in Fig. 6. Similar primes produced an advantage of 50 ms relative to the control condition. Different-lighting primes also produced a facilitation effect, of 34 ms ($p = .02$). This indicates that correct image-based information is not necessary for a facilitation effect. The image-based lighting effect—the difference between the same- and different-lighting conditions—was 17 ms and was marginally reliable ($F[1,21] = 3.62$, $p = .07$; $SE_{\mathrm{Difference}} = 9\,\mathrm{ms}$). This suggests that the change of image-based information in the different-lighting condition may

---

[4] The horizontal displacements were random in size from image to image within the photographic sequence and can be attributed to hand-done placements of negatives on the scanner by the developer.

reduce the facilitation effect. Data from subsequent experiments provide evidence of the reliability of this small effect when the position of primes and targets on the retina is similar.

Accuracy was high in each condition (above 97%) and did not differ between conditions.

3.3. Discussion

In Experiment 2 both the same- and different-lighting primes produced facilitation. The effect for different-lighting primes is interesting because it occurred even though there was a marked visual onset of shadow and light in the targets. This indicates that facilitation can occur when visual information in addition to the critical probes appears in targets, relative to the primes. Note that in similar prime conditions, the image remains constant with the exception of the onset of the critical probes.

The main reason for the lighting manipulation was to measure facilitation when the image-information changed drastically between prime and target. The results can be viewed in at least two ways. On one hand, there was a sizeable and reliable facilitation effect with different-lighting primes despite the image changes in that condition. This suggests that a substantial component of the prime-induced representation is abstract in nature, as expected given prior evidence for abstract spatial representations (e.g., Franklin & Tversky, 1990; Talyor & Tversky, 1992).

On the other hand, there was some evidence of a reduction in the facilitation effect when image-based information changed—a marginally reliable difference between the same-lighting (similar prime) and different-lighting conditions. Such an effect, if reliable, would support the idea that image-based information contributes to the scenic representation.

One potentially important factor is the lighting of the critical probe objects in individual targets. The strong lighting effects sometimes rendered the critical probes difficult to find, as intended. Consequently, mean reaction times for individual targets varied markedly. Were lighting effects stronger for some targets than others—e.g., for targets with critical probes that were difficult to find because of the lighting? The data were collapsed across prime condition and ordered by target RT, and then grouped into quartiles. Subjective examination of individual targets suggested that RT did increase as the visibility of the critical probes decreased. The results by quartile are shown in Table 1. Interestingly, the differences between similar primes and different-lighting primes were strong in the two faster quartiles and absent in the two slower quartiles. Thus, lighting effects were strongest with highly visible probes. Perhaps lighting effects were absent with difficult probes because the long processing time permits lighting-based differences to be overcome. Consequently, the following experiments were conducted with critical probes that were relatively easy to find. This case seems most representative of normal layout perception and it should be most sensitive to lighting effects.

To anticipate, the result were as follows. The small advantage for same-lighting over different-lighting replicates when the prime and target occupied similar

Table 1
Results for each quartile in Experiment 2: The control mean, facilitation effects with similar-lighting and different-lighting primes, and the difference between similar- and different-lighting conditions

Control mean	Facilitation effect		Lighting effect
	Similar-lighting	Different-lighting	(Similar – different)
Fastest quartile			
852	33	12	21
Second fastest			
951	58	18	40
Third fastest			
1070	58	62	−4
Slowest quartile			
1237	56	52	4

positions on the screen. However, when the prime position was shifted relative to the target in Experiment 5, the advantage disappeared, implying that the advantage is limited to certain conditions.

4. Experiment 3: Manipulation of scene-scope

The critical issue of the scope of scenic representations was addressed directly in this experiment. MSRT claims that scenic representations can be large in scope, covering a major portion of the visual field and facilitating spatial processing across that region. This claim is critical because alternative approaches are based on the idea that scene representations are limited constructions controlled by attention (e.g., Pylyshyn, 1999). In the alternative approaches, scene representations might be limited to one attended object, or perhaps several, depending on the assumptions made about attention. In any case, there should be a clear limit in the scope of the representation.

In the present experiment, if scenic representations are sharply limited in scope, then they should be highly effective when the relevant prime and target area is limited in scope, because all of the relevant area could be represented. Thus, there should be sizeable facilitation effects with small-scope scenes. However, when the relevant area is large in scope, limited representations should not be highly effective and facilitation should be markedly reduced.

Alternatively, if scenic representations can be broad in scope, as in MSRT, then the representations could handle small and large scope scenes. The representations could be about as effective with large-scope scenes as with small-scope scenes.

Scene scope was varied by using the same doll house images but marking regions of various sizes. Four groups of observers were assigned different scopes, ranging from very small scope (only a portion of one chair was relevant) to very large scope (the entire scene was relevant, as in Experiment 1). The relevant region was marked for all stimuli, throughout the experimental session.

In order to define scene scope, the concept of an object-surface-activity map was developed. The concept begins with the assumption that the significant components of scenes are the major visible objects and surfaces in the scene. To demarcate these scenic units, a conservative criterion of preserving continuous objects and surfaces was used, to avoid overestimating scope. As can be seen for layout 2 in Fig. 7A, there were eight units in the scene, bounded by black lines. Note that two units extend across occluding objects (the units being the back wall and the floor). There were also eight units in layout 1.

The second step was to specify activity within the scene by superimposing the critical probes that were used in the experiment. This is done for one large-scope layout in Fig. 7B. For each probe-pair, there were two individual probe occurrences, indicated by ovals, and one relation between the probes, indicated by the connecting line. Doubly-connected ovals participated in two probe-pairs. The map was then used to generate predictions.

For each layout in the large-scope condition (e.g., Fig. 7B), there was a total of 12 probe-pairs—i.e., 24 individual probe-occurrences plus 12 relations between probes. The probes were distributed across the eight units so that attending to only one unit or a few units would be effective only a small portion of the time. If the busiest unit in the scene was attended (unit 2 in Fig. 7A, the front of the house), 5 of 24 individual probes would occur there (one location is probed twice). However, a probe-pair relation never occurred within the unit. Therefore, a scene representation consisting of this unit would be ineffective most of the time. In fact, if the representation consisted of the four central units (2–5), only 8 of 24 probe occurrences and 1 of 12 relations would occur within the units. Thus, even a moderate scope scene representation would be ineffective much of the time in the large-scope condition (Experiment 3A).

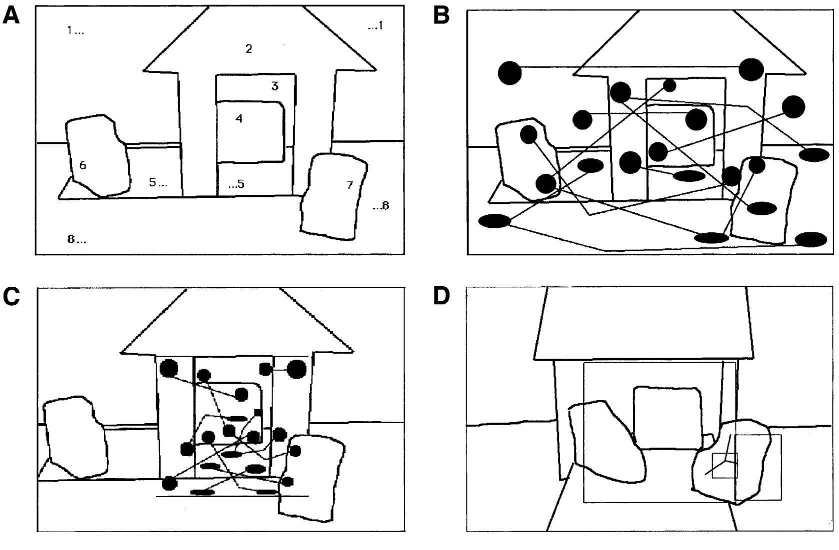


Fig. 7. (A) Units and (B and C) object-surface activity maps, (D) the relevant regions for each level of scope.

The scope was reduced to five or six units in Experiment 3B (*medium* scope). Layout 2 had six units, whose activity is illustrated in Fig. 7C. As can be seen, the probes were distributed across the six units, with no relations occurring within a single unit. Attending to several of these units would be ineffective most of the time. Layout 1 had five units, with activity distributed in a similar manner. In contrast, in Experiment 3C the scope was limited to about one unit (*small* scope), and in Experiment 3D scope was limited to a portion of one unit (*very small* scope; e.g., see Fig. 7D). A limited-scope representation could handle all of the relevant scene in these cases. Thus, the amount of facilitation should be maximal in the smaller scope cases, if facilitation depends on a sharply limited scene representation.

In contrast, the present proposal is that scene representations can be large in scope. Therefore, facilitation effects should be similar in size across scope conditions.

4.1. Method

There were four groups of observers, one for each level of scope illustrated in Fig. 7D. The scope of the relevant region was marked in all primes and in all targets. For each observer the three prime conditions (similar, different-lighting, and control) appeared equally often.

There was a total of 12 probe-pairs distributed across the relevant areas in each layout of Experiments 3A and 3B. The probe-pairs in Experiment 3A were the same as in Experiment 1. In Experiment 3C, the number of pairs was reduced to 8 per layout because of the greatly reduced scope, and in Experiment 3D the number was reduced to 4 per layout. As noted earlier, the probe-pair locations were the same for the two lighting-versions of each layout. Thus, a given similar prime predicted the same sets of probe-pair locations as its different-lighting mate.

The stimuli were the second edition versions of the images. This meant that there were no red colors in the scene except for the critical probes (and red outlines that marked scope). Several reddish objects in the first edition images were changed to purple in the second edition images. Also, the positions of corresponding images matched exactly in the second edition images.

Scope was marked by outline rectangles in Experiments 3B–3D, as illustrated in Figs. 7C and D. (In Experiment 3A, the entire image was relevant.) The red rectangles differed slightly in size and placement between the two layouts, because the rectangles were fit to the structure of the scenes. The same rectangles appeared in the control primes, and a different control prime was used for each layout, to match the scenic images exactly. Across the four scope conditions, the average dimensions in degrees were (horizontal by vertical): (3A) 23.5 × 16.1, (3B) 7.8 × 6.3, (3C) 3.1 × 4.5, and (3D) 2.0 × 1.7. There was a total of 576 trials in Experiments 3A and 3B, and 480 trials in Experiments 3C and 3D. The number of trials was reduced in the smaller scope conditions because there were fewer targets—each prime—target pair occurred four times in Experiments 3A and 3B, five times in Experiment 3C, and 10 times in Experiment 3D.

A total of 74 students participated. The data for three were discarded for long mean RTs (beyond the 3 standard deviation criterion), leaving the following *n*'s:

16 (14 female) in Experiment 3A, 20 (17 female) in Experiment 3B, 19 (16 female) in Experiment 3C, and 16 (12 female) in Experiment 3D.

4.2. Results

The facilitation effects for each sub-experiment are shown in Fig. 8, along with the relevant control means. Analysis of Variance was used to conduct planned comparisons, with scope (sub-experiment) and prime type as the main factors.

The first question was whether the similar prime facilitation effect differed as a function of scene-scope. The facilitation effect averaged 50 ms across scope, and there were no reliable differences in the size of the effect as a function of scope ($F[3, 67] < 1$, $MS_{\text{Error}} = 906$, for the interaction of scope and similar vs. control prime). When the most extreme groups were compared (large scope vs. small and very small scope), the facilitation effect did not vary ($F[1, 16] < 1$). The apparent tendency toward less facilitation in Experiment 3A may be due to error, because the same probe-pairs locations were used in Experiment 1 and the similar facilitation effect was 50 ms. In summary, the facilitation effect was roughly the same magnitude when the relevant scene-scope was quite large as when it was very small. This is consistent with the idea of broad-scope scene representations, but not with the idea of sharply limited-scope representations.

The second question concerns the lighting effect: Is there a reduction in facilitation when image-based information changes between prime and target? The facilitation effect averaged 17 ms less with different-lighting primes relative to similar primes and this difference was highly reliable ($F[1, 67] = 17.20$, $p < .001$, $MS_{\text{Error}} = 608$).

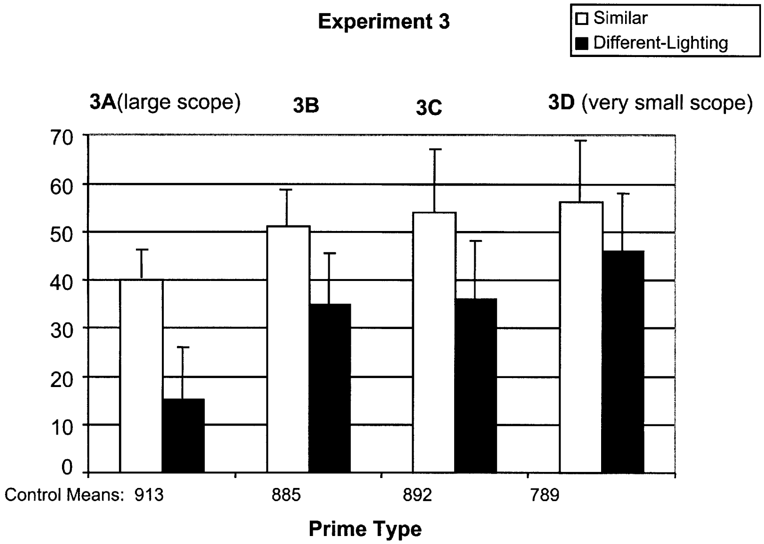


Fig. 8. Facilitation as a function of prime type, for each scope group in Experiment 3.

The lighting effect suggests that image-based information can be a functionally important component of the scenic representation.

There is some evidence in Fig. 8 that the lighting effect was stronger with large-scope scenes than the other conditions. However, the interaction between scope and lighting (similar- vs. different-lighting prime) was not reliable, $F[3,67] < 1$, $MS_{\text{Error}} = 607$. Even when the most extreme groups were used, the effect was not reliable ($F[1,67] = 1.61$, $p > .20$ for lighting × large vs. very small scope).

Accuracy levels were consistently high (above 96.8% in each condition). There were no effects in an Analysis of Variance assessing prime condition and scope.

4.3. Discussion

The similar prime facilitation effect was generally constant across marked differences in scenic scope, with large facilitation effects occurring even in the broad scope cases—with an average scope of 5.5 scenic units plus relations in the medium scope condition, and eight units plus relations in the large scope condition. This is consistent with the idea that scene representations can vary in scope from narrow to quite broad. There was no evidence that scene representations are sharply limited in scope. Note that the constancy of facilitation across scope cannot be attributed to general semantic factors that are constant across scope, because Experiment 1 indicates that specific information about layout is necessary for a strong facilitation effect.

Factors in addition to scope did vary between sub-experiments. As scope decreased, the spatial relations that were probed became finer-grained because there was less space to probe. (However, note that the number of probe-pairs was also reduced with scope.) In addition, the amount of uncertainty decreased somewhat with scope, as did mean RT. These factors might influence the size of facilitation effects, but they cannot be used to explain the lack of sharp limitations in scope. If representations were sharply limited, then limitations should have been observed in spite of these factors.

The results disambiguate the lighting effect in the previous experiment—the effect of image-based information was highly reliable in the present case. The effect suggests that some image-based information is used in scene representation, at least in certain conditions.

5. Experiment 4: Can a limited region be cued?

Experiment 4 was designed to provide a further test of the possibility that scene representations are limited in scope. Scenic scope was manipulated in a second way, by cuing a limited region in the scenes. The observers were instructed to focus on the middle portion of the scene—the medium scope region used in Experiment 3B. This region was marked clearly throughout the experiment and on 80% of the trials the region was probed. On the remaining trials, the probes appeared outside of the cued region. These probes were also outside of a larger oval region defined by an oval "spotlight" fit through the corners of the rectangular region.

If scene representations are sharply limited in scope, then the 5.5 scenic units of the cued region should contain enough information to exhaust representational capacity. The processing of probes inside the cued region may be facilitated, but the processing of probes outside the cued region should receive little facilitation from the scene representation. On the other hand, if scene representations can be broad in nature, then processing of the scene should result in a broad scope representation that facilitates processing in both cued and uncued regions.

An additional purpose of this experiment was to examine possible differences in the processing and representation of background and foreground information. Phenomenologically, foreground objects usually seem much more salient during scene perception than background information. In many models, a basic stage of object and scene processing is to segregate objects from the background, in order to focus processing effort on them. This should result in more efficient processing and a stronger representation of foreground information. In contrast, background information is often assumed to be a low priority for processing and to remain undifferentiated. Consistent with these ideas, Baylis and Cale (2001) found that figural shape representations can be primed but not otherwise equivalent ground regions. To begin examining foreground and background processing in the present context, probe-pair locations were varied within the cued region: Probe-pairs were located on foreground objects or on background surfaces. If there is a priority for foreground processing and representation, then there should be advantages for foreground probe-pairs in control RTs and in scenic priming effects.

5.1. Method

The three prime conditions (similar, different-lighting, and control) were crossed with three types of probe-pair locations. Of the 20 probe pairs for each layout, eight were *cued foreground* (located on furniture or the front house wall in the cued region), eight were *cued background* (located on the floors or back walls in the cued region), and four were *uncued* (located outside of the cued region). The object-surface-activity map for layout 1 is shown in Fig. 9. Each of the 20 targets for a given lighting/layout combination appeared equally often. The cued region was designated by a red outline rectangle in all primes and targets. Each of the 80 targets was preceded equally often by the three prime types, with each combination occurring twice for a total of 480 trails. The test trials were preceded by a relatively long practice period (96 trials) to provide exposure to the cueing probabilities. Observers were instructed that most of the probes would occur within the designated region and that they should focus their attention within that region. Thirty six observers participated (32 females).

5.2. Results

The facilitation effects for the scenic primes and each probe-pair location are shown in Fig. 10, along with the relevant control means. Analysis of Variance was

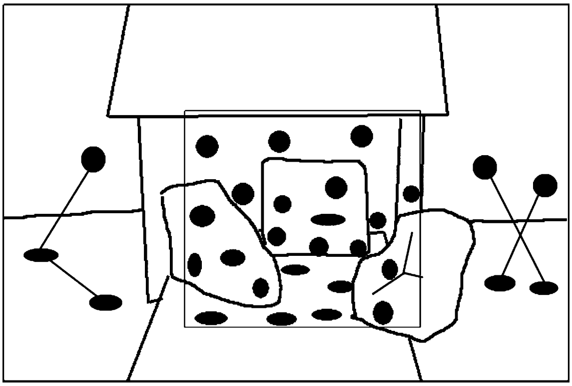


Fig. 9. Probes and relevant region for Experiment 5. Relations are indicated only for the uncued probes.

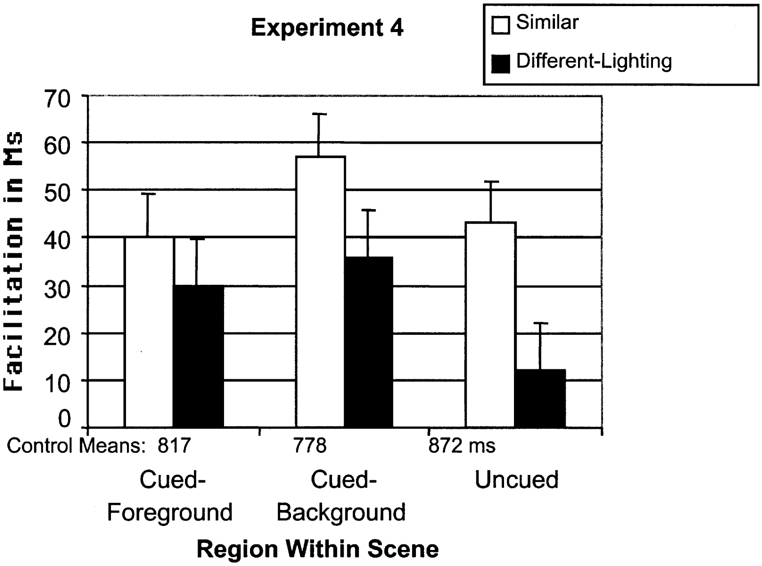


Fig. 10. Facilitation as a function of probe-location and prime type, in Experiment 4.

used to conduct planned comparisons, with prime type and probe-pair location as the factors.

The main question was whether the similar facilitation effect differed between the cued region and the uncued regions. The facilitation effect averaged 49 ms across the cued region and was similar (43 ms) in the uncued region ($F[1,35] < 1$, $MS_{\mathrm{Error}} = 1677$, for the interaction of cuing [cued foreground and cued background vs. uncued] and similar vs. control prime). Thus, the facilitation effect was approximately the same magnitude in the cued region and in outlying regions. This is consistent with the idea of broad-scope scene representations, but not with the idea of sharply limited-scope representations.

Lighting effects were also measured in this experiment. The changes in image-based information in the different-lighting prime condition caused a 20-ms reduction in the facilitation effect relative to the similar condition, $F(1,35) = 26.84$, $p < .001$, $MS_{\mathrm{Error}} = 879$. This is consistent with the claim that image-based information is important for scene representations.

This experiment also provides information about the processing of foreground and background locations within the cued region. The data for the control prime condition were examined first. These data reflect primarily bottom-up processing—i.e., with no effect of prime-induced representations. The result was that probe-pairs in background locations were processed 39 ms faster than those in foreground locations, $F(1,35) = 9.94$, $MS_{\mathrm{Error}} = 2710$. This is contrary to the expectation that foreground processing would be faster than background processing. The effects of a scenic prime-based representation were measured by the similar prime facilitation effect. There was a tendency for greater priming for probe-pairs in background locations than for foreground locations, although it was not reliable ($F[1,35] = 2.79$, $p = .10$, $MS_{\mathrm{Error}} = 865$, for the interaction of same vs. control prime and background vs. foreground cued locations). This tendency is opposite to what would be expected if foreground information was more prominent than background information in scenic representations.

There is a suggestion of a relationship between region and lighting effects in Fig. 10, with lighting effects being strongest in the background-cued region and in the uncued region. However, this was not a reliable effect ($F[2,70] = 1.54$, $p > .20$, $MS_{\mathrm{Error}} = 1234$, for the interaction of probe-type and same- vs. different-lighting).

Accuracy levels were consistently high (above 97% in each condition). There were no effects in an Analysis of Variance on prime type and probe location. Accuracy averaged 98% for each of the three probe location types.

5.3. Discussion

Similarly large facilitation effects occurred in the cued region and in uncued regions. This is again consistent with the idea of broad-scope scene representations.

Attention may have been distributed broadly in this experiment despite the instructions and probability manipulation. A broad attention-based representation of the scene would be consistent with the results, but inconsistent with the idea of sharply limited attentional representations.

The lighting-based effects were replicated in this experiment. This provides further evidence of the importance of image-based information within certain conditions.

The results for foreground and background probe locations have some interesting implications. The findings were that background probe-pairs were processed more efficiently than foreground pairs in the control condition, and that there was a tendency toward more prime-based facilitation for background pairs. These results are inconsistent with the idea, suggested by the phenomenology of scene perception, that foreground information is processed and represented more efficiently than background information. The results are also inconsistent with recent results with object priming (Baylis & Cale, 2001).

Some caution is warranted in interpreting the control-condition comparison because there were no controls on the difficulty of background and foreground probe pairs. It is possible that depth relations were more difficult to determine for foreground pairs, for example. (Some evidence against this possibility is the similarly high levels of accuracy with foreground and background probes, however.) Nevertheless, the results are clearly inconsistent with a strong priority for foreground information over background information, because a foreground-first order of processing should supersede metric differences in depth separation.

The tendency toward more facilitation with background pairs cannot be explained by probe-pair difficulty, because control performance was factored out (subtracted in the calculation of facilitation effects). Thus, this tendency is inconsistent with the hypothesis of an advantage in representation for foreground information.

The results are consistent with the idea that the overall shape of the environment is processed by a primitive geometric system (Cheng, 1986) that is separate from a more advanced system for processing object relations (see also Hermer-Vazquez, Spelke, & Katsnelson, 1999; Rensink, 2000a). The large-scope, geometric system could process background information. It may be highly efficient because it usually involves simple, uniform structures such as groundplanes and walls. In contrast, a more local, object-based system may process foreground information such as object relations. The geometric system may be more implicit than the object-based system, which appears to be associated with language and both evolutionary and ontogenetic maturity (see Hermer-Vazquez et al., 1999). The object-based system may account for the phenomenological impression that foreground is attended to. The present priming effects for both background and foreground locations imply that both systems are activated by primes during spatial relations processing.

6. Experiment 5: Invariance across image position?

MSRT claims that scenic representations are active over time, across events such as eye movements. Thus, despite the considerable changes in retinal inputs brought by such events, there should be some continuity in the scenic representation. However, in the previous experiments the primes and targets were always presented in the same screen locations. Could priming effects result from sensory integration between retinally organized inputs? There are several arguments against such explanations (see Section 1.6), and evidence against the idea of sensory persistence from prime to target (see Method of Experiment 1). Nevertheless, an important question remains: Are scene representations invariant across retinal image position? The specific issue examined in the present experiment concerns whether facilitation extends across changes in image position.

In one condition of the earlier study (Sanocki & Epstein, 1997), the position of primes on the monitor was shifted by a small amount relative to targets. There was only a very small decrement in the facilitation effect (8 ms, reliable only when

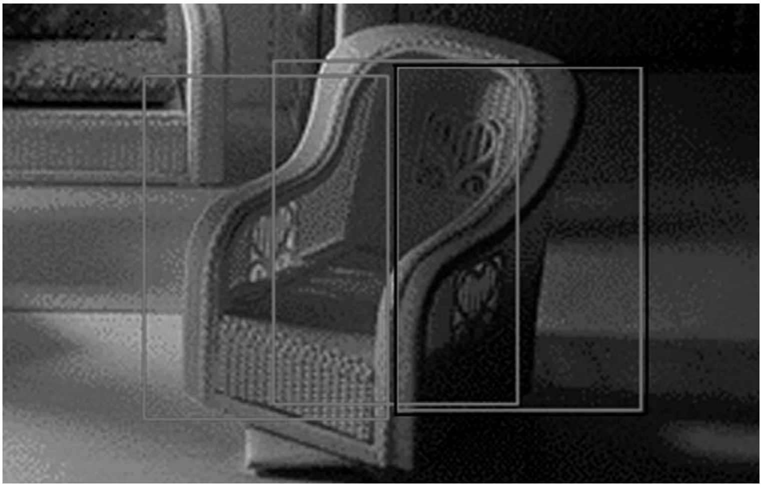


Fig. 11. A close-up of the chair from the target image, with a black outline marking the relevant region. The horizontal locations of the relevant region in the prime are shown as the gray outline. (The prime regions are also shifted vertically in this figure for clarity; in the experiment the relevant region moved only horizontally.)

three sub-experiments were pooled), suggesting that the representations were predominantly position-invariant. However, the scenes used were homogenous, consisting of a single topographical feature that was repeated across the scene. Such simple scenes may lend themselves to position-invariance. A much stronger test of invariance across image position would be provided by the present scenic images, which are complex and asymmetrical.

A second and related issue involves the image-based effects of changes in lighting direction. Analogous to the general facilitation effects, the lighting effects would be more significant if they survived changes in image position, because such a result would imply that image-based information contributes to a lasting and general scene representation.

In summary, Experiment 5 examined which aspects of the prime-based representations are effective across shifts in image position. Primes and targets appeared within the same screen window, but the images were either corresponding or shifted, with the prime image shifted to the left of the target image by an amount equivalent to one-half or one full chair-width (1.55 or 3.1°, see Fig. 11). The shifting effects were examined at two levels of scene scope—the small scope and the large scope from Experiment 3. The large shift size corresponded to the total width of the small scope area so that the shift completely changed the retinal image of the relevant scene.

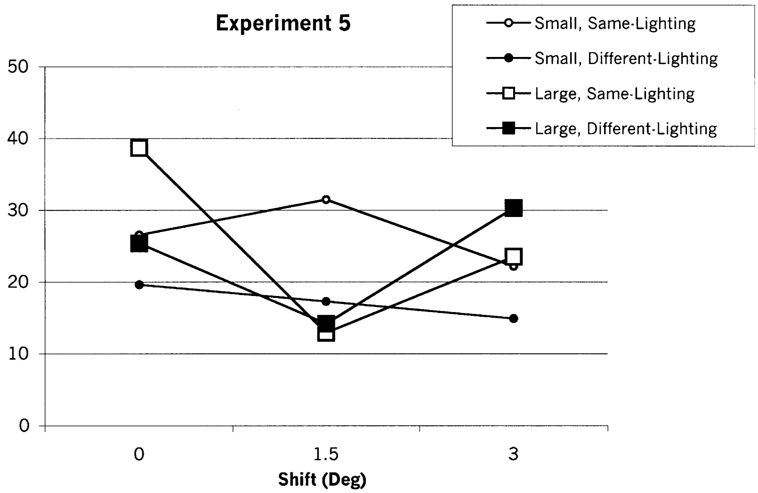
6.1. Method

The two groups of observers received different scopes defined by the relevant regions. Each observer received the similar, different-lighting, and control prime conditions, crossed with the three levels of prime-shift. The primes were shifted by moving the image to the left within the same screen window. Thus, the extreme left of the images was cut off. Background was duplicated and added to the far right of images, to fill the window. Control primes were the same as in Experiment 3 and their patterns (rectangles) were shifted by the same amounts as the scenic primes. There were 12 probe pairs per layout in the large scope condition (a total of 432 trials) and 8 probe pairs in the small scope condition (total of 288 trials). Sixty two observers participated (41 females) but the data for 6 were discarded for low accuracy or long overall RTs, leaving 26 participants in the small scope condition and 30 in the large scope condition.

6.2. Results

The facilitation effects as a function of amount of shift are shown in Fig. 12. In the analysis the factors were scope, prime type, and amount of shift. As in Experiment 3, the control mean was considerably longer in the large scope condition (965 ms) than in the small scope condition (750 ms; $F[1, 54] = 28.60$). However, the priming effects were generally similar across scope.

The main question was whether scenic-primes were effective across shifts of image location. Responses were slowed somewhat by image shifts, as can be seen in the control means reported in the figure, $F[2, 108] = 22.58$. The overall 26 ms advantage for similar primes relative to control primes was reliable, $F[2, 108] = 22.68$, $MS_{\text{Error}} = 1681$. Most important, the facilitation effect did not vary with amount



Control Means

Small:	734	746	770
Large:	963	956	976

Fig. 12. Facilitation effects in Experiment 5 as a function of the shift in prime position relative to the target position.

of prime shift ($F[2, 108] < 1$, $MS_{\text{Error}} = 1688$, for the interaction of shift and similar vs. control prime). Thus, the scenic representations were invariant with respect to image position.

Was there an effect of changes in the lighting direction? Overall, facilitation was 6 ms less with different-lighting than with similar primes, but this difference was not reliable ($F[1, 54] < 1$, $MS_{\text{Error}} = 1813$). The facilitation effect for different-lighting primes was reliable, $F[1, 54] = 16.22$, $MS_{\text{Error}} = 2119$. There were no interactions involving lighting condition.

Aside from the three main effects, the only other reliable effect in the analyses was an interaction of scope and amount of shift, $F[2, 108] = 3.37$, $p = .04$, $MS_{\text{Error}} = 1535$. Shifts were more detrimental in the smaller scope condition than in the larger scope condition, even though the overall RTs were faster in the smaller scope condition (and even though the absolute amounts of shift were equal). The interaction can be seen in the control means reported in the figure. With small scope scenes, the largest shift delayed responding by 36 ms, whereas the delay was only 14 ms with large scope scenes. The interaction of shift and scope occurred with control primes and with scenic primes; thus, the dependence of shift-difficulty on scope did not require a scenic prime-based representation. The effect is more general and may be related to the shifting of attention. The interaction suggests that the difficulty of processing a given amount of image shift depends on the size of the effective representation—with large scope representations, a shift of a given size is less detrimental than with small scope representations. However, note that large scenes were also processed more slowly overall than small scenes, so it is possible that the cause of the shifting difficulty is related to the fast processing times for the small scenes rather than their small scope. With large scenes there could be more time for adjusting to the shift because overall scene processing takes longer.

6.3. Discussion

Experiment 5 indicates that scenic-prime based representations survive substantial shifts in image position. The largest image-shifts completely changed the locations of objects—for example, the chair moved rightward one full chair width—yet the prime-based facilitation was approximately constant across these shifts. This suggests that scene representations can remain active as observers move their eyes throughout scenes.

The magnitude of the facilitation effects was smaller in the present experiment (26 ms for similar prime facilitation) than in the previous experiments (50 ms modal magnitude). The reason for the reduced effects is not clear. The reduction could be related to difficulties posed by shifting image positions, or perhaps to the unnaturalness of sudden image shifts (as opposed to shifts resulting from planned and controlled eye movements).

The effects of changes in lighting direction on facilitation in Experiments 2–4 were modest (17–20 ms) but generally reliable. In the present experiment, the effect was smaller (6 ms overall) and unreliable. Thus, the image-based effects of lighting were greatly reduced or absent in the present experiment. The image-based effects appear to be limited to when primes and targets consistently share the same position, as in Experiments 2–4. When primes and targets share position on only some trials, as in the present experiment, no reliable effects of image-based information were obtained. Whether the effects would occur when image position changes with intentional eye movements remains to be determined. It is possible that the image-based information is useful for intentional eye movements, perhaps for guidance or targeting. The use of image-based information may be optional, with use increasing when the conditions makes the use of image-based information advantageous (e.g., Experiments 2–4). In this view, the sudden image-shifts of Experiment 5 would be assumed to reduce the usefulness of image-based information.

An alternative possibility is that the image-based effects observed in the previous experiments are a sensory masking effect. Such an effect would be low level in nature and should not be altered strategically. There were several problems for a sensory masking explanation discussed in the Introduction. Nevertheless, the present results provide further evidence on the idea. The sensory masking explanation would entail that in Experiments 2–4, performance in the different-lighting conditions would reflect the positive effects of a scenic representation combined with the negative sensory masking effect. However, a sensory effect should also have occurred in the present experiment. Specifically, it should have caused a decrement for the different-lighting primes relative to similar primes in the no-shift condition (as in Experiments 2–4). Furthermore, a negative sensory effect should have occurred with both scenic prime-types when image position shifted because the non-corresponding prime and target images should interfere with each other. In contrast, such an effect would be absent in the control prime condition because that prime had negligible image structure. As a result, scenic prime-based facilitation should be reduced when image position shifted. None of these results occurred, providing strong evidence against the sensory masking account. Thus, in summary, the results seem to be best explained by a strategic change in the use of image-based information.

7. General discussion

The present experiments demonstrate that the on-line processing of spatial layout is fruitfully studied with information processing methods. The experiments begin to measure the representation and use of specific layout information that binds identity and location (Experiment 1), the scope of scene representations (all experiments), possible differences in processing of foreground and background information (Experiment 4), invariance across image location (Experiment 5), and the possible role of image-based information (Experiments 2–5). The results support four main conclusions and raise numerous further issues.

The first conclusion is that prime-induced scene representations facilitate subsequent spatial relations processing. Facilitory effects were obtained in each experiment, ranging as high as 50 ms. These effects occurred with clear, high quality natural images of scenes—not with impoverished stimuli. The results replicate the

Sanocki and Epstein (1997) findings and extend them by adding numerous control and comparison conditions.

Second, the results of Experiment 1 indicated that fairly specific differences in spatial layout contribute to the facilitation effect. Primes that were similar to the targets in most general respects, having the same gist and the same objects in generally similar arrangements, but differing in details of layout, were not as effective as primes identical to the target. These results imply that scene representations bind information about object identity and location.

Third, the facilitory effects were broad in scope, occurring across a complex scene in all relevant cases. There were no benefits of narrowing the relevant scene scope in Experiments 3 and 4. Indeed, Experiments 1 and 3A, conducted with the largest scope and the same set of 24 probe-pair and relations, suggest that scene representations had a scope as high as 8 object or surface units plus relations. Because specific layout information causes the facilitation effect (Experiment 1), the broad scope of facilitation cannot be attributed to general semantic factors (such as the primes predicting the scene category). The facilitation effects were generally similar across large and very small scope conditions, suggesting that scenic representations function in a similar manner across a fair range of scenic scopes.

Fourth, Experiment 5 provided evidence that the scene representations survive changes in image location. This result is consistent with the idea that scene representations contribute to the integration and active exploration of scenic information (e.g., Hochberg, 1978).

A fifth idea examined in these experiments was that image-based information contributes to scene representations, an idea that follows from mental models (e.g., Barsalou, 1999; Glenberg, 1997; Miller & Johnson-Laird, 1976). The prediction was that changes in image-based information, produced by changes in lighting direction, would disrupt scene representations. However, the effects of such changes were modest, even with marked changes (which affected critical probes directly in Experiment 2). The modest effects suggest that some image-based information can be used in certain conditions. At the same time, however, the substantial facilitation effects for different-lighting primes indicate that image-based information is not essential to the prime-based representation. The use of image-based information may be optional. The results are more consistent with an emphasis on symbolic representations that are abstracted away from image-based information.

7.1. Explaining scenic priming effects

Taken together, the present results indicate that scene representations are complex structures that have a significant impact on spatial processing. Explanations based on a strict division between bottom-up processing and limited attentional representations (e.g., Pylyshyn, 1999) do not conveniently explain the effects of large-scope representations. Explanations based on priming of loosely connected object representations do not conveniently explain the added benefits of appropriate spatial layout relative to different-layout primes (Experiments 1 and 1B). An early vision module (Pylyshyn, 1999) might be expanded to include primed spatial relations, but the resultant module would encompass much of vision.

A more natural account of the results follows from the idea of interaction between larger prime-based representations of the context and more local spatial processing of the target and critical relations. In the sections that follow, MSRT is explicated to provide an explanation of the present results. The explanation emphasizes relations between scenic entities and begins to integrate ideas from the literatures on the processing of familiar forms and figural goodness.

7.2. MSRT as a relational network

In this section, the theoretical language of structural description theory (e.g., Oden, 1979; Palmer, 1975, 1977) is used to instantiate a model consistent with MSRT, and to explain task performance. The model is based on the idea that experience with a scenic prime quickly activates a network representation involving multiple levels of information. The highest levels of the network consist of scene labels for the class of scenes and for particular scenes (e.g., "doll house" and "doll house layout 1," respectively). The lowest levels involve primitive symbols such as "edge" and "vertex" that result from intensive lower and intermediate level processing. Between these levels are the critical levels for the present purposes, which correspond to objects and surfaces and their components.

Major objects and surfaces are represented by major units. The major units are linked to and receive input from sub-networks of minor units representing their constituents and their properties. For example, the "wall" major unit is connected to sub-networks representing components such as windows and important properties such as its approximate boundaries. The sofa major unit is connected to sub-networks for its major parts and inter-part relations (e.g., Hummel & Biederman, 1992). However, whereas many models assume a hierarchical structure in which constituent symbols feed into only one major unit (e.g., Hummel & Biederman, 1992), the present model is based on the assumption of overlap between sub-networks for different major units. For example, when the sofa occludes the wall, the sub-networks for the wall and sofa share some units—e.g., the sofa's boundary is also an interruption in the visible portions of the wall.[5] Shared units are one source of interrelation within the scenic network.

A second source of interrelation in the network is spatial relations between major units. These relations are represented as links—e.g., the sofa linked to the wall by an "in front of" relation. These links specify ordinal information and perhaps some coarse metric information. Most nearby major units are assumed to be linked in this way, with linkage increasing with proximity. Linkage is also more likely between units embedded within the same larger unit (e.g., chairs within a house), but links can cross boundaries (McNamara, 1986). Linkages also vary with the frequency with

---

[5] The assumption is that for the wall, both the interruption in visible portions and the implied continuity of the wall behind the sofa are represented in the sub-network.

which the spatial relations are used in a situation (e.g., probed within an experiment). In summary, the prime scene is represented by a network with many interrelations. The relations between major units and between their sub-networks have the effect of binding the network together into a configuration. This includes the binding of object representations with information about layout.

Presentation of a target begins to activate a network similar to the prime-based network. However, when target information matches information in the prime-based network, the target information is integrated into the prime-based network. Thus, the prime-based representation becomes the target-based representation. At least some information in the target but not in the prime (e.g., the critical probes) can be new to the emerging network. The process of evaluating the prior representation for similarity to the current stimulus is similar to the review of object files (Kahneman et al., 1992).

Because of the high priority of the critical probes, presentation of the target initiates an attention-switching process in which attention is shifted to the critical probes and the major units and sub-networks surrounding them. This results in faster processing in those regions. Relations between the critical probes (close vs. far and left vs. right) are established within the network to determine the response. These relations correspond to depth cues such as occlusion and relations to the ground plane. The similar-prime facilitation effect arises because these relations are established relatively quickly when an appropriate prime-based network is already available. With control primes and different-layout primes, the same critical relations must be established but so must much of the target network that the relations are embedded within.

When the target information does not match the prime-based network, as with a different-layout prime, a target-based network is created that is separate from the prime-based network. The critical probe relations are established and a response is generated. The new network is created with little interference from the previous, prime-based network. This assumption is supported by the general similarity of performance between control prime conditions, where it is clear that a new target network must be created, and different-layout conditions (Experiments 1, 1B, and 1C) or different-scene conditions (Sanocki & Epstein, 1997, Experiments 2 and 3). One important detail is the threshold for creating a new target network rather than modifying the prime network: How much target-prime mismatch is necessary before a new network is created? This issue is related to the relational nature of the network.

Because any one major unit is involved in numerous relations in the network, modifying that unit's spatial relations within the network would involve some cost in processing time. When the same objects occur in clearly different arrangements in the prime and target, considerable processing would be necessary in order to change a prime-based network into a target-based network. Consequently, a new and independent target-based network is created when the target is presented. As a result, little facilitation occurs when the spatial layout is altered significantly (Experiment 1B) or when the objects are changed while many spatial relations were constant (Experiment 1C). However, if the entire scene is shifted (Experiment 5), most of the relations (sub-networks, their interrelations, and relations between major units) remain constant. In this case, the prime-based network can be modified into a target network relatively easily. Thus, facilitation (re-use of the prime-based network) depends on preservation of relations throughout the scene. The exact boundary conditions for preservation of the prime-based network are an important topic for further research.[6] However, some predictions can be made. For example, when the eyes move between prime and target, a prime-based network can be re-used when most of the scenic relations are preserved. When substantial new relations appear at the new eye position, there should be little facilitation. Thus, the idea of integration of information in a scene (e.g., Hochberg, 1978) is qualified and assumed to depend on preservation of relations. Also note that the facilitation effects for different-lighting primes, discussed shortly, suggest that differences in image-based information do not trigger creation of a new network.

Both large and small scope scenes can be represented in the networks, although the density of the network (the units and relations) presumably increases with a complex small-scope scene. Nevertheless, the processing advantages of an appropriate pre-existing prime-based network are similar across marked differences in scope (Experiment 3). Similarly, Experiment 4 supports the assumption that both background surfaces and foreground objects are included in the prime-based representation.

Based on the effects of lighting, the representation of image-based information is assumed to be optional. When prime and target positions are constant within trials, image-based information could be useful. For example, a brightly lit chair could be represented (in part) as a bright spatial landmark that is constant across prime and target. Brightness might be represented by links between the major chair unit and lower retinotopically organized levels that process image-based information. When prime and targets match on such properties, processing would be relatively fast. When the prime and targets mismatch on this type of information, the prime-based network would have to be modified and this would increase processing time, resulting in the small decrement for different-lighting primes. However, when image positions change suddenly (Experiment 5), information about lightness or darkness of objects would be less useful because the relation between lower levels and higher levels would change with image position. Consequently, image-based information would be left out of the scenic representations.

7.2.1. How is a broad scope representation maintained?

The network representation is complex because it includes numerous entities and relations between the entities—eight major units and their interrelations in the large scope conditions. This certainly is more information than the 3–4 objects that appears to be the limit of visual working memory for objects (e.g., Vogel, Woodman, & Luck, 2001). How could so much information be represented and maintained? The present answer to this question involves three main assertions. The first is that when

---

[6] In Experiment 1, there was a marginally reliable facilitation effect for the different-layout prime condition. The different-layout prime in this condition was generally similar to the target (e.g., both had a central house and the same ordinal relation between chairs), suggesting that when different layouts are similar enough, there may be some transfer of prime-based information, at least for some observers.

the literature on visual working memory (VWM) is examined, there is evidence for a high capacity memory for *layout*. The second assertion is that the scene representations examined in the present experiments are a type of VWM for layout. The third assertion is that interrelations underlie the high capacity of layout memory.

There is evidence in the literature that VWM can be considerably greater than four objects in capacity when layout information—the relative position of elements—is relevant. In classic studies of visual working memory, Phillips (1974) found accurate detection of changes of the position of a single square element within grids of up to 25 elements. Simons (1996) found that while memory for object properties was limited, memory for their layout was near the five-item ceiling of those studies (see also Rensink, 2000b). Brockmole, Wang, and Irwin (2000) used a dot-integration task to study visual working memory for the position of dots within matrices and found capacities beyond 10 positions. Jian, Olson, and Chun (2000) obtained several results consistent with the idea that the memory representation was configural in nature. However, in each case the high capacity could be explained by assuming that the elements were encoded as a single object with a particular shape—as an object configuration.

Motivated by the present findings of broad-scope representations and the prior VWM results, Sanocki, Sellers, and Mittelstadt (2001) created a VWM task with scenes that were much more complex than single object configurations. The scenes were generated from a doll house and yard divided into 12 regions. Layouts were generated by filling from 4 to 8 regions with furniture from a set of 24 pieces of furniture assigned to particular regions. The results indicated that the observers' memory capacity corresponded to 5.2 regions with over 10 pieces of furniture, as measured by a simple model (Pashler, 1988). This provides evidence of a fairly high capacity memory for scenic layout. Furthermore, memory performance was high for changes that altered the envelope or configuration of the scene and for internal changes that did not alter envelope. Thus, the memory representations were based on the layout of complex scenes, rather than any simple aspect of overall shape or configuration. As in the network model, the layouts can be assumed to be configural, involving many relations within the scene. Thus, the representations may be similar to those examined in the present experiments.

How might high-capacity layout configurations be maintained? Sanocki et al. (2001) hypothesized that the ability to encode multiple types of relations contributed to capacity. In their Experiment 3, they manipulated relatedness directly with dot matrices. In the high-relations, *regular* condition, stimulus arrays were generated by taking random subsets of 12–20 elements from a normal 5×5 array (horizontal and vertical rows and columns with constant spacing). A regular frame outlined the grid. In this condition, there was an abundance of relations such as collinearlity, various types of symmetry, and relations to the canonical axes of the frame and the environment. In the low-relations, irregular condition, the arrangement of dots was distorted by moving elements to create a haphazard and somewhat oblique configuration with irregular spacing and no alignment. In addition, the frame was rotated in the opposite direction from the group of elements. As a result, there were no dominant axes within the grid and the frame did not match the environmental axes.

Consequently, there should be less collinearity, less symmetry, and fewer relations involving axes.[7] The results were that memory performance was high in the regular condition, with an estimated capacity of 13.6 items, and significantly lower in the irregular condition (9.3 item capacity).

Thus, the present proposal is that encoded relations contribute to the high capacity of visual working memory and of scene representations (see also Jian et al., 2000). There probably are multiple types of relations, including spatial relations between adjacent items, gestalt relations such as collinearity and curvilinearity, and relations involving canonical axes or other salient reference frames. It is also possible that semantic relations contribute to scene memory. The importance of relations is well established in other literatures of visual perception. For example, the Gestalt principles are relational in nature and contemporary accounts of Gestalt organization rely upon mechanisms for encoding relations (e.g., Palmer, 1983). The encoding of relations is also believed to underlie the efficient processing of familiar visual forms such as letters and objects (e.g., Enns & Prinzmetal, 1984; Sanocki, 1991, 1993, 1999). The present proposal is that these principles extend to the representation of scenes.

7.3. How much information is represented in memory during scene perception?

The conclusion that extensive scene representations are activated and used during layout perception contrasts with conclusions about the extent of representations from other tasks. Evidence for very limited representations comes from tasks such as visual search, where previously viewed distractors may not be represented (e.g., Wolfe, Klempen, & Dahlen, 2000), and change detection tasks, where changes in a scene are often difficult to detect (e.g., Rensink et al., 1997; see also O'Regan, 1992). A number of factors may influence the extent and utility of scene representations in a given situation, and an understanding of these factors can provide insight into visuocognitive functioning. One potentially important and related pair of factors is the organization and familiarity of the scene. Cognitive functions such as maintaining a representation or detecting change should be more efficient within contexts that are better organized (e.g., Sanocki et al., 2001) or more familiar.

In the present context, there may be another way in which representations of layout are interesting and important—they may serve as useful configural representations. In many situations throughout evolution, it would seem advantageous to represent a scene-configuration in memory. For example, a castle guard could watch for predators or attacking warriors by monitoring a scene-configuration and attending to any change in shape, which may signal approaching danger. Also, when looking for a trail for walking through an undeveloped wood, it is critical to continually monitor the surrounding scene for openings that signal a possible trail. In each case, the shape of the entire scene is relevant, meaning that each region of the scene is

---

[7] Because of the propensity for humans to encode relations, the assumption was that relations were encoded in the irregular condition, but that there were *fewer* relations encoded than in the regular condition because of the irregularities.

potentially important. The fact that each region is potentially important may be a critical property of complex configural representations.

In the present experiments, most areas were potentially important because the spatial probes were distributed fairly evenly across the scene. This may have encouraged a configural representation.[8] Change detection can also be highly efficient when the stimuli are well-orgranized arrays of elements (e.g., Phillips, 1974; Sanocki et al., 2001). However, when the stimulus scenes are unfamiliar and possible changes vary along multiple dimensions (e.g., size, complexity, type of change), observers must attend to a variety of components at differing levels of scale. Therefore, there is no organization into similarly important regions. In visual search experiments, the target is much more important than the distractors, creating differences in region importance. In these later cases, broad configural representations would be neither appropriate nor adaptive.

Acknowledgments

This research was supported in part by Grant DBS-9213246 from the National Science Foundation. I thank many students in the Visual Cognition Lab who helped in conducting the experiments, and Eric Sellers, Jeff Mittelstadt, and Jennifer Sibley-Perone for the stimuli in Experiment 1. I also thank Amy Shelton, Ron Rensink, and an anonymous reviewers for helpful comments on previous versions.

Appendix A. Method and results of sensory persistence study

A grid of 24 rectangles (4 rows by 6 columns) was created by cutting dark regions from the scenes and placing them on the light beige background of the control stimulus. On each trial, 23 of the rectangles were presented, about half (10–15) in an initial 1-s display and the remainder in a second display that remained visible until the response. The inter-stimulus interval was 84 ms, as in the main experiments. Observers indicated whether the empty region was on the left or right half of the screen. If there was strong sensory persistence from the prime, then the two displays of filled regions could be integrated together, allowing observers to locate the one remaining, empty region with high accuracy.

Accuracy averaged 62.7%, ranging between 52.5 and 75.0% across the seven observers. The low level of performance is inconsistent with a sensory persistence explanation. The above-chance portion of performance can be accounted for by assuming that observers hold some blank regions from the first display in visual short term memory. The total percentage of correct responses can be decomposed into correct responses that occur when the critical blank region was held in memory, and correct

---

[8] However, scenes may lend themselves to configural representation in general; Experiment 4 suggested that the scene representations were not limited to the area that was demarcated.

responses that come from guessing (with a 50% probability of correctness). Using a simple model (e.g., Pashler, 1988), the capacity of this short term memory is estimated to be about three items. This is far smaller than the number of regions that should be available if sensory persistence underlied performance.

The model is as follows. Correct responses come from two sources. One is short term memory—trials in which the critical blank region is held in memory. Since only blank regions are critical, the first display limits the blank regions to about 13 (on average). If the capacity of memory is n, then on $n/13$ trials the critical item will be in memory. The other source of correct responses is guessing, on the remaining $((1-n)/13)$ trials. When applied to the present data, a value of $n = 3.3$ produces the accuracy rate of 62.7%. Thus, the results can be explained by assuming that several regions from prime are be held in short term memory under the display conditions of this study experiment.

References

Aks, D. J., & Enns, J. T. (1992). Visual search for direction of shading is influenced by apparent depth. *Perception & Psychophysics, 52*, 63–74.

Andersen, G. J., Hahn, S., & Saidpour, A. (2001). Static scene information and the perception of locomotion [Abstract]. *Journal of Vision, 1*(3), 2. Available: http://journalofvision.org/1/3/2, DOI 10.1167/1.3.2.

Barsalou, L. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22*, 577–609.

Baylis, G. C., & Cale, E. M. (2001). The figure has shape, but the ground does not: Evidence from a priming paradigm. *Journal of Experimental Psychology: Human Perception and Performance, 27*, 633–643.

Biederman, I. (1972). Perceiving real-world scenes. *Science, 177*, 77–80.

Biederman, I. (1981). On the semantics of a glance at a scene. In M. Kubovy, & J. R. Pomerantz (Eds.), *Perceptual organization* (pp. 213–253). Hillsdale, NJ: Lawrence Erlbaum.

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

Biederman, I., & Cooper, E. E. (1991). Priming contour-deleted images: Evidence for intermediate representations in visual object recognition. *Cognitive Psychology, 23*, 393–419.

Boyce, S. J., Pollatsek, A., & Rayner, K. (1989). Effect of background information on object identification. *Journal of Experimental Psychology: Human Perception and Performance, 15*, 556–566.

Brockmole, Wang, & Irwin, D. (2000). *Integration of percept and image*. Paper presented at the annual meeting of the Psychonomic Society, Houston, TX.

Cavanagh, P., & Leclerc, Y. G. (1989). Shape from shadows. *Journal of Experimental Psychology: Human Perception and Performance, 15*, 3–27.

Cheng, K. (1986). A purely geometric module in the rat's spatial representation. *Cognition, 23*, 149–178.

Chun, M. M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology, 36*, 28–71.

Cooper, L. A., & Lang, J. M. (1996). Imagery and visual-spatial representations. In E. L. Bjork, & R. A. Bjork (Eds.), *Memory* (pp. 129–164). San Diego: Academic Press.

DiLollo, V. (1980). Temporal integration in visual memory. *Journal of Experimental Psychology: General, 109*, 75–97.

Efron, R. (1970). Effect of stimulus duration on perceptual onset and offset latencies. *Perception & Psychophysics, 8*, 231–234.

Enns, J. T., & Prinzmetal, W. (1984). The role of redundancy in the object-line effect. *Perception & Psychophysics, 35*, 22–32.

Enns, J. T., & Rensink, R. A. (1991). Preattentive recovery of three-dimensional orientation from line drawings. *Psychological Review*.

Epstein, W., & Rogers, S. (1995). *Perception of space and motion*. San Diego, CA: Academic Press.

Franklin, N., & Tversky, B. (1990). Searching imagined environments. *Journal of Experimental Psychology: General, 119*, 63–76.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.

Glenberg, A. M. (1997). What memory is for. *Behavioral and Brain Sciences, 20*, 1–55.

Goodale, M. A. (1995). The cortical organization of visual perception and visuomotor control. In S. M. Kosslyn, & D. N. Osherson (Eds.), *Visual cognition: An invitation to cognitive science: Vol. 2* (2nd ed.., pp. 167–213). Cambridge, MA: MIT Press.

Gordon, R. D., & Irwin, D. E. (2000). The role of physical and conceptual properties in preserving object continuity. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 26*, 136–150.

Grimes, J. (1996). On the failure to detect changes in scenes across saccades. In K. Akins (Ed.), *Perception (Vancouver studies in cognitive science)* (Vol. 5). New York: Oxford University Press.

Henderson, J. M. (1994). Two representational systems in dynamic visual identification. *Journal of Experimental Psychology: General, 123*, 410–429.

Henderson, J. M. (1997). Transaccadic memory and integration during real-world object identification. *Psychological Science, 8*, 410–429.

Henderson, J. M., & Hollingworth, A. (1997). Eye movements during scene viewing: An overview. In G. Underwood (Ed.), *Eye guidance while reading and while watching dynamic scenes*. New York: Elsevier.

Henderson, L. (1977). Word recognition. In N. S. Sutherland (Ed.), *Tutorial essays in experimental psychology* (Vol. 1). Hillsdale, NJ: Lawrence Erlbuam.

Hermer-Vazquez, L., Spelke, E. S., & Katsnelson, A. S. (1999). Sources of flexibility in human cognition: Dual-task studies of space and language. *Cognitive Psychology, 39*, 3–36.

Hill, H., & Bruce, V. (1996). The effects of lighting on the perception of facial surfaces. *Journal of Experimental Psychology: Human Perception & Performance, 22*, 986–1004.

Hochberg, J. E. (1978). *Perception* (2nd ed). Englewood Cliffs, NJ: Prentice-Hall.

Hollingworth, A., & Henderson, J. M. (1998). Does consistent scene context facilitate object perception? *Journal of Experimental Psychology: General, 127*, 398–415.

Hollingworth, A., & Henderson, J. M. (1999). Object identification is isolated from scene semantic constraint: Evidence from object type and token discrimination. *Acta Psychologica, 102*, 319–343.

Hummel, J. E., & Biederman, I. (1992). Dynamic binding in a neural network for shape recognition. *Psychological Review, 99*, 480–517.

Intraub, H. (1989). Illusory conjunctions of forms, objects, and scenes during rapid serial visual search. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 15*, 98–109.

Intraub, H., Bodamer, J., & L (1993). Boundary extension: Fundamental aspect of pictorial representation or encoding artifact? *Journal of Experimental Psychology: Learning, Memory, & Cognition, 19*, 1387–1397.

Intraub, H., & Richardson, M. (1989). Wide-angle memories of close-up scenes. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 15*, 179–187.

Irwin, D. E. (1996). Integrating information across saccadic eye movements. *Current Directions in Psychological Science, 5*, 94–100.

Jian, Y., Olson, I. R., & Chun, M. M. (2000). Organization of visual short-term memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 26*, 683–702.

Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object-files: Object-specific integration of information. *Cognitive Psychology, 24*, 175–219.

Kleffner, D. A., & Ramachandran, V. S. (1992). On the perception of shape from shading. *Perception & Psychophysics, 52*, 18–36.

Kosslyn, S. M. (1987). Seeing and imagining in the cerebral hemispheres: A computational approach. *Psychological Review, 94*, 148–175.

Lewicka, P., Hill, T., & Czyzewska, M. (1992). Nonconscious acquisition of information. *American Psychologist, 47*, 796–801.

Logan, G. D. (1995). Linguistic and conceptual control of visual spatial attention. *Cognitive Psychology, 28*, 103–174.

Logan, G., & Zbrodoff, J. (1999). Selection for cognition: Cognitive constraints on visual spatial attention. *Visual Cognition, 6*, 55–81.

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

Massaro, D. W., & Sanocki, T. (1993). Visual information processing in reading. In D. Willows, R. Kruck, & E. Corcos (Eds.), *Visual processes in reading and reading disabilities* (pp. 139–161). Hillsdale, NJ: Lawrence Erlbaum.

McClelland, J. L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review, 86*, 287–330.

McClelland, J. L., & Miller, J. (1978). Structural factors in figure perception. *Perception & Psychophysics, 26*, 221–229.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review, 80*, 375–407.

McConkie, G. W., & Rayner, K. (1976). Indentifying the span of the effective stimulus in reading: Literature review and theories of reading. In H. Singer, & R. B. Ruddell (Eds.), *Theoretical models and processes in reading* (pp. 137–162). Newark, DE: International Reading Institute.

McNamara, T. P. (1986). Mental representations of spatial relations. *Cognitive Psychology, 18*, 87–121.

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology, 90*, 227–234.

Miller, G. A., & Johnson-Laird, P. N. (1976). *Language and perception*. Cambridge: Harvard University Press.

Neisser, U. (1967). *Cognitive psychology*. Englewood Cliffs, NJ: Prentice-Hall.

Neisser, U. (1994). Multiple systems: A new approach to cognitive theory. *European Journal of Cognitive Psychology, 6*, 225–241.

Oden, G. C. (1979). A fuzzy logical model of letter identification. *Journal of Experimental Psychology: Human Perception and Performance, 5*, 336–352.

Oliva, A., & Schyns, P. G. (1997). Coarse blobs or fine edges? Evidence that information diagnosticity changes the perception of complex visual stimuli. *Cognitive Psychology, 34*, 107–772.

O'Regan, J. K. (1992). Solving the "real" mysteries of visual perception: The world as an outside memory. *Canadien Journal of Psychology, 46*, 461–488.

Paap, K. R., Newsome, S. L., McDonald, J. E., & Schvaneveldt, R. W. (1982). An activation-verification model for letter and word recognition: The word-superiority effect. *Psychological Review, 89*, 573–594.

Palmer, S. E. (1975). Visual perception and world knowledge: Notes on a model of sensory-cognitive interaction. In D. A. Norman, & D. E. Rumelhart (Eds.), *Explorations in cognition* (pp. 279–307). San Fransisco: Freeman.

Palmer, S. E. (1977). Hierarchical structure in perceptual representation. *Cognitive Psychology, 9*, 441–474.

Palmer, S. E. (1983). The psychology of perceptual organization: A transformational approach. In J. Beck, B. Hope, & A. Rosenfeld (Eds.), *Human and maching vision*. New York: Academic Press.

Pashler, H. (1988). Familiarity and visual change detection. *Perception & Psychophysics, 44*, 369–378.

Pani, J. R. (1997). Descriptions of orientation in physical reasoning. *Current Directions in Psychological Science, 6*, 121–126.

Phillips, W. A. (1974). On the distinction between sensory storage and short-term visual memory. *Perception & Psychophysics, 16*, 283–290.

Potter, M. C. (1976). Short term conceptual memory for pictures. *Journal of Experimental Psychology: Human Perception and Performance, 2*, 509–522.

Pylyshyn, Z. (1999). Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. *Behavior and Brain Sciences, 22*, 341–423.

Reicher, G. M. (1969). Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology, 81*, 275–280.

Rensink, R. A. (2000a). The dynamic representation of scenes. *Visual Cognition, 7*, 17–42.

Rensink, R. A. (2000b). Visual search for change: A probe into the nature of attentional processing. *Visual Cognition, 7*, 345–376.

Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science, 8*, 368–373.

Rueckl, J. G., Mikolinski, M., Raveh, M., & Miner, C. S. (1997). Morphological priming, fragment completion, and connectionist networks. *Journal of Memory & Language, 36*, 382–405.

Rumelhart, D. E. (1977). Toward an interactive model of reading. In S. Dornic (Ed.), *Attention and performance IV*. Hillsdale, NJ: Erlbaum.

Samuel, A. G. (1997). Lexical activation produces potent phonemic percepts. *Cognitive Psychology, 32*, 97–127.

Sanocki, T. (1991). Effects of early common features on form recognition. *Perception & Psychophysics, 50*, 490–497.

Sanocki, T. (1993). Time course of object recognition: Evidence for a global-to-local contingency. *Journal of Experimental Psychology: Human Perception and Performance, 19*, 878–898.

Sanocki, T. (1999). Constructing structural descriptions. *Visual Cognition, 6*, 299–318.

Sanocki, T. (2001). Interaction of scale and time during object recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 27,290–302.

Sanocki, T., & Epstein, W. (1997). Priming spatial layout of scenes. *Psychological Science, 8*, 374–378.

Sanocki, T., Sellers, E., & Mittelstadt, J. (2001). High-capacity visual memory for layout [Abstract]. *Journal of Vision, 1*, 124. Available: http://journalofvision.org/1/3/124.

Sedgwick, H. A. (1986). Space perception. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance* (pp. 1–57). New York: Wiley.

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

Shelton, A. L., & McNamara, T. P. (2001). Systems of spatial reference in human memory. *Cognitive Psychology, 43*, 1–37.

Simons, D. J. (1996). In sight, out of mind: When object representations fail. *Psychological Science, 7*, 301–305.

Simons, D. J. (2000). Change blindness and visual memory. *Special Issue of Visual Cognition, 7*.

Srinivas, K. (1996). Contrast and illumination effects on explicit and implicit measures of memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 22*, 1123–1135.

Stiles-Davis, Kritchevsky, M., & Ursula, B. (1988). *Spatial cognition: Brain bases and development*. Hillsdale, NJ: Lawrence Erlbaum.

Stone, G. O., Vanhoy, M., & Van Orden, G. C. (1997). Perception is a two-way street: Feedforward and feedback phonology in visual word recognition. *Journal of Memory & Language, 36*, 337–359.

Talyor, H. A., & Tversky, B. (1992). Descriptions and depictions of environments. *Memory & Cognition, 20*, 483–496.

Tarr, M. J., Kersten, D., & Bulthoff, H. H. (1998). Why the visual recognition system might encode the effects of illumination. *Vision Research, 38*, 2259–2275.

Tarr, M. J., & Bulthoff, H. H. (1998). Image-based object recognition in man, monkey and machine. *Cognition, 67*, 1–20.

Treisman, A., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12*, 97–136.

Treisman, A., & Gormican, S. (1988). Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review, 95*, 15–48.

Tsotsos, J. K. (1990). Analyzing vision at the complexity level. *Behavioral and Brain Sciences, 13*, 469–4233.

Tversky, B., Kim, J., & Cohen, A. (1999). Mental models of spatial relations and transformations from language. In G. Rickheit, & C. Habel (Eds.), *Mental models in discourse processing and reasoning* (pp. 239–258). Amsterdam: Elsevier.

Ungerleider, L. G., & Mishkin, M. (1982). Two cortical systems. In D. J. Ingle, M. A. Goodales, & R. J. W. Mansfield (Eds.), *Analysis of visual behavior* (pp. 549–586). Cambridge, MA: MIT Press.

Vishton, P. M., & Cutting, J. E. (1995). Wayfinding, displacements, and mental maps: Velocity fields are not typically used to determine one's aimpoint. *Journal of Experimental Psychology: Human Perception & Performance, 21*, 978–995.

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2001). Storage of features, conjunctions, and objects in visual working memory. *Journal of Experimental Psychology: Human Perception & Performance, 27*, 92–114.

Williams, A., & Weisstein, N. (1978). Line segments are perceived better in coherent contexts than alone: An object-line effect. *Memory & Cognition, 6*, 5–17.

Wolfe, J. M., Klempen, N., & Dahlen, K. (2000). Postattentive vision. *Journal of Experimental Psychology: Human Perception and Performance, 26*, 693–705.

Zeki, S. (1993). *A vision of the brain*. Oxford: Blackwell Scientific Publications.

Further reading

Chun, M. M., & Phelps, E. A. (1999). Memory deficits for implicit contextual information in amnesic subjects with hippocampal damage. *Nature Neuroscience, 2*, 844–847.

Cutting, J. E., & Vishton, P. M. (1995). Perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth. In W. Epstein, & S. Rogers (Eds.), *Handbook of perception and cognition: Vol. 5, Perception of space and motion* (pp. 1–46). San Diego, CA: Academic Press.

Henderson, J. M., & Seifert, A. B. (1999). The influence of enantiomorphic transformation of transaccadic object integration. *Journal of Experimental Psychology: Human Perception and Performance, 25*, 243–255.

Jonides, J., & Mack, R. (1984). On the cost and benefit of cost and benefit. *Psychological Bulletin, 96*, 29–44.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390*, 279–281.

Palmer, S. E. (1978). Fundamental aspects of cognitive representation. In E. Rosch, & B. Lloyd (Eds.), *Cognition and categorization* (pp. 259–303). Hillsdale, NJ: Erlbaum.

Pollatsek, A., Rayner, K., & Collins, W. E. (1984). Integrating pictorial information across eye movements. *Journal of Experimental Psychology: General, 113*, 426–442.

Pylyshyn, Z. (1994). Some primitive mechanisms of spatial attention. *Cognition, 50*, 363–384.

Rensink, R. A. (1998). Early completion of occluded objects. *Vision Research, 38*, 2489–2505.