

Extrapolating spatial layout in scene representations

MONICA S. CASTELHANO

Queen's University, Kingston, Ontario, Canada

AND

ALEXANDER POLLATSEK

University of Massachusetts, Amherst, Massachusetts

Can the visual system extrapolate spatial layout of a scene to new viewpoints after a single view? In the present study, we examined this question by investigating the priming of spatial layout across depth rotations of the same scene (Sanocki & Epstein, 1997). Participants had to indicate which of two dots superimposed on objects in the target scene appeared closer to them in space. There was as much priming from a prime with a viewpoint that was 10° different from the test image as from a prime that was identical to the target; however, there was no reliable priming from larger differences in viewpoint. These results suggest that a scene's spatial layout can be extrapolated, but only to a limited extent.

It is clear that there is considerable information in a 2-D image of a scene about the 3-D spatial arrangement of the scene being depicted and, specifically, about the distances of objects in the scene from the location of the portrayer of the scene. A large number of studies have employed simple visual stimuli to explore which types of visual cues that convey depth information are automatically activated when a 2-D image of a visual stimulus is presented. For example, simplified stimuli have been used to demonstrate the basic monocular and binocular cues that lead to interpretations of their size, depth, and distance (for a review, see Palmer, 1999); however, few studies have examined the nature of spatial processing in real-world scenes.

Spatial information processing is especially relevant to the understanding of scene recognition. Both neural-based accounts and computational models use spatial layout as a central component of scene processing. Early studies of the neural processing of scenes by Epstein and colleagues (e.g., Epstein & Kanwisher, 1998) identified the parahippocampal place area (PPA; a region of the posterior medial temporal lobe) as a region that preferentially responds to images of scenes (vs. faces or objects) and, more specifically, to their spatial layout, as shown by a larger response to scenes than to individual buildings (e.g., a house). More recently, studies have found that the processing that occurs in the PPA may be closely linked to the structural geometry of a particular view. For instance, the PPA is sensitive to changes in viewpoint of a scene (Epstein, Graham, & Downing, 2003; Epstein, Higgins, & Thompson-Schill, 2005). Epstein et al. (2005) had participants judge whether two sequentially presented images were of the

same place. The images could represent two different places, the same place from different viewpoints, or the same place from the same viewpoint. The authors found a viewpoint-specific effect, in which a viewpoint change led to less adaptation than when there was no viewpoint change. Henderson, Larson, and Zhu (2008) also found that the structural geometry of a scene influenced the PPA. They compared the activation produced by close-up scenes (e.g., view of a desk in an office) and full scenes (e.g., the entire office). They found that the PPA's level of activation was stronger for the full scenes than for the close-up scenes, which suggests that the PPA is sensitive to the real-world size of the depicted image, even when the actual size of the image is held constant.

The spatial layout also has an important role in recent computational models of complex scene recognition. For example, the strong link between the semantics of a scene and its spatial qualities is important in the *spatial envelope theory* (Greene & Oliva, 2009; Oliva & Torralba, 2001, 2002; Ross & Oliva, 2010). According to this theory, a scene's semantic category can be derived from a small set of perceptual dimensions (e.g., naturalness, openness, roughness, expansion, ruggedness) representing the spatial structure of the scene. Once values for these dimensions are calculated, the model generates a multidimensional space. From the scenes projected together within this space, semantic categories emerge (e.g., streets, highways, coasts). The model is thus able to predict scene category solely on the basis of a holistic representation of the scene shape but guided by the perceptual dimensions above. Knowing how the visual system represents

M. S. Castelhana, monica.castelhana@queensu.ca

the 3-D shape of a scene (i.e., the spatial layout of the objects and the background), therefore, may also inform us how the visual system is representing complex scenes semantically.

One basic question that has yet to be answered is how well the visual system can extract and utilize the 3-D information from a single view of a scene. This question has been explored to some extent in studies examining how viewpoint dependent various aspects of scene representation are. Some have supported viewpoint dependence (Christou & Bühlhoff, 1999; Chua & Chun, 2003; Garsoffky, Huff, & Schwan, 2007; Garsoffky, Schwan, & Hesse, 2002; Nakatani, Pollatsek, & Johnson, 2002), and others have supported viewpoint invariance (Burgess, 2002; King, Burgess, Hartley, Vargha-Khadem, & O'Keefe, 2002). For instance, Christou and Bühlhoff had participants explore an attic (consisting of multiple rooms) from certain viewpoints. When participants were asked to recognize still images taken from this environment, scene recognition was highly dependent on whether the participant had any previous experience with that space from that viewpoint, thus showing a high degree of viewpoint dependence. In another study, Garsoffky et al. (2002) asked participants to study dynamic scenes depicting a soccer game. They also found that when participants were asked to recognize scenes taken from the same or from different viewpoints, performance was highly viewpoint dependent. These studies' findings led to the conclusion that representations of the scene and/or the space are viewpoint dependent and that the visual system does not extrapolate 3-D information to other viewpoints.

Other studies, however, have found that with more extended experience of the space, viewpoint-invariant performance in scenes is observed. In a recent study, Castelano, Pollatsek, and Rayner (2009) had participants study two images prior to an immediate memory test in which participants were asked to discriminate between old and new views of the same scene. The two study images were always from two different viewpoints, 40° apart. For the memory test, the distractor images were either an *interpolated* viewpoint that was 20° from each of the study images or an *extrapolated* viewpoint that differed by 20° from one of the study images and by 60° from the other. Participants found that interpolated test images were more difficult to reject than the extrapolated test images, even when similarity of the views was accounted for. This result supports the notion that the information from the two study images was, in some sense, averaged in memory so that the interpolated test view was harder to distinguish from the study images without the need for a viewpoint-invariant representation of the scene to be initially formed. Conversely, Waller, Friedman, and colleagues (Friedman & Waller, 2008; Waller, Friedman, Hodgson, & Greenauer, 2009) found that with extensive training from multiple views, a novel view that was between the trained views but not shown during training was more easily identified as the same scene than were the trained views. These studies suggest that, under certain circumstances, extrapolation

of scene information outside the known viewpoints can be observed; however, it is not clear from these studies what role spatial information in the scenes (especially depth information) played in the generalization, rather than memory for specific objects in the scene and their relative locations in the 2-D images.

In the present study, we investigated the generalizability of the spatial layout of a room from a single view using a different task than those typically used in the investigation of viewpoint dependence of scene representations. We used a spatial layout priming paradigm based on the one introduced by Sanocki and colleagues (Sanocki, 2003; Sanocki & Epstein, 1997; Sanocki & Sulman, 2009) that specifically probes depth information. In our task, participants were shown a *prime* scene followed by a *target* scene (the same scene from either the same visual angle or a new visual angle). The target scene had two dots superimposed on two probe objects, and participants simply had to respond (*left* or *right*) which of the two objects with the dots appeared closer to them in depth. By using a variation of this paradigm in which the prime and target scenes appear from different viewpoints, we expected to obtain a direct measure of how spatial information of a scene is extracted and generalized.

Given past research, how plausible is it that the visual system is capable of extracting 3-D information from a single view of a scene and of using that information to bootstrap processing of similar viewpoints? On the one hand, previous studies on object recognition have indicated that when an object is familiar, viewpoint invariance is observed (Bühlhoff, Edelman, & Tarr, 1995). If one can generalize from objects to scene, it is plausible that even though a specific scene is unfamiliar, people have plenty of experience with certain scenes (especially North Americans with the indoor environments that were used in the present study) and with their spatial layout, since they have often navigated through these types of spaces. Thus, the visual system may be able to apply the processes that appear to work for familiar objects to scenes to “anticipate” and reduce processing loads of the same scene from a different view (Hollingworth & Henderson, 2004; Intraub & Richardson, 1989). On the other hand, it may be that scene representations rely heavily on linking spatial relations of objects within the scene and with other objects in the scene (Castelano & Henderson, 2007; Hollingworth, 2009). Thus, if a change in viewpoint is large enough, the information from one scene view may not generalize to other views because the shift in relations between the objects is tantamount to viewing a new scene. In addition to their spatial relations, a simple inventory list of the objects contained in a 2-D image of a scene would also change because a change in viewpoint causes some objects to disappear and new ones to appear. This is another challenge to the visual system's ability to generalize spatial information in a scene from a single view to different views of the scene.

Previous studies have examined whether a translational shift of the scene image results in decreased spatial priming (Sanocki, 2003, Experiment 5; Sanocki & Epstein,

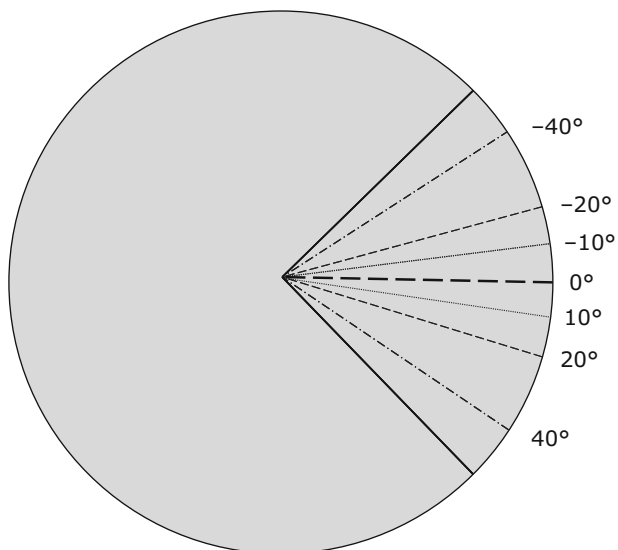


Figure 1. Bird's-eye view of the room and camera placement. The outer solid lines represent the walls of the room. Each scene was rendered from seven different viewpoints. The test image was always taken from the 0° position. The cameras at 10°, 20°, and 40° to the left and to the right of the test positions are depicted.

1997, Experiment 3). The results point to an invariant representation of the scene. However, translational shifts, especially when small, do not appreciably change the spatial relations of objects within the scene (especially depth relationships); these shifts mainly affect the relative position of objects with respect to retinal coordinates. As a result, it is difficult to conclude from these experiments how spatial layout information is extracted and whether the representation is invariant with respect to viewpoint

changes that result in changing the relative position of objects within the scene.

EXPERIMENT 1

The present study investigated how well the visual system can extrapolate depth information from a single view of a scene. As indicated above, this question was addressed by priming participants with an image of a scene and testing the extent to which the prime facilitated judgments of depth information when there was a change of viewpoint between the prime scene image and the target scene image.

Method

Participants. Ninety undergraduate students participated for credit toward an introductory psychology course. All had normal or corrected-to-normal vision.

Stimuli and Apparatus. Stimuli consisted of fifty 3-D rendered images of naturalistic scenes, created with DataBecker HomeDesign 5.0, that represented mostly indoor scenes (with a few outdoor scenes, such as a backyard and playground). These scenes (which subtended a visual angle of 40.6° horizontally and 31.2° vertically) were displayed in grayscale at a resolution of 800 × 600 pixels on a 21-in. monitor with a refresh rate of 100 Hz. Each scene was rendered from seven different viewpoints: -40°, -20°, -10°, 0°, +10°, +20°, and +40° (the test image was always presented at the 0° viewpoint). Each scene was rendered from these seven different viewpoints (see Figure 1 for a diagram of camera positions). The 0° viewpoint was used for the test image and represented the viewpoint position that allowed for the greatest number of viewpoint changes in both rotation to the left and rotation to the right. Figure 2 shows an example stimulus in each of seven different viewpoints.

Design and Procedure. Participants were shown a *preview* image immediately followed by a *test* image (Figure 3). The preview images could be a mask¹ (*control* condition), the same scene from the same viewpoint (*identical* condition), or the same scene from a viewpoint that was ±10° away (*10° difference* conditions), ±20° away (*20° dif-*

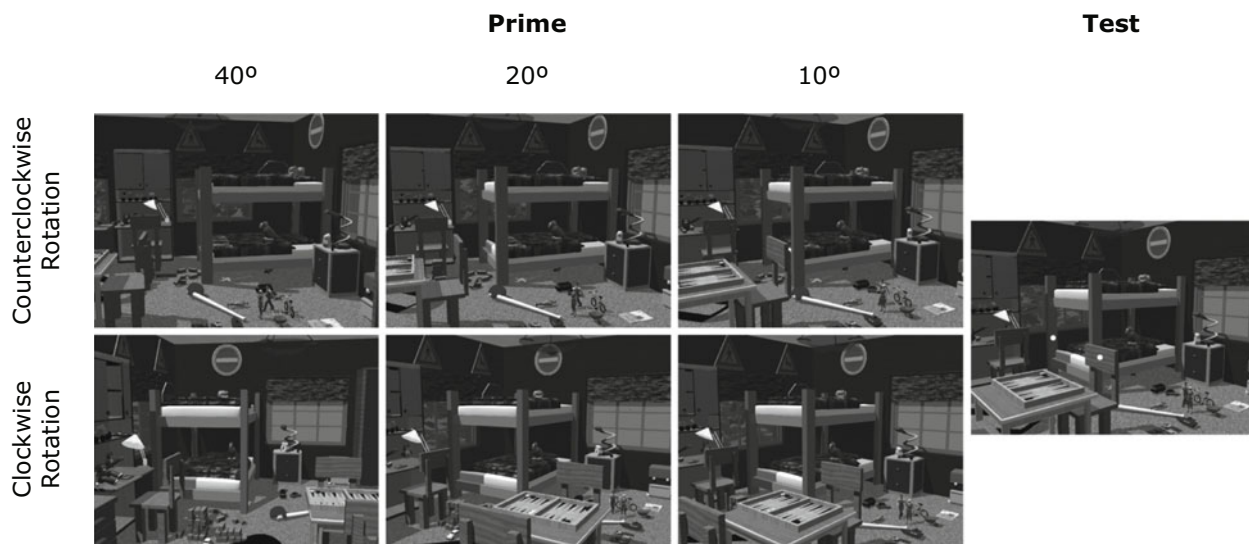


Figure 2. Example scene in each of the viewpoint prime conditions and in the test condition. For each of the viewpoint prime conditions, there were two possible images used (clockwise and counterclockwise rotations). These were counterbalanced across participants. For the test image, two red dots were placed on two objects (represented in the image as large white dots), and participants had to indicate (left or right) which probed object appeared closer to them in space. The probe dots were presented in red in the experiments but are depicted here as white dots for ease of viewing.

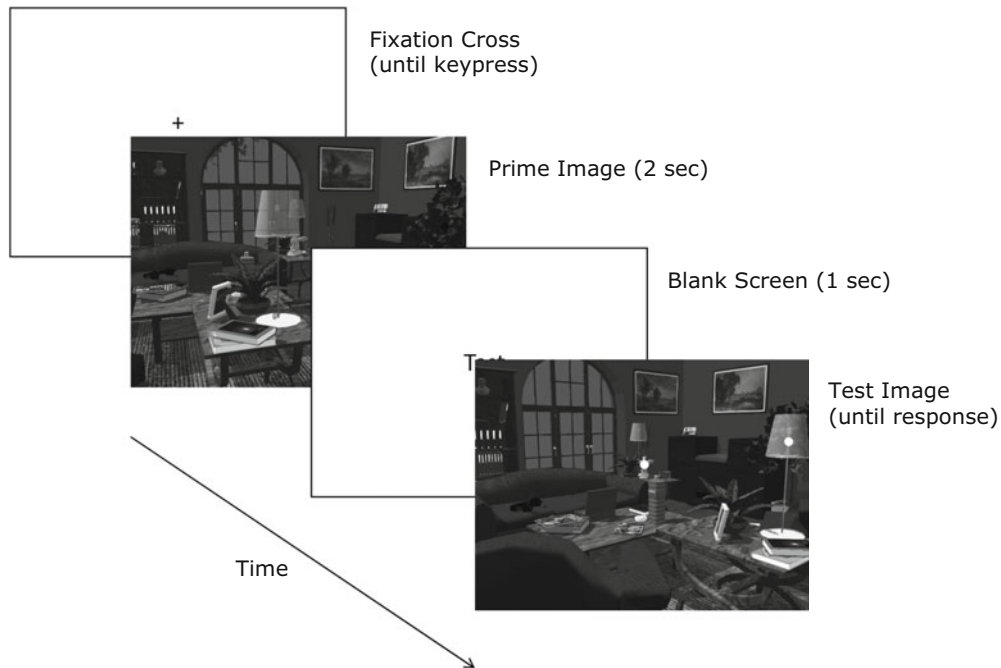


Figure 3. Sequence of images seen for each trial.

ference conditions), or $\pm 40^\circ$ away (40° difference conditions). The five preview conditions were counterbalanced over the scenes across participants. For each of the different viewpoint conditions, clockwise and counterclockwise rotations were counterbalanced as well (for example, a participant would see half the scenes being presented at a 20° difference rotated clockwise and the other half at a 20° difference rotated counterclockwise). Participants were instructed to study the preview image of the scene in order to help them respond to the test image, in which they would indicate which dot (left or right) was closer to them in space. On each test image, two red dots were placed on either two surfaces or two objects in such a way that one of the dots was clearly more in the foreground than the other. To keep error rates low, the dots were kept close enough to each other that (subjectively) no large shift of attention or of the eyes was necessary to compare the two. Participants were instructed to judge the dots as if they were an observer standing in the room at the viewing position of the camera that recorded the image.

There were 50 experimental trials (10 trials in each preview condition) preceded by 15 practice trials. None of the stimuli shown in the practice trials were used in experimental trials, and each experimental scene was viewed on only 1 trial. There was no break between the practice and experimental trials.

At the beginning of each trial, a black fixation cross was shown at the center of a white background. Participants were instructed to look at this fixation cross and to press a key when they were ready to begin a trial. On each trial, the preview image was shown for 2 sec. The 2-sec study time was longer than that used in most spatial layout priming studies (Sanocki, 2003; Sanocki & Epstein, 1997; Sanocki & Sulman, 2009); however, we opted for a longer study time in order to increase the chances that priming would occur; that is, it may take appreciable time for these spatial relations to extrapolate over viewpoint changes (Castelhano et al., 2009; Shepard & Metzler, 1971). The preview image was followed by the word "Test," which was displayed for an additional 1 sec at the center of the screen to indicate that the next image would require a response. The test image was then displayed until the participant responded. Participants responded which of the two dots (left or right) appeared closer to them in space. They responded with the left key on a button box to indicate the dot on the left and with the corresponding right button to

indicate the dot on the right. Participants were instructed to respond as quickly as possible while maintaining a high accuracy rate. No feedback was given. The experiment took 15–20 min to complete.

Results

Six participants were excluded for having accuracy rates at chance or lower. In addition, because we were interested in the initial processing of the test scene (and not in responses that reflected processing that involved problem solving), 10 participants were excluded for having average response times (RTs) greater than 2,000 msec.² The accuracy and mean RT for each condition for the remaining 74 participants are reported in Table 1. The accuracies were quite high (96%), with no difference among the preview conditions ($F < 1$); thus, the focus here is on the RT data.

An omnibus ANOVA showed that the preview type had a significant effect on the RT measure [$F(4,292) = 3.06$, $p < .05$]. Furthermore, a planned contrast showed that the 58-msec difference between the control and identical conditions was significant [$t(73) = 2.09$, $p < .04$]. Of greater theoretical interest was the significant 65-msec difference between the control condition and the 10° difference conditions [$t(73) = 2.84$, $p < .01$]. Neither the 12-msec difference between the control condition and the 20° difference conditions nor the 3-msec difference between the control condition and the 40° difference conditions was close to significant (both $ts < 1$). In addition, we compared the identical condition with the other conditions and found that responses were significantly faster than those in the 40° difference conditions [$t(73) = 2.21$, $p < .01$] and marginally significantly faster than those in the 20° difference conditions [$t(73) = 1.93$, $p = .056$], but not faster than those in the 10° difference conditions ($t < 1$).

Table 1
Mean Accuracy (Percentages) and Response Times (in Milliseconds)
to the Test Image by Preview Condition for Experiment 1

Condition	Accuracy		Response Time		Difference in Response Time From Control	Supplementary Analysis		
	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>		Response Time		Difference in Response Time From Control
						<i>M</i>	<i>SE</i>	
Control	96	1.1	1,161	43	–	–	–	
Identical	96	1.2	1,103	37	58	–	–	
10°	97	0.1	1,096	39	65	1,090	36	
20°	96	0.1	1,149	44	11	1,158	46	
40°	96	0.1	1,164	43	–3	1,162	45	

Supplementary analyses. There was a mediating variable that we thought would be of interest to explore. Upon inspecting the primes, we discovered that the target objects were not in all the prime stimuli; therefore, it is possible that the limited extent to which spatial layout was extrapolated was due to the target absence in the prime. As a result, we conducted an analysis using only those scenes in which both target objects were present in both prime scenes. Means are presented on the right-hand side of Table 1. The priming effects from this analysis were quite similar to those in the main analysis. The priming effects at 10°, 20°, and 40° were 73 msec [$t(79) = 2.21$, $p < .05$], 5 msec ($t < 1$), and 1 msec ($t < 1$), respectively. Thus, we do not think that the lack of a consistent priming effect at either 20° or 40° was due to the absence of target objects in the image.³

Discussion

First, the results from Experiment 1 replicate earlier findings that there is a clear benefit from the identical prime in reducing the time to make a spatial judgment. It is worth emphasizing that the information from the prime did help in making the judgments: One could have made the case, a priori, that the best strategy was to ignore the prime, given that the prime was physically different from the test image 80% of the time.

Second, the results suggest that spatial relations of surfaces and objects within the scene representation can be extrapolated to neighboring viewpoints, but only to a limited extent (10°). The sudden drop in performance at 20° from the original viewpoint suggests that after a limited change in viewpoint, a new scene representation is required. There are some caveats with this conclusion, however. First, the scenes used in Experiment 1 were much more cluttered and complex than those used in earlier studies (Sanocki, 2003; Sanocki & Epstein, 1997); thus, it is possible that the limited generalizability resulted from the makeup of these scenes. In addition, the target objects and surfaces that were tested with the dot probes were not always present in the prime scenes, especially in the more extreme viewpoint-change conditions (e.g., 20° and 40°). Although the supplementary analysis supported the original conclusion that generalization was seen only for small viewpoint changes (10°) and not at all for slightly larger ones (e.g., 20°), it is not clear whether the lack of a priming effect is merely a case of reduced power, be-

cause the supplementary analysis removed data points. As a result, we ran a second experiment to address these two concerns: (1) scene clutter and complexity and (2) target presence in the prime scenes.

EXPERIMENT 2

In Experiment 2, we employed a modified stimulus set. As indicated above, the scenes used in Experiment 1 were complex and had many more objects that could disappear or appear as the scene was rotated; thus, the limited generalization of spatial layout observed could have been due to either of these problems. In Experiment 2, the scenes were simplified by removing most of the content objects. In addition, all prime images included both of the target objects that were subsequently probed in order to increase power and to reduce the need to engage in problem solving (i.e., nonperceptual processing).

Method

Participants. Eighty undergraduate students participated for credit toward an introductory psychology course. All had normal or corrected-to-normal vision, and none had participated in Experiment 1.

Stimuli and Apparatus. Twenty-two new scenes were added (from the same computer-generated scene database) for a total of 72 scenes. The other 50 images were the same as those used in Experiment 1, except that they were greatly simplified by removing many of the content objects and thus reducing clutter. An example of the simplified scenes is shown in Figure 4. The apparatus was the same as in Experiment 1.

Design and Procedure. The design and procedure were very similar to those of Experiment 1, with a few exceptions. First, the 40° difference conditions were not included in Experiment 2. Thus, the preview image was a mask (control condition), the same scene from the same viewpoint (identical condition), or the same scene from a viewpoint that was $\pm 10^\circ$ away (10° difference conditions) or $\pm 20^\circ$ away (20° difference conditions). There were thus four preview conditions (18 scenes per condition). Second, the dots were always placed on surfaces and objects that were visible in both the preview image and the test image. Otherwise, the procedure and task were identical to those in Experiment 1.

Results

Two participants were excluded for having accuracies at chance or lower. In addition, 6 participants were excluded for having average RTs greater than 1,800 msec.⁴ The accuracy and mean RT for each condition for the remaining 72 participants are reported in Table 2. On average, ac-



Figure 4. Example scenes from Experiment 1 and Experiment 2 depicted in the test image. The probe dots were presented in red in the experiments but are depicted here as white dots for ease of viewing.

curacies were quite high (98%), with no significant differences among the preview conditions ($F < 1$); thus, as was the case in Experiment 1, the focus of our analyses was on the RT data.

An omnibus ANOVA showed that the preview type had a significant effect on the RT measure [$F(4,292) = 5.41$, $p < .01$]. Furthermore, a planned contrast showed that the 57-msec difference between the control condition and identical condition was significant [$t(73) = 2.09$, $p < .04$]. Of greater theoretical interest was the significant 66-msec difference between the control condition and the 10° difference conditions [$t(73) = 2.84$, $p < .01$]. The -2-msec difference between the control condition and the 20° difference conditions was not significant (both t s < 1). In addition, we compared the identical condition with the other conditions and found that there was a

significant difference from the 20° difference conditions [$t(73) = 2.60$, $p < .02$], but not from the 10° difference conditions ($t < 1$).

Discussion

The results of Experiment 2 replicated those of Experiment 1; they showed that the spatial layout of the scene could be extrapolated from a prime shifted 10° but not from one shifted 20°. These results indicate that the limited generalizability in Experiment 1 was not due to high scene complexity or to the absence of the target objects and surfaces in some of the primes. In addition, although the 20° prime produced a small, insignificant priming effect in Experiment 1, the 20° prime produced a very small negative effect in Experiment 2. It therefore seems unlikely that the failure to obtain a significant effect at 20°

Table 2
Mean Accuracy (Percentages) and Response Times
(in Milliseconds) to the Test Image by Preview Condition
for Experiment 2

Condition	Accuracy		Response Time		Difference in Response Time From Control
	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	
Control	99	0.04	1,118	42	–
Identical	98	0.05	1,061	36	57
10°	98	0.05	1,052	36	66
20°	99	0.03	1,121	40	–2

in Experiment 1 was due to a lack of power in the experimental design; instead, it was likely due to there being virtually no priming effect at this angular displacement. Thus, it seems that the scene representation formed from the prime has some information about object and surface spatial relations that facilitate processing—but only if the difference in viewpoint angle is relatively small. We obviously cannot determine from these two experiments exactly where the generalization ends; it is somewhere between 10° and 20°. The exact answer may depend on various factors, such as the specific scene layouts.

GENERAL DISCUSSION

The major results from the two experiments are quite clear: (1) There were clear priming effects from an identical prime, replicating the results of Sanocki and colleagues (Sanocki, 2003; Sanocki & Epstein, 1997; Sanocki & Sulman, 2009); and (2) this priming effect generalized but only to primes that differed by 10° from the test scene. Furthermore, Experiment 2 demonstrated that this limited generalization occurred for even relatively simple, “uncluttered” scenes and when the target locations were present in both the prime and test scenes.

The obvious question is: What type of information is maintained from the prime stimulus that aids in making the distance judgment in the test stimulus? It seems unlikely that what was retained from the prime is a full 3-D model of the scene geometry. If such a representation was retained, that is, one would have expected significant priming at all rotation angles, possibly with RTs decreasing linearly as the viewpoint change increased as a result of something like a “mental rotation” process (Shepard & Metzler, 1971). Another unlikely possibility is that what was retained from the prime was an ordinal representation of how far various objects and surfaces were from the viewer (i.e., a recoding of the distance information into a more symbolic form). This also does not seem very plausible, because the placement of the dots was not predictable, so an ordinal list of distances of potential objects and surfaces would have to be fairly long. Moreover, if this ordinal representation was preserved, there should have been relatively equal priming in the 0°, 10° difference, and 20° difference conditions (where the ordinal information about the depth of the objects would have been the same), and this was clearly not what occurred.

Instead, we think that the most plausible representation of the information of the prime that is maintained and utilized is not very different from Marr’s (1982) “2 1/2-D sketch.” That is, the representation appears to be “viewer centered” and to contain more than an ordinal list of the distances that the objects and surfaces are from the viewer (or at least some of those distances, perhaps constrained by visual short-term memory limitations). The fact that the priming effect with a 10° change in viewpoint is virtually the same as with an identical scene prime, however, suggests that neither the distances nor the absolute *x*-, *y*-coordinates of the features in the prime are encoded veridically. This is consistent with Castelhana and Henderson’s (2007) finding that initial scene representations are not metric but instead contain relational information in a more abstract way that can withstand changes in scale.

However, the present study indicates that this relational information from the prime scene can be applied to making the distance judgments in the target scene only for small changes in viewpoint. One possible reason for this limitation is that the greater the angular displacement between prime and test images, the more the relational information changes. Another possibility, which seems more likely, is that the greater the angular displacement between prime and test images, the more both the absolute locations of the objects in the image and the distance between them changes. That is, there may be a “race” between use of the bottom-up information from the test image and the more top-down priming effects coming from the information in memory from the prime. It may be that the application of the top-down relational information is slowed down quite substantially the greater the location mismatch is between the 2-D representations in the prime and target images. The work showing that this priming is translation invariant (Sanocki, 2003; Sanocki & Epstein, 1997) suggests that if this hypothesis is correct, the mismatch is in relative, rather than absolute, locations of objects.

Another possibility is that more general or categorical spatial information could be encoded with a general idea of objects being either in the foreground or in the background. Since the viewpoint changes are rotations around a single point at the back of the room, the objects in the foreground would undergo greater change in the 2-D image (with the objects coming into and falling out of view) than would those objects further back. Thus, if there is a general coding for objects in the foreground, these would become increasingly invalid as the angle of rotation increases (as outlined above).

Regardless of the correctness of the hypotheses for why there is such limited generalization across angular displacements, the data coincide with earlier studies that conclude that viewpoint-dependent information about scenes is difficult to apply from one scene viewpoint to another (Castelhana et al., 2009; Christou & Bühlhoff, 1999; Garsoffky et al., 2007; Garsoffky et al., 2002; Nakatani et al., 2002). These data further indicate that the information abstracted from the prime that facilitates processing of the test image is unlikely to be any automatic application of 3-D knowledge about the scene.

AUTHOR NOTE

This work was supported by a grant from the Microsoft Corporation to Keith Rayner and by grants from the Natural Sciences and Engineering Research Council of Canada and the Advisory Research Committee of Queen's University to M.S.C. The authors thank Catherine Seohyun Jee, Effie Pereira, and Angelita Wong for help with data collection. Correspondence concerning this article should be addressed to M. S. Castelhana, Department of Psychology, Queen's University, 62 Arch Street, Kingston, ON, K7L 3N6 Canada (e-mail: monica.castelhana@queensu.ca).

REFERENCES

- BÜLTHOFF, H. H., EDELMAN, S. Y., & TARR, M. J. (1995). How are three-dimensional objects represented in the brain? *Cerebral Cortex*, *5*, 247-260.
- BURGESS, N. (2002). The hippocampus, space, and viewpoints in episodic memory. *Quarterly Journal of Experimental Psychology*, *55A*, 1057-1080.
- CASTELHANO, M. S., & HENDERSON, J. M. (2007). Initial scene representations facilitate eye movement guidance in visual search. *Journal of Experimental Psychology: Human Perception & Performance*, *33*, 753-763.
- CASTELHANO, M. S., POLLATSEK, A., & RAYNER, K. (2009). Integration of multiple views of scenes. *Perception & Psychophysics*, *71*, 490-502.
- CHRISTOU, C. G., & BÜLTHOFF, H. H. (1999). View dependence in scene recognition after active learning. *Memory & Cognition*, *27*, 996-1007.
- CHUA, K.-P., & CHUN, M. M. (2003). Implicit scene learning is viewpoint dependent. *Perception & Psychophysics*, *65*, 72-80.
- EPSTEIN, R. [A.], GRAHAM, K. S., & DOWNING, P. E. (2003). Viewpoint-specific scene representations in human parahippocampal cortex. *Neuron*, *37*, 865-876.
- EPSTEIN, R. A., HIGGINS, J. S., & THOMPSON-SCHILL, S. L. (2005). Learning places from views: Variation in scene processing as a function of experience and navigational ability. *Journal of Cognitive Neuroscience*, *17*, 73-83.
- EPSTEIN, R. [A.], & KANWISHER, N. (1998). A cortical representation of the local visual environment. *Nature*, *392*, 598-601.
- FRIEDMAN, A., & WALLER, D. (2008). View combination in scene recognition. *Memory & Cognition*, *36*, 467-478.
- GARSOFFKY, B., HUFF, M., & SCHWAN, S. (2007). Changing viewpoints during dynamic events. *Perception*, *36*, 366-374.
- GARSOFFKY, B., SCHWAN, S., & HESSE, F. W. (2002). Viewpoint dependency in the recognition of dynamic scenes. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *28*, 1035-1050.
- GREENE, M. R., & OLIVA, A. (2009). Recognition of natural scenes from global properties: Seeing the forest without representing the trees. *Cognitive Psychology*, *58*, 137-176.
- HENDERSON, J. M., LARSON, C. L., & ZHU, D. C. (2008). Full scenes produce more activation than close-up scenes and scene-diagnostic objects in parahippocampal and retrosplenial cortex: An fMRI study. *Brain & Cognition*, *66*, 40-49.
- HOLLINGWORTH, A. (2009). Two forms of scene memory guide visual search: Memory for scene context and memory for the binding of target object to scene location. *Visual Cognition*, *17*, 273-291.
- HOLLINGWORTH, A., & HENDERSON, J. M. (2004). Sustained change blindness to incremental scene rotation: A dissociation between explicit change detection and visual memory. *Perception & Psychophysics*, *66*, 800-807.
- INTRAUB, H., & RICHARDSON, M. (1989). Wide-angle memories of close-up scenes. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *15*, 179-187.
- KING, J. A., BURGESS, N., HARTLEY, T., VARGHA-KHADEM, F., & O'KEEFE, J. (2002). Human hippocampus and viewpoint dependence in spatial memory. *Hippocampus*, *12*, 811-820.
- MARR, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco: Freeman.
- NAKATANI, C., POLLATSEK, A., & JOHNSON, S. H. (2002). Viewpoint-dependent recognition of scenes. *Quarterly Journal of Experimental Psychology*, *55A*, 115-119.
- OLIVA, A., & TORRALBA, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision*, *42*, 145-175.
- OLIVA, A., & TORRALBA, A. (2002). Scene-centered description from spatial envelope properties. In *Proceedings of the Second Workshop on Biologically Motivated Computer Vision* (pp. 263-272). Berlin: Springer.
- PALMER, S. E. (1999). *Vision science: Photons to phenomenology*. Cambridge, MA: MIT Press.
- ROSS, M. G., & OLIVA, A. (2010). Estimating perception of scene layout properties from global image features. *Journal of Vision*, *10*(1, Art. 2), 1-25.
- SANOCKI, T. (2003). Representation and perception of spatial layout. *Cognitive Psychology*, *47*, 43-86.
- SANOCKI, T., & EPSTEIN, W. (1997). Priming spatial layout of scenes. *Psychological Science*, *8*, 374-378.
- SANOCKI, T., & SULMAN, N. (2009). Priming of simple and complex scene layout: Rapid function from the intermediate level. *Journal of Experimental Psychology: Human Perception & Performance*, *35*, 735-749.
- SHEPARD, R. N., & METZLER, J. (1971). Mental rotation of three-dimensional objects. *Science*, *171*, 701-703.
- WALLER, D., FRIEDMAN, A., HODGSON, E., & GREENAUER, N. (2009). Learning scenes from multiple views: Novel views can be recognized more efficiently than learned views. *Memory & Cognition*, *37*, 90-99.

NOTES

1. The mask was composed of scrambled sections taken from all the scenes that were shown in the experiment.
2. This cutoff is similar to that of an earlier study (Castelhana et al., 2009), and the remaining RTs were closer to the RTs that were found with the same paradigm in earlier studies (Sanocki, 2003; Sanocki & Epstein, 1997).
3. Note that in the 40° difference conditions there were only 12 scenes in which both targets were present in the prime, so the estimate of the priming effect was not very reliable.
4. The cutoff is equal to approximately 2 standard deviations from the mean. On average, participants were faster to respond in Experiment 2 than in Experiment 1.

(Manuscript received February 10, 2010;
revision accepted for publication May 4, 2010.)