

Scene context influences without scene gist: Eye movements guided by spatial associations in visual search

Monica S. Castelhana · Chelsea Heaven

Published online: 20 May 2011
© Psychonomic Society, Inc. 2011

Abstract Although the use of semantic information about the world seems ubiquitous in every task we perform, it is not clear whether we rely on a scene's semantic information to guide attention when searching for something in a specific scene context (e.g., keys in one's living room). To address this question, we compared contribution of a scene's semantic information (i.e., scene gist) versus learned spatial associations between objects and context. Using the flash-preview–moving-window paradigm Castelhana and Henderson (*Journal of Experimental Psychology: Human Perception and Performance* 33:753–763, 2007), participants searched for target objects that were placed in either consistent or inconsistent locations and were semantically consistent or inconsistent with the scene gist. The results showed that learned spatial associations were used to guide search even in inconsistent contexts, providing evidence that scene context can affect search performance without consistent scene gist information. We discuss the results in terms of hierarchical organization of top-down influences of scene context.

How past experience impacts current processing is an issue that comes up in many areas of research, and one area of research in which this is central is scene processing. Whereas early research focused discussion on the influence of scene schemas and frames (see Friedman, 1979), more recent research tends to focus on the influence of scene gist (Henderson & Hollingworth, 1999; Oliva, 2005). Scene gist is thought to include semantic expectations about what objects belong and where they are likely to be found (Biederman, Mezzanotte, & Rabinowitz, 1982; Castelhana & Henderson, 2008; Henderson & Hollingworth, 1999; Oliva, 2005). When a scene is processed, these expectations have an effect on subsequent

processing. For instance, Biederman et al. found that with certain violations of expectations (e.g., size, position, probability, support), identification of objects became more difficult.

One question that remains is whether knowledge arising from scene gist also affects visual search. Henderson, Weeks, and Hollingworth (1999) found that target objects that were semantically consistent with scene gist were located more quickly than inconsistent ones. They concluded that attention could be directed to the target more rapidly when scene gist could be used to identify likely target locations. Other studies (Malcolm & Henderson, 2009; Neider & Zelinsky, 2006; Zelinsky & Schmidt, 2009) had participants look for semantically consistent targets but manipulated their locations within scenes. Findings showed that attention is directed effectively when targets are in their expected locations, but not when in unexpected or inconsistent locations. So, matching the target with scene gist seems to be an important component of how scene context influences attention allocation in search.

Although researchers often point to scene gist as the source for knowledge of where objects belong in a scene, there is reason to believe that where an object is found and whether it belongs are separate stores of knowledge. Across a number of fields, research has outlined the influence of knowledge based on learned spatial associations (Chamizo, 2003; Chun & Jiang, 1999; Gillner & Mallot, 1998). For instance, research has shown that in both humans and rats, learned spatial associations play an important role across a number of tasks, including navigation, landmark acquisition, and learning of arbitrary spatial configurations (Chamizo, 2003; Gillner & Mallot, 1998; Sturtz, Kelly, & Brown, 2010). In addition, using repeated visual search of letter arrays, researchers have shown that spatial associations between the target and the context, the surrounding elements, or the response can markedly improve search performance (Chun & Jiang, 1999; Jiang & Wagner, 2004; Kunar, Flusberg, Horowitz, & Wolfe, 2007). Thus, on the basis of these results, it seems reasonable that the learned spatial associations between the

M. S. Castelhana (✉) · C. Heaven
Department of Psychology, Queen's University,
62 Arch St., Kingston, ON K7L 3N6, Canada
e-mail: monica.castelhana@queensu.ca

object and its context should remain useful even when scene gist (defined as whether the object belongs in the scene) does not. In the present study, we investigated whether spatial associations are influential in situations in which knowledge of scene gist does not easily apply.

In order to have a clear understanding of how current incoming perceptual information is processed, it is necessary to know how relevant knowledge is stored and accessed. Up until now, little research has explored the organization of different types of knowledge arising from previous experience with scenes. Now, the two sources of information from past experiences outlined above (scene gist and spatial associations) can be thought of as separate knowledge systems working in parallel, as systems that are integrated and duly influenced by one another, or as nested systems in which the knowledge is hierarchically organized (e.g., scene gist information leads to notions of the spatial arrangement of objects). We explored these possibilities in the present study.

We investigated the influence of scene gist, learned spatial associations, and how these sources of information relate. Participants were asked to search for target objects that were either consistent or inconsistent with the scene gist and were placed in consistent or inconsistent locations within these scenes. We considered two alternatives for how these two sources of information are related.

The first possibility is the one most supported by the discussion in the field (see Tatler, 2009). It posits that context influences are organized hierarchically, where retrieving scene gist leads to other expectations, such as knowledge of spatial associations. Thus, when an object is searched for within an inconsistent scene context, learned spatial associations would have no effect on performance because they would not apply to an object that is not semantically associated with the scene.

The second possibility is that scene gist and spatial associations are separate sources of knowledge about a scene context. This notion is not currently discussed in the scene-processing literature, but is supported by studies in animal cognition, human learning, and navigation. In this case, when an object is searched for within an inconsistent scene context, an object that is located in a consistent spatial location (e.g., a cookbook on a bathroom counter) would be found earlier than one located in an inconsistent spatial location (e.g., a cookbook next to a toilet).

Method

Participants Sixteen Queen's University undergraduates participated for course credit or \$10/h, and each had normal or corrected-to-normal vision (excluding eyeglasses).

Stimuli and apparatus The stimuli were 48 color photographs of indoor scenes (800 × 600 pixels). For each, a

semantically consistent and inconsistent target was selected and placed in a consistent and inconsistent location (see Fig. 1a). For semantically inconsistent targets, each image was paired with an image from a different category, and targets were swapped using Adobe Photoshop CS3. For semantically consistent targets, each image was paired with an image from the same category, and targets were swapped (i.e., a teapot and a toaster in a kitchen).¹

Each target was placed in a consistent and an inconsistent location in each consistent and inconsistent scene, away from the image center (see Fig. 2).² Violation of other scene context expectations, such as size or support (see Biederman et al., 1982) were excluded by adjusting targets in size and placing them on plausible surfaces. Thus, each image had four search scenes: inconsistent-target-consistent-location, inconsistent-target-inconsistent-location, consistent-target-consistent-location, and consistent-target-inconsistent-location. The scene previews were created by excluding the target objects from each original scene. The target previews were centered target images on a gray background, with the category name in black text.

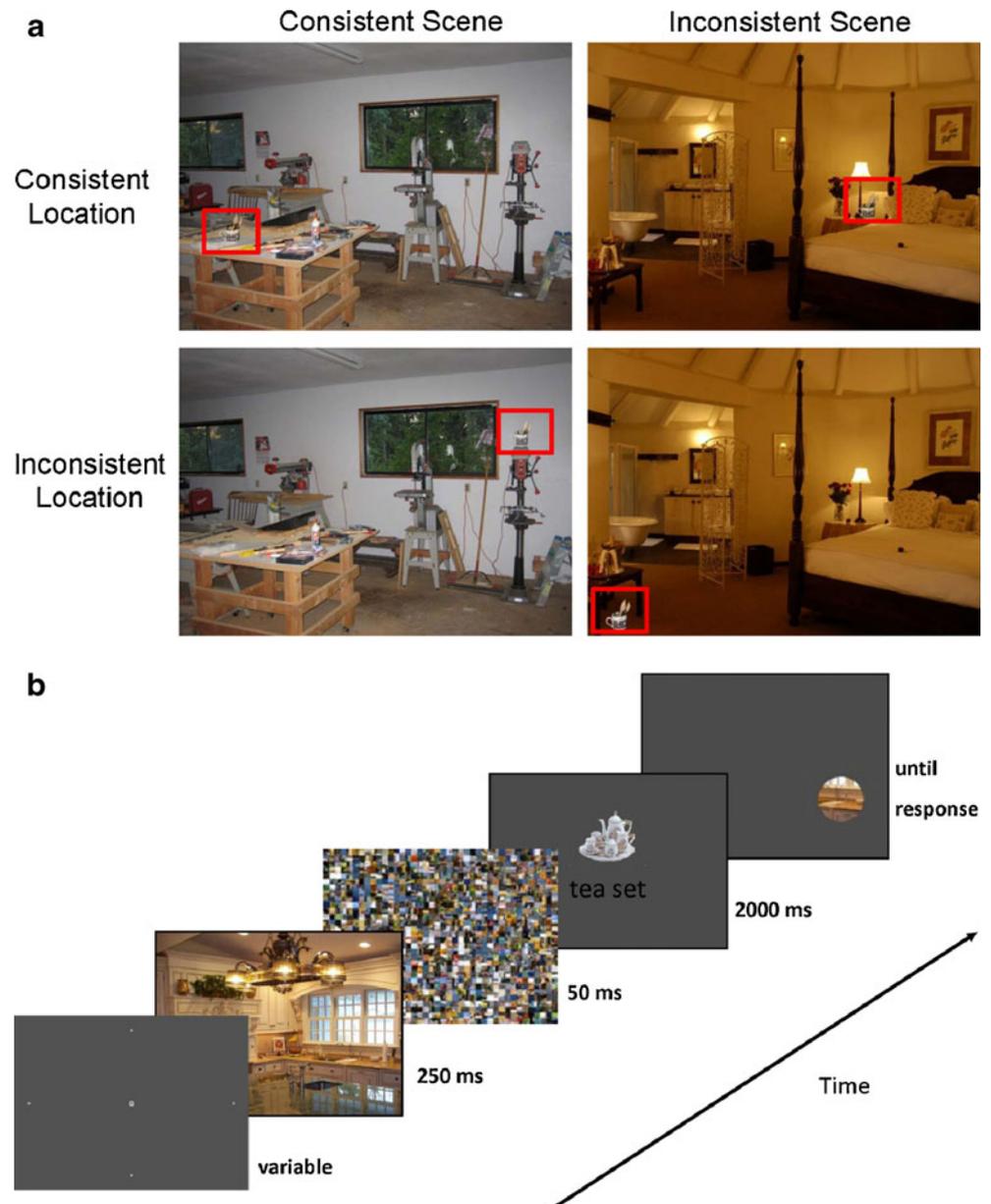
Participants' eyes were tracked using an EyeLink 2000 (SR Research) sampling at 1000 Hz. The stimuli were displayed on a 21-in. CRT monitor at a refresh rate of 100 Hz. Scenes subtended 38.1° × 28.6° of visual angle, and on average targets subtended 2.59° along their longest axis.

Procedure Participants were seated 60 cm from the monitor on a headrest and were instructed to look for a target object in a photograph. Each participant was calibrated at the start of the experiment, and calibration was considered accurate when estimated fixation positions fell, on average, within ~0.4° for all points. Calibration was checked at the start of each trial, and recalibration took place when accuracy decreased below the established standards. For the experiment, we used the flash-preview-moving-window (FPMW) paradigm (Castelhano & Henderson, 2007; see Fig. 1b). For each trial, participants were presented the preview scene for 250 ms, the mask for 50 ms, and the target preview for 2,000 ms. The search scene would then onset, and

¹ This trading of semantically consistent objects controlled for any cut-and-paste effects from inserting targets into a scene that might affect object processing.

² A norming study ($N = 24$) was run to select inconsistent/consistent locations for each target in each scene type. Each participant was presented with the 48 scenes and the name of the inconsistent/consistent target in the scene. The participant was then asked to select locations that they would "look first" or "look last" by choosing one of four locations (4AFC). The locations were selected on the basis of plausible surfaces across the scene so that support violations would not occur (see Biederman et al., 1982). Half of the trials were "look first" questions, and trial type was randomized. For each image, most likely and least likely options with the highest number of responses were selected.

Fig. 1 a Example of search scenes for each scene consistency and location consistency condition. The target is a mug of paint brushes, and its location is highlighted in both the workshop and bedroom scenes. **b** Example screen shots of one trial in the present experiment. Participants fixated a center point on a calibration screen, viewed the preview scene for 250 ms, followed by a visual mask for 50 ms, followed by word naming of the target object for 2,000 ms. The search scene was then shown through a 2° radius circular window centered at fixation until response or for a maximum of 20 s



participants searched for the target through a moving window (2° radius) tied to their current fixation. The search scene was visible only through the window, while the rest of the screen was masked with a gray screen. By using the FPMW paradigm, we allowed participants to rely only on their expectations and knowledge of the scene context but eliminated any effects of immediately available peripheral information (e.g., detection of target features or other related objects). The search scene was displayed until the participant pressed a button or until 20 s had elapsed.

After 6 practice trials, participants completed 48 experimental trials. The scenes in each experimental condition were presented in random order for each participant. Each scene was viewed only once, but as a result of target trading, each target was presented twice—once as a

semantically consistent object and another time as a semantically inconsistent object. There was always a target in the search scene, but the target was never present in the preview scene. The experiment took ~30 min to complete.

Results

In addition to accuracy and response time (RT), we examined eye movements in order to distinguish between the initial search for the target and the subsequent identification of the target, once it was fixated (Castelhano & Heaven, 2010; Castelhano, Pollatsek, & Cave, 2008; Malcolm & Henderson, 2009).

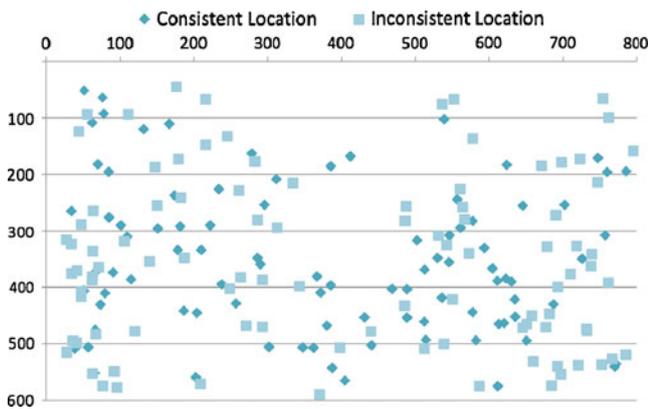


Fig. 2 Spatial distribution of target locations for the consistent and inconsistent locations. Each point represents the central point of the target object. No target objects were placed in the center of the screen to avoid ceiling effects during search. Target locations were normed and were chosen on the basis of physical constraints of the images, so as not to introduce other anomalous effects (e.g., size, occlusion, etc.)

Accuracy and response times Mean accuracy and RTs are presented in Table 1. For the accuracy measure, a two-way repeated measures analysis of variance (ANOVA) revealed an effect of location consistency, $F(1, 15) = 12.87, p < .01$, and an effect of scene consistency, $F(1, 15) = 6.53, p < .05$. There was no significant interaction, $F(1, 15) < 1, n.s.$ Because we were interested in mechanisms that led to a successful search, we examined only correct response in the remaining analyses. For the RT measure, a two-way repeated measures ANOVA revealed that there was a main effect of location consistency, $F(1, 15) = 26.21, p < .001$, but no main effect of scene consistency and no significant interaction.³

Latency to target Mean target latencies and number of fixations are presented in Fig. 3. The target latency reflects the initial search for the target from the onset of the search until the first fixation on the target (excluding the first fixation). A two-way repeated measures ANOVA showed that there was a main effect of location consistency, $F(1, 15) = 25.72, p < .001$, but no effect of scene consistency, $F(1, 15) < 1, n.s.$ However, there was a significant interaction, $F(1, 15) = 8.33, p < .05$. Planned comparisons revealed that when participants searched through a consistent or an inconsistent scene, search for a consistent location was faster than that for an inconsistent location, $t(15) = 5.31, p < .01$, and $t(15) = 2.81, p < .05$, respectively. When participants looked for a target in a consistent location, consistent scenes still produced shorter

³ The lack of an effect of scene consistency may seem at odds with previous studies; however, the scene consistency factor includes trials on which the scene category is consistent but the placement of the object is not. So, these trials include more performance variability than has been found in previous studies.

latencies than did inconsistent scenes, $t(15) = 3.42, p < .01$, but not for an inconsistent location, $t(15) = 1.21, n.s.$ ⁴

Although highly correlated with target latency, the number of fixations to the target allows for a direct measure of the efficiency with which probable target locations are selected. A two-way repeated measures ANOVA on number of fixations made to the target revealed that there was a main effect of location consistency, $F(1, 15) = 26.10, p < .001$, but no effect of scene consistency, $F(1, 15) = 2.54, n.s.$ However, there was a significant interaction, $F(1, 15) = 7.51, p < .001$. Planned comparisons revealed that when participants searched through a consistent scene or an inconsistent scene, a consistent location needed fewer fixations than did an inconsistent location, $t(15) = 5.54, p < .01$, and $t(15) = 2.85, p < .05$, respectively. When participants looked for a target in a consistent location, consistent scenes required fewer fixations than did inconsistent scenes, $t(15) = 3.68, p < .01$, but not for an inconsistent location, $t(15) = 0.67, n.s.$

First fixation duration Mean first fixation durations and first gaze durations are presented in Fig. 4. The first fixation duration on an object is an indicator of the initial object recognition (Henderson, 1992; Rayner & Pollatsek, 1992), and we measured whether scene consistency or location would have an effect, using a two-way repeated measures ANOVA. Analyses showed that there was a main effect of location consistency, $F(1, 15) = 26.10, p < .001$, and a marginal effect of scene consistency, $F(1, 15) = 3.63, p = .076$, but no interaction, $F(1, 15) = 3.14, n.s.$ Planned comparisons revealed that there was no difference between consistent and inconsistent locations when the target was being verified in a consistent scene, $t(15) = 0.036, n.s.$, but there was an advantage for the consistent location in an inconsistent scene, $t(15) = 2.43, p < .05$. When a consistent and an inconsistent scene were compared for the consistent location, there was no difference, $t(15) = 0.37, n.s.$, but fixation durations were shorter for the consistent scene than for the inconsistent scene for the inconsistent location, $t(15) = 2.64, p < .05$. This was a surprising finding that will be discussed further.

First gaze duration The first gaze duration on an object reflects the subsequent processing time on the target within the first glance and is calculated as the sum of all fixations on the target before the eyes move to another region of the scene or the trial ends. A two-way repeated measures ANOVA revealed

⁴ We found that for consistent scenes, targets were still located faster. This may be related to the fact that there is less variability in probable locations with consistent scenes than with inconsistent scenes. When we examined average variability for consistent and inconsistent targets in the location selection in our norming study, we indeed found that this was the case (consistent variance, 1.09; inconsistent variance, 1.42; $t(47) = 2.18, p < .05$).

Table 1 Mean accuracy and response times as a function of scene consistency and location consistency

	Consistent Scene				Inconsistent Scene			
	Consistent Region		Inconsistent Region		Consistent Region		Inconsistent Region	
	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>
Accuracy (%)	82	3	71	3	87	3	78	3
Response time (ms)	5,812	438	8,479	465	6,171	340	7,824	383

that there was a main effect of scene consistency, $F(1, 15) = 7.15$, $p < .05$, but no effect of location, $F(1, 15) = 1.50$, n.s., and no interaction, $F(1, 15) < 1$, n.s. Planned comparisons revealed that there was no difference between consistent and inconsistent locations when the target was being verified in a consistent scene, $t(15) = 0.93$, n.s., but, unlike the first fixation duration, there was no longer an advantage for the consistent location over the inconsistent location in an inconsistent scene, $t(15) = 0.88$, n.s. When comparing the consistent scene and inconsistent scene, there was also no difference for a consistent location, $t(15) = 1.07$, n.s., but gaze

durations were shorter for the consistent scene for an inconsistent location, $t(15) = 2.22$, $p < .05$. So, unlike the first fixation duration measure above, the first gaze duration shows the emergence of the consistent scene effect found in earlier studies (Friedman, 1979; Henderson et al., 1999).

Discussion

In the present study, we investigated how scene gist and learned spatial associations influence search performance.

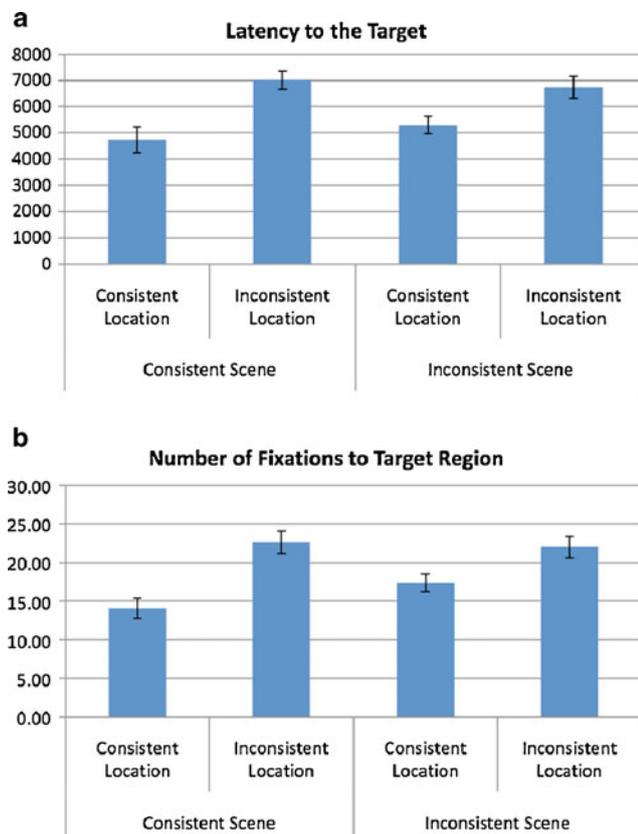


Fig. 3 Mean guidance eye movement measures across the scene consistency and location consistency conditions. **a** Latency to the target object. **b** Number of fixations to the target object. Error bars represent standard errors of the means. See the text for details

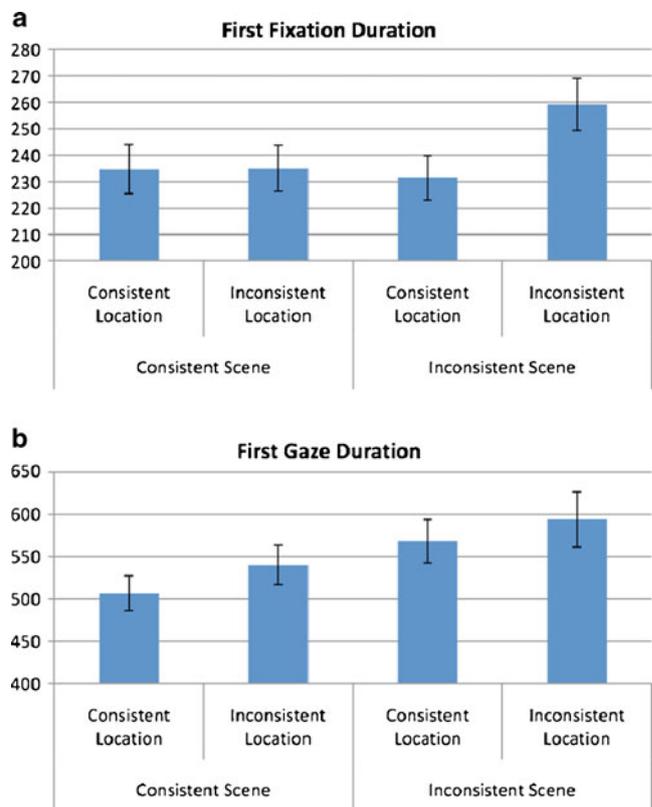


Fig. 4 Mean object identification eye movement measures across the scene consistency and location consistency conditions. **a** First fixation duration. **b** First gaze duration. Error bars represent standard errors of the means. See the text for details

The results provided evidence that learned spatial associations could affect search performance even when there was no direct link between the target and the scene. The results showed that consistent locations led to shorter search times, shorter latencies to the target, and fewer fixations. In fact, eye movement measures reflecting guidance of attention (latency and number of fixations to target) revealed only a small difference between the consistent and inconsistent scenes with consistent locations. Thus, it seems that even when scene gist does not directly apply, the spatial associations that are extracted between an object and the scene can be exploited. This is counter to the general notion that scene gist is the vehicle by which all relevant past experience with a context is retrieved.

It should be noted here that we are not suggesting that semantic information is not useful in the guidance of search but, rather, that spatial associations can play an important role. For instance, the results from the present study are consistent with previous findings that semantic information alone does not improve visual search in scenes (Castelhano & Heaven, 2010; Castelhano & Henderson, 2007). In a recent study, Zelinsky and Schmidt (2009) also showed that specifying the spatial location of a target that has no a priori associations with any scene region improved search by narrowing likely locations of the target. Taken together with the present study, these findings suggest that a scene's spatial layout and its association with various objects may play a considerable role in attentional guidance independently of whether that object is associated with that scene.

Consistent with previous studies (Friedman, 1979; Henderson et al., 1999), we also found that once it was fixated, semantically consistent scene context affected how easily the target was identified (as shown by the first gaze duration measure). Interestingly, spatial relations also affected identification of the target, but only for the early measure of processing (first fixation duration). We found that the first fixation durations were longer in inconsistent scenes (as has been found in the past), but not when the target was in a consistent location. The reason for this early help from a consistent location in an inconsistent scene is unclear, but we hypothesize that a similar scene location may have similar visual properties that may assist with the initial parsing of the object from the background. For instance, if an object typically appears on a counter, perhaps seeing the same object on the same type of surface (e.g., a table) would be easier to parse than if it was on a different type of surface (e.g., a lamp shade).

One question that arises from the present study is how spatial associations between scenes and objects are represented. One possible form is though learned visual associations between a target object and the visual properties of its immediate surroundings (e.g., surface textures; see Oliva & Torralba, 2007). For instance, in a

recent computational model of eye movements during search in real-world scenes, Ehinger, Hidalgo-Sotelo, Torralba, and Oliva (2009) operationalized scene context as the learned association between target placement and the scene's visual features (e.g., scale, layout, and viewpoint). Thus, context effects arose from the learned association between the scene and the target's typical location (across a number of exemplars). Ehinger et al. found that these learned associations could account for search performance. In this way, search is guided by information other than scene gist.

Another question that arises from the present study is whether it is reasonable to assume that associations between scene gist and target objects are stored and accessed independently of associations between spatial locations and target objects. However, neurocognitive studies may provide some insight. Previous studies have suggested that spatial processing and related semantic information may be processed in separate regions. Epstein and colleagues (Epstein, Graham, & Downing, 2003; Epstein & Kanwisher, 1998) showed that an area of the parahippocampal area (parahippocampal place area [PPA]) preferentially responds to pictures of scenes and topographical landmarks. They concluded that the PPA encodes scene layout and general geographical features of a space. In another set of studies, Bar and colleagues (Aminoff, Gronau, & Bar, 2007; Bar, 2004) found that a different region of the parahippocampal cortex (anterior to the PPA) responds to semantic associations between co-occurring objects, as well as scenes and objects. They concluded that this parahippocampal area is devoted to processing relational information, at both a semantic level and a spatial level. Thus, it seems that although both semantic information and spatial relational information form parts of scene knowledge, they may be independent stores of knowledge that have differing effects on behavior (although we should note that this distinction is still under some debate; see Epstein & Ward, 2010)

The present study has larger implications for the architecture of how learned knowledge from scene contexts is organized, retrieved, and applied. A mainstay of the literature is that scene gist is the vehicle by which all relevant past experience is retrieved (Tatler, 2009). Outside consistent scene gist information, information about where objects are typically located is inaccessible, and thus, the assumption is that visual search will always take longer for target objects presented in inconsistent scene contexts. For instance, Biederman et al. (1982) explored different violations by scene context of object identification. They found that certain violations had a greater effect on performance, such as probability (i.e., likelihood of object presence in scenes). However, how probability related to position violations was never explored. In the present study, we explored position and probability (which we refer to as

scene gist and learned spatial associations) and found (in Biederman's terms) that the position information can be useful even when probability is not. This may be related to the idea that search can be based on object features, regardless of the scene in which an object appears. In a recent computational model, Kanan, Tong, Zhang, and Cottrell (2009) demonstrated that search can be guided by object appearance, irrespective of information derived from scene gist. The present study calls into question the assumption that all relevant information from past experiences is organized according to scene gist and suggests that other types of information can be used to guide search outside the influence of scene gist.

Upon further consideration, it seems reasonable that the spatial associations between objects and surfaces be accessible in novel environments, even when knowing the scene gist is not useful. One can easily see how trying to use information that was reliable in the past would lead to appropriate strategies even in situations that are unusual. In the present study, it meant looking in places that, to some extent, matched some knowledge of object locations. Thus, although linked, it may be that scene gist and learned spatial associations are not as intertwined as previously thought.

Author Note The authors would like to thank Effie Pereira for help with data collection and Carrick Williams, Kristin Weingartner, Sian Beilock, and Kevin Munhall for helpful discussions and comments on earlier versions of the manuscript. This work was supported by grants from the Natural Sciences and Engineering Research Council of Canada, Canada Foundation for Innovation, and the Advisory Research Committee of Queen's University to MSC.

References

- Aminoff, E., Gronau, N., & Bar, M. (2007). The parahippocampal cortex mediates spatial and nonspatial associations. *Cerebral Cortex*, *27*, 1493–1503.
- Bar, M. (2004). Visual objects in context. *Nature Reviews. Neuroscience*, *5*, 617–629.
- Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: Detecting and judging objects undergoing relational violations. *Cognitive Psychology*, *14*, 143–177.
- Castelhano, M. S., & Heaven, C. (2010). The relative contribution of scene context and target features to visual search in scenes. *Attention, Perception, & Psychophysics*, *72*, 1283–1297.
- Castelhano, M. S., & Henderson, J. M. (2007). Initial scene representations facilitate eye movement guidance in visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *33*, 753–763.
- Castelhano, M. S., & Henderson, J. M. (2008). The influence of color on perception of scene gist. *Journal of Experimental Psychology: Human Perception and Performance*, *34*, 660–675.
- Castelhano, M. S., Pollatsek, A., & Cave, K. (2008). Typicality aids search for an unspecified target, but only in identification and not in attentional guidance. *Psychonomic Bulletin & Review*, *15*, 795–801.
- Chamizo, V. D. (2003). Acquisition of knowledge about spatial location: Assessing the generality of the mechanism of learning. *The Quarterly Journal of Experimental Psychology*, *56B*, 102–113.
- Chun, M. M., & Jiang, Y. (1999). Top-down attentional guidance based on implicit learning of visual covariation. *Psychological Science*, *10*, 360–365.
- Ehinger, K., Hidalgo-Sotelo, B., Torralba, A., & Oliva, A. (2009). Modelling search for people in 900 scenes: A combined source model of eye guidance. *Visual Cognition*, *17*, 945–978.
- Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature*, *392*, 598–601.
- Epstein, R., & Ward, E. J. (2010). How reliable are visual context effects in the parahippocampal place area? *Cerebral Cortex*, *20*, 294–303.
- Epstein, R., Graham, K. S., & Downing, P. E. (2003). Viewpoint-specific scene representations in human parahippocampal cortex. *Neuron*, *37*, 865–876.
- Friedman, A. (1979). Framing pictures: The role of knowledge in automatized encoding and memory for gist. *Journal of Experimental Psychology: General*, *108*, 316–355.
- Gillner, S., & Mallot, H. A. (1998). Navigation and acquisition of spatial knowledge in a virtual maze. *Journal of Cognitive Neuroscience*, *10*, 445–463.
- Henderson, J. M. (1992). Object identification in context: The visual processing of natural scenes. *Canadian Journal of Psychology*, *46*, 319–341.
- Henderson, J. M., & Hollingworth, A. (1999). High-level scene perception. *Annual Review of Psychology*, *50*, 243–271.
- Henderson, J. M., Weeks, P. A., & Hollingworth, A. (1999). The effects of semantic consistency on eye movements during complex scene viewing. *Journal of Experimental Psychology: Human Perception and Performance*, *25*, 210–228.
- Jiang, Y., & Wagner, L. C. (2004). What is learned in spatial contextual cueing—Configuration or individual locations? *Perception & Psychophysics*, *66*, 454–463.
- Kanan, C., Tong, M. H., Zhang, L., & Cottrell, G. W. (2009). SUN: Top-down saliency using natural statistics. *Visual Cognition*, *17*, 979–1003.
- Kunar, M. A., Flusberg, S., Horowitz, T. S., & Wolfe, J. M. (2007). Does contextual cuing guide the deployment of attention? *Journal of Experimental Psychology: Human Perception and Performance*, *33*, 816–828.
- Malcolm, G. L., & Henderson, J. M. (2009). The effects of target template specificity on visual search in real-world scenes: Evidence from eye movements. *Journal of Vision*, *9*(11, Art. 8), 1–13.
- Neider, M. B., & Zelinsky, G. J. (2006). Scene context guides eye movements during search. *Vision Research*, *46*, 614–621.
- Oliva, A. (2005). Gist of the scene. In L. Itti, G. Rees, & J. K. Tsotsos (Eds.), *Encyclopedia of neurobiology of attention* (pp. 251–256). San Diego: Elsevier.
- Oliva, A., & Torralba, A. (2007). The role of context in object recognition. *Trends in Cognitive Sciences*, *11*, 520–527.
- Rayner, K., & Pollatsek, A. (1992). Eye movements and scene perception. *Canadian Journal of Psychology*, *46*, 342–376.
- Sturz, B. R., Kelly, D. M., & Brown, M. F. (2010). Facilitation of learning spatial relations among locations by visual cues: Generality across spatial configurations. *Animal Cognition*, *13*, 341–349.
- Tatler, B. W. (2009). Current understanding of eye guidance. *Visual Cognition*, *17*, 777–789.
- Zelinsky, G. J., & Schmidt, J. (2009). An effect of referential scene constraint on search implies scene segmentation. *Visual Cognition*, *17*, 1004–1028.