**Journal of Experimental Psychology: Learning, Memory, and Cognition**

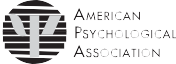
**Examining the Hierarchical Nature of Scene Representations in Memory**

Monica S. Castelhano, Suzette Fernandes, and Jordan Theriault

Online First Publication, October 25, 2018. <http://dx.doi.org/10.1037/xlm0000660>

CITATION

Castelhano, M. S., Fernandes, S., & Theriault, J. (2018, October 25). Examining the Hierarchical Nature of Scene Representations in Memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. Advance online publication. <http://dx.doi.org/10.1037/xlm0000660>



Journal of Experimental Psychology: Learning, Memory, and Cognition

© 2018 American Psychological Association

0278-7393/18/$12.00 <http://dx.doi.org/10.1037/xlm0000660>

Examining the Hierarchical Nature of Scene Representations in Memory

# Monica S. Castelhano and Suzette Fernandes

Queen’s University

# Jordan Theriault

Boston College

How are scene representations stored in memory? Researchers have often posited that scene represen- tations have a hierarchical structure with background elements providing a scaffold for more detailed foreground elements. To further investigate scene representation and the role of background and foreground information, we introduced a new stimulus set: chimera scenes, which have the central block of objects belonging to one scene category (foreground), and the surrounding structure belonging to another (background). We used a contextual cueing paradigm and emphasized the relative importance of each by having the target placed on either the background or foreground. In a transfer block, we found that though changes to the background were highly detrimental to search performance for background targets, search performance was only slightly affected by changes to either the foreground or the background for foreground targets. These results indicate that rather than a fixed hierarchy, the structure of scene representations are more aptly captured by a parallel model that stores information flexibly.

*Keywords:* scene processing, memory representation, visual search, contextual cueing

In any given environment, there are various forms of high-level contextual information, such as the fact that silence is a funda- mental feature of libraries, as well as individual components of that environment, such as the books in a library. Scene representations and the various levels of information have been conceptualized in multiple ways across studies. We can think of these as a collection of visual and semantic details that together make up the larger contextual knowledge. Alternatively, we can think of the high- level context information as the scaffolding that provides an or- ganizing principle for integrating the contents and visual details of a scene. In this study, we investigated how visual scene informa- tion is stored in memory, contrasting the relative importance of the contextual background information with importance of the objects the scene contains.

Several studies have shown that our capacity to hold scene information in memory is substantial. Classic memory studies have demonstrated a high capacity for storing visual information about hundreds and even thousands of visual images [(Nickerson, 1965,](#_bookmark16) [1968;](#_bookmark18) [Shepard, 1967;](#_bookmark39) [Standing, 1973;](#_bookmark43) [Standing, Conezio, &](#_bookmark45)

Monica S. Castelhano and Suzette Fernandes, Department of Psychol- ogy, Queen’s University; Jordan Theriault, Department of Psychology, Boston College.

We thank Ellen O’Donoghue, Karolina Krzys´, Mollie Roy, Elysée Kukwabantu, and Shane Kennedy for their help with creating the stimuli, pilot testing the experiment, and collecting data. This work was supported by funding from the Natural Sciences and Engineering Research Council of Canada, Canadian Foundation for Innovation, and Ontario Ministry of Research and Innovation to Monica S. Castelhano.

Correspondence concerning this article should be addressed to Monica

S. Castelhano, Department of Psychology, Queen’s University, 62 Arch Street, Kingston, ON K7L 3N6, Canada. E-mail: [monica.castelhano@](mailto:monica.castelhano@queensu.ca) [queensu.ca](mailto:monica.castelhano@queensu.ca)

[Haber, 1970).](#_bookmark45) These studies reveal that even with complex visual information, we are able to distinguish previously viewed scenes from novel ones at high performance levels (ranging from 92% to 98%) in a simple discrimination task. Further, [Standing et al.](#_bookmark45) [(1970)](#_bookmark45) showed that when participants were asked to distinguish the mirror-reversed images, performance remained relatively high (88%). These researchers concluded that a considerable amount of detailed visual information is being stored in memory.

More recently, [Konkle, Brady, Alvarez, and Oliva (2010)](#_bookmark44) in- vestigated the degree to which visual details of scenes are stored in visual long-term memory. They investigated whether object dif- ferences across images within scene categories would cause inter- ference in memory performance. In their study, participants stud- ied over 2,000 photographs from 128 different scene categories. For each scene category, the number of images presented during learning varied (from 1 to 64). The study replicated seminal studies on memory capacity for visual images with a high level of memory performance overall (96%). However, accuracy decreased as more exemplars were present in a given category (from 84% to 76% for four to 64 categories, respectively). The decrease was relatively minor considering that for each doubling in the number of exem- plars, there was an approximate 2% decrease in performance. The researchers posited that in order for specific images within a category to be distinguished, a significant amount of visual detail must be stored. The authors concluded that visual details such as the appearance of individual objects play a larger role in memory retrieval for scenes than its categorical name.

There has also been a great deal of research examining how scene details are encoded using the contextual cueing paradigm. The contextual cueing paradigm was first introduced by [Chun and](#_bookmark19) [Jiang (1998),](#_bookmark19) who found that participants were faster finding letter targets in repeating arrays than in novel arrays. They referred to this improved performance as the *contextual cueing effect*. [Brock-](#_bookmark12) [mole and Henderson (2006b)](#_bookmark12) extended the contextual cueing effect

This document is copyrighted by the American Psychological Association or one of its allied publishers.

This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

1

to real-world scenes and found similar results with faster perfor- mance when searching in a repeated scene. Furthermore, they found that compared with arrays, the effect was larger (�1 to 2 s) and learning occurred faster (within the first two to three blocks) with scenes. They concluded that the scene’s semantic information played a key role in target location retrieval. [Brockmole, Castel-](#_bookmark9) [hano, and Henderson (2006)](#_bookmark9) examined whether the information being used to retrieve target location was based on the larger contextual (global) information or specific object (local) details. They found that search performance was faster when the global, contextual information of a scene was repeated, compared with the local, object information. They concluded that targets in scenes were more strongly associated with a scene’s global context com- pared with objects immediately surrounding the target.

This document is copyrighted by the American Psychological Association or one of its allied publishers.

This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

[Brooks, Rasmussen, and Hollingworth (2010)](#_bookmark13) further investi- gated how global scene information had an effect on the encoding of local scene information. Brooks et al. proposed that targets are associated to a subregion of the scene. They embedded a “game board” (an array of dumbbell shapes) within a scene and asked participants to search for a target object within it. The game board served as a functional subregion within the global context of the scene in which the target appeared. In one experiment, participants saw repeated (and novel) pairings of targets, game boards, and scenes, and across learning trials showed a contextual cueing effect. Then, in a transfer test, changing either the game board or the scene eliminated the contextual cueing effect. However, in another experiment, they examined whether participants developed a contextual cueing effect when the target location and scene were repeated, but the board varied. In this case, changing the game board (but not the scene) eliminated the contextual cueing effect. In sum, they found that a functional subregion can be associated with global scene context during learning, which indicates that global context is necessary to retrieve information about the sub- region and target. However, if global scene context was not paired with the subregion and target during learning, then the scene-target association was not learned independently.

On the basis of their findings, [Brooks et al. (2010)](#_bookmark13) suggested a hierarchical model of scene representations. In this model, a global context can be broken down into nested subregions. The model is based on a series of navigation studies that demonstrated that memory for larger environments is organized hierarchically [(Hirtle](#_bookmark40)

[& Jonides, 1985;](#_bookmark40) [McNamara, 1986;](#_bookmark48) [McNamara, Hardy, & Hirtle,](#_bookmark50) [1989).](#_bookmark50) These navigation studies examined memory for spatial relations in subregions of a larger environment. The subregions were distinct clusters defined by their content objects and were also nested within the larger environment. They found that navi- gation and recall of the details of the subregions depended on the observer’s ability to retrieve information about the larger environ- ment. Concerning scenes, the application of such a model would mean that the retrieval of information in any one part of the scene (e.g., contents on a desk) would depend on the recognition of the background or larger contextual information (e.g., a specific of- fice).

An alternative explanation of scene representations can be de- rived from what we call *parallel memory models*. In these models, information encoding varies in strength (i.e., some information is weighted differently from others), but there is no hierarchical scene structure. These variations in strength ultimately lead to differences in how easily some information is retrieved relative to

others. By varying the relative strength of the global or local information in a scene, a parallel memory model could result in global information being easier to retrieved and used in a search task. One of the more prominent memory models that falls in this category would be MINERVA2 [(Hintzman, 1984,](#_bookmark34) [1988;](#_bookmark35) [Hintzman](#_bookmark37)

[& Curran, 1994).](#_bookmark37) It was created to account for seemingly categor- ical differences in memory studies within a single system. In this model, individual episodic memories are encoded as traces made up of feature vectors that contain detailed information. Critically, retrieval relies on both an adequate memory retrieval cue and the fidelity of the originally encoded information. Depending on the probe, information retrieved can be from a specific instance in time (a single episodic memory trace), or it can reflect a more general construct which is formed from a sum of information drawn from similar experiences (a composite memory). With regard to scenes, studies have demonstrated that gross visual features and spatial layout are strong cues for retrieving general information about scenes [(Greene & Oliva, 2009a;](#_bookmark26) [Oliva & Torralba, 2001);](#_bookmark23) how- ever, retrieving specific details from a scene has proven to be more difficult [(Grimes, 1996;](#_bookmark29) [Rensink, O’Regan, & Clark, 1997;](#_bookmark36) [Si-](#_bookmark41) [mons & Levin, 1997).](#_bookmark41)

The contextual cueing effect, as described in previous studies, is predicted by both hierarchical and parallel memory models, yet critically, these models differ in how information is accessed. In the hierarchical model, global information must be retrieved first, as local information is embedded within it. In the parallel memory model, global and local information are accessed independently; global information will likely be retrieved more quickly than the local information, based on the retrieval cue and relative global and local information weightings. However, retrieving global in- formation is not required to access local information. Thus for a parallel memory model, all else being equal, information about a specific scene subregion should not rely on the recall of the larger context. Here, we examine which of these two models best cap- tures how information from scenes is structured.

In the present study, we investigated how the foreground and background of a scene are stored in memory and contrast the two theoretical hypotheses of visual memory described in the preced- ing text: (1) the hierarchical memory model and (2) the parallel memory model. In the current study, we defined the *background* as the large boundary defining elements such as walls, floors, ceil- ings, and main structures that define the space [(Oliva, 2005;](#_bookmark21) [Oliva](#_bookmark23)

[& Torralba, 2001).](#_bookmark23) We defined the *foreground* as the content objects large and small that are moveable to some extent within the scene, is closer in depth to the viewer’s viewpoint, and is encap- sulated by the space defined by the background [(Henderson &](#_bookmark31) [Hollingworth, 1999;](#_bookmark31) [Oliva, 2005).](#_bookmark21) With this qualitative distinction, we can examine how they are stored in memory individually and in relation to each other.

Across two experiments, we used a contextual cueing paradigm to examine whether disrupting either global or local scene infor- mation would detrimentally affect the retrieval of learned target locations. Participants were presented with scenes that had fore- ground and backgrounds that were either semantically related or unrelated (i.e., chimera scenes, see [Figure](#_bookmark0) 1). Targets appeared in either the foreground or the background. This manipulation al- lowed us to investigate whether information in the foreground is encoded separately from information in the background.

*Figure 1.* Two example stimuli used in both experiments. The top panel shows the normal version of the scenes and the bottom panel shows the chimera version of the scenes. For all scenes, the images were roughly divided into two thirds background and one third foreground. Two images from separate categories were matched and the foreground objects were swapped to create the chimera version for each background. See the online article for the color version of this figure.

In each experiment, there were three phases: (1) learning phase, (2) transfer phase, and (3) memory phase. In the learning phase, participants searched scenes for a target letter (*T*), which could appear in either the foreground or the background. Par- ticipants completed trial blocks that contained both novel and repeated scenes. In repeated scenes, the target object always appeared in the same location (only the orientation of the target was randomly changed). The Transfer phase took place during the last block of trials and the repeated scenes were changed according to the transfer conditions. In the literature, the back- ground and foreground elements are often referred to as global and local scene information, respectively [(Greene & Oliva,](#_bookmark26) [2009a;](#_bookmark26) [Oliva & Torralba, 2006).](#_bookmark25) Here, we use the *global* and *local* labels to refer to the type of changes that were made to the scenes during the transfer phase so as not to confuse them with the manipulation of the target placement. There were three transfer conditions: (1) the foreground of the scene was re- placed with a new foreground, whereas the background re- mained unchanged (*local change*); (2) the background of the scene was replaced with a new background, whereas the fore- ground remained unchanged (*global change*); or (3) no changes made to the repeated trial (*no change*). We examined response times in each phase and across conditions. The experiment requires two manipulations in relation to the background and foreground of the scene: in the target placement and in the transfer conditions. For clarity, we henceforth use global and local to refer transfer condition changes (global refers to change

This document is copyrighted by the American Psychological Association or one of its allied publishers.

This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

to the background and local refers to change to the foreground). In the memory phase, participants were tested on their memory for the scene images. Participants indicated whether repeated, previously viewed novel, and foil scenes were old or new. Additionally, participants were asked to indicate the specific location of the target object when they indicated the scene was old.

If the hierarchical memory model is correct, then retrieving local scene information requires that global information be accessible. In this case, when global information is changed in the transfer phase (global change), search times should be slowed to rates observed for novel scenes. However, if the parallel process model hypothesis is correct, and information retrieval is not nested—then local scene information is retrievable regardless of whether global information is altered. In this case, when either local or global information is replaced in the transfer phase, search times should be slower than for repeated scenes (the retrieval cue is an imperfect match for the stored representation) but still effective and faster than novel scenes.

## Experiment 1

**Method**

**Participants.** Forty Queen’s University undergraduates par- ticipated in the experiment and were compensated either $10/hr or course credit for their participation. All participants had normal or

corrected-to-normal vision. The study was reviewed and cleared by the Queen’s University General Research Ethics Board (GPSYC-398 – 07).

This document is copyrighted by the American Psychological Association or one of its allied publishers.

This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

**Materials and apparatus.** Stimuli were computer-generated scenes created using Data Becker Home Design software, Version

5.0 (Data Becker International, Düsseldorf, Germany). There were 60 scenes in total ranging across a number of rooms and spaces within and around a house (e.g., kitchen, living room, garage, backyard). Each scene was divided into a background and fore- ground components.1 All scenes were constructed using the same template such that they were built into the corner with two adjoin- ing walls to provide the main definition of the space and depth of the scene.2 When measured from a bird’s eye rendering of the images, the depth of the foreground information (from the camera position to the back wall occupied on average 52% (13% standard deviation) of the distance. When rendered as two-dimensional images, the pixel content for the foreground vs. background was roughly equated (51% foreground, 49% background).

To address the question whether the foreground region aligned with how people viewed images, we had a separate group of participants (*N* = 42) judge the foreground region for each image (normal version of each scene). To assess, we created five versions of the foreground region for each image (original size, 25% and 50% larger, 25% and 50% smaller, see [Figure](#_bookmark1) 2).

For each trial, we presented three regions at a time such that there were three possible response configurations in total. Across configu- rations, the original was presented as either the smallest region of the three (Configuration 1), the medium size region (Configuration 2), or the largest region (Configuration 3). Participants were instructed to choose the area that best corresponded to their notion of foreground. Participants responded with the number of the region, which corre- sponded the largest (1), medium (2), and smallest (3) region (as labeled in the rightmost column of [Figure](#_bookmark1) 2).

We found that across participants, the original region was se- lected at a higher rate than other regions, regardless of response configuration (see [Figure](#_bookmark2) 3). The pattern was reflected in the statistical analysis, for which we collapsed across response con- figuration and compared each of the regions to the original one. We found that, across the board, the original was chosen a signif- icantly greater proportion of times (>50%: *t*[41] = 3.87, *p* <

.001; >25%: *t*[41] = 9.03, *p* < .001; <25%: *t*[41] = 10.11, *p* <

.001; <50%: *t*[41] = 5.52, *p* < .001).

From these scenes, we generated chimera scenes in which the foreground component from one scene was switched with another (see [Figure](#_bookmark0) 1). To do this, we paired up images from different semantic categories and produced two images per scene. For instance, with a living room and kitchen, switched the foregrounds of each to produce the chimera version of each, in which the foreground of the living room was placed in the kitchen and the foreground of the kitchen was placed in the living room.

Each scene contained a single gray letter (*T*) presented in nine-point Arial font that was rotated either to the right or to the left. The target had one of two possible locations: background or foreground that corresponded with the scene regions defined above. The use of a letter ensured that the placement of the target was not predicted by previous knowledge of object placement and scene context [(Brockmole et al.,](#_bookmark9) [2006).](#_bookmark9) Because of the limited number of chimera scenes, we also used a number of filler scenes as novel scenes during the learning blocks

(randomly chosen from 250 images created in with the same soft- ware).

The experiment was run on a Standard Dell computer using MATLAB R2014a and PsychToolBox3 [(Brainard, 1997;](#_bookmark10) [Pelli, 1997).](#_bookmark32) The monitor was 21" diagonally, and images subtended 38.1° X 26.6° at a resolution of 800 X 600 pixels with a refresh rate of 100 Hz. Head position was not restricted and participants sat approximately 60 cm from the monitor.

**Procedure.** Participants were instructed to search for a small gray capital *T* hidden in the scene and to indicate the direction of the *T* (top rotated to the right or to the left) by pressing the corresponding button marked with left and right pointing arrows on the keyboard. The response button and search time were re- corded, and the scene was displayed until participants responded or until 20 s had elapsed. Participants were asked to keep their fingers as close as possible to the response keys to minimize the time needed to respond once they had found the target. No mention of repeating stimuli was made. The experiment consisted of a learn- ing phase, transfer phase and memory phase, and are explained in the following text in more detail.

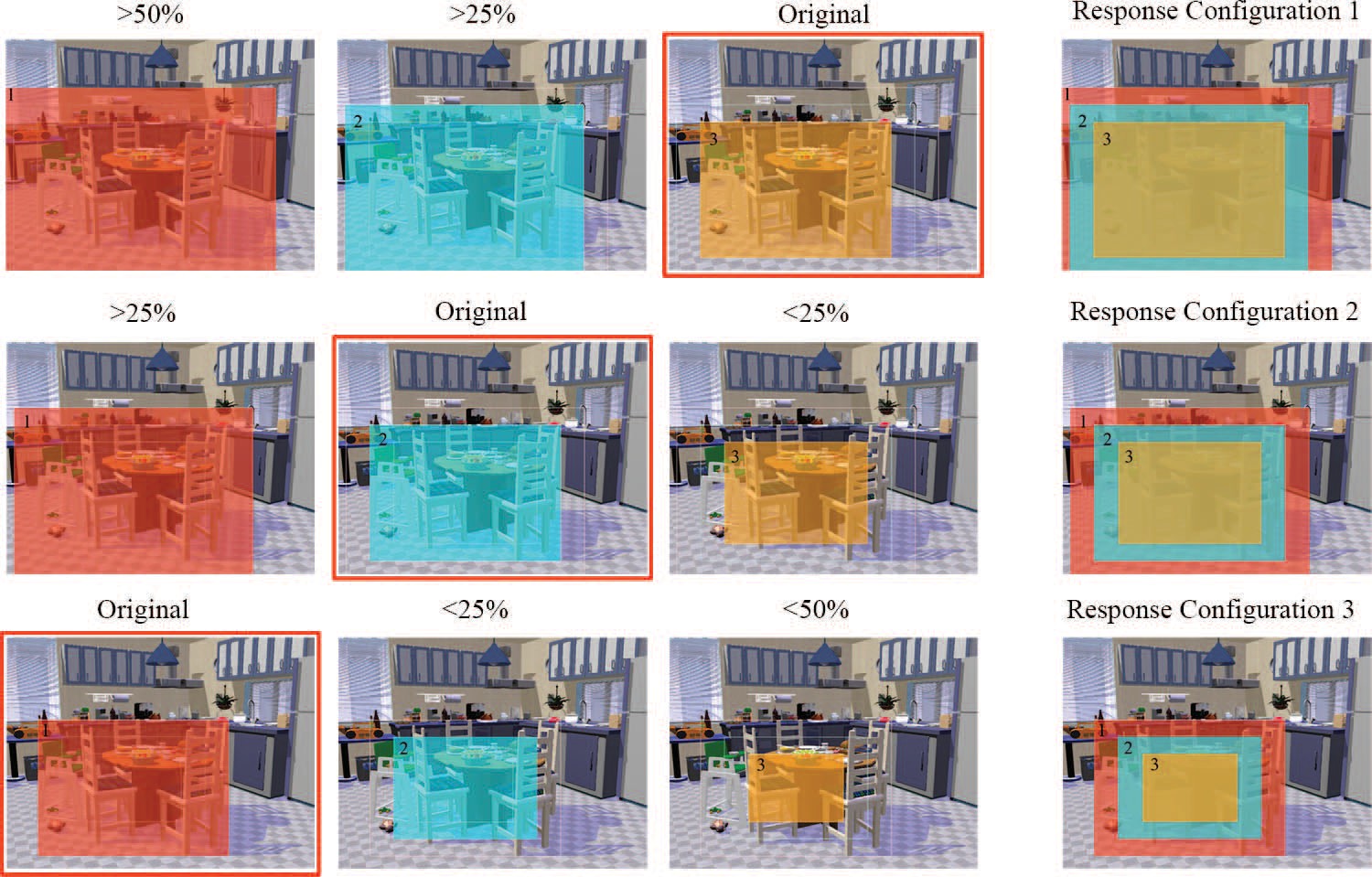
***Learning phase.*** All trials began with the presentation of a black fixation cue centered on a gray screen. After one second, the fixation cue was replaced with a scene. The experimental program determined target locations for each participant. It was randomly located within the background or foreground of the image, according to the target place condition. The target position was randomly set for novel scenes, and maintain at the same *x–y* position for the repeated scenes across all blocks. In repeated arrays, only the direction of the target letter (i.e., right or left) changed across trial blocks.

Participants completed nine learning phase blocks of 24 trials each. Each block contained 12 novel and 12 repeated scenes (learning condition) presented in a random order. For the repeated scenes, scene condition (normal vs. chimera) was equally distributed (12 pairs of scenes were selected and counterbalanced across participants). For novel scenes, there was an uneven split between the scene conditions as we used filler scenes. In each block, there were fewer chimera than normal novel scenes (three and nine, respectively, or 30 and 210 in total, respectively). This was done so that none of the chimera scenes used as repeated scenes overlapped in content (background or fore- ground) with the novel scenes. Participants were given a break at the conclusion of the second, fourth, sixth, and eighth blocks. Otherwise, the next block began immediately without any indication to the participant.

***Transfer phase.*** The transfer phase was run as an additional 10th block of trials after the learning phase was completed. Partici- pants were not given any indication that this block was different from

1 A separate group of participants (*N* = 20) rated the interpretability of each region (background only or foreground only) using a category veri- fication task. Although backgrounds were categorized significantly more accurately, *t*(18) = 4.26, *p* < .01, than foregrounds, both were categorized at a high level of accuracy (background: 95%; foreground: 89%).

2 All scenes were created using the same template whereby they were built into the corner with two adjoining walls, which provided the main definition of the space and depth of the scene. Thus, the indoor and outdoor scenes had a similar shape and depth.



*Figure 2.* Examples of the three response configurations used in the foreground judgment experiment. For each, three regions are presented: large (red/dark grey), medium (blue/medium grey), and small (orange/light grey). Please note that we have increased the intensity of the colors and the reduced the transparency for illustrative purposes. In each configuration, the original foreground region is presented (highlighted in red/grey box) as the smallest, medium, and largest regions relative to the other two regions presented. Participants saw one of these configurations for each scene, but saw all configurations across scenes viewed (fully counterbal- anced across participants and scenes). See the online article for the color version of this figure.

the previous ones. There were 12 novel trials and 12 repeated trials.3 Repeated trials were in one of three transfer conditions: (1) the foreground of the scene was replaced with a new foreground, but the background remained unchanged (local change), (2) the background of the scene was replaced with a new background, but the foreground remained unchanged (global change), or (3) no changes made to the repeated trial (no change). Presentation order was randomized and scene condition was balanced across transfer conditions and across participants. Across all transfer conditions, the *x*–*y* coordinates of the target remained unchanged from the learning trials.

This document is copyrighted by the American Psychological Association or one of its allied publishers.

This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

***Memory phase.*** The memory phase was administered on the same participants immediately following the transfer phase after a brief break (�2 min). Participants completed 36 test trials: previously repeated scenes from Block 10 (12), previously viewed novel scenes (12), or were new foil images that had not been previously viewed (12). All test images were equally distributed between scene condi- tions (normal vs. chimera). The repeated images tested were taken from each transfer condition equally (four per condition). Presentation order was randomized.

For each scene image, participants indicated whether they had previously seen any part of the image by pressing a button on the keyboard corresponding to “yes” or “no.” The target was not present in the image. If they responded “yes,” a cursor appeared on the screen and participants indicated the target location. Once they clicked on a location, the trial ended and the next trial began automatically. If they responded “no,” the next trial began immediately. Accuracy for the target location was calculated as the Euclidean distance (in pixels, as viewing distance was not controlled) from the clicked location to the actual location of the target.

## Results and Discussion

**Analysis.** Measures were examined using repeated-measures analysis of variance (ANOVA). When applicable, we controlled for violations of the assumption of sphericity by using the Greenhouse-Geisser correction to degrees of freedom. For all planned comparisons and post hoc contrasts, alpha was adjusted such that the family wise error did not exceed .05. The specific alpha used is reported below prior to the report of the tests.

**Learning phase.** We examined whether there was a contex- tual cueing effect in search times across the learning phase with a 9 (Block) X 2 (novel and repeated learning conditions) within- subjects ANOVA.4 Means for each block by learning condition are presented in [Figure 4,](#_bookmark3) Panel A. The effect of block was significant, *F*(4.26, 166.28) = 27.15, *p* < .001, ll2 = .41, when Greenhouse- Geisser corrected (= .53), as was the effect of learning condition,

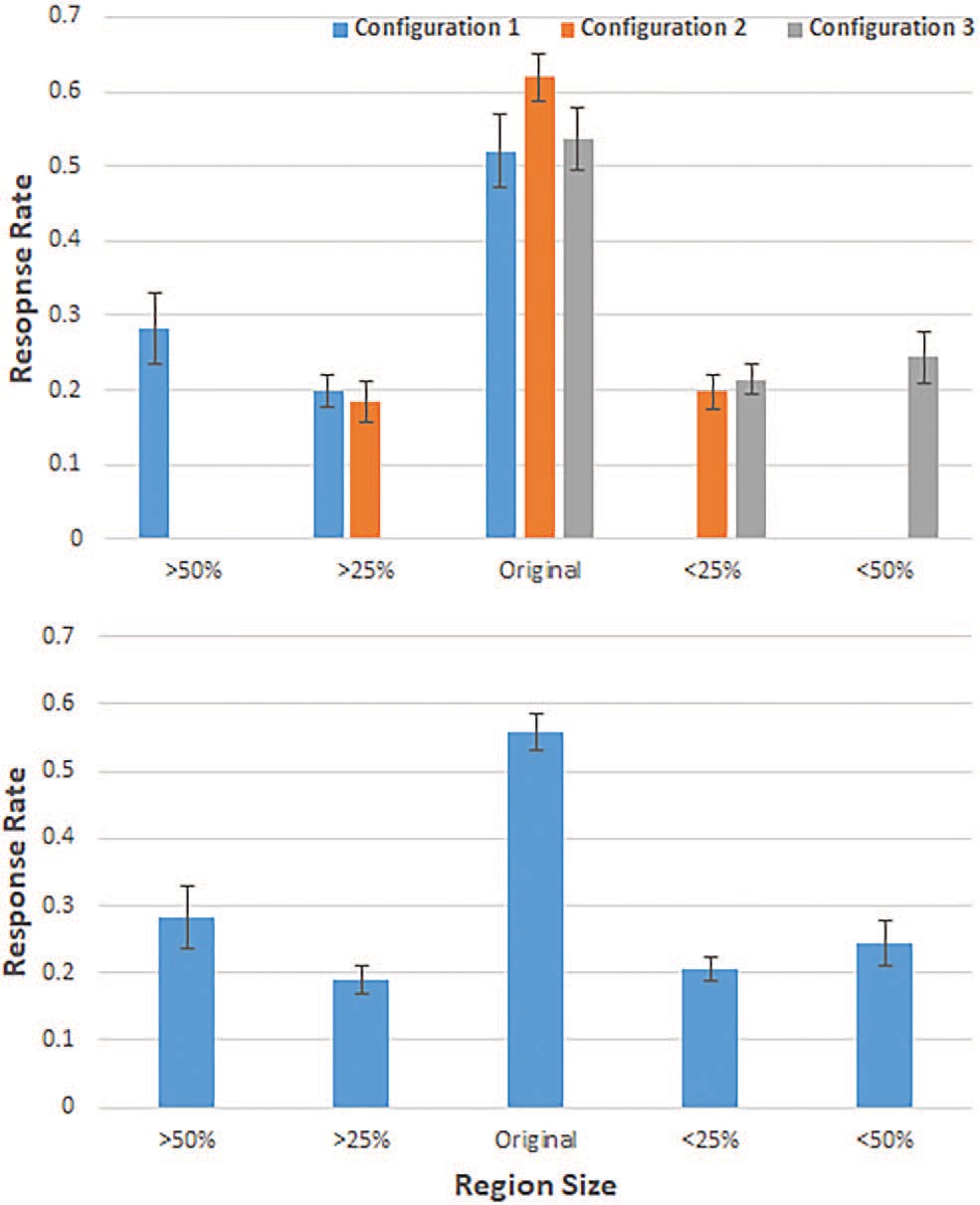
p

3 Because of the limited number of chimera scenes, across transfer and target placement conditions, we had two items/conditions. This rather low number of data points per condition was offset by increasing the number of participants per condition (40 participants in the present study compared with eight with four items/condition in [Brockmole & Henderson, 2006a,](#_bookmark14) [2006b),](#_bookmark12) using a completely within-subjects design, and adopting conser- vative analyses.

4 The semantic interpretation of the Chimera scenes was outside the scope of the current study. Rather, we were interested in how the change in the contents (background or foreground) would affect retrieval of target location. Chimera scenes were used as a tool for changing the respective region in each scene. Thus, in all analyses reported image type was collapsed to study the effects of changing foreground and background information.

This document is copyrighted by the American Psychological Association or one of its allied publishers.

This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.



*Figure 3.* Response rates for each region size, in each response config- uration (top) and collapsed across response configuration (bottom). Error bars reflect standard error of the mean. See the online article for the color version of this figure.

*F*(1, 39) = 143.80, *p* < .001, ll2 = .79, and their interaction, *F*(8,

p

312) = 22.91, *p* < .001, ll2 = .37. Planned paired sample t tests confirmed the contextual cueing effect (a = .00568): there were no significant differences in search times between novel and repeated arrays in Blocks 1 and 3; but in Blocks 2 and 4 through 9, repeated arrays response times were significantly faster than were those for novel arrays (all *p*s < .001).

p

To examine whether there was an effect of target location on learning, we divided the repeated condition above by target loca- tion. We then conducted a 9 (Block) X 2 (Repeated Foreground and Repeated Background) within-subjects ANOVA. The learning patterns for foreground and background target locations were not significantly different. Although there was a main effect of block, *F*(3.27, 127.34) = 61.80, *p* < .001, ll2 = .61, when Greenhouse- Geisser corrected (= .41), there was no main effect of target location, *F*(1, 39) = 2.84, *p* = .1, ll2 = .07, and no significant interaction between the target location conditions across blocks, *F*(4.59, 178.85) = 1.12, *p* = .35, ll2 = .03, when Greenhouse- Geisser corrected (= .57; see [Figure 4,](#_bookmark3) Panel B). Thus, it seems learning was equivalent across target locations. After establishing that search times improved in repeated arrays relative to novel arrays and did not differ by target location, we analyzed the contributions of background and foreground information in the transfer phase.

*p*

*p*

*p*

**Transfer phase.** We conducted a 2 X 4 repeated-measures ANOVA that examined target placement (foreground vs. back-

ground) by transfer condition (no change, global change, local change, and novel) on search time. Results revealed a significant difference in search time across transfer conditions, *F*(3, 126) = 7.6, *p* < .001, ll2 = .17, across target placement, *F*(3, 126) = 63.24, *p* < .001, ll2 = .63, and a marginally significant interaction, *F*(3, 126) = 2.3, *p* = .08, ll2 = .06. Means for each transfer condition by target placement condition are presented in [Figure 5.](#_bookmark4)

To examine whether retrieval of the local information of a scene is dependent on having access to the global information, we conducted planned comparisons between the two key transfer conditions (local change and global change) and the two control transfer conditions (no change and novel) for each of the target locations (eight comparisons in total; a = .00625).

p

p

p

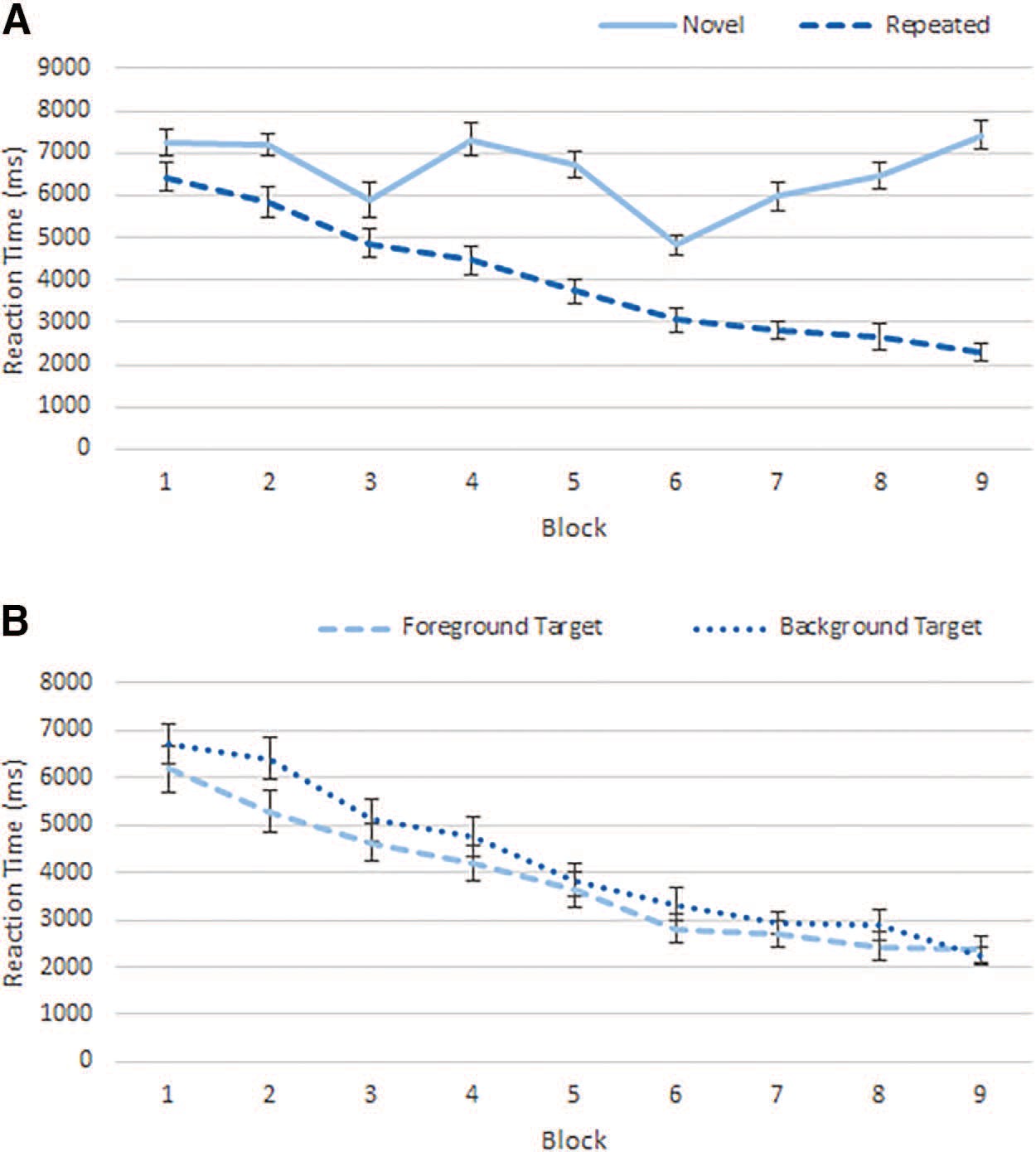
The hierarchical model predicts that retrieving target location is dependent on retrieval of its associated global context. If correct, then the contextual cueing effect should remain for the local change condition (where the global information would remain constant). Consistent with this, first we found local change scenes showed significantly faster search times than novel scenes for both the foreground and background targets, *t*(39) = -8.37, *p* < .001, *d* = 2.68, and *t*(39) = -6.75, *p* < .001, *d* = 2.16, respectively. Second, there was no significant difference between the local change and no change conditions for the foreground and back- ground targets, *t*(39) = 1.36, *p* = .18, *d* = 0.43, and *t*(39) = -2.2, *p* = .03, *d* = 0.7, respectively.

The hierarchical model also predicts that local context can only guide attention to the target when paired with its associated global [context (Brooks et al., 2010).](#_bookmark13) If correct, then the contextual cueing effect should be disrupted in the Global Change condition (where the global information would change). Because the memory re- trieval of either background or foreground target locations is dependent on the global context, any changes to the global infor- mation will result in the scene being functionally a novel scene. Here we found mixed results. For the background targets, the pattern of results was consistent with this hypothesis: Global change scenes had significantly slower search times than no change scenes, *t*(39) = -3.11, *p* = .004, *d* = 1.0, and did not differ significantly than novel scenes, *t*(39) = -1.3, *p* = .2, *d* =

0.41. However in contrast to the hierarchical prediction, we found for foreground objects, global change scenes showed significantly faster search times than novel scenes, *t*(39) = -3.4, *p* = .002, *d* = 1.09, and did not significantly differ from the no change scenes, *t*(39) = -2.43, *p* = .02, *d* = 0.78. Thus, for the background target, which emphasize the global context of the image, global changes eliminated contextual cueing. However, for foreground targets, which emphasize a functional subcomponent of the scene, global changes had a small effect on the contextual cueing effect.

**Memory phase.** For the memory test, we looked at memory for the scene identity (do you remember seeing this scene either as a whole or in part) and memory for target position (indicate with the cursor the approximate position of the target in this picture). We look at performance in each part of the test in turn. Means are presented in [Figure 6.](#_bookmark5)

For the scene identification test, participants were tested on their memory for scenes as they appeared in the transfer condition; thus, they were instructed to reply “yes” if all or part of the scene seemed familiar to them (proportion of “yes” responses to foils was .075, indicating that all participants could perform this task well). We conducted a 2 X 4 ANOVA examining target placement



*Figure 4.* Mean reaction times across (A) Block X Learning Condition and (B) Block X Target Location Condition for repeated trials only. Error bars represent standard error of the mean. See the online article for the color version of this figure.

(background and foreground) and transfer condition (no change, global change, local change, and novel) on proportion “yes” re- sponses. We found that the transfer condition had an effect, *F*(3,

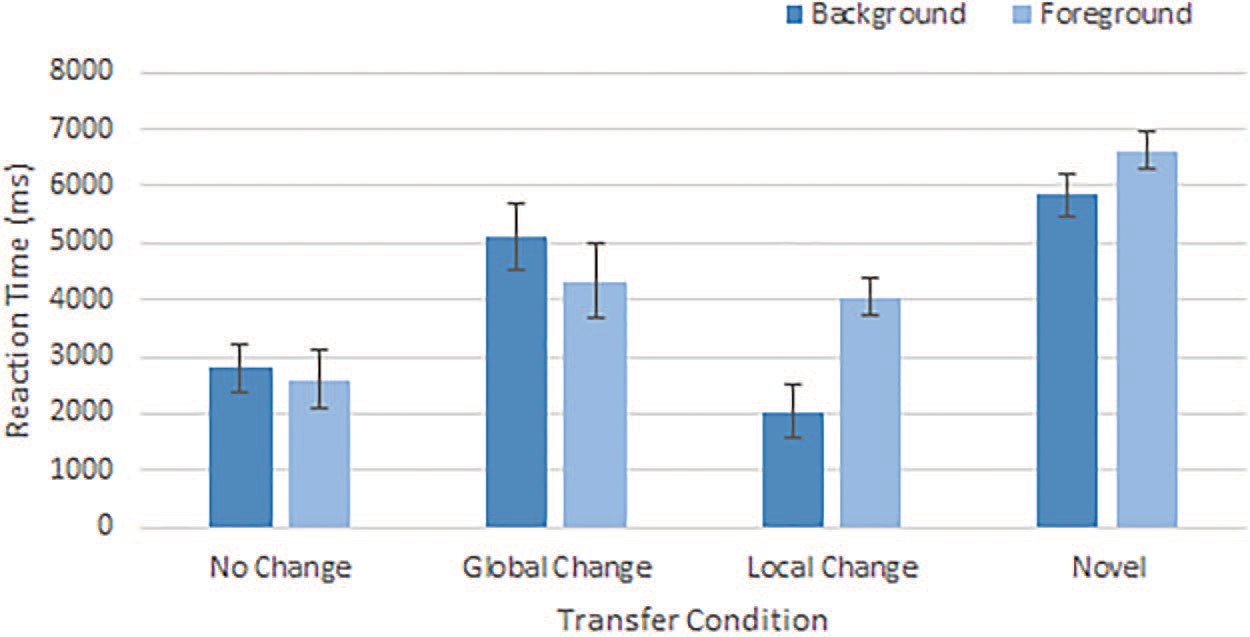
111) = 85.93, *p* < .001, ll2 = .7, but the placement of the target

p

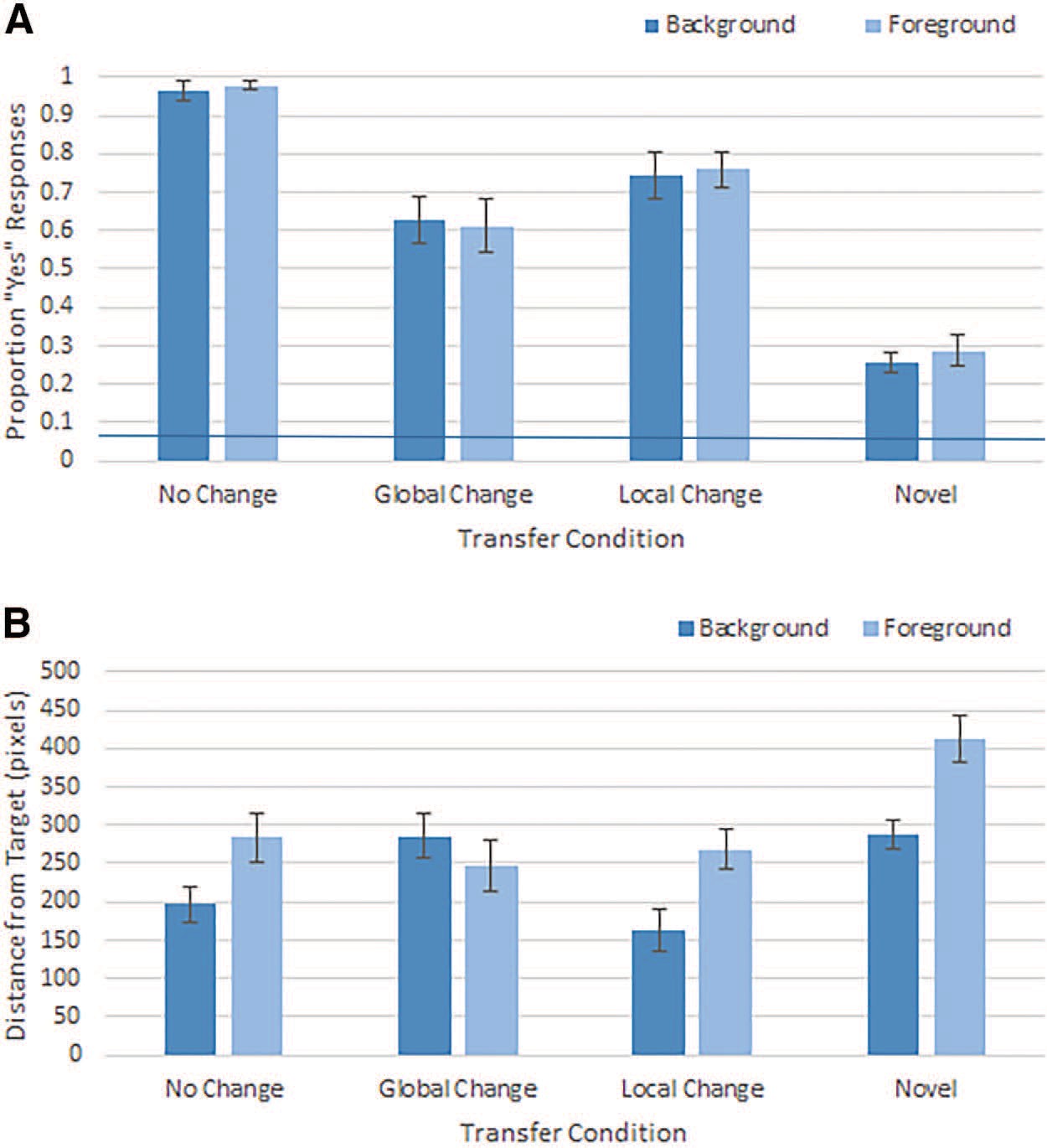
had no effect, *F*(1, 37) < 1, *p* > .9, and there was no interaction, *F*(3, 111) < 1, *p* > .7. When we collapsed across target placement and examined how performance fared across transfer conditions (a = .008), we found that accuracy was significantly better than

This document is copyrighted by the American Psychological Association or one of its allied publishers.

This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.



*Figure 5.* Mean reaction times across Transfer Condition X Target Place Condition in Experiment 1. Error bars represent standard error of the mean. See the online article for the color version of this figure.



*Figure 6.* Mean memory performance for both test questions in Experiment 1. (A) Proportion “yes” responses for Scene Identity Memory Question X Transfer and Target Place Conditions (horizontal bar represents false alarm rate of .075). (B) Average Distance to Target Position X Transfer and Target Place Conditions. Error bars represent standard error of the mean. See the online article for the color version of this figure.

the novel condition for both the local change, *t*(39) = 10.87, *p* <

This document is copyrighted by the American Psychological Association or one of its allied publishers.

This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

.001, *d* = 2.14, and global change, *t*(39) = 8.96, *p* < .001, *d* = 1.58, conditions. In addition, accuracy was significantly worse for the no change condition than both the local change, *t*(39) = 5.09, *p* < .001, *d* = 1.07, and global change, *t*(39) = 8.11, *p* < .001, *d* = 1.74, conditions, but these did not significantly differ from each other, *t*(39) = -2.42, *p* = .02, *d* = 0.50.

For the scenes that participants positively responded to in the identification test, we examined the recall accuracy of target lo- cations. Overall, the memory for the exact position of the target was poor (average 265 pixels, which corresponds to �12° at a viewing distance of 60 cm). In a 2 X 4 ANOVA, we found a marginal effect of target placement, *F*(1, 13) = 4.55, *p* = .053, ll2 = .26, with better overall memory for background targets than foreground targets. We also found a significant effect of transfer condition, *F*(3, 39) = 5.50, *p* = .003, ll2 = .30, but no interaction, *F*(1.92, 25.06) = 2.10, *p* = .15. We examined how performance varied across transfer conditions and targets (a= .005) and found that target location memory for background targets followed the same pattern: When information in close proximity to the target was consistent, memory performance was better than when it was changed. The local change condition was significantly better than

p

p

for the novel, *t*(39) = -3.06, *p* = .004, *d* = .84, and did not differ from the no change condition, *t*(39) = .39, *p* > .7. However, the reverse was true for the global change condition, which showed marginally worse performance than the no change condition, *t*(39) = -2.87, *p* = .007, *d* = .66, and no difference from the novel condition, *t*(39) = -.70, *p* = .5. For the foreground targets, memory for both local and global change conditions were better than the novel images, *t*s(39) > -3.22, *p*s < .001, *d*s > .89, but did not differ from each other, *t*s(39) < .73, *p*s > .47.

On the basis of the memory test results, performance largely depends on the information being probed (scene identity vs. target location). For scene identification, performance was higher when background information was consistent (local change) and re- mained relatively high when foreground information was consis- tent (global change). For target location, the pattern of perfor- mance was parallel to the contextual cueing effects across transfer conditions. Interestingly, responses were more accurate overall when the target appeared in the background, although it is unclear why.

Concerning the contextual cueing effect, the results seem to be compatible in some way with each of the models. For the hierar-

chical memory model, we would predict that there would be an asymmetric pattern with a large disruption for global change and very little disruption with a local change. For the parallel memory model, we would predict that the type of change is irrelevant as any consistent information from learning to transfer could act as an effective cue to retrieve target location. Consistent with the parallel memory model, foreground targets were associated with both the local and global context as the contextual cueing effect was re- tained to some extent with either a local or a global change. However, the pattern of the background target was consistent with the hierarchical memory model. Results in this case were asym- metrical in that the contextual cueing effect was retained when global information was consistent, but disrupted with the global information was changed. Despite this apparent contradiction in the overall pattern of results, differences in how background versus foreground information is encoded could explain these differences. We further discuss each model and its respective implementation

significant, *F*(4.58, 174.15) = 17.12, *p* < .001, ll2 = .31, when Greenhouse-Geisser corrected (= .57). There was also an effect of learning condition, *F*(1, 38) = 136.2, *p* < .001, ll2 = .78, and a two-way interaction between learning and block, *F*(6.13, 232.85) = 10.32, *p* < .001, ll2 = .21, when Greenhouse-Geisser corrected (= .77). The effect of group was not significant, *F*(1,

38) = .04, *p* > .25, ll2 = .001, but there was a significant Block X Group interaction, *F*(8, 304) = 2.81, *p* = .005, ll2 = .07, and a significant Group X Learning X Block three-way interaction, *F*(8, 304) = 4.09, *p* < .001, ll2 = .1.

p

p

p

p

p

p

When analyzed separately for each group, we found the same overall pattern of results. For the background target group, there was a significant effect of Block, *F*(2.97, 56.45) = 8.76, *p* < .001, ll2 = .31, when Greenhouse-Geisser corrected (= .37), and learning, *F*(1, 19) = 69.12, *p* < .001, ll2 = .78, as well as an interaction, *F*(8, 152) = 8.33, *p* < .001, ll2 = .31. The same was true for the foreground target group, with a significant effect of

p

p

p

2

of prioritized information in the General Discussion.

block, *F*(4.38, 83.29) = 10.86, *p* < .001, llp

= .36, when

In Experiment 1, all participants searched for background and foreground objects equally often, and thus, neither area was pri- oritized across experimental trials. In Experiment 2, we examined whether the global and local information of a scene could be disassociated and differentially prioritized. Separate groups of participants saw the target in either the foreground or the back- ground region across a majority of trials. If the hierarchical mem- ory model is correct, then changing the background of a scene should have a detrimental effect on contextual cueing, even when the participant is encouraged to focus on the foreground. If the parallel memory model is correct, then changing the background of the scene should have a small effect on contextual cueing when a majority of targets are in the foreground of the scene.

## Experiment 2

**Method**

**Participants.** Forty Queen’s University undergraduates par- ticipated in Experiment 2 (none had participated in Experiment 1) and were compensated either $10/hr or course credit for their participation. All participants had normal or corrected-to-normal vision. The study was reviewed and cleared by the Queen’s Uni- versity General Research Ethics Board.

Greenhouse-Geisser corrected (= .55), and learning, *F*(1, 19) =

67.19, *p* < .001, ll2 = .78, as well as a significant interaction, *F*(4.95, 94.05) = 6.59, *p* < .001, ll2 = .26, when Greenhouse- Geisser corrected (= .62). Planned paired-sample *t* tests for each group confirmed the contextual cueing effect and delineated the different pattern between groups (a = .003). For the background target group, there were no significant differences in search be- tween novel and repeated arrays in Blocks 1 through 3, but in Blocks 4 through 9 repeated arrays response times were signifi- cantly faster than were those for novel arrays (all *p*s < .001). For the foreground target group, there were no significant differences in search between novel and repeated arrays in Blocks 1 through 3 and Block 6, but in Blocks 4 and 5 and 7 and 9, repeated arrays response times were significantly faster than were those for novel arrays (all *p*s < .001). Thus, the three-way interaction arose from the slightly differing patterns between the learning conditions across blocks; however, both groups did show evidence of a contextual cueing effect.

p

p

**Transfer phase.** The effects of the transfer were first analyzed with a 2 X 4 mixed-measures ANOVA that examined group (foreground target vs. background target) and transfer conditions (novel, no change, local change, and global change). The ANOVA revealed a significant difference in search time across transfer conditions, *F*(3, 114) = 30.2, *p* < .001, ll2 = .44, and a significant

p

p

**Materials and apparatus.** Stimuli were identical to those

interaction, *F*(3, 114) = 7.91, *p* < .001, ll2 = .17, but no signif-

used in Experiment 1.

This document is copyrighted by the American Psychological Association or one of its allied publishers.

This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

**Procedure.** The procedure was similar to Experiment 1, with the following exceptions. The participants searched for targets that were either largely placed in the foreground (80% in each block) or in the background (80%). The participants were not explicitly instructed on the placement of the targets, but anecdotally all participants reported that they noticed this pattern by the end of the experiment.

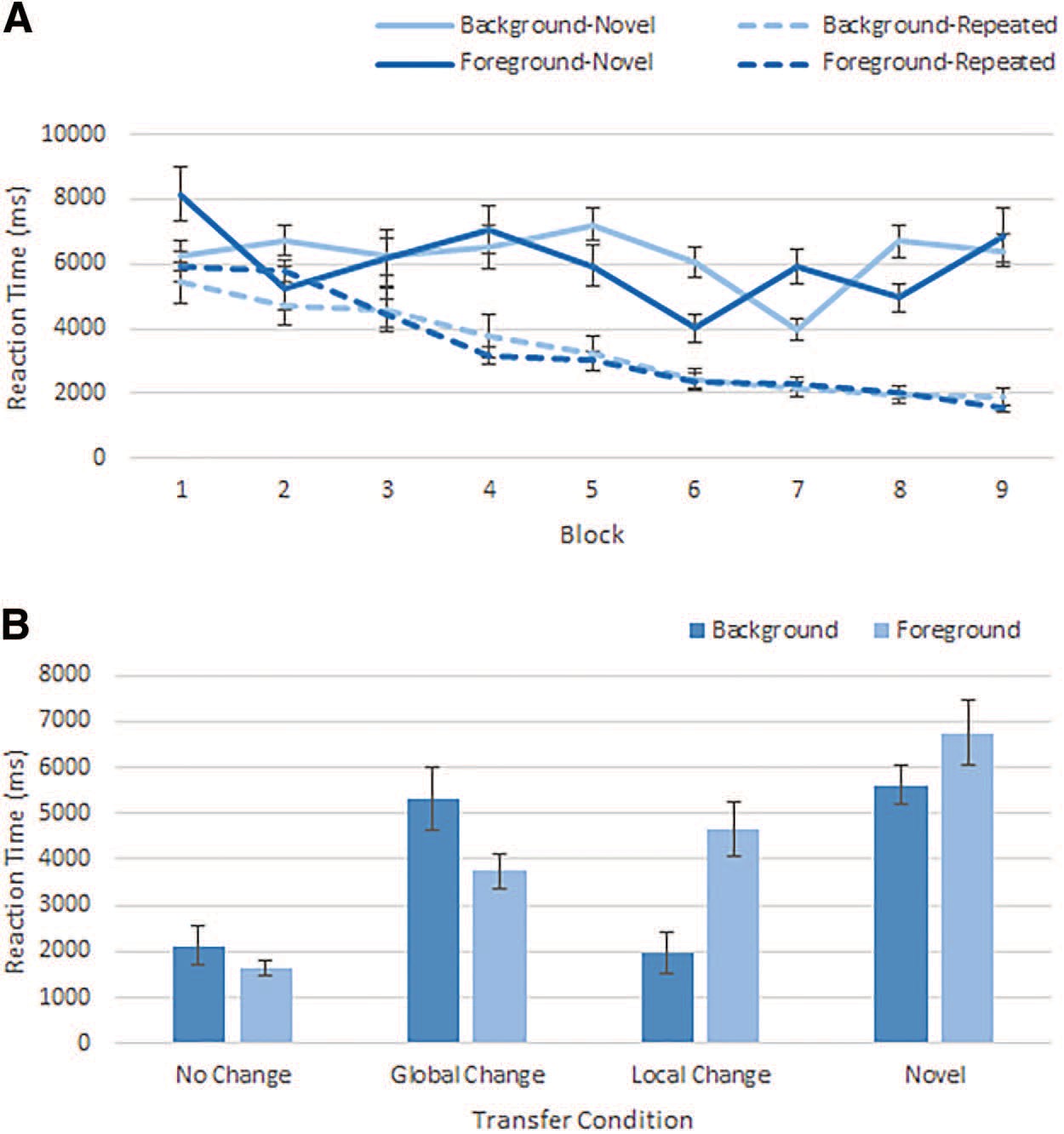
## Results and Discussion

**Learning phase.** We examined whether there was a contex- tual cueing effect in search times across the learning phase with a 2 (Group) X 2 (Learning Condition) X 9 (Block) mixed-measures ANOVA. Means for each block by learning condition for each group are presented in [Figure 7,](#_bookmark6) Panel A. The effect of block was

icant main effect of group, *F*(1, 38) = 1.07, *p* > .25, ll2 = .03. Means for each block by learning condition for each group are presented in [Figure 7](#_bookmark6) (Panel B).

We further examined how target placement (background vs. foreground group) modulated the effect of the transfer conditions by conducting planned comparisons between the two key transfer conditions (local change and global change) and the two control transfer conditions (no change and novel) for each of the target groups (a = .00625). If focusing on either background or fore- ground targets helped encode the visual information immediately surrounding the target, then the contextual cueing effect should only be most affected when that background or foreground infor- mation is changed, respectively for each target placement. How- ever, if global context is a precursor for retrieving foreground information, as the hierarchical memory model predicts, then foreground targets should no longer show the contextual cueing

p



*Figure 7.* Mean reaction times across (A) Block X Target Place Group and Learning Condition and (B) Transfer Condition X Target Place Group Condition in Experiment 2. Error bars represent standard error of the mean. See the online article for the color version of this figure.

This document is copyrighted by the American Psychological Association or one of its allied publishers.

This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

effect when either the background or the foreground information is changed. Alternatively, if both are associated with the target loca- tion to a greater or lesser degree, then when participants are focused on the foreground, as the parallel memory model predicts, the effect should persist regardless of the information changed (i.e., Experiment 2 should show the same pattern as Experiment 1). First, we examined the effect of changing the foreground infor- mation (local change). When participants were focused on back- ground targets, search times for local change were significantly faster than novel scenes, *t*(19) = -6.49, *p* < .001, *d* = 2.98, but not significantly different from no change condition, *t*(19) = .28, *p* > .25, *d* = 0.12. Thus, for the background target group, chang- ing the foreground had no effect on contextual cueing. Participants focused on the foreground produced a different pattern of results. Search times for local change were significantly slower than no change scenes, *t*(19) = -4.66, *p* < .001, *d* = 2.14, and were marginally faster than novel scenes, *t*(19) = -2.77, *p* = .01, *d* =

1.27. It should be noted that in both these cases, the effect size was quite large [(Cohen, 1977).](#_bookmark20) Thus, for the foreground target group, changing the foreground slowed search, but participants were nonetheless able to recover useful information from the global context.

Second, we examined the effect of changing the background information (global change). If the hierarchical memory model is

correct, and global context is necessary to guide attention to its paired local context, then global change scenes should impair the contextual cueing effect for both groups. For the background target group, search times for global change scenes were significantly slower than no change scenes, *t*(19) = -4, *p* = .001, *d* = 1.84, and did not differ significantly from novel scenes, *t*(19) = -.42, *p* >

.25, *d* = 0.19. In contrast, for the foreground target group, search times for global change scenes were significantly faster than novel scenes, *t*(19) = -3.77, *p* = .001, *d* = 1.73, and significantly slower than no change scenes, *t*(19) = -4.84, *p* < .001, *d* = 2.22. Depending on the groups’ focus, different patterns of results emerged. When participants were focused on the foreground, then both foreground and background information were used to recover some information about the target location. We found that chang- ing the foreground (local change) or background (global change) disrupted, but did not eliminate the contextual cueing effect. This was not the case for participants focused on the background, where only the background information produced a contextual cueing effect. It is interesting that in this case, the foreground information could not be used to recover any useful information about the background target location. We discuss this pattern further in the general discussion, but it suggests that each scene region is en-

coded with differing levels of importance.

The results from Experiment 2 had two overall patterns that are consistent with those from Experiment 1: (1) The background targets clearly showed differing patterns of retrieval compared with those located in the foreground, and (2) foreground informa- tion stored in functional subregions can be retrieved (to some extent) independently of the availability of the global context. The first pattern is consistent with a hierarchical model of memory in that the background has a special status that is largely independent of the foreground information. However, this pattern of results could also be consistent with a parallel model in which the back-

mate position of the target in this picture). Means are presented in [Figure 8.](#_bookmark7)

For the scene identification test, we examined the proportion of “yes” responses to each condition (proportion of “yes” responses to foils was .06). In a 2 X 4 mixed ANOVA, we analyzed group (background target and foreground target) and transfer conditions (no change, global change, local change, and novel). We found that a significant effect of transfer, *F*(3, 114) = 70.37, *p* < .001, llp = .65, and a significant interaction between group and transfer

2

2

condition, *F*(3, 114) = 10.32, *p* < .001, llp = .21, but no effect of

ground information has a higher weighting than foreground infor-

mation. The second pattern of results is of more theoretical inter- est. This pattern is not consistent with the basic premise of the hierarchical model of memory, in which the retrieval of functional subregions is dependent on the availability of the background context. This pattern is, however, consistent with a parallel model of memory in which the information is retrievable, although to a lesser extent due to the mismatch of the cue with the memory. We further discuss the assumptions of each model in the General Discussion.

**Memory phase.** Again, for the memory test, we looked at memory for the scene identity (either as a whole or in part) and memory for target position (indicate with the cursor the approxi-

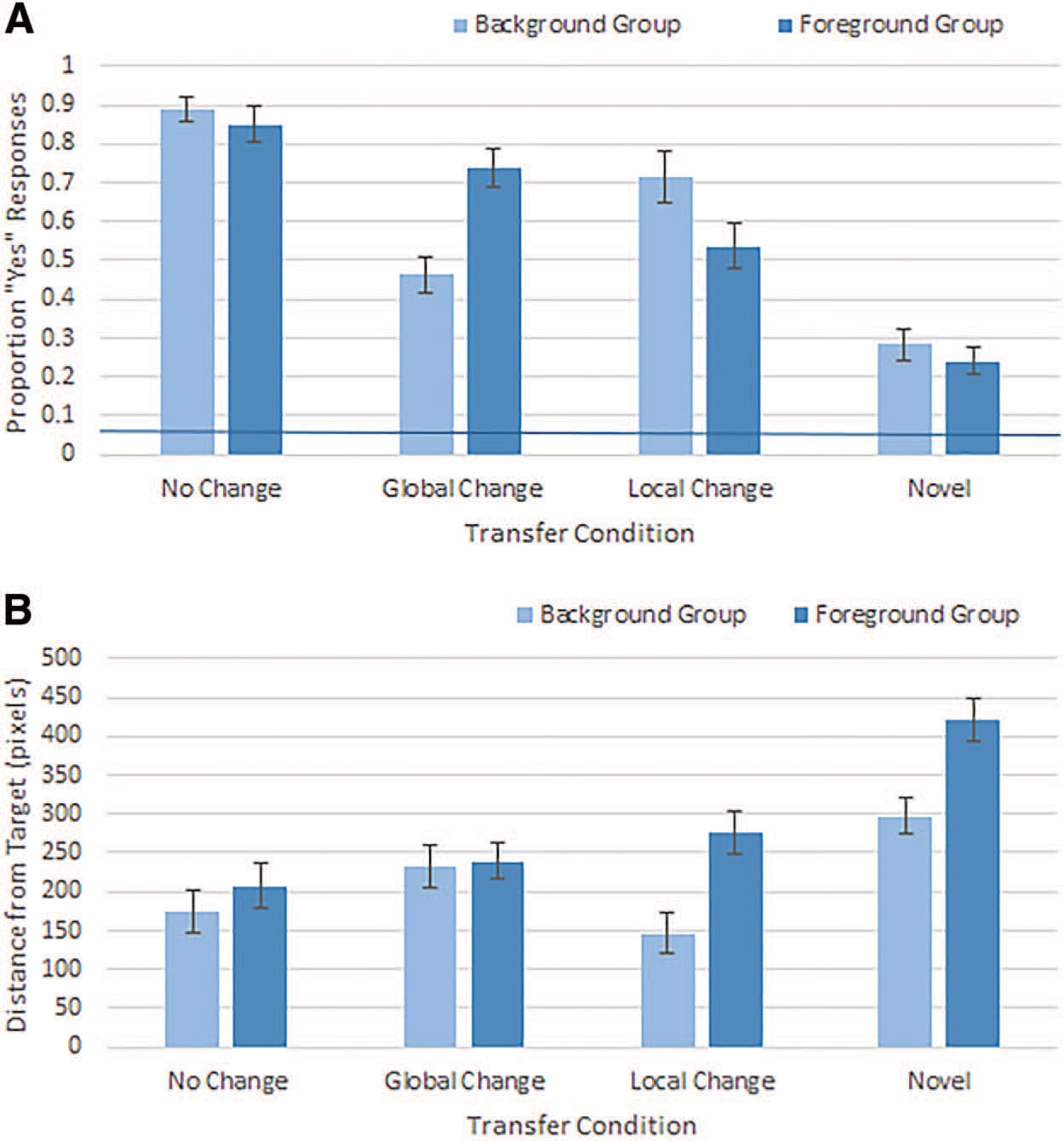
group, *F*(1, 38) < 1, *p* > .9.

For each group, we conducted four planned comparisons to contrast the control transfer conditions (no change and novel) with the experimental transfer conditions (global and local change; a=

.00625). When we examined the pattern for the background target group, we found that memory performance was significantly lower for global change than the no change, *t*(19) = 8.23, *p* < .001, *d* = 2.37, and it was not significantly different from the Novel condi- tion novel condition, *t*(19) = 2.58, *p* = .018, *d* = .92. For the local change, there was no difference from the no change, *t*(19) = 2.90, *p* = .009, *d* = .65, and it was significantly better than novel, *t*(19) = 5.33, *p* < .001, *d* = 1.74.

This document is copyrighted by the American Psychological Association or one of its allied publishers.

This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.



*Figure 8.* Mean memory performance for both test questions in Experiment 2. (A) Proportion “yes” responses for Scene Identity Memory Question X Transfer and Group Conditions (horizontal bar represents false alarm rate of .06). (B) Average Distance to Target Position X Transfer and Group Conditions. Error bars represent standard error of the mean. See the online article for the color version of this figure.

When we examined the pattern for the foreground target group, we found memory performance significantly lower for local change than the no change, *t*(19) = 5.78, *p* < .001, *d* = 2.37), but better than novel condition, *t*(19) = 5.63, *p* < .001, *d* = 1.42. For the global change, there was no difference from the no change, *t*(19) = 2.13, *p* = .046, and it was significantly better than novel, *t*(19) = 9.44, *p* < .001, *d* = 2.57. Overall, this pattern mimicked the results from the contextual cueing effect and those found in Experiment 1, with the proximity of scene information to the target location providing the most support for memory re- trieval. What is clear is that though there is no apparent association between the foreground with the background targets, there is an association between both the foreground and background with the foreground targets.

This document is copyrighted by the American Psychological Association or one of its allied publishers.

This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

We also looked at the accuracy of memory for exact target position for scenes that participants reported recognizing. We examined distance from target location in a 2 X 4 mixed ANOVA with group (foreground and background) and transfer condition (no change, global change, local change, and novel). Similar to Experiment 1, we found a main effect of group, *F*(1, 32) = 4.64, *p* = .039, ll2 = .13, with better overall memory for background targets than foreground targets. We also found a significant effect of transfer condition, *F*(2.53, 75.82) = 17.73, *p* < .001, ll2 = .37, when Greenhouse-Geisser corrected (= .84), and a significant interaction, *F*(3, 96) = 5.00, *p* = .003, ll2 = .14. We examined how performance varied across transfer conditions and targets (a = .005). For background targets, target location memory fol- lowed a similar pattern to that seen for the contextual cueing effect: When information in close proximity to the target was maintained, memory performance was better than when it was changed. The local change condition was significantly better than novel condition, *t*(15) = -3.70, *p* = .002, *d* = 1.29, and did not differ from the no change condition, *t*(18) = .39, *p* > .7. However, although the global change condition performance was numeri- cally lower than the no change condition, it was not significantly different, *t*(19) = -1.95, *p* = .07, *d* = .63, and not significantly different from the novel condition, *t*(16) = -1.70, *p* > .1. For the foreground targets, both the local change, *t*(15) = -4.67, *p* <

p

p

p

.001, *d* = 1.08, and global change, *t*(16) = -5.25, *p* < .001, *d* = 1.60, conditions were significantly better than the novel condition; neither the local change condition, *t*(39) = 2.38, *p* = .02, *d* = .55, nor the global change condition, *t*(39) = 1.25, *p* > .2, *d* = .28, differed from the no change condition.

The results from the memory test show that performance largely depended on the information being probed (scene identity vs. target location). For scene identification, having the background information present seemed to improve performance, but perfor- mance was still relatively high when performance was based on the foreground information. For target location, the placement of the target seemed to matter more. Consistent with Experiment 1, responses were overall more accurate when the target appeared on a background element.

## General Discussion

In the present study, we investigated the structure of scene representations in memory. We contrasted two opposing views of the structure: hierarchical memory models versus parallel memory models. On the one hand, hierarchical models posit that informa-

tion is structured in a nested fashion and retrieval of more detailed information is dependent on the retrieval of context information [(Brooks et al., 2010;](#_bookmark13) [Hirtle & Jonides, 1985;](#_bookmark40) [McNamara, 1986;](#_bookmark48) [McNamara et al., 1989).](#_bookmark50) On the other hand, parallel models posit information is stored such that the retrieval of one piece of infor- mation is not dependent on another, but information may be stored with differing strengths (or weights; [Hintzman, 1984,](#_bookmark34) [1988;](#_bookmark35) [Hintz-](#_bookmark37) [man & Curran, 1994).](#_bookmark37) We investigated which of the two models best captures how scene information in memory is structured, as well as how different cues subsequently affect information re- trieved.

To investigate the two alternative memory models, we examined whether the *scene context* (defined as the scene background), *local scene subregions* (defined as the scene foreground), or both drive learned associations with a target. We tested this question using stimuli that contained semantically distinct subregions of the scene (chimera scene; e.g., background of a kitchen and foreground of a bedroom). Using a contextual cueing paradigm, participants searched for target letters placed either in the background or foreground of the scenes across two experiments. Importantly, we found that where the target letter was placed had an effect on how information was retrieved.

For the background targets, we found that learned associations were dependent on the immediately surrounding visual informa- tion and were not affected by information in the foreground. In both experiments, the contextual cuing effect for the background targets was wholly dependent on being able to recognize the relevant scene background information. When scene background information was inconsistent (global change transfer condition), there was no preservation of the contextual cueing effect. This pattern of results indicates that information provided by the fore- ground information was either not encoded or not a sufficient cue for the retrieval of the background target location.

In contrast, for the foreground targets, we found that when the information immediately surrounding the target was changed (lo- cal change transfer condition), the contextual cueing effect was not completely abolished: some information about the target location was retrievable from the background context information. Previous studies have shown similar effects, where the ability to extract information from scene context to locate the target [(Brockmole et](#_bookmark9) [al., 2006;](#_bookmark9) [Brockmole & Henderson, 2006b).](#_bookmark12) Along similar lines, when the surrounding background information was changed (global change transfer condition), we found a similarly decreased but not altogether abolished contextual cueing effect. The finding that a changed background can permit some retrieval of the target information indicates that not all information about the target is routed through information about the larger scene context.

The benefit from both the background and foreground infor- mation for foreground targets is in contrast to previous studies that have found that changing the background context impedes retrieval of target location and abolishes the contextual cueing effect [(Brooks et al., 2010).](#_bookmark13) However, it is consistent with the notion that information from the foreground is associated with the target to some degree [(Brockmole et al., 2006;](#_bookmark9) [Rosenbaum](#_bookmark38)

[& Jiang, 2013).](#_bookmark38) In past studies, these preserved cueing effects of the foreground elements were largely found only when the larger context proved to be nonpredictive of the target location. One important difference between the current study and previ- ous studies is the presence of a strong semantic cue from the

local, foreground information. A comparison of findings from visual arrays and more naturalistic scenes suggest that semantic cues may play an important role in the differences found across studies.

## Mechanisms of Retrieval

Studies using the contextual cueing paradigm with scenes have found different pattern of results from those using contextual cueing paradigm with visual arrays of letters [(Brockmole et al.,](#_bookmark9) [2006;](#_bookmark9) [Brockmole & Henderson, 2006a,](#_bookmark14) [2006b).](#_bookmark12) [Table 1](#_bookmark8) provides a summary of the differences in effect sizes across studies exam- ining scenes and arrays. Research into the contextual cueing effect involving visual arrays has shown relatively small contextual cueing effects and implicated local context as driving the effect [(Brady & Chun, 2007;](#_bookmark11) [Jiang & Wagner, 2004;](#_bookmark42) [Olson & Chun,](#_bookmark27) [2002).](#_bookmark27) Conversely, studies involving scenes have produced longer search times and correspondingly a larger benefit from the con- textual cueing effect [(Brockmole et al., 2006;](#_bookmark9) [Brockmole & Hen-](#_bookmark14) [derson, 2006a,](#_bookmark14) [2006b),](#_bookmark12) as well as implicated global context as driving the effect [(Brockmole et al., 2006).](#_bookmark9) [Rosenbaum and Jiang](#_bookmark38) [(2013)](#_bookmark38) also examined this contradiction between scenes and arrays and had participants search for a target among distractors that were overlaid on a natural scene image. They found that when the scene was predictive, a contextual cueing effect was driven by the scene image, but when the scene was not predictive, the array was used to retrieve target location. Consistent with previous findings, the magnitude of the effect in the [Rosenbaum and Jiang (2013)](#_bookmark38) study changed according to the type of information used to retrieve target location (see [Table](#_bookmark8) 1), with a larger effect when it was based on the scene and a smaller effect when based on the array infor- mation. [Brooks et al. (2010)](#_bookmark13) operationalized their subregion as an array of symbols and produced a contextual cueing effect of the same magnitude typically observed in array studies. This suggests that the main source of information for retrieving the target posi- tion came from the array embedded in the scene context.

In the case of the current study, we found that for foreground targets both the background and foreground were effective in conveying target information as changing one or the other served to reduce but not abolish the effect. This was not the case for the background targets, where the information in the background was exclusively responsible for the effect. At first, it may seem that the

results for the background targets are most consistent with a hierarchical model, whereas the results from the foreground targets are most consistent with a parallel model. However, we contend that together the results are most consistent with a parallel model. Further, we would posit that the way in which information across the scene is stored is flexible, depending on task constraints. In addition, the larger context or background is stored with a greater level of importance than the foreground information. We examine each of these conjectures in turn.

## Flexible Representations

Previous studies have shown a pattern of results that is either consistent with either a parallel [(Brockmole et al., 2006)](#_bookmark9) or a hierarchical memory structure [(Brooks et al., 2010;](#_bookmark13) [Rosenbaum &](#_bookmark38) [Jiang, 2013).](#_bookmark38) Interestingly, the results from the present study have elements of both. We found that changing visual information at the location of the target (whether background or foreground) reduced the contextual cueing effect. So, at the very least, we can say that memory for the specific visual details surrounding the target are an important part of the memory for the target’s placement [(Brady &](#_bookmark11) [Chun, 2007).](#_bookmark11) This meant that the global change was most disrup- tive for the background target and the local change for the fore- ground target. The interesting and flexible part of the representa- tion seems to be for the information presented at a greater distance from the target (both visual and semantic) had different levels of encoding depending on the position of the target. Thus, it seems that the extent of the scene context associated with the target depends on whether it is placed in foreground or the background. The link between a foreground target and the foreground and background of the scene suggests different weighting of the infor- mation for foreground and background information. This is clear from a number of studies, including the present study. For exam- ple, [Rosenbaum and Jiang (2013)](#_bookmark38) suggested scene context plays a much greater role in cueing target location. Thus, even when the array is predictive, the learned association is between the scene context and the target location. They refer to this pattern as *overshadowing*, where the scene context is such an exceptionally effective cue that any other predictive information is simply not learned. A similar pattern of effects was found in other studies [(Brockmole et al., 2006;](#_bookmark9) [Brooks et al., 2010).](#_bookmark13) However, overshad-

Table 1

*Novel and Repeated Search Times and Contextual Cueing Effect for Studies Across Different Stimulus Types*

Study Stimulus type

Novel search (ms)

Repeated search (ms)

Contextual cueing effecta

This document is copyrighted by the American Psychological Association or one of its allied publishers.

This article is intended solely for the personal use of the individual [user and is not to be](#_bookmark9) disseminated broadly.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| [Chun and Jiang (1998)](#_bookmark19) | Array of letters | �880 | �800 | �9% |
| [Olson and Chun (2002)](#_bookmark27) | Array of letters | �2,300 | �2,060 | �10% |
| [Jiang and Wagner (2004)](#_bookmark42) | Array of letters | �1,250 | �1,125 | �10% |
| [Brockmole and Henderson (2006a)](#_bookmark14) | Real-world scenes | �3,500 | �1,150 | �67% |
| [Brockmole and Henderson (2006b)](#_bookmark12) | Real-world scenes | �3,700 | �1,000 | �72% |
| [Brockmole, Castelhano, and Henderson (2006)](#_bookmark9) | Naturalistic scenes | �3,375 | �1,375 | �59% |
| [Brooks, Rasmussen, and Hollingworth (2010)](#_bookmark13) | Array subregion embedded in a scene | �850 | �800 | �6% |
| [Rosenbaum and Jiang (2013)](#_bookmark38) | Nonpredictive array on predictive scene | �2,300 | �1,500 | �30% |
|  | Predictive array on nonpredictive scene | �2,400 | �2,200 | �8% |
| Present Experiment 1 | Semantically distinct regions of a scene | �7,400 | �2,300 | �69% |
| Present Experiment 2 | Semantically distinct regions of a scene | �6,600 | �1,700 | �74% |

a Percentage difference between repeated and novel search times in the final learning block.

owing may also reflect a difference in the strength or weighting of the context information in memory.

This document is copyrighted by the American Psychological Association or one of its allied publishers.

This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Much like [Rosenbaum and Jiang (2013),](#_bookmark38) we found that there is a greater weight on background information overall. In Experiment 2 especially, even when the background was of little consequence to the task (foreground group), changing it still affected perfor- mance. Similarly, [Brockmole et al. (2006)](#_bookmark9) found a benefit from local visual information when global context was nonpredictive, but the benefit was reduced compared with when global context remained consistent and local context varied. Taken together, these results suggest that rather than a fixed nested structure that requires global information be accessed prior to accessing local informa- tion, performance reflects a parallel model that represents fore- ground and background information to greater or lesser degrees.

The pattern of results in the current study could be explained through a parallel model with differential weighting of foreground and background information. On the one hand, the foreground and background may be weighted according to the placement of the target, and so the weights of the larger context versus subregions would depend on task. On the other hand, there is a higher weighting of the scene context, regardless of task. In the case of the background targets, the background is heavily weighted be- cause of the importance of context and because it is where the target is located. Other areas of the scene are weakly encoded, if at all. For the foreground targets, the background is encoded because of the importance of context, whereas the foreground is encoded because it is the location of the target. We are currently developing a simulation of such a model for a future study.

## The Role of Context

The role of background context in memory is different from that of the foreground or scene subregions as has been made clear in this and in other studies [(Brooks et al., 2010;](#_bookmark13) [Navon, 1977;](#_bookmark15) [Rosenbaum & Jiang, 2013).](#_bookmark38) There are various reasons for this difference between how information from each source is stored and retrieved. One possibility is ease—larger environmental infor- mation is simply more readily available early on and therefore is easier to associate with the target. We know from a number of studies examining scene gist perception that context information is available before the first saccade is launched [(Castelhano & Hen-](#_bookmark17) [derson, 2008;](#_bookmark17) [Greene & Oliva, 2009b;](#_bookmark28) [Potter, 1976).](#_bookmark33) When asso- ciations are made between the target location and information from the stimuli, the background information is simply known earlier and for longer and thus has a better chance of being encoded. Thus, it may be easier to associate target location with the larger scene context because more detailed information simply takes longer to acquire.

Another possibility is its importance—scene context provides a framework and a set of assumptions that can assist subsequent processing. The advantage of knowing scene context early and its effect on processing has been shown with other types of processing [(Davenport & Potter, 2004;](#_bookmark22) [Friedman, 1979;](#_bookmark24) [Palmer, 1975).](#_bookmark30) Thus, having scene context information available rapidly may speak to its deliberate evolutionary and ecological value. Although there is no doubt that acquiring scene context information is quite rapid, it nonetheless is computationally expensive [(Malcolm, Nuthmann, &](#_bookmark46) [Schyns, 2014),](#_bookmark46) and only more recently have we seen an improve- ment in the rates of scene categorization in computational models

[(Oliva & Torralba, 2001;](#_bookmark23) [Xiao, Hays, Ehinger, & Oliva, 2010;](#_bookmark47) [Zhou, Khosla, Lapedriza, & Oliva, 2014).](#_bookmark49) The apparent ease with which this information is gathered may mean that overall, context plays an important role across a number of high-level processes. Another more fundamental reason for the importance of context may simply reflect a basic processing mechanism, often referred to as the global precedence effect [(Navon, 1977).](#_bookmark15) Commonly studied with smaller letters configured to resemble larger letters, research- ers have found across a number of different tasks that with these stimuli the global information dominates processing, especially when speeded responses are required. Similarly, the word domi- nance effect reflects the same pattern of results, where the identi- fication of a letter is superior when it is presented within a word than when it is presented in isolation. At its core, it seems that the global information may play an important role in processing and memory regardless of task or stimulus type and would be an

interesting question to address in future research.

The importance of context abounds in the literature, however the current study also points to a system that is not completely reliant on accessibility to the context to be able to use information from subregions. Although this is so, the finding from the present study that subregions can be recognized independently of the context should not be taken as a dismissal of the utility of global context, and that it has been shown to be exclusively responsible for contextual cueing in other experiments underscores its importance [(Brockmole et al., 2006;](#_bookmark9) [Brockmole & Henderson, 2006a,](#_bookmark14) [2006b).](#_bookmark12)

## Conclusion

In summary, we believe that the current findings help to resolve the ambiguity surrounding the representations of scenes in mem- ory. A parallel model could provide a mechanism to retrieve information from any semantically rich region, rather than oblig- atorily needing the global scene context within which the original target association was learned. This finding illustrates that seman- tic content is a necessary feature used in our memory for scenes. When omitted, the patterns may not reflect a fundamental source of information that our visual system can rely on in any search beyond the laboratory.

## References

Brady, T. F., & Chun, M. M. (2007). Spatial constraints on learning in visual search: Modeling contextual cuing. *Journal of Experimental Psy- chology: Human Perception and Performance, 33,* 798 – 815. [http://dx](http://dx.doi.org/10.1037/0096-1523.33.4.798)

[.doi.org/10.1037/0096-1523.33.4.798](http://dx.doi.org/10.1037/0096-1523.33.4.798)

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10,*

433– 436. <http://dx.doi.org/10.1163/156856897X00357>

Brockmole, J. R., Castelhano, M. S., & Henderson, J. M. (2006). Contex- tual cueing in naturalistic scenes: Global and local contexts. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32,* 699 – 706. <http://dx.doi.org/10.1037/0278-7393.32.4.699>

Brockmole, J. R., & Henderson, J. M. (2006a). Recognition and attention guidance during contextual cueing in real-world scenes: Evidence from eye movements. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 59,* 1177–1187. [http://dx.doi.org/10.1080/](http://dx.doi.org/10.1080/17470210600665996)

[17470210600665996](http://dx.doi.org/10.1080/17470210600665996)

Brockmole, J. R., & Henderson, J. M. (2006b). Using real-world scenes as contextual cues for search. *Visual Cognition, 13,* 99 –108. [http://dx.doi](http://dx.doi.org/10.1080/13506280500165188)

[.org/10.1080/13506280500165188](http://dx.doi.org/10.1080/13506280500165188)

Brooks, D. I., Rasmussen, I. P., & Hollingworth, A. (2010). The nesting of search contexts within natural scenes: Evidence from contextual cuing.

*Journal of Experimental Psychology: Human Perception and* *Perfor- mance, 36,* 1406 –1418. <http://dx.doi.org/10.1037/a0019257>

This document is copyrighted by the American Psychological Association or one of its allied publishers.

This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Castelhano, M. S., & Henderson, J. M. (2008). The influence of color on the perception of scene gist. *Journal of Experimental Psychology:* *Hu- man Perception and Performance, 34,* 660 – 675. [http://dx.doi.org/10](http://dx.doi.org/10.1037/0096-1523.34.3.660)

[.1037/0096-1523.34.3.660](http://dx.doi.org/10.1037/0096-1523.34.3.660)

Chun, M. M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology,* *36,* 28 –71. <http://dx.doi.org/10.1006/cogp.1998.0681>

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences*. Retrieved from [http://www.sciencedirect.com/science/book/97801](http://www.sciencedirect.com/science/book/9780121790608)  [21790608](http://www.sciencedirect.com/science/book/9780121790608)

Davenport, J. L., & Potter, M. C. (2004). Scene consistency in object and background perception. *Psychological Science, 15,* 559 –564. [http://dx](http://dx.doi.org/10.1111/j.0956-7976.2004.00719.x)

[.doi.org/10.1111/j.0956-7976.2004.00719.x](http://dx.doi.org/10.1111/j.0956-7976.2004.00719.x)

Friedman, A. (1979). Framing pictures: The role of knowledge in autom- atized encoding and memory for gist. *Journal of Experimental Psychol- ogy: General, 108,* 316 –355. [http://dx.doi.org/10.1037/0096-3445.108](http://dx.doi.org/10.1037/0096-3445.108.3.316)

[.3.316](http://dx.doi.org/10.1037/0096-3445.108.3.316)

Greene, M. R., & Oliva, A. (2009a). Recognition of natural scenes from global properties: Seeing the forest without representing the trees. *Cog- nitive Psychology, 58,* 137–176. [http://dx.doi.org/10.1016/j.cogpsych](http://dx.doi.org/10.1016/j.cogpsych.2008.06.001)

[.2008.06.001](http://dx.doi.org/10.1016/j.cogpsych.2008.06.001)

Greene, M. R., & Oliva, A. (2009b). The briefest of glances: The time course of natural scene understanding. *Psychological Science, 20,* 464 – 472. <http://dx.doi.org/10.1111/j.1467-9280.2009.02316.x>

Grimes, J. (1996). On the failure to detect changes in scenes across saccades. In K. A. Akins (Ed.), *Perception. Vancouver studies in cog-* *nitive science* (pp. 89 –110). New York, NY: Oxford University Press. Henderson, J. M., & Hollingworth, A. (1999). High-level scene perception.

*Annual Review of Psychology, 50,* 243–271. [http://dx.doi.org/10.1146/](http://dx.doi.org/10.1146/annurev.psych.50.1.243) [annurev.psych.50.1.243](http://dx.doi.org/10.1146/annurev.psych.50.1.243)

Hintzman, D. L. (1984). MINERVA 2: A simulation model of human memory. *Behavior Research Methods, Instruments, & Computers, 16,* 96 –101. <http://dx.doi.org/10.3758/BF03202365>

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review, 95,* 528 –551. <http://dx.doi.org/10.1037/0033-295X.95.4.528>

Hintzman, D. L., & Curran, T. (1994). Retrieval dynamics of recognition and frequency judgments: Evidence for separate processes of familiarity and recall. *Journal of Memory and Language, 33,* 1–18. [http://dx.doi](http://dx.doi.org/10.1006/jmla.1994.1001)

[.org/10.1006/jmla.1994.1001](http://dx.doi.org/10.1006/jmla.1994.1001)

Hirtle, S. C., & Jonides, J. (1985). Evidence of hierarchies in cognitive maps. *Memory & Cognition, 13,* 208 –217. [http://dx.doi.org/10.3758/](http://dx.doi.org/10.3758/BF03197683) [BF03197683](http://dx.doi.org/10.3758/BF03197683)

Jiang, Y., & Wagner, L. C. (2004). What is learned in spatial contextual cuing—Configuration or individual locations? *Perception & Psycho-* *physics, 66,* 454 – 463. <http://dx.doi.org/10.3758/BF03194893>

Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010). Scene memory is more detailed than you think: The role of categories in visual long-term memory. *Psychological Science, 21,* 1551–1556. [http://dx.doi](http://dx.doi.org/10.1177/0956797610385359)

[.org/10.1177/0956797610385359](http://dx.doi.org/10.1177/0956797610385359)

Malcolm, G. L., Nuthmann, A., & Schyns, P. G. (2014). Beyond gist: Strategic and incremental information accumulation for scene categori- zation. *Psychological Science, 25,* 1087–1097. [http://dx.doi.org/10](http://dx.doi.org/10.1177/0956797614522816)

[.1177/0956797614522816](http://dx.doi.org/10.1177/0956797614522816)

McNamara, T. P. (1986). Mental representations of spatial relations. *Cog- nitive Psychology, 18,* 87–121. [http://dx.doi.org/10.1016/0010-0285(86)](http://dx.doi.org/10.1016/0010-0285%2886%2990016-2)

[90016-2](http://dx.doi.org/10.1016/0010-0285%2886%2990016-2)

McNamara, T. P., Hardy, J. K., & Hirtle, S. C. (1989). Subjective hierar- chies in spatial memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 211–227. [http://dx.doi.org/10.1037/0278-](http://dx.doi.org/10.1037/0278-7393.15.2.211)

[7393.15.2.211](http://dx.doi.org/10.1037/0278-7393.15.2.211)

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology, 9,* 353–383. Retrieved from<http://www.sciencedirect.com/science/article/pii/0010028577900123>

Nickerson, R. S. (1965). Short-term memory for complex meaningful visual configurations: A demonstration of capacity. *Canadian Journal of Psychology, 19,* 155–160. <http://dx.doi.org/10.1037/h0082899>

Nickerson, R. S. (1968). A note on long-term recognition memory for pictorial material. *Psychonomic Science, 11,* 58 –58. [http://dx.doi.org/10](http://dx.doi.org/10.3758/BF03330991)

[.3758/BF03330991](http://dx.doi.org/10.3758/BF03330991)

Oliva, A. (2005). Gist of the scene. In L. Itti, G. Rees, & J. K. Tsotsos (Eds.), *The Encyclopedia of Neurobiology of Attention* (pp. 251–256). San Diego, CA: Elsevier. Retrieved from [http://cvcl.mit.edu/papers/](http://cvcl.mit.edu/papers/oliva04.pdf) [oliva04.pdf](http://cvcl.mit.edu/papers/oliva04.pdf)

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. Retrieved from <http://link.springer.com/> article/10.1023/A:1011139631724

Oliva, A., & Torralba, A. (2006). Building the gist of a scene: The role of global image features in recognition. *Progress in Brain Research, 155,* 23–36. [http://dx.doi.org/10.1016/S0079-6123(06)55002-2](http://dx.doi.org/10.1016/S0079-6123%2806%2955002-2)

Olson, I. I. R., & Chun, M. M. M. (2002). Perceptual constraints on implicit learning of spatial context. *Visual Cognition, 9,* [273–302. http://](http://dx.doi.org/10.1080/13506280042000162) [dx.doi.org/10.1080/13506280042000162](http://dx.doi.org/10.1080/13506280042000162)

Palmer, T. E. (1975). The effects of contextual scenes on the identification of objects. *Memory & Cognition, 3,* 519 –526. [http://dx.doi.org/10.3758/](http://dx.doi.org/10.3758/BF03197524) [BF03197524](http://dx.doi.org/10.3758/BF03197524)

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10,* 437– 442. [http://](http://dx.doi.org/10.1163/156856897X00366) [dx.doi.org/10.1163/156856897X00366](http://dx.doi.org/10.1163/156856897X00366)

Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory, 2,* 509 – 522. <http://dx.doi.org/10.1037/0278-7393.2.5.509>

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. [http://dx.doi.org/10.1111/j.1467-9280.1997](http://dx.doi.org/10.1111/j.1467-9280.1997.tb00427.x)

[.tb00427.x](http://dx.doi.org/10.1111/j.1467-9280.1997.tb00427.x)

Rosenbaum, G. M., & Jiang, Y. V. (2013). Interaction between scene- based and array-based contextual cueing. *Attention, Perception & Psy- chophysics, 75,* 888 – 899. <http://dx.doi.org/10.3758/s13414-013-0446-9> Shepard, R. N. (1967). Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior, 6,* 156 –163.

[http://dx.doi.org/10.1016/S0022-5371(67)80067-7](http://dx.doi.org/10.1016/S0022-5371%2867%2980067-7)

Simons, D., & Levin, D. (1997). Change blindness. *Trends in Cognitive Sciences, 1,* 261–267. Retrieved from [http://www.sciencedirect.com/](http://www.sciencedirect.com/science/article/pii/S1364661397010802) [science/article/pii/S1364661397010802](http://www.sciencedirect.com/science/article/pii/S1364661397010802)

Standing, L. (1973). Learning 10,000 pictures. *The Quarterly Journal of Experimental Psychology, 25,* 207–222. [http://dx.doi.org/10.1080/](http://dx.doi.org/10.1080/14640747308400340)

[14640747308400340](http://dx.doi.org/10.1080/14640747308400340)

Standing, L., Conezio, J., & Haber, R. N. (1970). Perception and memory for pictures: Single-trial learning of 2500 visual stimuli. *Psychonomic Science, 19,* 73–74. <http://dx.doi.org/10.3758/BF03337426>

Xiao, J., Hays, J., Ehinger, K., & Oliva, A. (2010). *Sun database: Large- scale scene recognition from abbey to zoo*. Retrieved from [http://](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5539970) [ieeexplore.ieee.org/xpls/abs\_all.jsp?arnumber=5539970](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5539970)

Zhou, B., Khosla, A., Lapedriza, A., & Oliva, A. (2014). *Object detectors emerge in deep scene CNNs*. Retrieved from [http://arxiv.org/abs/1412](http://arxiv.org/abs/1412.6856)

[.6856](http://arxiv.org/abs/1412.6856)

Received July 19, 2016

Revision received July 3, 2018

Accepted July 3, 2018 •