Acta Psychologica 169 (2016) 100–108

Contents lists available at ScienceDirect

Acta Psychologica

journal homepage: www.elsevier.com/locate/actpsy

Gist in time: Scene semantics and structure enhance recall of searched objects

Emilie L. Josephs [a,*,1], Dejan Draschkow [d,1], Jeremy M. Wolfe [b,c], Melissa L.-H. Võ [d]

[a] Cognitive and Neural Organization Lab, Harvard University, Cambridge, MA, USA
[b] Visual Attention Lab, Brigham and Women's Hospital, Boston, MA, USA
[c] Harvard Medical School, Cambridge, MA, USA
[d] Scene Grammar Lab, Johann Wolfgang Goethe-Universität, Frankfurt, Germany

ARTICLE INFO

Article history:
Received 18 August 2015
Received in revised form 24 February 2016
Accepted 20 May 2016
Available online 3 June 2016

Keywords:
Repeated search
Scene perception
Object recall
Scene gist
Integration time
Task effects
Incidental memory
Semantic guidance

ABSTRACT

Previous work has shown that recall of objects that are incidentally encountered as targets in visual search is better than recall of objects that have been intentionally memorized (Draschkow, Wolfe, & Võ, 2014). However, this counter-intuitive result is not seen when these tasks are performed with non-scene stimuli. The goal of the current paper is to determine what features of search in a scene contribute to higher recall rates when compared to a memorization task. In each of four experiments, we compare the free recall rate for target objects following a search to the rate following a memorization task. Across the experiments, the stimuli include progressively more scene-related information. Experiment 1 provides the spatial relations between objects. Experiment 2 adds relative size and depth of objects. Experiments 3 and 4 include scene layout and semantic information. We find that search leads to better recall than explicit memorization in cases where scene layout and semantic information are present, as long as the participant has ample time (2500 ms) to integrate this information with knowledge about the target object (Exp. 4). These results suggest that the integration of scene and target information not only leads to more efficient search, but can also contribute to stronger memory representations than intentional memorization.

1. Introduction

Suppose that you wanted to learn what objects were present in a room with the goal of being able to recall those objects later. One way to do this would be to examine the room, intentionally trying to memorize the objects. However, explicit memorization is not the only way to encode information about objects encountered in the world. We also acquire memory representations for objects incidentally when, for example, we search for an object without an explicit instruction to memorize (Castelhano & Henderson, 2005; Draschkow & Võ, 2016; Hout & Goldinger, 2010; Hout & Goldinger, 2012; Howard, Pharaon, Körner, Smith, & Gilchrist, 2011; Olejarczyk, Luke, & Henderson, 2014; Võ, Schneider, & Matthias, 2008).

Are intentionally memorized and incidentally encountered objects encoded differently? This seems likely, given that different tasks (e.g. search and memorization) require different interactions with the same stimuli. For instance, Rothkopf, Ballard, and Hayhoe (2007) showed task-specific deployment of attention to different parts of the same scene or objects. They found that the proportion of fixations landing on obstacles compared to targets changes depending on whether the observer is told to avoid the obstacles or collect the targets. Additionally, the authors demonstrated that different areas of the objects are selected for fixation in each of these tasks. Task-specific effects on eye movements are also evident in saccade lengths, which are longer in search tasks than in free viewing, and in the local image characteristics of the areas selected for fixation (Tatler, Baddeley, & Vincent, 2006). In addition to such differences in the deployment of eye movements, task-specific cognitive requirements also seem to cause differences in the extraction of information from a single fixation. For example, Tatler and Tatler (2013) found that task irrelevant objects received the same number of fixations when participants were told to memorize all the objects in the scene as when they were told to memorize a specific subset of objects (e.g. only objects used to make tea), yet object memory was higher in the first case.

Do these task-dependent modulations of attention and information extraction produce differences in recall between objects that have been memorized and objects that were searched? In the work of Võ and Wolfe (2012), participants searched for objects in scenes. Participants located targets in scenes more quickly if they had searched for them in a previous block. They showed no such improvement if they had been familiarized with the scene in other ways, such as searching

* Corresponding author at: Cognitive and Neural Organization Lab, William James Hall, 720, 33 Kirkland St, Cambridge, MA 02138, USA.
E-mail address: ejosephs@g.harvard.edu (E.L. Josephs).
[1] authors contributed equally.

http://dx.doi.org/10.1016/j.actpsy.2016.05.013
0001-6918/

for letters superimposed on the targets, exploring the scene for 30 s to determine if a man or woman decorated the room or by explicitly trying to memorize the scene prior to the search. A follow-up by Hollingworth (2012) also demonstrated that a previous search for a target speeds a subsequent search more than memorizing object locations or evaluating semantic properties of the scene. It seems that while looking *at* an object in a scene certainly creates a memory trace for it, looking *for* an object can build memory representations that can be used to find the same object again more efficiently (Võ & Wolfe, 2012; for a review see Võ & Wolfe, 2015).

These studies suggested differences between the representations formed incidentally during search and those formed intentionally during memorization. However, they did not measure recall explicitly, instead inferring target memorization from improved reaction times. In order to more directly compare representations created during search with those created during memory, Draschkow et al. (2014) used a free recall task. In their study, participants performed one block where they searched for objects in photographs of scenes, and another where they memorized objects in a different set of scenes. Each block was followed by a free recall test in which participants were asked to draw all the objects they could remember. A comparison of the average number of drawn targets revealed better recall following the search block than following the memorization block.

The first two experiments in the Draschkow et al. (2014) paper showed a substantial effect of task on recall rates of target objects. However, these results left open the question of whether all search tasks, regardless of stimulus set, lead to stronger memory representations, or whether this effect is only present when search is being performed in a naturalistic scene. Searches through scenes rather than displays of isolated items have been shown to make use of a very rich array of semantic scene information (for a review see Henderson, 2007; Wolfe, Võ, Evans, & Greene, 2011). Such recruitment of information may strengthen the representation of the searched objects over and beyond that of objects in non-scene contexts. In Experiment 3, Draschkow et al. (2014) tested the role of the scene content of the images by repeating their experiment using non-scene stimuli. They created images with unique textures as backgrounds (folds of fabric, droplets of water, a field of clover leaves), upon which they placed images of the objects that had been designated as targets in the previous experiments (see Fig. 1). The thumbnail images consisted of isolated exemplars of each of the original targets. These thumbnails were evenly distributed on the background images, removing any meaningful spatial relationships between the objects. Repeating the experiment with these non-scene stimuli abolished the original results: searched objects no longer showed a recall benefit over intentionally memorized objects.

These results indicated that simply searching for objects does not always build stronger representations than simply memorizing them. Performing the search in a meaningful, semantically rich scene seems to be important. However, the Draschkow et al. (2014) study could not specify why the effect was only observed in scenes. One possibility is that scenes are highly information-rich displays relative to randomly-organized collections of objects. In the process of transforming the scenes into non-scene stimuli in the Draschkow et al. (2014) study, objects were dissociated from their backgrounds, made uniform in size, and were placed at random locations on the screen. As will be described below, each of these sources of information has been shown to play a role in facilitating object perception or guiding search in scenes (Bar, 2004; Biederman, Mezzanotte, & Rabinowitz, 1982; Castelhano & Heaven, 2011; Torralba, Oliva, Castelhano, & Henderson, 2006). It is possible that accessing one or more of these sources of information during search for a target could create a stronger memory representation for that object than memorizing it.

Relationships Between Objects: One source of information exploited during real-world searches is the learned regularity in object grouping. Coffee mugs can be reliably found in proximity to coffee makers, pens are often found next to notebooks. Indeed, Castelhano and Heaven (2011) found that objects in their correct spatial grouping are easier to find and recognize than those in incongruent spatial groupings, even if the identity of the scene is made ambiguous by the presence of incongruent objects. Furthermore, ambiguous drawn object are more easily recognized if they are grouped with related objects (Bar, 2004). In general, objects in probable locations are better recognized than the

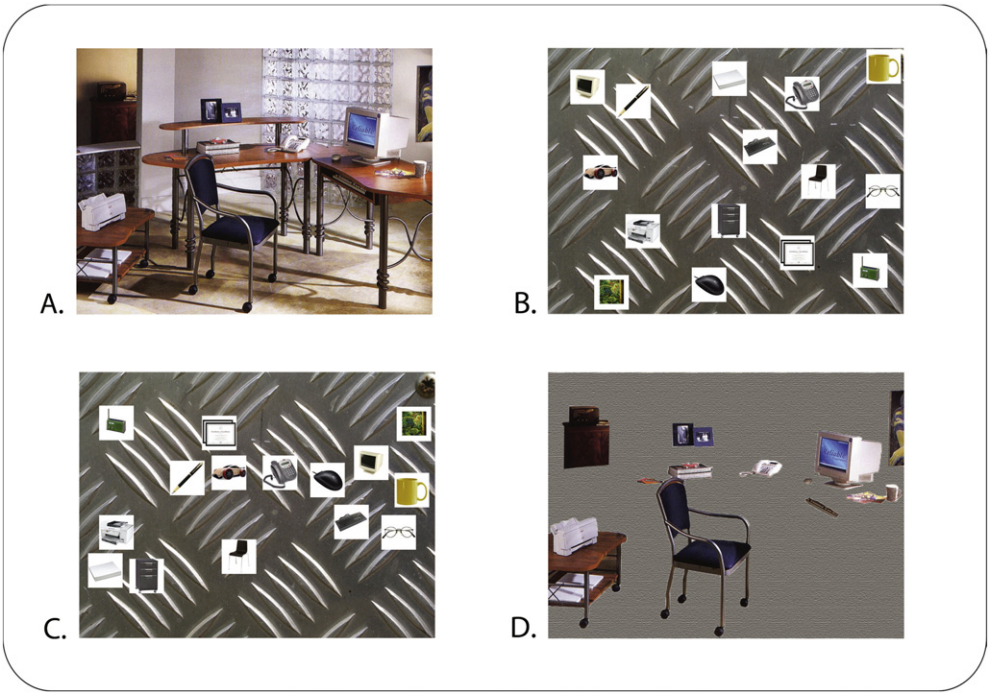

Fig. 1. A–B: Images used in Draschkow et al. (2014). B was created by finding new examplars of the targets selected in A and placing them on a non-scene background. C–D: images created for the current study. C was created by moving the objects from B into the spatial relationships they had in A. Notice how now the computer, keyboard and mouse are located near each other, as we would expect in real world situations. Stimulus D was created from stimulus A using image editing software to remove the background.

same objects presented in unexpected object-to-object or scene contexts (e.g. Biederman et al., 1982; Davenport & Potter, 2004; Gronau, Neta, & Bar, 2008). In addition to mere proximity, there are other object-based features that help specify the spatial relationship between objects, such as relative size, relative distance from the viewer, and orientation. Biederman et al. (1982) showed that inconsistently sized objects are missed more frequently in searches through line drawings of scenes. Furthermore, ambiguous line drawings of objects are easier to recognize when correctly oriented relative to a non-ambiguous object (Bar & Ullman, 1996), and target objects are recognized more accurately when they are oriented toward a related object in a functionally consistent way (e.g.: the pitcher is tilted toward the glass rather than away from it; Green & Hummel, 2006). All of these relationships were absent from the Draschkow et al. (2014) non-scene stimuli. In the present study, we incrementally add them back in to assess their role in boosting object recall following search in scenes.

Global scene information: Besides object-based information, another important source of information relevant to searches in natural scenes is the global scene information. These global features contribute to the gist of the scene, providing rapidly-accessible information about the scene's basic level category, its functional affordances, the objects that belong within it, and their typical locations (Greene & Oliva, 2009; Oliva, 2005; Oliva & Torralba, 2001). According to some models, such scene gist information is integrated with top-down task-specific information at an early stage and directs attention to potential target locations (Ehinger, Hidalgo-Sotelo, Torralba, & Oliva, 2009; Torralba et al., 2006). In support of this model, Võ and Henderson (2010) found that increasing the amount of time dedicated to integrating these two sources of information can improve search efficiency, even under conditions where scene information is degraded. Thus, by integrating the target identity with gist information, a store of structural and semantic knowledge is recruited to guide fixations. This recruitment might contribute to better recall upon later testing. In the memorization task in our previous work, the objects were framed with a red square, so participants never needed to perform a search. This may have reduced the availability of gist information in the memorization condition.

It is important to note here that scene-related effects are known to play a role in object memorization, in the absence of a search task. Both working memory and long-term memory have been shown to benefit from scene-related contextual factors, such as background information, correct spatial relations among objects, and semantic and functional relations among stimuli (Brewer & Treyens, 1981; Hollingworth, 2006; Kaiser, Stein, & Peelen, 2015; Gronau & Shachar, 2014). However, it is possible that the recruitment of such factors during search, which requires more engagement with the structural and semantic content, will cause better recall.

The current study aims to determine which of the sources of information mentioned above, contributes to the recall advantage for objects encoded incidentally during search in a scene compared to intentionally memorized objects. Isolating the type of information that drives the task-dependence of recall in natural scenes may allow us to better understand how information extraction and encoding differs between memorization and search tasks. Over a series of four experiments, our approach will be to incrementally introduce these sources of information into non-scene images and measure their effect on object recall following search and memorization tasks.

In Experiment 1, we tested recall of objects in non-scene stimuli in which thumbnails of objects drawn from a scene category were placed in their expected groupings on non-scene backgrounds. In Experiment 2, we further increased the spatial congruence between these objects by re-introducing other object-to-object spatial relationships such as orientation, relative size, and relative depth by using targets cut out of the original scenes rather than thumbnails. In Experiment 3, we added global gist information to these cut-out objects by adding a flash-preview of the full scene before initiating search or memorization for the same stimuli that were used in Experiment 2. In Experiment 4, we increased the integration time between the preview and the search display to increase recruitment of gist information in the performance of the task (see Võ & Henderson, 2010). To anticipate, recall was significantly modulated by task only in cases where gist information was provided and in which there was ample time to integrate this information with the identity of the target. In this condition, searched objects were remembered at a higher rate than memorized objects; in all other conditions, recall rates did not differ.

2. General methods

2.1. Participants

All participants in these experiments had 20/25 vision or better, normal color vision, and had no history of eye or muscular disorders. They gave informed consent according to the guidelines of the Brigham and Women's Hospital. Ten participants were tested in each experiment, in accordance with the Draschkow et al. (2014) experiments.

2.2. Apparatus

Participants viewed stimuli on a 19″ Mitsubishi Diamond Pro 991 TXM CRT monitor, and the experiment was built and run in Experiment Builder (SR Research, Canada). Eye movements were recorded with an EyeLink1000 desktop eye-tracker (SR Research, Canada) at a sampling rate of 1000 Hz. The position of one eye was tracked while viewing was binocular.

2.3. Design

The design for the current series of experiments (Fig. 2) was based on Draschkow et al. (2014). There were two blocks in each experiment: a search block and a memorization block, the order of which was counterbalanced across participants. In the search block, a trial started with the presentation of a word in white text superimposed on a scene for 750 ms, which identified the target object. Following this, the scene was presented alone and the participant was instructed to search for the designated target as fast as possible. Since each experiment used a different set of experimental displays, details about these scenes will be given in individual sections below. Fig. 1 gives examples of the different types of scenes. When observers located the target, they fixated it and pressed a button on a game pad. The trial ended with the button press, and the next trial immediately began with the presentation of the next search display together with the following target word. The memorization block was designed to be as similar as possible to the search block, while eliminating the need to perform a search. Thus, a trial in the memorization block would also start with the presentation of the target word for 750 ms superimposed on the scene, but upon the offset of the cue word, the target object was immediately framed by a red square. The participants were instructed to carefully memorize as much as they could of the display, paying particular attention to the target object. The scene with the framed object was displayed for 3 s before the next trial started with the presentation of the next scene with its target word. There were five different scenes in each block, each containing 10 targets that were queried one at a time in random order over the course of the experiment. Thus, each block contained 50 trials. The order of the trials was randomly determined, so consecutive searches were not necessarily performed in the same display.

Following each block, the participants performed a free recall test. They were told to expect a test after the memorization block, but were misdirected so they would not expect a test following the search block (i.e. by pretending it was a different experiment with a very similar design), in order to avoid the employment of memorization strategies during search. The order of the blocks was counterbalanced across subjects. The test consisted of a drawing task in which participants were given 5

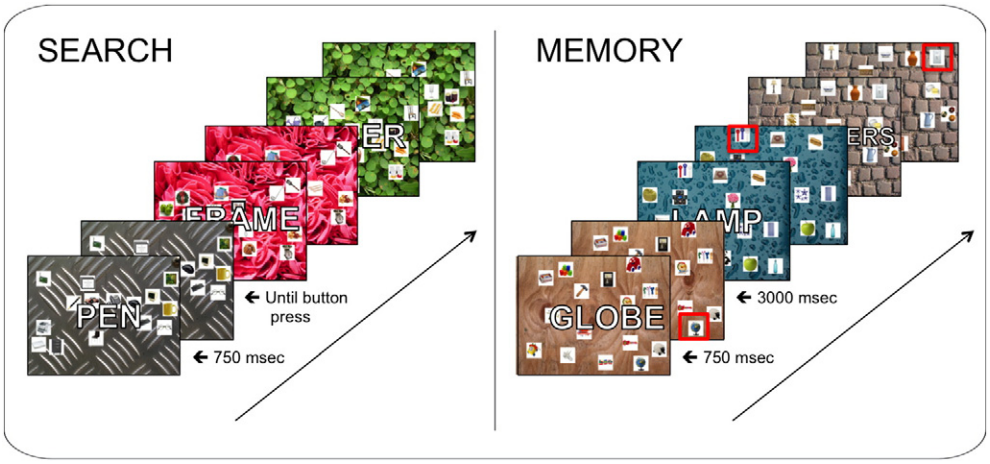

Fig. 2. Methods for Experiments 1 and 2. Note that Experiment 1 used stimuli as depicted in panel C from Fig. 1, and Experiment 2 used stimuli as illustrated in panel D. At the start of a trial, the scene and the target name appeared simultaneously and remained present for 750 ms, after which the target word disappeared. In the search condition, participants indicated with a key press when they found the target, then the next trial started. In the memory condition, the target was framed for 3 s and they tried to memorize it. Scenes were presented in random order.

sheets of paper, one for each of the 5 displays in the block they had just completed, and instructed to draw as much as they could remember from each scene, or if they preferred, to write the names of the targets in their approximate locations. They were given one sheet at a time, and had unlimited time to fill the sheet, at which point they gave it to the experimenter, who handed them the next blank sheet. Participants could draw the scenes in any order, but were told that once they handed a sheet to an experimenter, they could not come back to it. However, they were provided with a place on each sheet to record objects that they remembered but that did not belong to the current scene. Objects correctly named here were included in the count of recalled targets. Participants were also provided with a list of five names, referring to the background in the five images to help them organize their recall. Since the background changed in each experiment, this list differed slightly across experiments: Experiment 1 referred to the texture that made up the background, Experiment 2 referred to the color, and Experiments 3 and 4 referred to the scene category. Overall, the experiment took approximately 45 min.

2.4. Data analysis

Performance on the tasks was evaluated by comparing the mean percentage of targets recalled. We obtained this by counting the number of targets each participant correctly recalled, dividing by the total number of targets in the block, and averaging this value across participants. We counted targets as correctly recalled if they were drawn, if they were correctly named, or if they were incorrectly named, but with enough similarity to the target objects (e.g.: substituting "dish" for "plate").

Raw eye tracking data was parsed into events using SR Research DataViewer. The experiment began with a 9-point calibration procedure, and drift checks were performed every ten trials throughout the experiment. Interest areas were defined in the scenes as the rectangular area that just encompassed the object. Gaze durations on each target were obtained by summing together the durations of all the fixations in that object's interest area throughout the entire block. Targets that were not fixated during the search task were excluded from eye-movement analysis (in Exp1 < 2% & in all further Experiments < 1% of the data).

For analysis of the effect of Task (Search vs. Memorization) on recall performance, linear mixed-effects models were run using the lme4 package (Bates, Maechler, Bolker, & Walker, 2014) in the R statistical programming environment (R Development Core Team, 2015). We chose the LMM approach as it allows between-subject and between-item variance to be estimated simultaneously and thus yields potential advantages over traditional F1\F2 analysis of variance (for a discussion see Baayen, Davidson, & Bates, 2008; Kliegl, Wei, Dambacher, Yan & Zhou, 2010; Kliegl, Masson & Richter, 2010). Where necessary, *p*-values were obtained by likelihood ratio tests of the full model with the effect in question against the model without the effect in question (i.e., model comparisons). The factor Task entered the analyses as a sum contrast (1 vs. −1). Therefore, the intercept estimates the grand mean of the dependent variable and the regression coefficients estimate the difference between factor levels. Task (Search vs. Memorization) and gaze duration were treated as a fixed effect. By including gaze duration as a fixed effect, we were able to consider effects of Task on memory performance beyond those resulting from mere differences in fixation time on each object. Intercepts for subjects and items (scenes), as well as by-subject random slopes for the effect of Task were included in the model as random factors. Following inspection of the distribution and residuals, gaze duration was log-transformed in order to meet LMM assumptions.

3. Experiment 1: testing the role of spatial associations

In the original Draschkow et al. (2014) study, searched objects showed a recall benefit over memorized objects, but only when the task was performed with natural scenes as stimuli. Repeating the experiment with isolated thumbnails of objects, randomly distributed on a non-scene background caused this recall benefit to disappear. One of the sources of information lost in the non-scene displays was the relative location of objects. Castelhano and Heaven (2011) showed that the relative position of objects impacts the speed of search, with targets located in congruent grouping with related objects being found more quickly (see also Malcolm & Henderson, 2010; Võ & Henderson, 2011). If extracting this kind of information during search contributes to stronger memory representations of the target, we should see better recall for searched than memorized objects. In Experiment 1 of the current study, we re-introduced the expected spatial relationships between objects (e.g. keyboard below computer screen) to test their contribution to object recall.

3.1. Stimuli and design

Ten displays were created for this experiment. Each was based on the scene stimuli used in Draschkow et al. (2014), which consisted of full-color pictures of indoor scenes (kitchen, bathroom, bedroom, etc). For the current experiment, we selected fifteen critical objects from the scenes (the same set as in the Draschkow et al. (2014) study), and found isolated image equivalents of each. Thus, if there was a teapot

in the scene, we found an image of a similar teapot on a blank background, without the distortions and occlusions that often accompany the 2D image of a 3D object in a real scene. These images were resized to subtend approximately the same visual angle, and placed on a 256 × 256 pixel white background (See Fig. 1, Panel C). The resulting thumbnails were then placed onto unique non-scene homogenous backgrounds (folds of fabric, droplets of water, a field of clover leaves, etc) to create 10 distinct experimental displays. Ten of these objects would be queried as targets, the other five were distractors. All of the targets from a given scene were kept together during their transfer to the non-scene, and no targets from other scenes were added or substituted. Crucially, the placement of the objects on the new backgrounds was not random: object locations were selected to match their spatial arrangement in the original scene stimuli, preserving expected spatial relationships between objects. In all other respects, the experiment was conducted as described in the General Methods section.

3.2. Results and discussion

In Experiment 1, there was no difference in recall rates between the search and memory conditions, $\beta = 0.056$, SE = 0.105, $z = -0.566$, $p = 0.57$, with 32% of the object recalled following memorization (standard error of the mean, SEM = 2.1) and 28% recalled following search (SEM = 2.2), (see Fig. 3 & Table 1). The LMM showed a marginal effect of gaze durations on memory performance, $\beta = 0.469$, SE = 0.270, $z = 1.738$, $p = 0.08$, with somewhat better recall for objects fixated longer. This in line with findings by Hollingworth and Henderson (2002) showing that objects that receive longer gaze durations were better encoded for a change detection task.

Retaining the expected spatial grouping of objects in Experiment 1 failed to strengthen memory representations of searched objects relative to memorized objects. This suggests that the property of scenes that contributes to the recall benefit cannot be captured simply by the grouping of objects, and that the mechanisms that contribute to the faster processing of congruently grouped objects (e.g. Bar, 2004; Castelhano & Heaven, 2011) do not support better encoding of searched objects. In Experiment 1, the targets had all been resized to the same size, removing cues such as relative size or depth, which are important in determining spatial relationship, and by finding new tokens of the original objects, we removed any information about the orientation of the objects. Furthermore, each individual object was surrounded by a white square, which had the effect of visually sundering the objects from each other, presenting them as separate units, which may interfere

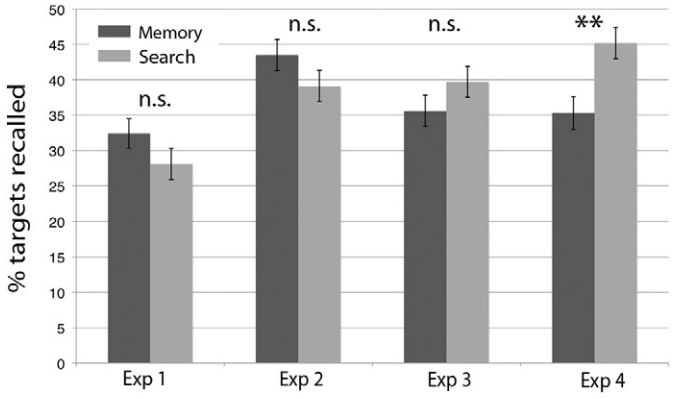

Fig. 3. Mean percentage of targets recalled after each block for each of the four experiments. In Experiment 1, there was no difference in object recall between memorization and search blocks. In Experiment 2, in spite of a significant increase in overall recall from Experiment 1, there was still no difference between the two conditions. In Experiment 3, there was a trend for higher recall rates after a search block. In Experiment 4, recall following the search block was significantly higher than following memorization. Error bars reflect standard errors of the mean.

with the formation of associations between neighboring objects. In Experiment 2, we preserve these image features and test their effect on target recall.

4. Experiment 2: strengthening spatial associations

In Experiment 2, we tested the role of object-to-object spatial associations beyond mere proximity in producing task-dependent differences in recall. In addition to preserving object groupings, stimuli in this experiment incorporated information about the spatial relationships between the objects, such as the relative size, depth and orientation of the targets. Each of these features has been shown to contribute to faster or more accurate object recognition (e.g., Biederman et al., 1982; Green & Hummel, 2006). If observers draw on such information to aid search, it may contribute to stronger representations for searched objects.

4.1. Stimuli and design

Ten new displays were created for this experiment. In order to preserve the location and relative size of the objects, we took the 10 full scenes from Draschkow et al. (2014), and used Adobe Photoshop CS4 to remove all image parts, except the target objects (See Fig. 1, Panel D). The cut-out objects were placed on a lightly textured background of a single solid color. Each scene background had a unique color. Taking the objects directly from the source scene had the effect of preserving not only position and size, but also additional depth and layout cues such as orientation and illumination, providing a much stronger sense of the relative locations of the targets. In all other respects, the experiment was conducted as described in the General Methods section.

4.2. Results and discussion

The presence of size, depth, and orientation information produced an overall boost in target recall in Experiment 2 as compared to Experiment 1: a 2-way between-subjects ANOVA shows a main effect of experiment, $F(1.18) = 6.81$, $p = 0.02$, with an average recall rate of 30% in Experiment 1 and 42% in Experiment 2 (Fig. 3 & Table 1). There was no between-experiment main effect of Task and no interaction. In Experiment 2, there was no difference in target recall rates between the search and memory conditions, $\beta = 0.074$, SE = 0.078, $z = 0.952$, $p = 0.34$. Following memorization, 44% of the objects were recalled (SEM = 2.2), and 39% of the targets were recalled following search (SEM = 2.2). Memory performance was marginally influenced by gaze durations, $\beta = 0.398$, SE = 0.211, $z = 1.887$, $p = 0.06$.

The inclusion of more information about the spatial relationships between objects caused an overall increase in target recall between Experiment 1 and Experiment 2. It seems that in general, targets are easier to remember when the relationships between them become more available. However, refining the spatial relationships between the objects did not lead to differences in recall between the search and memorization conditions. This result suggests that information distributed more globally in the image of the scene might be responsible for the advantage of search over memorization. Such a finding would be in line with results from Võ and Schneider (2010), who showed that search efficiency in rendered scenes could be improved by preceding the search with a brief preview consisting either of the identical scene or the same scene with the target objects removed (i.e. the scenes still contained extensive spatial and semantic information, but no information about the visual forms of the targets). No such facilitation was found when the preview contained only objects on textured backgrounds (analogous to our stimuli in Experiment 2). In Experiments 3 and 4, we explore the role of briefly previewed scene information in causing task-dependent recall.

Table 1
Summary of object recall including results from Draschkow et al. (2014) as well as results from the present experiment. Overall, with an increasing amount of information, the proportion of recalled objects increases. Furthermore, search produces better recall than intentional memorization when the tasks are performed in real world scenes or when non-scene images are supplemented by a flash preview of a complete scene and participants have 2500 ms to integrate target information with scene information (see Võ & Henderson, 2010). This suggests that integrating scene and target information in the first 2500 ms of a search can cause stronger memory representations than intentional memorization.

Experiment	Stimulus	%recalled objects search	% recalled objects memorization	Effect of task?
Draschkow et al. (2014); Exp. 2	Full scenes	44.6	31.6	yes (*t*-test)
Draschkow et al. (2014); Exp. 3	Randomly placed thumbnails of objects on non-scene backgrounds	19.6	24.2	no (*t*-test)
Experiment 1	Thumbnails of objects in their expected spatial relations on non-scene backgrounds	28.1	32.4	no (LMM)
Experiment 2	Objects photoshopped from scenes and placed on uniform background	39.1	43.6	no (LMM)
Experiment 3	Flash preview of full scene, 500 ms integration time, photoshopped scene	39.7	35.6	no (LMM)
Experiment 4	Flash preview of full scene, 2500 ms integration time, photoshopped scene	45.3	35.2	yes (LMM)

5. Experiment 3: integrating gist with target information

Experiments 1 and 2 show that while expected spatial associations between objects can facilitate processing and increase overall recall for target objects; those associations do not contribute to the difference in recall between searched and memorized objects. Another possible source of information that could have boosted recall for searched targets in the study by Draschkow et al. (2014) was information present in the background of the image. Such information contributes to the rapidly accessible gist of the scene, which includes knowledge about the global features of the current scene as well as semantic knowledge about that scene category in general. During search, this information is integrated with knowledge about the target object in order to guide attention to the relevant objects (Torralba et al., 2006; for a review see Wolfe et al., 2011). The structure of the scene may be less relevant and less useful in the memorization conditions, where integration is less beneficial for the task at hand, as finding the object is not necessary. In Experiment 3, we sought to test the role of global gist information in producing a recall benefit for searched objects. We introduced gist information to the incomplete scenes from Experiment 2 by preceding each collection of cut-out objects with a 250 ms flash preview of the complete scene from which the objects had been extracted. Such a design exposes participants briefly to the gist of the scene, while requiring that they perform the task in the same displays as the previous experiment. This allows us to explore how the presence or absence of global scene information changes the object representation formed in response to a given image.

5.1. Stimuli and design

Experiment 3 employed a variant of the flash preview paradigm (e.g. Castelhano & Henderson, 2007; Hillstrom, Scholey, Liversedge, & Benson, 2012; Võ & Henderson, 2010, 2011). The full, unaltered scene served as the preview, and was shown for 250 ms. Following this, the target word appeared for 750 ms against a blank grey background, and then the screen was blank for 500 ms before the experimental display was presented (Fig. 4). This prevented the scene preview from perceptually merging with the incomplete experimental display. The displays in this experiment consisted of the modified scenes used in Experiment 2 (Fig. 1, Panel D). In all other respects, Experiment 3 resembled the two previous experiments.

5.2. Results and discussion

Adding gist information in the form of a flash preview did not meaningfully increase recall relative to the same display without a preview: a 2-way ANOVA comparing recall in Exp2 and Exp3 showed no effects of experiment, $F(1, 18) = 0.41$, $p = 0.05$, no effect of condition, $F(1, 18) = 0.01$, $p = 0.91$, and no interaction ($F(1, 18) = 3.27$, $p = 0.087$.

In Experiment 3, there was no significant difference in recall between the two conditions, $\beta = -0.123$, SE $= 0.086$, $z = -1.431$, $p = 0.15$, although we did see a reversal of the previous trend: mean recall for searched targets (40%, SEM $= 2.2$) was higher than for memorized targets (36%, SEM $= 2.1$) (Fig. 3 & Table 1). Memory performance

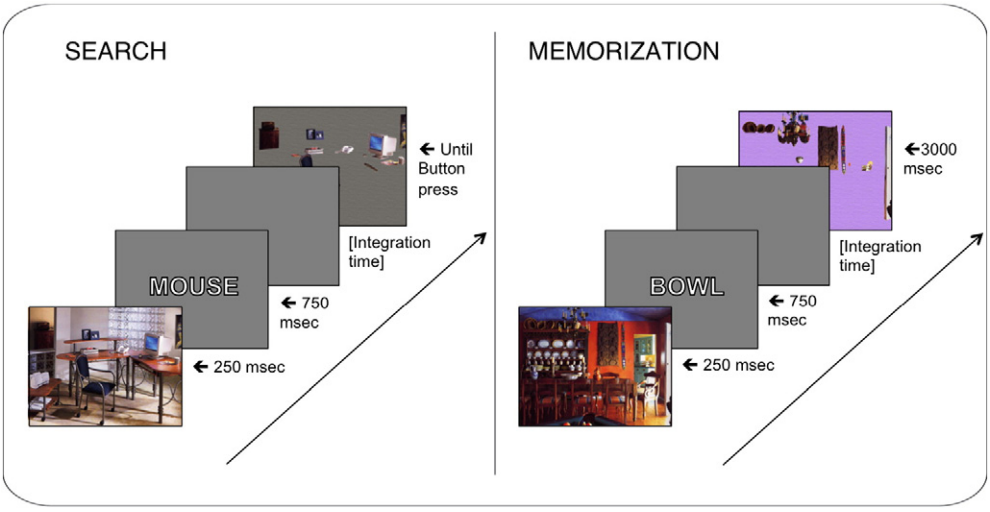

Fig. 4. Methods for Experiments 3 and 4. The only difference between the two experiments was the duration of a blank screen between the presentation of the target word and the experimental display: in Experiment 3, it was present for 500 ms, and for Experiment 4, it was present for 2500 ms. Trials started with a flash preview of a scene, followed by a blank screen with the target name for 750 ms and then a blank screen for 500 ms (Exp.3) or 2500 ms (Exp.4). After this interval, the task was the same as in the previous experiments.

was not influenced by gaze durations, $\beta = 0.244$, SE = 0.247, z = 0.989, $p = 0.32$.

In Experiments 1 and 2, the trend was for better recall in the memory condition, although the effect was non-significant. In this experiment, the trend reverses, with the small, non-significant effect favoring the search condition, suggesting that while scene information may strengthen object recall, the effect under the current circumstances is too weak to reach significance. Previous work has shown that information from the preview may take time to accumulate. Specifically, Võ and Henderson (2010) showed that a very briefly flashed preview (as short as 50 ms) can lead to significantly faster searches, provided that enough time was given to integrate scene gist and target identity information before the search was initiated. This interval between the presentation of the scene preview plus target word, and the initiation of the search (i.e. the integration time) is thought to allow the integration of gist information from the initial scene representation with semantic knowledge about the target in order to compute likely target locations. Such recruitment would contribute to a richer representation of the target object, which could contribute to a stronger memory trace. Thus, in Experiment 4, we extended the integration time.

6. Experiment 4: increasing integration time

In Experiment 3, supplementing non-scene images with global scene information caused a trend for better recall for searched objects than memorized objects, but this result was not significant, possibly because there was insufficient time to integrate this additional information. In Experiment 4, we increased the integration time from 500 ms to 2500 ms in order to boost the integration of target and scene gist information.

6.1. Stimuli and design

With the exception of the increase in integration time from 500 ms to 2500 ms, the stimuli and procedure of Experiment 4 was identical to Experiment 3 (Fig. 4).

6.2. Results and discussion

As can be seen in Fig. 3 and Table 1, Experiment 4 restored the search advantage seen in Draschkow et al. (2014). Memory performance was significantly influenced by Task, with higher recall rates in the search condition (45%, SEM = 2.2) compared to the memory condition (35%, SEM = 2.3), $\beta = -0.250$, SE = 0.075, z = −3.339, $p < 0.01$ (Fig. 3, Table 1). Target gaze durations significantly predicted the recall of items, independent of encoding condition, $\beta = 0.760$, SE = 0.254, z = 2.987, $p < 0.01$.

The results from Experiment 4 indicate that when there is ample time to integrate information from a flash preview with knowledge about the target object, searching for targets in a scene leads to higher recall rates than intentionally memorizing them. This strongly suggests that the recruitment of knowledge about the global and semantic properties of a scene in relation to a target not only leads to quick and efficient searches (Castelhano & Henderson, 2007; Torralba et al., 2006), but also strengthens memory representations for the target object relative to the representation created when no search is required.

7. General discussion

The current series of experiments were designed to investigate the claim made by Draschkow et al. (2014) that searching for objects can lead to better recall than intentionally memorizing them. The authors demonstrated that this effect was only present when the initial encoding involved objects embedded in scenes. Memorization and search did not produce different recall rates when the encoding tasks were performed in non-scene images with target objects of uniform size randomly scattered on a textured background. In order to understand what properties of scenes might be contributing to this scene-specific recall pattern, we gradually added scene information to non-scene displays and tested subsequent recall in four experiments. We included proximity grouping between objects in Experiment 1, but this fairly minimal structure did not boost recall for searched targets. In Experiment 2, we enriched the spatial relationships between objects by adding depth and relative size information. Again, recall following the search and memory tasks did not differ. Recall rates were higher than in Experiment 1, suggesting that more realistic spatial relations do boost memory regardless of task. Experiments 3 and 4 showed that adding a flash preview of the full scene to the non-scene stimulus boosts recall for searched targets over memorized targets, but this was statisticallysignificant only in cases where there was sufficient integration time (2500 ms) between the presentation of the preview plus target name and the appearance of the search display.

Is merely finding an object enough to boost memory? The Attentional Boost Effect (Swallow & Jiang, 2010) shows that recognition memory for a target scene (as measured in an AFC task) is better for scenes that were presented concurrently with a target identification task. This hypothesis holds that the positive detection of a target might open an "attentional gate" that facilitates encoding of all task relevant stimuli present on the screen in that moment, and since our search task required the identification of a target, but the memory task did not, it is possible that this accounts for our results. However, we only found a recall benefit after search when sufficient scene information was present, which is not predicted by the Attentional Boost Effect. Thus, the Attentional Boost Effect alone cannot explain our findings.

Do people simply have better memory for objects that they looked at longer? Work by Hollingworth and Henderson (2002) suggests that objects that are fixated longer are indeed better remembered. By nature, search tasks and memorization tasks require different kinds of processing, which in our study was reflected in occasional differences in gaze durations between the two tasks. However, by including gaze duration as a fixed effect in our linear mixed effects modeling approach, we are able to show a significant effect of Task on recall rates, above and beyond any effect of differences in gaze durations between the two tasks.

Why might a lag after the presentation of scene and target information and before the initiation of the search lead to better recall? We know that search in scenes recruits a rich store of pre-existing knowledge about the scene category (for a review see Wolfe et al., 2011). However, using a flash-preview moving-window paradigm Võ and Henderson (2010) showed that this information is not instantaneously available to guide search to a specific target. In their experiment, participants searched for targets in scenes in a display that was obscured except for a small circle around the center of fixation. Such a display led to longer than normal search times, but searches could be improved by adding a flash preview of the scene. Importantly, information from this preview needed time to accumulate: in the Võ and Henderson (2010) experiment, a 250 ms preview followed by 3000 ms of integration time led to much larger improvements in searches than 500 ms of integration time. They further showed that adding the lag right after the preview but before the target word did not show such benefits. Thus the lag allowed the participant to integrate bottom-up information about the previewed scene with top-down knowledge about the target in order to compute likely target locations. At small lags, such integration is cut short, and existing cognitive structures are recruited to a lesser extent, perhaps leading to weaker memory traces.

If adding a long integration time allowed the whole-scene preview to become an effective prime, might it not do so for an object-only preview (resembling the images from Experiment 2)? Prior data indicate that this would not be the case. Võ and Schneider (2010) showed that previews consisting of the full scene (as in Experiments 3 and 4 of the present paper) significantly improved search times relative to a control condition. A preview of only the structural components, i.e. the layout without objects, of the scene also produced significant improvement.

However, previews consisting only of spatially and functionally aligned objects displayed on a neutral background — similar to the displays used in our Experiment 2 — did not produce improvement even though the ISI from preview onset to search display onset was 2000 ms. Though the task here is somewhat different, it seems likely that preview of isolated objects alone would not produce the search benefit found with whole scene previews. In converging work, searches for targets in scenes, and other tasks like estimating which of two objects is closer, are not effectively primed by images that are merely conceptually identical to the target image (Castelhano & Henderson, 2007; Võ & Schneider, 2010). If there is ample time to integrate bottom-up scene information with top-down knowledge, it is the coherent, global structure of the scene that seems to be crucial in recruiting the scene knowledge that can benefit search, and with it boost target object memory.

The hypothesis that search actively integrates knowledge about the scene with knowledge about the target is in line with the contextual guidance model of search (Torralba et al., 2006), in which bottom-up information about the scene and top-down knowledge about the scene's category, the target parameters and the task requirements interact early in the search process to guide eye movements. The current experiment extends this theory to suggest that this integration can have consequences besides more efficient object search, such as strengthened object representations and better object recall after search completion.

Our results are analogous to findings from the education and cognitive psychology literature that show better recall for test items (usually lists of words or passages from a text) if a single study period is followed by a quiz than if it is followed by the opportunity for more study (LaPorte & Voss, 1975; Roediger & Karpicke, 2006; Zaromb & Roediger, 2010). Our search task could be thought of as the quiz ("Where is the item?"), while memorization is obviously similar to further study. Our results show that searching for objects only boosts recall when sufficient information is present and can be integrated with the target information. This might be considered similar to the finding, in the quiz studies, that the strongest retention effects are observed when quiz questions encourage participants to engage with prior knowledge about the learned items, and form connections between this knowledge and the new material (King, 1994). For instance, retention is increased if the quiz requires more thorough synthesis of the learned information, with short-answer questions leading to better recall than multiple choice questions (Kang, McDermott, & Roediger, 2007). At present, we simply note the analogies between the literature on the role of quizzes in the recall of vocabulary or text content and our work on the role of search in memorization of objects in scenes. It is possible that these similarities are more than skin deep (see Võ & Wolfe, 2013 for a different parallel between processing of scenes and processing of linguistic material) but further research would be required to make that case.

While we have shown an important role for the integration of scene and target information, the current study cannot distinguish whether general scene semantic information (i.e. basic level category) or structural information about the specific scene was most important in this integration, since both were present in the scene preview. However, other studies have shown that while general semantic information does help guide eye movements during search, it is not as effective as specific scene information (Castelhano & Heaven, 2010; Castelhano & Henderson, 2007). In addition, the current design does not indicate whether the computations that occur during the integration time alone are sufficient to cause better recall for searched targets, or whether the time spent searching for the targets is also necessary to cement later target recall.

This paper is not the first to report strong memory representations for searched objects compared to intentionally memorized objects (e.g. Draschkow et al., 2014; Hollingworth, 2012; Võ & Wolfe, 2012), but it is the first to test a possible mechanism for the effect. We propose that the search for targets in scenes integrates scene semantic and scene structure information with knowledge about the target object, which not only benefits search, but also leads to stronger memory representations of objects in naturalistic scenes. These results may advance our understanding of how information extraction and encoding from naturalistic scenes differs between memorization and search tasks.

Acknowledgments

This work was funded by NEI EY017001, and ONR N000141010278 to JMW and Deutsche Forschungsgemeinschaft (DFG) Grant VO 1683/2-1 to MLV. Many thanks to Sage Boettcher and Avi Aizenman for their help in data collection and discussions.

References

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412.
Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience*, *5*(8), 617–629.
Bar, M., & Ullman, S. (1996). Spatial context in recognition. *Perception*, *25*, 343–352.
Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1–7. (URL) http://CRAN.Rproject.org/package= lme4
Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: Detecting and judging objects undergoing relational violations. *Cognitive Psychology*, *14*(2), 143–177.
Brewer, W. F., & Treyens, J. C. (1981). Role of schemata in memory for places. *Cognitive Psychology*, *13*(2), 207–230.
Castelhano, M. S., & Heaven, C. (2010). The relative contribution of scene context and target features to visual search in scenes. *Attention, Perception, & Psychophysics*, *72*(5), 1283–1297.
Castelhano, M. S., & Heaven, C. (2011). Scene context influences without scene gist: Eye movements guided by spatial associations in visual search. *Psychonomic Bulletin & Review*, *18*(5), 890–896.
Castelhano, M. S., & Henderson, J. M. (2005). Incidental visual memory for objects in scenes. *Visual Cognition: Special Issue on Real-World Scene Perception*, *12*, 1017–1040.
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*(4), 753.
Davenport, J. L., & Potter, M. C. (2004). Scene consistency in object and background perception. *Psychological Science*, *15*(8), 559–564.
Draschkow, D., Wolfe, J. M., & Võ, M. L. -H. (2014). Seek and you shall remember: Scene semantics interact with visual search to build better memories. *Journal of Vision*, *14*(8), 10.
Draschkow, D., & Võ, M. L. -H. (2016). Of "what" and "where" in a natural search task: Active object handling supports object location memory beyond the object's identity. *Attention, Perception & Psychophysics*. http://dx.doi.org/10.3758/s13414-016-1111-x.
Ehinger, K. A., 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*(6–7), 945–978.
Green, C., & Hummel, J. H. (2006). Familiar interacting object pairs are perceptually grouped. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(5), 1107.
Greene, M. R., & Oliva, A. (2009). The briefest of glances: The time course of natural scene understanding. *Psychological Science*, *20*(4), 464–472.
Gronau, N., Neta, M., & Bar, M. (2008). Integrated contextual representation for objects' identities and their locations. *Journal of Cognitive Neuroscience*, *20*(3), 371–388.
Gronau, N., & Shachar, M. (2014). Contextual integration of visual objects necessitates attention. *Attention, Perception, & Psychophysics*, *76*(3), 695–714.
Henderson, J. M. (2007). Regarding scenes. *Current Directions in Psychological Science*, *16*(4), 219–222.
Hillstrom, A. P., Scholey, H., Liversedge, S. P., & Benson, V. (2012). The effect of the first glimpse at a scene on eye movements during search. *Psychonomic Bulletin & Review*, *19*, 204–210.
Hollingworth, A. (2006). Scene and position specificity in visual memory for objects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(1), 58.
Hollingworth, A. (2012). Task specificity and the influence of memory on visual search: Comment on Võ and Wolfe (2012). *Journal of Experimental Psychology: Human Perception and Performance*, *38*(6), 1596–1603.
Hollingworth, A., & Henderson, J. M. (2002). Accurate visual memory for previously attended objects in natural scenes. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(1), 113.
Hout, M. C., & Goldinger, S. D. (2010). Learning in repeated visual search. *Attention, Perception, & Psychophysics*, *72*(5), 1267–1282.
Hout, M. C., & Goldinger, S. D. (2012). Incidental learning speeds visual search by lowering response thresholds, not by improving efficiency: Evidence from eye movements. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(1), 90–112.
Howard, C. J., Pharaon, R. G., Körner, C., Smith, A. D., & Gilchrist, I. D. (2011). Visual search in the real world: Evidence for the formation of distractor representations. *Perception*, *40*, 1143–1153.
Kaiser, D., Stein, T., & Peelen, M. V. (2015). Real-world spatial regularities affect visual working memory for objects. *Psychonomic Bulletin & Review*, *22*, 1784–1790.

Kang, S. H. K., McDermott, K. B., & Roediger, H. L. (2007). Test format and corrective feedback modify the effect of testing on long-term retention. *European Journal of Cognitive Psychology*, *19*, 528–558.
King, A. (1994). Guiding knowledge construction in the classroom: Effects of teaching children how to question and how to explain. *American Educational Research Journal*, *31*, 338–368.
Kliegl, R., Masson, M. E., & Richter, E. M. (2010a). A linear mixed model analysis of masked repetition priming. *Visual Cognition*, *18*(5), 655–681.
Kliegl, R., Wei, P., Dambacher, M., Yan, M., & Zhou, X. (2010b). Experimental effects and individual differences in linear mixed models: Estimating the relationship between spatial, object, and attraction effects in visual attention. *Frontiers in Psychology*, *1*.
Laporte, R. E., & Voss, J. F. (1975). Retention of prose materials as a function of postacquisition testing. *Journal of Educational Psychology*, *67*, 259–266.
Malcolm, G. L., & Henderson, J. M. (2010). Combining top-down processes to guide eye movements during real-world scene search. *Journal of Vision*, *10*(2), 4.
Olejarczyk, J. H., Luke, S. G., & Henderson, J. M. (2014). Incidental memory for parts of scenes from eye movements. *Visual Cognition*, *22*(7), 975–995.
Oliva, A. (2005). Gist of the scene. *Neurobiology of attention*, *696*(64), 251–258.
Oliva, A., & Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision*, *42*(3), 145–175.
R Core Team (2015). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing, (URL http://www.R-project.org/).
Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*, 249–255.
Rothkopf, C. A., Ballard, D. H., & Hayhoe, M. M. (2007). Task and context determine where you look. *Journal of Vision*, *7*(14), 16.
Swallow, K. M., & Jiang, Y. V. (2010). The attentional boost effect: Transient increases in attention to one task enhance performance in a second task. *Cognition*, *115*(1), 118–132.
Tatler, B. W., & Tatler, S. L. (2013). The influence of instructions on object memory in a real-world setting. *Journal of Vision*, *13*(2), 5.
Tatler, B. W., Baddeley, R. J., & Vincent, B. T. (2006). The long and the short of it: Spatial statistics at fixation vary with saccade amplitude and task. *Vision Research*, *46*(12), 1857–1862.
Torralba, A., Oliva, A., Castelhano, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. *Psychological Review*, *113*(4), 766.
Võ, M. L. -H., & Henderson, J. M. (2010). The time course of initial scene processing for guidance of eye movements when searching natural scenes. *Journal of Vision*, *10*(3), 14 (1–13).
Võ, M. L. -H., & Henderson, J. M. (2011). Object–scene inconsistencies do not capture gaze: Evidence from the flash-preview moving-window paradigm. *Attention, Perception, & Psychophysics*, *73*(6), 1742–1753.
Võ, M. L. -H., & Schneider, W. X. (2010). A glimpse is not a glimpse: Differential processing of flashed scene previews leads to differential target search benefits. *Visual Cognition*, *18*(2), 171–200.
Võ, M. L. -H., & Wolfe, J. M. (2012). When does repeated search in scenes involve memory? Looking at versus looking for objects in scenes. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(1), 23–41.
Võ, M. L. -H., & Wolfe, J. M. (2013). Differential ERP signatures elicited by semantic and syntactic processing in scenes. *Psychological Science*, *24*(9), 1816–1823. http://dx.doi.org/10.1177/0956797613476955.
Võ, M. L. -H., & Wolfe, J. M. (2015). The role of memory for visual search in scenes. *Annals of the New York Academy of Sciences*, *1339*, 72–81.
Võ, M. L. -H., Schneider, W. X., & Matthias, E. (2008). Transsaccadic scene memory revisited: A 'theory of visual attention (TVA)' based approach to recognition memory and confidence for objects in naturalistic scenes. *Journal of Eye Movement Research*, *2*(2), 7 (1–13).
Wolfe, J. M., Võ, M. L. -H., Evans, K. K., & Greene, M. R. (2011). Visual search in scenes involves selective and non-selective pathways. *Trends in Cognitive Science*, *15*(2), 77–84.
Zaromb, F. M., & Roediger, H. L. (2010). The testing effect in free recall is associated with enhanced organizational processes. *Memory & Cognition*, *38*, 995–1008.