Published in final edited form as:

Cognition. 2013 February ; 126(2): 198–212. doi:10.1016/j.cognition.2012.09.017.

The interplay of episodic and semantic memory in guiding repeated search in scenes

Melissa L.-H. Võ and Jeremy M. Wolfe

Visual Attention Lab, Harvard Medical School, Brigham and Women's Hospital

Abstract

It seems intuitive to think that previous exposure or interaction with an environment should make it easier to search through it and, no doubt, this is true in many real-world situations. However, in a recent study, we demonstrated that previous exposure to a scene does not necessarily speed search within that scene. For instance, when observers performed as many as fifteen searches for different objects in the same, unchanging scene, the speed of search did not decrease much over the course of these multiple searches (Võ & Wolfe, 2012). Only when observers were asked to search for the same object again did search become considerably faster. We argued that our naturalistic scenes provided such strong “semantic” guidance — e.g., knowing that a faucet is usually located near a sink — that guidance by incidental episodic memory — having seen *that* faucet previously — was rendered less useful. Here, we directly manipulated the availability of semantic information provided by a scene. By monitoring observers’ eye movements, we found a tight coupling of semantic and episodic memory guidance: Decreasing the availability of semantic information increases the use of episodic memory to guide search. These findings have broad implications regarding the use of memory during search in general and particularly during search in naturalistic scenes.

Keywords

Repeated and multiple search; Scene perception; Eye movements; Semantic memory; Episodic memory; Search guidance

1. Introduction

We constantly interact with a complex environment that is predictable and variable at the same time. For example, we know that corkscrews generally rest on surfaces. They are usually found in kitchens rather than bedrooms and they most often inhabit the same drawer as the rest of the silverware does. Our knowledge of these regularities can be considered a form of semantic memory. These semantic regularities are probabilistic. Your search for the corkscrew could easily go astray at a friend’s party if one of the guests deposited it in a low probability location. Under those circumstances, other cues would guide search. You might try to retrieve a vague memory of having seen the corkscrew on your friend’s piano or you would go ahead and search for a small, shiny object with a corkscrew-like shape. When do

© 2012 Elsevier B.V. All rights reserved.

Corresponding Author: Melissa Le-Hoa Võ, Visual Attention Lab, Harvard Medical School, 64 Sidney Street, Suite 170, Cambridge, MA 02139, USA, Phone: 001-617-768-8812, mlvo@search.bwh.harvard.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

we rely on guidance by probabilistic, semantic scene knowledge and when might we rely on episodic memory for a specific, previously noted location of that particular object?

In a recent study, we demonstrated that repeatedly searching for multiple different objects in the same, unchanging scene does not dramatically speed search despite the observer's increasing familiarity with the scene (Võ & Wolfe, 2012). Neither previewing nor memorizing each scene for 30 seconds produced marked benefits on subsequent object search. These results seem to run counter to our intuition that increased familiarity should improve the efficiency of search. In that case, we argued that search in our naturalistic scenes was guided by powerful scene semantics and that this strong semantic guidance minimized the usefulness of episodic memory in search guidance. In the present paper, we manipulate the availability of semantic information in a scene in order to investigate the circumstances under which episodic memory will guide search.

1.1. Sources of guidance during search in naturalistic scenes

1.1.1. Feature guidance—From experiments using very simple displays, we know that there is limited set of attributes that can be used to guide search. If you are looking for the large, red, tilted, moving vertical line, you can guide your attention toward the size, color, orientation, and motion of the items in a display. The idea of guidance by a limited set of basic attributes (somewhere between one and two dozen) can be called 'classic guided search' (see Wolfe, 2007; Wolfe & Horowitz, 2004). Search for the corkscrew would be aided by knowledge that it is shiny, has a very distinct shape, and is usually not bigger than your fist. Schmidt and Zelinsky (2009), for example, found that when targets were described using text, more descriptive cues led to faster searches than did less descriptive ones. More precise knowledge about *this* corkscrew's visual features, e.g. via a picture cue, would further speed search (Castelhano & Heaven, 2010; Castelhano, Pollatsek, & Cave, 2008; Malcolm & Henderson, 2009; Vickery, King, & Jiang, 2005). Similarly, Wolfe et al. (2011, Exp.6) showed that when searching for an object twice in the same scene, search benefits were partially driven by learning to associate object specific features with the target word.

1.1.2. Semantic guidance—When targets are embedded in scenes, rather than in random arrays of items, search can draw on the rich information provided by the scene itself, in addition to any feature guidance. Over the course of our lifetime, we have learned to use the regularities encountered in our visual world to aid search. For instance, we learn to associate types of objects, e.g. any kind of toothbrush, to locations in certain types of scenes, e.g. a sink in any kind of bathroom scene. Thus, in addition to guidance by basic features, scenes offer "semantic" guidance, i.e. guidance by the structure and meaning of scenes. Semantic guidance allows drawing on a rich knowledge base — also referred to as sets of scene priors — readily accessible from even short glimpses of a scene (e.g., Castelhano & Henderson, 2007; Droll & Eckstein, 2008; Ehinger, Hidalgo-Sotelo, Torralba, & Oliva, 2009; Hidalgo-Sotelo, Oliva, & Torralba, 2005; Torralba, Oliva, Castelhano, & Henderson, 2006; Võ & Henderson, 2010). Semantic knowledge can be provided by the scene background and by specific, diagnostic objects in the scene. Diagnostic objects are those that by themselves strongly imply a certain scene category and/or the presence of other objects nearby. Thus, a toilet implies a bathroom and a table might imply nearby chairs. Semantic guidance, based on inter-object relationships within a scene, seems to be strong enough to guide search even when the background of a scenes is missing from a search display (see Wolfe et al., 2011, Exp. 5). The scene background provides its own information — like surface structures that objects might rest on — especially, when object-to-object relationships are weak. Thus, unlike a random display of isolated objects, a real scene itself can actually tell you where some objects are more likely to be found.

1.1.3 Episodic memory guidance—Contextual cueing studies have shown that even meaningless “scenes”, in the form of repeated display configurations, can be learned in very short periods of time, with very simple items, and without observers’ explicit awareness that they have been repeatedly exposed to the same target-distractor arrangements (Chun & Jiang, 1998; for a review see Chun & Turk-Browne, 2008). Observed benefits may result from more efficient allocation of attention to subsets of the visual display that most likely contains the target item or, perhaps, from enhanced decision processes (Kunar et al., 2007). Classic contextual cueing paradigms provide evidence that the location of a particular target exemplar can be associated with a particular search array and thus might be taken as rather pure evidence that, even in the absence of semantic guidance, episodic memory for previous exposures to a scene can improve search.

When searching through real-world scenes, however, associations of targets to their context are often more abstract. For instance, we seem to be able to exploit relational contingencies that emerge across different scenes of the same category suggesting that statistical regularities abstracted across a range of stimuli are governed by semantic expectations (Brockmole & Võ, 2010). Further, search in naturalistic scenes seems to be biased to associate target locations to more global rather than local contexts (e.g., Brockmole & Henderson, 2006a; Brockmole, Castelano, & Henderson, 2006; Brooks, Rasmussen, & Hollingworth, 2010). Unlike the usually implicit target-context associations in artificial displays, episodic memory for target-scene associations in real-world scenes tends to be explicit (Brockmole & Henderson, 2006b).

There is ample evidence that we have massive memory for objects (Brady, Konkle, Alvarez, & Oliva 2008; Hollingworth, 2004; Konkle, Brady, Alvarez & Oliva, 2010; Tatler & Melcher, 2007) as well as scenes (Konkle, et al., 2010; Standing, 1973). Previously fixated (and therefore attended) objects embedded in scenes can be retained in visual long-term memory for hours or even days (e.g., Hollingworth, 2004; Hollingworth & Henderson, 2002; Hollingworth, Williams, & Henderson, 2001; for a review see Hollingworth, 2006). Even incidental fixations on objects during search improve subsequent recognition memory (e.g., Castelano & Henderson, 2005; Hout & Goldinger, 2010; Võ, Schneider, & Matthias, 2008; Williams, Henderson, & Zacks, 2005). Moreover, Hollingworth (2009) showed that previewing a scene benefitted subsequent search through it. The effect of a preview increased with longer preview durations, consistent with evidence that observers accumulate visual scene information over the course of scene viewing (Hollingworth, 2004; Hollingworth & Henderson, 2002; Melcher, 2006; Tatler et al., 2003, 2005).

To summarize, search in scenes can, in principle, benefit from a rich set of guiding sources: Feature, semantic, and episodic. Under which circumstances is one source of guidance prioritized over another?

1.2. Repeated Search

In order to locate targets more efficiently, it seems reasonable to maintain memory representations of previously attended items to avoid resampling of already rejected distractors (or their locations). Accordingly, previous work has shown that memory *can* indeed guide search (e.g., Boot et al., 2004; Gilchrist & Harvey, 2000; Hout & Goldinger, 2010, 2012; Howard, Pharaon, Körner, Smith, & Gilchrist, 2011; Klein & MacInnes, 1999; Körner & Gilchrist, 2007, 2008; Kristjansson, 2000; Peterson et al., 2007; Peterson et al., 2001; Solman & Smilek, 2010). Interestingly, however, Wolfe, Klempen, and Dahlen (2000) conducted a series of experiments in which observers repeatedly searched through an unchanging set of letters and found that search efficiency did not change even after hundreds of searches through the same small set of letters. They argued that, while observers clearly had essentially perfect memory for the repeatedly searched display, in this case, access to

that memory was relatively slow, making it more efficient to rely on purely visual search. If it is worthwhile, participants can use memory even with very simple displays of this sort. Thus, in a subsequent experiment, participants searched repeatedly through 18 letters of which only 6 were ever queried. Search became more efficient because the participants learned and remembered which items could possibly be targets. Within that “functional set size” (Neider & Zelinsky, 2008) participants seemed to perform a visual search on each trial, but they could use a form of scene memory to restrict that search to the six relevant items (Kunar, Flusberg, & Wolfe, 2008). Oliva, Wolfe, and Arsenio (2004) found similar results with small sets of realistic objects in scenes. Using a panoramic search display that would only allow observers to search a subpart of the whole scene on any given trial, they found that given a choice, participants appeared to search the display *de novo* rather than relying on memory. However, as in the earlier experiment, participants had quickly realized that only a small set of objects, displayed in the scene, was ever task relevant. This dramatically reduced the functional set size of objects that would need to be searched throughout the experiment (see also Neider and Zelinsky, 2008).

In real scenes, Wolfe, Alvarez, Rosenholtz, and Kuzmova (2011) observed very little improvement in RT across multiple searches for different objects in the same unchanging scene. It is not possible to truly count the “objects” in a real scene (Is that book an “object”? What about the title of the book or the letter “T” in the title?). However, Wolfe et al (2011) did estimate the functional set size and found it to be quite small; certainly far smaller than any count of the actual objects. They argued that on the first search for any of the 15 objects in the scene, semantic guidance was very effectively reducing the functional set size.

Though semantic guidance may have massively reduced the effective set size on the first search, there was still room for improvement in the second search for the same object where RTs were significantly reduced. Wolfe et al. identified two sources of improvement. First, Os remembered that a target word like “apple” was associated with a particular depiction of an apple in the scene, permitting more specific guidance the second time. In addition, episodic memory for the previous search might guide gaze to the location where that specific apple had been found in that specific scene on first exposure. To test the degree to which the second search benefit was merely due to the use of word cues, Wolfe et al. (2011, Exp. 6) replicated their experiment using exact picture cues starting with the first exposure. They found that the exact search target improved performance on the first search but did not eliminate the large advantage of the second search. Thus, the improvement in search times on the second search for the same target is probably a combination of memory for the specific visual features of the target and episodic memory for the location of a previous search target in a specific scene. In Võ and Wolfe (2012), we tracked the eyes during repeated search in scenes. This allowed us to visualize the different forms of guidance described above: the restriction of search to semantically plausible locations on the first search for an object and the further restriction to a subset of those locations on subsequent searches for the same item.

1.3. The rationale for the present study

In Võ and Wolfe (2012), we argued that the abundance of semantic guidance in real-world scenes dramatically reduces the usefulness of other forms of guidance even if the information is demonstrably available. While episodic memory for previous search targets showed great search benefits when the same object had to be found again, we were intrigued by how little guidance seemed to be based on episodic scene representations of incidentally viewed objects (Note: it is possible to see the presence of some episodic guidance as a proverbial glass half-full; see Hollingworth, *in press*). We certainly would not deny that episodic memory representations were incidentally generated during repeated search (see Castelano & Henderson, 2005; Võ et al., 2008) nor that these memory representations can,

in principle, guide search. The hypothesis we test in this paper is that search speed is governed by the most effective and reliable form of guidance available. If the target is the only red item in a field of green, it will matter little that you know where that item was on a previous search (episodic) or where it usually lies in such a scene (semantic). If the scene quickly tells you where the target must be, there will be less use for a more slowly accessed memory of where it was on the last trial. In order to test this proposed relationship between semantic and episodic guidance, we manipulated scenes in such a way that the use of semantic guidance was greatly impeded. We then tracked participants' eye movements while they repeatedly searched the same scenes to see whether search improved with repeated exposure, indicating increased usage of episodic memory to guide search.

2. Methods

2.1. Participants

Overall we tested 60 participants in a between-subject design, fifteen observers per experimental condition (Group 1: Mean Age = 25, SD = 5, 9 female; Group 2: M = 24 SD = 6, 12 female; Group 3: M = 22, SD = 7 female; Group 4: M = 24, SD = 6, 10 female). All were paid volunteers who had given informed consent. Each had at least 20/25 visual acuity and normal color vision as assessed by the Ishihara test.

2.2. Stimulus Material and Experimental Design

Ten full-color, 3D rendered images of real-world scenes were presented across all conditions of this study. An additional image was used for practice trials. Images were created to contain 15 singleton targets, i.e., only one object resembling the target would be present in each scene. Scenes were displayed on a 19-inch computer screen (resolution 1024 × 768 pixel, 100 Hz) subtending visual angles of 37° (horizontal) and 30° (vertical) at a viewing distance of 65 cm.

Across participant groups, four versions of the ten scenes were created by modulating target object placement (semantically consistent vs. inconsistent) and scene background (present vs. absent) (examples of the 4 experimental conditions can be seen in Figures 1a–d): In the “Consistent-With-Background” condition, the image depicted a full scene with the 15 target objects placed at probable locations in the scene (see Figure 1a). The images in the “Inconsistent-With-Background” condition also depicted full scenes, but the 15 target objects were rearranged such that they were placed in highly improbable locations within the scene, e.g. a toilet on top of a washing machine (see Figure 1b). The same scenes *without* scene background were presented in the “Consistent/Inconsistent-Without-Background” conditions. That is, the same objects were shown isolated on a uniform background and either preserved probable object-to-object relations (Figure 1c) or did not (Figure 1d). The average object eccentricity did not differ between consistent and inconsistent placements: Consistent: M = 10.60, SD = 3.68 degree of visual angle; Inconsistent: M = 10.92, SD = 4.15 degree of visual angle; $t(149) < 1$.

As an indicator that memory was or was not guiding search, we analyzed search performance as a function of multiple searches for different objects in the same scene (Epoch 1: searches 1–5 vs. Epoch 2: searches 6–10 vs. Epoch 3: searches 11–15) as well as target repetition (Block 1 vs. Block 2) as within-subject factors. As we will show, the effects of trial number seem to be non-linear, with the largest effects over the first few trials. To capture this, we split the searches into early, middle, and late which, however, should not imply that the epochs have any special status beyond that. The order of object searches was Latin-square randomized such that across 15 participants each of the 150 objects was equally often the target in each of the three search epochs.

2.3. Apparatus

Eye movements were recorded with an EyeLink 1000 desktop mount system (SR Research, Canada) at a sampling rate of 1000 Hz. Viewing was binocular but only the position of the right eye was tracked. Experimental sessions were carried out on a computer running OS Windows XP. Stimulus presentation and response recording were controlled by Experiment Builder (SR, Research, Canada).

2.4. Procedure

Each experiment was preceded by a randomized 9-point calibration and validation procedure. Before each new scene, a drift correction was applied or - if necessary - a recalibration was performed.

In all experiments, participants would search, one after another, for 15 different objects in the same unchanging scene. Each search constitutes a different trial. At the start of each search trial, the target object was defined by a target word presented in the center of the scene (font: Courier, font size: 25, color: white with black borders, duration: 750 ms). Participants were instructed to search for the object as fast as possible and, once found, to press a button of a joystick while fixating the object. This triggered auditory feedback (high-pitch = correct, low-pitch = incorrect). The same scene remained continuously visible for 15 search trials (see Figure 2). It was then replaced by another scene for the next 15 trials, and so on for 150 trials (10 scenes). These 150 trials constitute one block. A drift check (or recalibration if necessary) was performed between each of the 10 different scenes.

In a second block, the same 150 trials were repeated. However, the order of scenes was randomized and the order of targets within a scene was also randomized. Participants were not told in advance that they would search for the same objects again in a subsequent block of the experiment.

The practice trials at the beginning of each experiment were not included in the final analyses. Each experiment lasted for about 20 minutes.

2.5. Data analysis

Raw data were preprocessed using EyeLink Data Viewer (SR Research, Canada). The interest area for each target object was defined by a rectangular box that was large enough to encompass that object. A search was deemed successful when the observers pressed the button while fixating the target's interest area or having fixated the interest area less than 500 ms prior to or after the button press. Unsuccessful searches were regarded as error trials as were trials with response times of more than 10 seconds [in all experiments error rates were less than 12%]. Incidental gaze durations on objects were calculated by summing up the time spent fixating an object's interest area throughout multiple searches up to the point where the critical object became a search target.

A different group of participants was assigned to each of the four experimental conditions. Results were analyzed with an ANOVA that included object position (consistent vs. inconsistent) and background (background present vs. background absent) as between-subject factors. Planned between-subject comparisons were conducted using Welch Two Sample Tests. We were especially interested in the interaction of the scene manipulations and the degree of memory guidance. We therefore included search epoch (Epoch 1: searches 1–5 vs. Epoch 2: searches 6–10 vs. Epoch 3: searches 11–15) and target repetition (Block 1 vs. Block 2) as within-subject factors.

3. Results

In Võ and Wolfe (2012), we claimed that strong semantic guidance during repeated search in scenes minimizes the utility of guidance by episodic memory representations established throughout scene viewing. The main aim of the following analyses is therefore to measure the degree of memory guidance as a function of varying degrees of available semantic guidance. During search, we can distinguish between episodic memory that was acquired incidentally when looking at distractors and episodic memory for a previous search of the same target. As surrogates of incidental episodic memory guidance, we measured 1) RT benefits across multiple searches in Block 1 and 2) correlations between incidental gaze durations on objects and subsequent RTs. Episodic memory guidance for repeated searches of previous targets was measured by 3) RT benefits when the same objects became search targets again in Block 2 compared to Block 1. The effects of our experimental conditions on error rates are reported at the end of this section.

3.1. Gaze distributions as a function of object position and background manipulations

Records of eye movements offer insights to the patterns of search in the different experimental conditions. We first analyzed distractor fixations as a function of object placement and background presence. An ANOVA based on percent fixated distractors per trial with object position and background condition as between-subject factors showed that a higher percentage of distractor objects was fixated when the objects in a scene were placed inconsistently, $F(1,56) = 53.55, p < .01, \eta^2 = .48$ (see Figure 3). While there was only a trend for a main effect of background, $F(1,56) = 3.06, p = .08, \eta^2 = .05$, background and object position interacted significantly, $F(1,56) = 4.92, p < .01, \eta^2 = .08$ presumably because the presence or absence of a scene background only affected guidance for inconsistently placed objects.

To visualize the modulation of gaze distributions as a function of object position and background, Figure 4 shows four exemplar heatmaps based on gaze distributions summed over all participants during the first search for a specific target object – in this case a soap dispenser in a bathroom – in each of the four experimental conditions. We take the degree of scene coverage as an indicator of the strength of guidance: The less coverage, the more guidance. The colors of the heat maps are based on summed gaze durations during search across all participants. The warmer the color, the longer was the scene location looked at. Both the heatmaps and the percent scene coverage values are based on gaze distributions of all participants assuming a 1° visual angle fovea for each fixation.

Together with the analysis of distractor fixations, these exemplar heat maps show a few interesting search characteristics: First, objects whose placement is constrained in a scene produce very restricted, highly guided search (see Figure 4a). In this particular case, gaze is directly guided to the target soap dispenser, which is commonly located near the sink, faucet etc, without the need to look at any other area of the scene. That is, all 15 subjects directly targeted the soap dish. This speaks to a strong role for covert attention preceding overt eye movement since you can't look for a soapdish, next to a sink, until you locate the sink. Interestingly, taking away the scene background (including the sink) did not impede search. Obviously, the preserved object-to-object relations (the soap dispenser near the faucet and the mirror) provided sufficient information to efficiently guide search (see Figure 4c). Second, semantic guidance is much diminished in scenes where objects are not placed in usual locations (e.g. the faucet near the window, the toilet paper near the ceiling), as seen in the tripled increase in the area covered by the eyes in Figure 4b compared to 4a. Indeed, in Figure 3b, we see that in the condition where the whole scene is displayed, gaze actually seems to be misguided to locations that would be probable if this were a consistent scene.

This leads to even a wider range of fixations for inconsistently placed objects in complete scenes as compared to isolated objects (18% in Figure 3b vs. 11% in 3d).

In sum, our manipulations of object positions and scene background seem to have modulated the degree of contextual guidance available during initial searches. The following analyses aim at identifying how these manipulations affected the degree of memory guidance during repeated search.

3.2. Effects of object positions and scene background information on repeated search

Figure 5 shows mean RTs as a function of object position and scene background across blocks and search epochs (mean time to first target fixations are shown in Table 1). To break down this 4-factorial analysis, we start by looking at the effect of object position and scene background on the degree of memory guidance across the first 15 initial searches within a scene restricting analyses to RTs in Block 1.

As can be seen in Table 1, we found similar effects for the “time to first target fixation”, which is measured from trial start to the initial fixation of a target interest area and conveys an additional measure of search efficiency that does not include the time taken to decide whether the fixated object is a target or not.

3.2.1. RTs for object position and background manipulations across epochs in Block 1—

We found a main effect of object position consistency (mean RT consistent 1388 ms vs. inconsistent: 1888 ms, $F(1,56) = 20.29$, $p < .01$, $\eta^2 = .27$). The main effect of the presence of background was also significant ($F(1,56) = 13.23$, $p < .01$, $\eta^2 = .19$). Perhaps counter-intuitively, RTs were shorter on average when the background was absent (1436 ms) than when it was present (1840 ms). This was probably due to the decreased visual complexity of the scene and facilitated figure-ground segmentation when the 15 target objects were presented isolated on a uniform background. Background and consistency factors did not interact, $F < 1$.

Interestingly, even without scene context, the positioning of target objects modulated search behavior. A planned contrast showed that isolated objects that preserved object-to-object relations were found faster (consistent, no background: 1227 ms) than when object-to-object relations were removed by scrambling objects (inconsistent, no background: 1664 ms), $t(87) = 4.73$, $p < .01$. This speaks to a substantial contribution of object-to-object relationships in scene search and warrants further investigation.

In addition to main effects of object consistency and background presence, we observed a main effect of search epoch ($F(2,112) = 30.09$, $p < .01$, $\eta^2 = .35$) with RTs decreasing as a function of epoch. Object consistency and search epoch interacted, $F(2,112) = 3.43$, $p < .05$, $\eta^2 = .07$, reflecting a greater RT decrease across epochs for inconsistent object placement (red lines) compared to consistent placement (green). No other interactions reached significance, object position $\times$ background: $F < 1$; background $\times$ epoch: $F(2,112) = 1.16$, $p = .32$, $\eta^2 = .02$; object position $\times$ background $\times$ epoch: $F < 1$.

In order to follow up on the effects observed for search epochs, we calculated RT differences between Epoch 1 – Epoch 2 as well as Epoch 2 – Epoch 3 and submitted these RT differences to an ANOVA with object position and background as between-subject factors.

As can be seen in Figure 6, the search benefit from Epoch 1 (object searches 1–5 within the same scene) to Epoch 2 (object searches 6–10 within the same scene), was greater for targets located in inconsistent compared to consistent positions, $F(1,56) = 4.91$, $p < .05$.

Background manipulation and interactions did not reach significance, $F(1,56) = 2.18, p = .14$ and $F(1,56) = 1.74, p = .19$, respectively. Planned contrasts showed significant increase in search benefits from Epoch 1 to 2 for the “Inconsistent-With-Background” condition compared to the “Consistent-With-Background” condition, $t(23) = 2.44, p < .05$, while inconsistently placed objects in scenes without background (“Inconsistent-Without-Background” condition), did not show significant search benefits from Epoch 1 to 2 compared to consistent object placements (“Consistent-Without-Background” condition), $t < 1$.

No further increase in search benefits was observed between Epoch 2 and 3 (see Figure 7), i.e., none of the main effects, nor the interaction reached significance, object placement: $F < 1$, background: $F(1,56) = 1.28, p = .26$, object placement $\times$ background: $F < 1$.

In sum, while RTs across all conditions decreased to some degree across the 15 searches, the benefit from searching the same scene multiple times was most pronounced for scenes where search targets were located in inconsistent locations within a complete scene – the situation where one might have expected episodic memory to be of the greatest benefit. Also, search benefits were initially stronger and then became weaker as search for the 15 objects continued.

3.2.2. RTs as a function of incidental gaze durations—Returning to Figure 3, suppose that an observer fixated the trashcan on the way to finding the soap dispenser. Would some memory for that fixation reduce the RT in the search for the trashcan on a later trial when the trashcan becomes the target? If so, there should be a negative correlation between the cumulative time observers spent incidentally fixating a distractor object and the RT measured once that object becomes a target. We plotted the RTs of every target search by every observer against the sum of all fixation durations on each object prior to it becoming the search target. Five data points with gaze durations greater than 10 s were excluded. The results are seen in Figure 8.

The correlation is significant only when items are placed in inconsistent locations on a visible background (“Consistent-With-Background” condition: $r = -.08, p < .001$, Fig. 8a). The other three correlations did not reach significance (all $ps > 0.1$). This suggests that episodic memory for fixated distractors is most useful when the guidance by consistent semantic scene information is not available but where the scene is present, perhaps to give a better spatial anchor to the memory for the inconsistent distractor positions.

3.2.3. RT benefits from Block 1 to Block 2—In order to investigate the degree to which episodic memory for search targets is able to guide a second search for the same target after many intervening trials, we next analyzed search benefits from Block 1 to Block 2 with greater search benefits indicating greater use of memory guidance in Block 2.

Going back to Figure 4, it is obvious that the second search for an object in a scene is substantially faster than the first (avg RTs: 1017 vs. 1638 ms, $F(1,56) = 354, p < .01, p\eta^2 = .86$). We also found main effects of object position, $F(1,56) = 14.12, p < .01, p\eta^2 = .20$, and background presence, $F(1,56) = 14.26, p < .01, p\eta^2 = .20$. Further, target repetition interacted with object position, $F(1,56) = 20.24, p < .01, p\eta^2 = .27$ and as three-way interaction with both object position and background, $F(1,56) = 4.38, p < .05, p\eta^2 = .07$.

As can be seen in Figure 9, the decrease in RT from Block 1 to Block 2 was greater for inconsistently placed objects than for consistently placed objects, $F(1,56) = 20.24, p < .01$. There was no main effect of the background manipulation, $F(1,56) = 2.33, p = .13$, but

Background and Consistency interacted significantly, $F(1,56) = 4.38, p < .05$, because the benefit was largest for the inconsistent, background present condition.

All conditions showed RT benefits from Block 1 to Block 2 in that they all differed significantly from 0. As in our earlier work (Võ & Wolfe, 2012), we believe that finding a search target within a scene creates strong binding between the object and the scene it is display within. However, the largest increase in search benefits was seen for objects positioned in inconsistent locations within complete scenes ("Inconsistent-With-Background condition" = 889 ms), with nearly double the benefit seen for the same scene with consistent object placements ("Consistent-With-Background condition" = 455 ms), $t(67) = 5.93, p < .01$. Search benefits from Block 1 to Block 2 were also greater for inconsistent vs. consistent objects in scenes that lacked scene background, $t(86) = 2.47, p < .05$, which again implies that preserved object-to-object relations play an important role in guiding search even when the objects are presented without scene background.

Figure 10 shows gaze distributions during search for the soap dispenser in Block 2. Little can change in the consistent conditions, where semantic guidance already directed the eyes to the target. In the inconsistent conditions, especially with the background present, episodic memory could and did reduce the search space from Block 1 to Block 2 (compare with Fig. 3).

In line with RT data, we found greater RT benefits across blocks for scenes with inconsistently placed objects and especially those where the scene background was present. In these seemingly normal scenes, semantic expectations might have been strong, but were substantially violated by the inconsistent placement of target objects. Increased search benefits from the first search of an object to its second search after many intervening trials imply that memory guidance is strongest in scenes that provide not only little, but rather detrimental semantic guidance.

3.2.4. Error Rates—Error rates included trials in which the target was not was fixated during the key response as well as trials in which the target was not found within 10 seconds. In an ANOVA on the error rates, we found a main effect of object consistency, $F(1,56) = 8.21, p < .01, \eta^2 = .13$, showing that, averaged across blocks, inconsistent objects produced modestly more errors than consistent ones (10% vs. 7%). Target repetition also significantly reduced error rates from 10% in Block 1 to 6% in Block 2, $F(2,112) = 50.50, p < .01, \eta^2 = .47$. In addition, object consistency and Block interacted significantly, $F(2,112) = 5.85, p < .05, \eta^2 = .09$. The background manipulation did not reach significance, $F < 1$, and it did not interact with object position, $F(2,56) = 2.18, p = .14, \eta^2 = .03$. No other interactions reached significance, all $F_s < 1$.

3.2.5. Object-specific context effects—The degree to which an object is contextually constrained plays a major role in applying scene knowledge. For instance, a toothbrush is highly constrained to the vicinity of a sink and most probably found in a bathroom, while a book can pretty much rest on any horizontal surface and is not strongly bound to a certain scene category. These inter-object differences in semantic guidance might provide extra evidence that strongly constrained objects rely less on episodic memory during repeated search.

To test this hypothesis, we conducted an additional experiment in which a new group of 15 subjects was presented with background-only versions of all ten scenes in which the 15 objects were missing. Upon presentation of a target word (750ms) they were asked to indicate with a mouse click where they thought this target *ought to be* if it were present. The rationale was that more contextually constrained objects would yield a tighter cluster of

suggested target positions than less contextually constrained objects. In addition, for highly constrained objects the difference between the actual location in the search scene and the averaged click location should be smallest. These two measures of contextual constraint were highly correlated ($r=.73$). Using this latter measure, we categorized the top 25 objects as highly constrained and the bottom 25 objects as least constrained. We then performed the same analyses, discussed above, on these two subsets of data. When comparing the strongly constrained subset of objects in Figure 11a with the weakly constrained objects in Figure 11b, one can first of all see that search times in block 1 are higher for weak than for strongly constrained objects (weak: 1185ms vs. strong: 1161ms), which implies that our categorization reflects degrees of search guidance. In addition, the difference between consistent and inconsistent placements of strongly constrained objects (1303ms) is much more pronounced than the consistent-inconsistent difference for weakly constrained objects (157ms).

As a measure of the degree of episodic memory guidance, we submitted RT benefits from block 1 to block 2 to an ANOVA with strongly vs. weakly constrained as within-subject factor and object placement as between-subject factor. We found main effects of both object placement, $F(1,56) = 26.96, p < .01, \eta^2 = .23$, and constraint, $F(1,56) = 8.34, p < .01, \eta^2 = .05$, and an interaction, $F(1,56) = 7.53, p < .01, \eta^2 = .05$. As can be seen in Figure 12, strongly constrained objects showed little search benefits from block 1 to block 2 (223ms) when positioned correctly, while search benefits almost tripled for weakly constrained objects (630ms). Further, inconsistent placement of strongly constrained objects resulted in greater search benefits from block 1 to block 2 (incon: 1091ms), while object placement had much less of an effect for weakly constrained objects (con: 639 ms vs. incon: 858ms).

4. Discussion

It seems self-evident that memory plays a role in search tasks. Memory might operate within a search by preventing you from revisiting rejected distractors or it might operate in various ways across repeated searches, as in the tasks described in this paper. In all of these cases, accessing that memory takes some amount of time and there are 29 circumstances where targets appear to be found by de novo visual search before memory has a chance to have much of an effect. Horowitz and Wolfe (1998) found that, in random arrays of letters, there was no difference in search efficiency between dynamic displays in which all distractors were randomly replotted every 100 ms and standard, static displays. This suggests that Os were sampling the display with replacement in both cases, since rejected distractors could not be marked in the dynamic displays. Across searches, Wolfe, Klempe, and Dahlen (2000) reported no improvement in search efficiency when Os searched repeatedly through the same display of letters. These are cases where the speed of what we might call amnesic search is such that memory simply has no chance to play a major role. This should not be taken to mean that memory does not have a role at all. Under the right circumstances, memory can discourage revisitation of previously attended distractors (e.g., Boot et al., 2004; Klein & MacInnes, 1999; Kristjansson, 2000; Peterson et al., 2007; Peterson et al., 2001; Takeda & Yagi, 2000). Episodic memory can also guide search more efficiently to previously fixated targets — though with limited capacity (e.g., Gilchrist & Harvey, 2000; Körner & Gilchrist, 2007, 2008).

Sometimes, memory for stored scene representations can guide subsequent search (e.g., Castelano & Henderson, 2007; Hollingworth, 2009; Võ & Henderson, 2010; Võ, Schneider, & Matthias, 2008), and, as discussed earlier, memory for targets previously found in scenes can drastically reduce search times when the same object in the same scene is searched for again (e.g., Võ & Wolfe, 2012; Wolfe et al., 2011). At other times, as in studies of multiple searches for different objects, memory may not play a major role in

search even though it is clearly available (e.g., Kunar et al., 2008; Oliva, Wolfe, & Arsenio, 2004; Võ & Wolfe, 2012; Wolfe et al., 2000, 2011).

We argue that the available forms of guidance race against each other for the ability to shape the deployment of attention. These forms of guidance would include semantic memory, episodic memory, and target templates (search images). Even if all are available, the fastest, strongest signal can dominate and hide the contribution of other forms of guidance. Thus, when some form of memory appears not to have a role, it need not be because relevant memories do not exist. Rather, we suspect that the search was completed before that memory had an opportunity to influence search time. In support of this notion, Oliva and colleagues (2004) showed that they could encourage the use of memory guidance, but given a choice between using memory and simply searching again, observers seemed to choose visual search. Based on the paradigm used here, we would not deny that incidental episodic memory for distractors exists, only that this form of memory loses to other forms of guidance, in this case semantic memory.

The aim of the present study, therefore, was not to argue *for* or *against* the role of memory per se but rather to gain a better understanding under which circumstances different forms of memory play a role in guiding search.

4.1. The interplay of episodic and semantic memory guidance in repeated scene search

There are at least two sorts of memory at play in visual search through repeated scenes. Semantic memory, built up over long experience, tells us that, in general, pictures hang on walls and toasters are found on kitchen counters. Episodic memory, built up in those cases where one remembers the episode during which a certain object-scene relation was encoded, tells us that *this* toaster was at *this* location in *this* scene. We can further distinguish between episodic memory that was acquired incidentally when looking at distractors and episodic memory for a previous search of the same target. Both forms of episodic memory are affected by the degree of semantic guidance provided by the scene. In Võ and Wolfe (2012), we proposed that there was only a small role for episodic memory guidance in multiple searches through real-world scenes because semantic memory was so abundant and effective (for reviews see Henderson & Ferreira, 2003; Wolfe, Võ, Evans, & Greene, 2011).

However, our visual world is not always as orderly as we might want it to be. Some scenes provide more semantic guidance than others. For example, compare searching for a pen on your office desk to searching for a knife on your dinner table. In the latter case, the semantics of knives on dinner tables are sufficiently constrained that there may be no added benefit from accessing episodic memory representations to retrieve where you last saw *this* knife on *this* table. In contrast, while your pen may be semantically constrained to most likely lie on your desk somewhere, it might be useful to recover the episodic trace that specifies where exactly the pen last fell from your hand.

In this study, we manipulated the semantics of the scene in order to test the hypothesis that episodic memory guidance — stemming both from incidental distractor fixations and previous target search episodes — gains importance when semantic memory guidance becomes weak or misleading. We employed two ways of modulating semantic guidance: 1) Rearranging objects such that they are placed in improbable locations, and 2) taking away the scene background information. We found that inconsistent object placement impeded search as seen in longer RTs and greater coverage of the visual display by the eyes (see Figures 3a and 3b). This shows that our manipulation of scene semantics worked, as desired. Moving on to the central questions of this paper, we found increased guidance by episodic memory when objects were inconsistently placed within a scene. This was indicated by larger RT benefits in repeated search for different objects within the same scene and also when objects were searched for a second time in a later block of trials. Further, the same

pattern of results was seen in the time to first fixation (see Table 1) — an eye movement measure that is often used to specifically indicate guidance to targets — which further supports the notion that inconsistent object placement directly affected search guidance. In addition, longer incidental gaze durations on distractors were associated with shorter RTs when a distractor became a target, but only in the condition where objects were placed at inconsistent locations in a scene.

Moreover, we found that even in scenes with generally consistent object placements, semantic guidance can differ greatly from object to object. We found that highly constrained objects, e.g. toothbrushes, show less episodic memory guidance than objects that are less constrained, e.g. cups. These results are consistent with the hypothesis that episodic memory becomes more useful as a source of guidance when semantic memory becomes less useful.

Interestingly, the absence of a scene's background produced shorter search times. The lack of scene information was obviously counteracted by reduced visual clutter, easier figure-ground segmentation, and, one imagines, a smaller set size. The fact that the consistent-inconsistent difference remained — even when the background was removed — highlights the importance of object-to-object relations for semantic guidance. Thus, the spatial arrangement of individual objects is an integral part of semantic guidance in scenes. Accordingly, taking away the scene background only led to increased memory-based search — as seen in increased search benefits from Block 1 to Block 2 — when object-to-object relations were eliminated.

4.2. Weighing the costs and benefits of memory guidance in repeated visual search

The general rationale is that memory guidance during search gains importance especially when other sources of guidance are either unavailable (e.g. the lack of semantic guidance in simple letter displays), too time or energy consuming (e.g. requiring large eye movements), or simply detrimental (e.g. when targets are misplaced). Imagine that the pen that you are looking for is the only bright red item on your desk. Your search will be guided by the feature “red”, not your episodic memory of where you had put it last time you used it. Similarly, Gibson, Li, Skow, Brown, and Cooke (2000), who used a multiple target search task, concluded that while memory guides search in some scene contexts, its utilization depends on the relative costs and benefits involved.

Time and space are important factors in the use of episodic memory in search. In arrays of randomly placed elements (hence, displays with no role for semantic memory), the strongest evidence *for* memory guidance during repeated search has come from studies using search displays with spread-out, isolated objects on uniform backgrounds that not only lack semantic guidance, but also necessitate visual orienting actions like eye movements (e.g., Hout & Goldinger, 2010; Körner & Gilchrist, 2007, 2008; Solman & Smilek, 2010; Takeda & Yagi, 2000) or even hand movements (Howard et al., 2011). In these cases, the “costs of searching”, i.e. moving your eyes and/or hands, again would be great enough to render episodic memory a useful source of guidance. Evidence *against* a strong role of memory during repeated search, on the other hand, has mainly come from studies that used very simple search displays that did not require eye movements (e.g., Kunar et al. 2008; Wolfe et al., 2000; Wolfe et al, 2002). In the latter cases, the visual displays were easily searchable without the need for eye movements. In Kunar et al. (Exp.5, 2008), letters presented at further eccentricity were even increased in size to compensate for the loss of visual acuity. In search displays like these, the cost of memorizing the display might outweigh its benefits, since it may be faster and less error prone to simply search again rather than to retrieve and reactivate stored memory traces. In line with this reasoning, Solman and Smilek (2010) showed that memory is preferentially devoted to those items least accessible to perceptual search.

Oliva and colleagues (2004) asked whether visual search had a mandatory priority over memory search. They used a variation of the repeated search task in which participants subsequently searched parts of a scene that was larger than the current field of view (a “panoramic search display”), which should have biased participants to engage in memory guided search. Participants indeed used memory to guide search in that they were able to restrict their search to only the few task relevant objects. This was possible, since the number of possible targets in their repeated search study was very limited (14 compared to 150 objects in our study). The rest of the objects depicted in the scenes were never targets and could therefore be readily ignored. However, they did not bother to access precise memory for each object location. Instead, they searched over and over through the correctly remembered set of task relevant objects.

In the study presented here, when confronted with search displays containing inconsistently placed objects, guidance from other sources was sufficiently limited that observers were induced to use episodic memory representations to guide repeated search. Not only was semantic guidance weak, target objects were scattered across the whole image and participants were explicitly told to move their eyes to the target location before making a key response. Searching the whole display over and over again would have been too costly. This adds to the evidence that the degree to which we use memory to guide search is determined by calculations of the costs and benefits of using memory to guide search.

Finally, we might ask why we found modest reductions in RT over 15 repeated searches for different items in the same scene while Võ and Wolfe (2012) did not. It may be relevant that the two studies used different materials. The present study used 3D rendered scenes to allow us to move objects at will, as did Hollingworth (in press). The Võ and Wolfe study, however, used photographs. Rendered scenes, while becoming more and more realistic, are still created artificially in that every object has to be intentionally placed to make up a scene, while photographs are based on preexisting real-world environments. This might alter semantic guidance in ways we yet have to understand.

5. Conclusion

Because we constantly have to locate and interact with objects in our environment, we have evolved to become highly flexible searchers. The results presented here, demonstrate how the nature of the search environment differentially biases search towards using semantic or episodic guidance. We conclude that the degree to which any given visual search will be guided by memory mainly depends on the costs and benefits for the use of one source of guidance over others.

Acknowledgments

This work was supported by grants to MV (DFG: VO 1683/1-1) and JMW (NEI EY017001, ONR N000141010278). We also want to thank Corbin Cunningham, Erica Kreindel, and Ashley Sherman for assistance in data acquisition.

References

- Boot WR, McCarley JM, Kramer AF, Peterson MS. Automatic and intentional memory processes in visual search. *Psychonomic Bulletin & Review*. 2004; 5:854–861. [PubMed: 15732694]
- Brady TF, Konkle T, Alvarez GA, Oliva A. Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Science U S A*. 2008; 105(38):14325–14329.

- Brockmole JR, Henderson JM. Recognition and attention guidance during contextual cuing in real-world scenes: Evidence from eye movements. *Quarterly Journal of Experimental Psychology*. 2006a; 59:1177–1187.
- Brockmole JR, Henderson JM. Using real-world scenes as contextual cues for search. *Visual Cognition*. 2006b; 13:99–108.
- Brockmole JR, Võ MLH. Semantic memory for contextual regularities within and across scene categories: Evidence from eye movements. *Attention, Perception & Psychophysics*. 2010; 72(7): 1803–1813.
- Brockmole JR, Castelhana MS, Henderson JM. Contextual Cueing in Naturalistic Scenes: Global and Local Contexts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2006; 32(4):699–706.
- Brooks DI, Rasmussen IP, Hollingworth A. The nesting of search contexts within natural scenes: Evidence from contextual cuing. *Journal of Experimental Psychology: Human Perception and Performance*. 2010; 36(6):1406–1418. [PubMed: 20731525]
- Castelhana MS, Heaven C. The relative contribution of scene context and target features to visual search in scenes. *Attention, Perception & Psychophysics*. 2010; 72(5):1283–1297.
- Castelhana M, Henderson JM. Incidental visual memory for objects in scenes. *Visual Cognition*. 2005; 12:1017–1040.
- Castelhana MS, Pollatsek A, Cave KR. Typicality aids search for an unspecified target, but only in identification and not in attentional guidance. *Psychonomic Bulletin & Review*. 2008; 15:795–801. [PubMed: 18792506]
- Chun MM, Jiang Y. Contextual cuing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology*. 1998; 36:28–71. [PubMed: 9679076]
- Chun, MM.; Turk-Browne, NB. Associative learning mechanisms in vision. In: Luck, SJ.; Hollingworth, AR., editors. *Visual memory*. Oxford: Oxford University Press; 2008. p. 209-245.
- Foulsham T, Underwood G. How does the purpose of inspection influence the potency of visual salience in scene perception? *Perception*. 2007; 36:1123–1138. [PubMed: 17972478]
- Gibson BS, Li L, Skow E, Brown K, Cooke L. Searching for one versus two identical targets: When visual search has a memory. *Psychological Science*. 2000; 11(4):324–327. [PubMed: 11273393]
- Gilchrist ID, Harvey M. Refixation frequency and memory mechanisms in visual search. *Current Biology*. 2000; 10:1209–1212. [PubMed: 11050390]
- Henderson, J.; Ferreira, F. Scene perception for psycholinguists. The interface of language. In: Henderson, JM.; Ferreira, F., editors. *The interface of language, vision, and action: Eye movements and the visual world*. NY: Psychology Press; 2003. p. 1-58.2004
- Henderson JM, Malcolm GL, Schandl C. Searching in the dark: Cognitive relevance drives attention in real-world scenes. *Psychonomic Bulletin & Review*. 2009; 16:850–856. [PubMed: 19815788]
- Henderson, JM.; Brockmole, JR.; Castelhana, MS.; Mack, M. Visual saliency does not account for eye movements during visual search in real-world scenes. In: van Gompel, R.; Fischer, M.; Murray, W.; Hill, R., editors. *Eye movements: A window on mind and brain*. Oxford: Elsevier; 2007. p. 537-562.
- Hollingworth A. Task Specificity and the Influence of Memory on Visual Search: Commentary on Võ and Wolfe. *Journal of Experimental Psychology: Human Perception and Performance*. 2012 in press.
- Hollingworth A. Two forms of scene memory guide visual search: Memory for scene context and memory for the binding of target object to scene location. *Visual Cognition*. 2009; 17:273–291.
- Hollingworth A. Visual memory for natural scenes: Evidence from change detection and visual search. *Visual Cognition*. 2006; 14:781–807.
- Hollingworth A. Constructing visual representations of natural scenes: The roles of short- and long-term visual memory. *Journal of Experimental Psychology: Human Perception and Performance*. 2004; 30:519–537. [PubMed: 15161384]
- Hollingworth A, Henderson JM. Accurate visual memory for previously attended objects in natural scenes. *Journal of Experimental Psychology: Human Perception and Performance*. 2002; 28:113–136.

- Hollingworth A, Williams CC, Henderson JM. To see and remember: Visually specific information is retained in memory from previously attended objects in natural scenes. *Psychonomic Bulletin & Review*. 2001; 8:761–768. [PubMed: 11848597]
- Horowitz TS, Wolfe JM. Search for multiple targets: Remember the targets, forget the search. *Perception and Psychophysics*. 2001; 63(2):272–285. [PubMed: 11281102]
- Horowitz TS, Wolfe JM. Visual search has no memory. *Nature*. 1998 Aug; 394(6):575–577. [PubMed: 9707117]
- Hout MC, Goldinger SD. 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*. 2012; 38:90–112. [PubMed: 21574743]
- Hout MC, Goldinger SD. Learning in repeated visual search. *Attention, Perception & Psychophysics*. 2010; 72:1267–1282.
- Howard CJ, Pharaon RG, Körner C, Smith AD, Gilchrist ID. Visual search in the real world: Evidence for the formation of distractor representations. *Perception*. 2011; 40:1143–1153. [PubMed: 22308885]
- Klein RM, MacInnes WJ. Inhibition of return is a foraging facilitator in visual search. *Psychological Science*. 1999; 10(4):346–352.
- Konkle T, Brady TF, Alvarez GA, Oliva A. Scene memory is more detailed than you think: the role of scene categories in visual long-term memory. *Psychological Science*. 2010; 21 (11):1551–1556. [PubMed: 20921574]
- Körner C, Gilchrist ID. Memory processes in multiple-target visual search. *Psychological Research*. 2008; 72:99–105. [PubMed: 17021837]
- Körner C, Gilchrist ID. Finding a new target in an old display: Evidence for a memory recency effect in visual search. *Psychonomic Bulletin & Review*. 2007; 14 (5):846–851. [PubMed: 18087948]
- Kristjansson A. In search of remembrance: Evidence for memory in visual search. *Psychological Science*. 2000; 11:328–332. [PubMed: 11273394]
- Kunar MA, Flusberg SJ, Wolfe JM. The role of memory and restricted context in repeated visual search. *Perception and Psychophysics*. 2008; 70(2):314–328. [PubMed: 18372752]
- Kunar MA, Flusberg SJ, Horowitz TS, Wolfe JM. Does Contextual Cueing Guide the Deployment of Attention? *J Exp Psychol Hum Percept Perform*. 2007; 33(4):816–828. [PubMed: 17683230]
- Malcolm GL, Henderson JM. The effects of target template specificity on visual search in real-world scenes: Evidence from eye movements. *Journal of Vision*. 2009; 9(11):Art. 8.
- Melcher D. Accumulation and persistence of memory for natural scenes. *Journal of Vision*. 2006; 6(1): 8–17. [PubMed: 16489855]
- Neider MB, Zelinsky GJ. Exploring set size effects in scenes: Identifying the objects of search. *Visual Cognition*. 2008; 16:1–10.
- O'Regan KJ. Solving the “real” mysteries of visual perception: The world as an outside memory. *Canadian Journal of Psychology*. 1992; 46:461–488. [PubMed: 1486554]
- Peterson MS, Beck MR, Vomela M. Visual search is guided by prospective and retrospective memory. *Perception & Psychophysics*. 2007; 69 (1):123–135. [PubMed: 17515222]
- Peterson MS, Kramer AF, Wang RF, Irwin DE, McCarley JS. Visual search has memory. *Psychological Science*. 2001; 12:287–292. [PubMed: 11476094]
- Schmidt J, Zelinsky GJ. Search guidance is proportional to the categorical specificity of a target cue. *Quarterly Journal of Experimental Psychology*. 2009; 62:1904–1914.
- Solman GJF, Smilek D. Item-specific location memory in visual search. *Vision Research*. 2010; 50:2430–2438. [PubMed: 20837043]
- Standing L. Learning 10,000 pictures. *Quarterly Journal of Experimental Psychology*. 1973; 25:207–222. [PubMed: 4515818]
- Takeda Y, Yagi A. Inhibitory tagging in visual search can be found if search stimuli remain visible. *Perception & Psychophysics*. 2000; 62(5):927–934. [PubMed: 10997039]
- Tatler BW, Melcher D. Pictures in mind: Initial encoding of object properties varies with the realism of the scene stimulus. *Perception*. 2007; 36:1715–1729. [PubMed: 18283923]

- Tatler BW, Gilchrist ID, Land MF. Visual memory for objects in natural scenes: From fixations to object files. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*. 2005; 58A(5):931–960.
- Tatler BW, Gilchrist ID, Rustad J. The time course of abstract visual representation. *Perception*. 2003; 32(5):579–592. [PubMed: 12854644]
- Torralba A, Oliva A, Castelano MS, Henderson JM. Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. *Psychological Review*. 2006; 113:766–786. [PubMed: 17014302]
- Vickery TJ, King L-W, Jiang Y. Setting up the target template in visual search. *Journal of Vision*. 2005; 5(1):Art. 8.
- Võ ML-H, Henderson JM. The Time Course of Initial Scene Processing for Guidance of Eye Movements When Searching Natural Scenes. *Journal of Vision*. 2010; 10(3):14, 1–13. [PubMed: 20377291]
- Võ MLH, Wolfe JM. When does repeated search in scenes involve memory? Looking at versus looking for objects in scenes. *Journal of Experimental Psychology: Human Perception and Performance*. 2012; 38(1):23–41. [PubMed: 21688939]
- Võ ML-H, Schneider WX, Matthias E. 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*. 2008; 2(2):7, 1–13.
- Williams CC, Henderson JM, Zacks RT. Incidental visual memory for targets and distractors in visual search. *Perception & Psychophysics*. 2005; 67:816–827. [PubMed: 16334054]
- Wolfe JM, Klempen N, Dahlen K. Post-attentive vision. *Journal of Experimental Psychology: Human Perception & Performance*. 2000; 26(2):693–716. [PubMed: 10811170]
- Wolfe JM, Alvarez GA, Rosenholtz RE, Kuzmova YI. Visual search for arbitrary objects in real scenes. *Attention, Perception, and Psychophysics*. 2011; 73:1650–1671.
- Wolfe, JM. Guided Search 4.0: current Progress with a model of visual search. In: Gray, W., editor. *Integrated Models of Cognitive Systems*. Oxford University Press; 2007. p. 99-119.
- Wolfe JM, Horowitz TS. What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience*. 2004; 5:495–501.

Highlights

- Search in scenes can be guided by its “semantic” properties or “episodic” memory.
- We directly manipulated the availability of semantic information in scenes.
- Eye movements indicated the contributions of episodic and semantic guidance.
- Episodic guidance increased when the scene’s semantic information decreased.
- We argue that observers tend to use the most reliable information to speed search.

Figure 1.

Example scenes as a function of object position (green = consistent, red = inconsistent) and background manipulations (solid lines = background present, dotted lines = background absent): a) Normal depiction of a scene with background and objects placed in probable locations, b) Complete scene with misplaced objects, c) and d) show scenes a) and b) without background.

Figure 2.
Target sequence for all participant groups.

Figure 3. Mean percentage of distractor fixations per search as a function object position (green = consistent, red = inconsistent) and background manipulations (solid lines = background present, dotted lines = background absent) [bars depict standard error].

Figure 4.

Gaze distributions from all participants during the first search for a specific target object (Block 1) – in this case a soap dispenser in a bathroom – as a function of object position (green = consistent, red = inconsistent) and background manipulations (solid lines = background present, dotted lines = background absent). Heat maps show fixated scene regions. The warmer the color, the longer was the scene location looked at. Percent values indicate scene coverage.

Figure 5. Mean RTs in Block 1 (left graph) and Block 2 (right graph) as a function object position (green = consistent, red = inconsistent), background manipulations (solid lines = background present, dotted lines = background absent) and search stage from Epoch 1 to Epoch 3 [bars depict standard errors].

Figure 6. Mean RT differences between Epoch 1 and 2 of Block 1 as a function object position (green = consistent, red = inconsistent) and background manipulations (solid lines = background present, dotted lines = background absent) [bars depict standard error].

Figure 7. Mean RT differences between Epoch 2 and 3 of Block 1 as a function object position (green = consistent, red = inconsistent) and background manipulations (solid lines = background present, dotted lines = background absent) [bars depict standard error].

Figure 8.

RTs for all searches across all participants plotted against the amount of time that a target was incidentally fixated as distractor on previous trials as a function of our 4 experimental conditions: a) “Consistent-With-Background” condition, b) “Inconsistent-With-Background” condition, c) “Consistent-Without-Background” condition, and d) “Inconsistent-Without-Background” condition.

Figure 9. Mean RT differences between Block 1 and 2 as a function object position (green = consistent, red = inconsistent) and background manipulations (solid lines = background present, dotted lines = background absent) [bars depict standard error].

Figure 10.

Gaze distributions from all participants during search for the soap dispenser in Block 2 as a function of object position and background manipulations. Heat maps show fixated scene regions. The warmer the color, the longer was the scene location looked at. Percent values indicate scene coverage (the difference in coverage between Block 2 and Block 1 is given in brackets).

Figure 11.

Mean RTs in for searches of A) strongly constrained objects and B) weakly constrained objects. For each subset of objects we present RTs in Block 1 (left graphs) and Block 2 (right graphs) as a function object position (green = consistent, red = inconsistent), background manipulations (solid lines = background present, dotted lines = background absent) and search stage from Epoch 1 to Epoch 3 [bars depict standard errors].

Figure 12.

Mean RT differences between Block 1 and 2 as a function object position (green = consistent, red = inconsistent) and object constraint (solid lines = strongly constrained, dotted lines = weakly constrained) [bars depict standard error].

Table 1

Mean time to first target fixation values in ms as a function of object position by background manipulations (Consistent with Background, Inconsistent with Background, Consistent without Background, Inconsistent without Background) and search stage from Epoch 1 to Epoch 3 [standard errors]. In addition, the table includes search benefits for search across epochs (Epoch1–2 and Epoch2–3) and blocks (Block1–2).

	Blocks	Search Epochs			Search Benefits		
		Epoch 1	Epoch 2	Epoch 3	Epoch1–2	Epoch2–3	Block1–2
Consistent with Background	Bl. 1	846 [103]	763 [99]	753 [74]	83	10	250
	Bl. 2	597 [89]	509 [91]	507 [88]	89	2	
Inconsistent with Background	Bl. 1	1184 [69]	944 [78]	1013 [56]	240	–69	496
	Bl. 2	619 [53]	513 [46]	522 [26]	106	–9	
Consistent without Background	Bl. 1	513 [38]	441 [31]	417 [37]	72	25	214
	Bl. 2	284 [36]	243 [29]	204 [20]	41	39	
Inconsistent without Background	Bl. 1	814 [46]	746 [59]	690 [75]	68	56	302
	Bl. 2	490 [47]	431 [49]	419 [56]	60	12	