Ann. N.Y. Acad. Sci. ISSN 0077-8923

ANNALS OF THE NEW YORK ACADEMY OF SCIENCES
Issue: *Competitive Visual Processing Across Space and Time*

The role of memory for visual search in scenes

Melissa Le-Hoa Võ[1] and Jeremy M. Wolfe[2]

[1]Scene Grammar Lab, Department of Cognitive Psychology, Goethe University Frankfurt, Frankfurt, Germany. [2]Visual Attention Lab, Harvard Medical School, Boston, Massachusetts

Address for correspondence: Melissa Le-Hoa Võ, Scene Grammar Lab, Department of Cognitive Psychology, Goethe University Frankfurt, Grüneburgplatz 1, 60323 Frankfurt am Main, Germany. mlvo@psych.uni-frankfurt.de

Many daily activities involve looking for something. The ease with which these searches are performed often allows one to forget that searching represents complex interactions between visual attention and memory. Although a clear understanding exists of how search efficiency will be influenced by visual features of targets and their surrounding distractors or by the number of items in the display, the role of memory in search is less well understood. Contextual cueing studies have shown that implicit memory for repeated item configurations can facilitate search in artificial displays. When searching more naturalistic environments, other forms of memory come into play. For instance, semantic memory provides useful information about which objects are typically found where within a scene, and episodic scene memory provides information about where a particular object was seen the last time a particular scene was viewed. In this paper, we will review work on these topics, with special emphasis on the role of memory in guiding search in organized, real-world scenes.

Keywords: memory; visual search; scene perception; eye movements

Introduction

Suppose that you are assembling what you need to cook dinner. This involves searching the cabinets and refrigerator for the food and equipment required. Memory could enter this activity in many ways. First, you need to remember what it is that you are looking for, including the features of that target object (e.g., I am looking for a red, round tomato). In a real scene, such as the kitchen, search will be additionally guided by knowledge about such scenes (Where are tomatoes typically found?), and perhaps, by specific memories for the current scene (Where did I put the tomatoes when I came home?)—memories that were existent before the search. As search progresses, more online memories could aid search by preventing perseveration on red, round non-tomatoes or by speeding subsequent search for the knife that was noted but passed over while looking for the tomato. You might be faster still when, having dispatched the tomato, you look for the knife again, in order to cut the onion. Each of these forms of memory has been investigated. Here, we will review the role of memory in visual search—focusing on guidance by spatial as opposed to feature memory—because particular learned spatial relations of objects play a key role in searches of organized, real-world scenes.

Search guidance in artificial displays

Some of the earliest work in the modern literature on visual search involved scenes,[1,2] but the bulk of the literature consists of studies that used arrays of fairly simple stimuli characterized by only a limited set of visual features, such as color and orientations, and generally placed randomly on blank backgrounds (reviewed in Refs. 3 and 4). These studies have shown that a limited set of target attributes can be used to guide attention toward candidate targets, even in meaningless displays.[5] For instance, if one is looking for a small, blue, moving vertical line, one can guide attention toward the target size, color, motion, and orientation. This idea of guidance by a limited set of basic attributes is the basis of what can be called the classic *guided search* (GS) model.[6–8]

What is the role of memory in search through simple displays? There are several aspects of

doi: 10.1111/nyas.12667

memory that could be relevant. Undoubtedly, there is memory for the target template (e.g., the small, blue, moving vertical line)—obviously essential for disregarding distractors and recognizing the target. More precise templates produce more efficient guidance of search,[9,10] with exact visual representations of the target serving to prime search.[11] As will be discussed in the next sections, the course of a current search could be assisted by memory for where you have already searched to avoid revisiting already inspected distractor locations. In searches with multiple targets, it would be of obvious value to have memory for targets that have already been found. In the case of repeated searches through the same display, it would be valuable if memory for prior searches informed the current search. Each of these types of memory-guided search has been studied in simple, artificial displays. We will briefly review the most important findings regarding inhibition of return (IOR), contextual cueing, and repeated searches in simple displays. We will then turn to the role of memory in more naturalistic search settings. To foreshadow an important conclusion, there is a distinction between the existence of a form of memory and its utility in a visual search. As will be discussed, there are multiple instances where a memory that could be used in search is not used because other processes guide attention to the target more efficiently.

The role of memory in search

Inhibition of return

IOR is a delay in shifting attention back to recently attended locations. The phenomenon was first reported by Posner and Cohen[12] in the context of cueing paradigms and was subsequently reported in visual search.[13,14] IOR would be of obvious utility in visual search if it could prevent attention from returning to rejected distractors. Many models of search, including early versions of GS, assumed that visual search processes were sampling "without replacement" from the display. Wolfe and Pokorny[15] failed to replicate the original[13] finding as, indeed, did Klein.[16] However, subsequent work showed that IOR in search was a reliable effect if the inhibited stimuli remained visible.[17] Thus, IOR exists in search, but does it support sampling without replacement? Using random arrays of letters, Horowitz and Wolfe[18] found that there was no difference in search efficiency between dynamic displays in which all distractors were randomly replotted every 100 ms and standard, static displays, suggesting that observers were sampling the display with replacement in both cases, because rejected distractors could not be marked in the dynamic displays. Horowitz and Wolfe[18] used this and other results to argue that "visual search has no memory," at least, no memory for rejected distractors. They speculated that the structure of the world makes it unnecessary to build fully elaborated visual representations and that "amnesia may be an efficient strategy for a visual system."

Subsequent work (reviewed in Ref. 19) suggests that the truth may lie between perfect IOR and an absence of useful IOR. Saccadic eye movements tend to be directed away from the last fixation location during visual search,[14,20] and saccades aimed back to the previous fixation location are preceded by longer fixations than saccades away from the previous fixation location.[14,21,22] Perhaps the best description of the role of IOR in search is that from Klein and MacInnes,[14] who refer to it as a "foraging facilitator." Although it seems clear that observers cannot use inhibition to mark every rejected distractor, it is plausible to assume that memory during search serves to prevent perseveration on single salient items.[14]

Although IOR is used in models that try to mimic natural viewing behavior,[23,24] recent studies have found evidence against IOR in real-world scenes.[25,26] Rather than an inhibition of the previous fixation location, Smith and Henderson,[27] for example, reported "facilitation of return" during scene viewing. The probability of refixating the last location was greater than or equal to other distance-matched locations, providing evidence against the view that IOR drives attention through a scene by decreasing the probability of return. The authors argue that the latency effects that have been attributed to IOR could be attributed to saccadic momentum (i.e., the tendency for saccades to continue the trajectory of the last saccade) rather than memory involvement. As an analogy, consider reading: readers do not revisit earlier text at random, not because of IOR but because of a rule that moves the eyes and attention down the line of text and on to the next line.

Repeated search of artificial displays

When revisiting a previously experienced context, the brain automatically generates predictions about the items that should appear in that context.[28] In contextual cueing studies, repeated exposure to the same meaningless arrays of items speeds search without observers' explicit awareness that they have been repeatedly exposed to the same target–distractor arrangements[29] (for a review, see Ref. 30). Although this implicit memory shortens reaction times, there is some question as to whether contextual cueing is a form of guidance or whether it simply facilitates responses.[31] In a different paradigm, Wolfe *et al.*[32] had observers repeatedly search through the same small sets of letters over hundreds of trials. Reaction times became faster, but there was no improvement in search efficiency even though observers clearly memorized the sets of three or six letters. Kunar *et al.*[33] explained the lack of memory benefit by showing that it took longer to access the memory than to simply search the display again. In this case, memory exists but conveys no benefit. Such memories can guide search, even in artificial displays, when they are given a chance, typically by slowing the search or providing an adequate preview time. For example, Solman and Smilek[34] used target eccentricity and item discriminability to modulate search difficulty. They showed that the more difficult the search, the more memory came into play.

Typically, memory from repeated search can be beneficial when stimuli are more complex, eye movements are required, and search is more demanding.[35–37] In contrast, memory from repeated search yields little benefit in tasks where simple displays support faster search.[32,33] Again, a unifying principle appears to be that having a memory is not the same as using that memory.

From arrays to scenes

Most of the studies we have discussed so far have investigated the contribution of memory in search through meaningless arrays of items. However, outside the laboratory, searches are more likely to occur in meaningful, structured scenes that are typically richer and more complex than laboratory search displays. Yet, searches in those scenes often feel relatively effortless—if we think about them at all. Searching for a red tilted bar among other colored, oriented items on a screen can be more demanding than searching for a sponge in an otherwise cluttered kitchen.[38] What ingredients of a scene allow for their efficient processing and how do those factors interact with the roles of memory?

Scene meaning

It is the meaning and rule-governed composition of the visual environment that gives those scenes their advantage in search tasks over random arrays of items (for a review, see Ref. 38). In classic contextual cueing paradigms that use letter arrays, the cueing effect develops over dozens of repetitions and tends to be quite small, with magnitude less than 100 ms[29] (for a review, see Ref. 39). In comparison, when real-world scenes are used, only four repetitions may be needed to produce a cueing effect of more than 2000 ms.[40] One could propose that the difference is due to the purely visual properties of scenes, but Brockmole *et al.*[41] used chess boards to show the contribution of a display's meaning to contextual cueing effects. As with many stimuli, the meaningfulness of chess boards varies with the expertise of the observer, whereas the visual features can remain fixed. Brockmole *et al.*[41] found that, when actual games were displayed, search benefits for repeated boards were four times greater for chess experts than for novices. Search by chess experts was guided by their ability to interpret the meaning of the chess scene.

Scene grammar

In the real world, there is virtually never the opportunity to process exactly the same visual input twice. Nevertheless, under most circumstances, the ability to understand and interact with new scenes and new views of old scenes is seemingly effortless. Like the ability to produce and understand an endless set of novel utterances, this ability cries out for explanation. A portion of our competence in this realm is because of an implicit knowledge of some subset of the many rules that govern the surrounding natural and manmade world (for a review, see Ref. 42). For instance, as part of this *scene grammar*, it is known that physical objects tend to rest on surfaces instead of floating in mid-air and that two objects cannot coexist in exactly the same place. It is also known that certain objects often co-occur in space: knives tend to be close to forks, toothpaste near toothbrushes, and keyboards near computer screens. Even their relative spatial relations can be constrained, in that, for example, knives are usually found to the right of forks and keyboards below screens. Generic knowledge of this scene grammar is stored in long-term memory and can be flexibly applied even in unfamiliar settings to find, for example, forks and keyboards (see also the "cognitive relevance framework"[43]).

In the 1980s, Biederman *et al.*[44,45] demonstrated that objects violating our generic knowledge of the world are more difficult to identify when presented briefly in an inconsistent scene context. Although this initial perceptual account of incongruent objects has been disputed[46] (e.g., by controlling response biases), Biederman's groundbreaking taxonomy that described various relations between objects and their surroundings still inspires work today. Biederman suggested that "something roughly analogous to what may be needed to account for the comprehension of sentences is required to account for the speed and accuracy of the comprehension of scenes never experienced before."[44] By analogy, he later categorized the various object–scene violations as either "semantic" or "syntactic" and showed that violating the relationship of an object to its surroundings impeded its visual perception and identification.[45] More recently, Võ *et al.* have referred to knowledge regarding what objects tend to be found where within a scene, as semantic and syntactic scene knowledge, respectively.[47–49] In an event-related potential (ERP) study, Võ and Wolfe[49] found a clear dissociation between semantic and syntactic processing: semantic inconsistencies (e.g., a mailbox in a bedroom) produced negative deflections in the N300–N400 time window, whereas mild syntactic inconsistencies (e.g., slippers on the bed) elicited a late positivity resembling the P600 found for syntactic inconsistencies in sentence processing. Extreme syntactic violations (e.g., a hovering beer bottle defying gravity) were associated with earlier perceptual processing difficulties reflected in the N300 response, but failed to produce a P600 effect. Moreover, they showed that this kind of semantic and syntactic processing in scenes elicits very similar brain responses to semantic and syntactic processing in language, suggesting that there might be some commonality in the mechanisms for processing meaning and structure across a wide variety of cognitive tasks.[50]

Semantic and syntactic guidance during visual search in scenes

If search is indeed unexpectedly efficient in scenes as compared to random arrays, it is likely that this can be attributed to guidance by the semantic and syntactic information that scenes possess and that random arrays do not. That is, when looking for a coffee mug in an office, the office imposes a set of constraints on the possible locations of the mug that are not imposed if looking for the image of a coffee mug in an array of objects on a screen. Quantitatively comparing the efficiency of search in scenes and arrays is difficult because search efficiency has usually been calculated with respect to the slope of the function relating reaction time to set size (i.e., the number of items within a display). Although set size is a straightforward concept in an array of objects, the set size of a real-world scene is simply not meaningfully calculable in any absolute sense. Consider a forest scene: Is that one forest, dozens of trees, or thousands of leaves? At the very least, the set size will be task dependent. Thus, Neider and Zelinsky[51] proposed that, when searching for objects in scenes, only a subset of all possible items is ever relevant for the current object search. In the example of the forest, leaves are not items if the task is to look for trees. Importantly, this functional set size can be dramatically reduced on the basis of strong "scene priors." For example, semantic and syntactic constraints would strongly restrict the items that could be coffee mugs in an office. The contextual guidance model presented by Torralba *et al.*[52] demonstrates the power of contextual guidance on the prediction of eye movements in real-world scenes. Thus, unlike a random display of isolated objects, a real scene itself can actually inform where a target is likely to be found and, consequently, where to direct one's attention and one's eyes.

Gist and nonselective processing

An important feature of scene priors is that they can be based on a rapidly acquired gist of the scene that does not require recognition of each object in the scene. Comprehension of scene gist only takes a brief glimpse.[53–55] It can be inferred from the layout of basic feature information without the need to cleanly segment that information into individual objects.[56] Accordingly, accuracy in scene recognition is not substantially affected by the number of objects in a scene or by blur (for a review, see Ref. 57). In their contextual guidance model, Torralba *et al.*[52] propose that an image is analyzed in two parallel pathways: the local and global pathways. Both of these pathways share a first stage in which the image is filtered by a set of multiscale-oriented filters. The local representation analyzes each spatial location independently and is used to compute local salience and to perform object recognition. The global pathway, however, represents the entire image holistically, on the basis of global scene statistics. This pathway supports ultra-rapid scene categorization,[56] which, in turn, activates stored scene priors, allowing shifts of attention and the eyes to locations that have a high probability of containing the search target. Thus, a key feature of the model is the interaction of local and global processing within the first glimpse in order to rapidly narrow down the search area to those parts of the scene that most probably contain the target (for a review, see Ref. 58).

Gaze guidance during search from a glimpse of a scene

The ability to use a short glimpse of a scene to guide eye movements during search has been demonstrated in a number of studies using the flash-preview moving-window paradigm.[48,59–61] In this paradigm, participants are first presented with a brief preview of the search scene, followed by presentation of a target word indicating which object they will be looking for. The scene is then presented again for search, but participants are only able to explore the scene through a gaze-contingent window that reveals only a small area of the scene tied to the current fixation location. Therefore, this paradigm allows isolation of the effect of the initial scene glimpse from the processing that takes place during later stages of scene viewing. The flash-preview moving-window paradigm shows that even a 50-ms glimpse of a scene can guide search as long as sufficient time is subsequently available to combine prior knowledge with the current visual input.[60] These results emphasize the constructive nature of the scene representations that are used to guide search and imply that the active control of human gaze in naturalistic scenes draws not only on currently available visual input but is also strongly influenced by the priors-based knowledge of scene grammar that has been learned over time.

All studies using the flash-preview moving-window paradigm report substantial preview effects. Thus, episodic memory of a particular scene preview can guide search. Note, however, that episodic memory in this particular case is assisting an unusually and artificially difficult search. The flash-preview moving-window paradigm restricts search to a window contingent on eye movements. This limits normal online scene processing and therefore increases dependence on the scene preview. Hillstrom *et al.*[62] directly tested whether a scene preview is beneficial when the scene is fully visible during search. They found that in full visibility previewing reduced solely the distance to the target of the second fixation but not subsequent eye movements, which then appeared to be guided by online information rather than episodic memory of the scene preview. Similarly, Võ and Wolfe[63] showed that previewing a scene for several seconds did not substantially speed subsequent unrestricted search in the same scene. They argued that the utility of episodic memory guidance is limited when the scene is fully visible during search and when semantic guidance is available (but see Ref. 64).

The role of memory during repeated search in naturalistic scenes

Although a glimpse of a scene can efficiently guide attention to the most probable target locations, one usually does not constantly jump from one scene to another. More typically, one tends to look for several objects within the same environment, sometimes repeating search for the same item. Returning to a cooking example, intuition conveys that the more time spent in a friend's kitchen, the more easily objects being looked for will be found, because over time an episodic memory representation of the scene will have been created. This intuition is backed by the finding that there is massive memory for objects,[65] as well as scenes.[66] Previously fixated (and thus attended) objects embedded in scenes can be retained in memory for hours or even days[67–69] (for a review, see Ref. 70). Thus, having incidentally looked at the knife while searching for the tomatoes, it seems reasonable to assume that the massive memory for objects and scenes would speed subsequent search for the knife. However, this expected repeated search benefit is not what is seen in the data.

In a study by Wolfe *et al.*,[38] participants searched repeatedly through the same indoor scenes for different objects and observed very little improvement in reaction times across multiple searches. Instead of using episodic memory to guide search, participants seemed to simply search *de novo* each time. Võ and Wolfe[63] replicated these findings in several eye-tracking experiments. In both studies, reaction times were markedly reduced when observers searched for a specific object for the second or third time, but looking at the knife while searching for the tomatoes did not speed search for the knife. Like repeated search through small letter arrays, repeated search in scenes is a situation where memory clearly exists but does not necessarily aid search. Võ and Wolfe[63] argue that powerful semantic and syntactic guidance in real-world scenes diminishes the usefulness of episodic memory because, as before, attention could be directed to the target before this memory could be used effectively. Oliva *et al.*[71] arrived at a similar conclusion after using a panoramic search display that would only allow observers to search a subpart of the whole scene from trial to trial. Again, participants appeared to search the display rather than rely on guidance by episodic memory. This even seems to be the case in environments where participants perform physical searches by moving their bodies in virtual space.[72] Hollingworth,[64] however, found that search is indeed speeded after familiarization with a new scene. Differences in trial sequence, type of feedback (auditory versus visual), and stimulus material (photographs versus 3D-rendered models of scenes) might have altered semantic guidance in ways yet to be explained, possibly accounting for the differences in these experimental results. While becoming more and more realistic, these scene models 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 yet to be understood.

Võ and Wolfe[73] further reasoned that episodic memory should have a more prominent role if other scene guidance failed. They presented participants with search displays containing inconsistently placed objects. For example, the soap was not placed on the sink but high up on the bathroom wall. In addition to weakening semantic guidance, objects were randomly scattered across the whole image rather than their usual accumulation on a few surfaces, such as tables, counters, and shelves. In this chaotic scene, searching the whole display over and over again would have been too costly, and consequently, episodic memory did speed repeated search (Fig. 1). Thus, the manipulation of semantic information in scene search had effects similar to those found when Solman and Smilek[34] manipulated search difficulty.

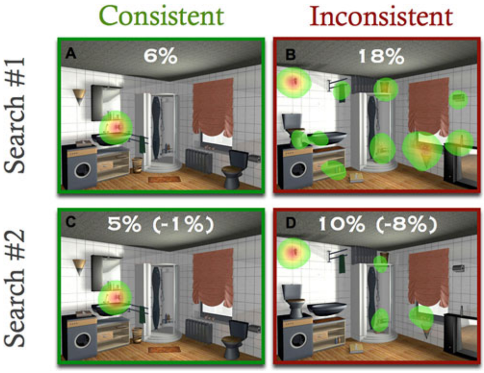

Figure 1. Fixation heatmaps that indicate (A) search space when looking for soap in the bathroom for the first time when placed in a consistent location and (B) the first time when placed in an inconsistent location (B). (C) Search space after several hundred trials looking for the soap again for a second time when placed in a consistent location or (D) when placed in an inconsistent location. The percentage indicates eye movement coverage of the scene and the percentage in brackets indicates reduction of search space from search #1 minus search #2. Data replotted from Ref. 68.

Increased visual difficulty pushes observers to use memory in search. What about actual effort? In a recent study, Solman and Kingstone[74] tested whether energetic costs modulate the use of memory. For this purpose, participants performed visual searches requiring either eye or head movements, posing lower or higher energetic costs. They reported greater use of memory in the more energy-demanding head-contingent search. Ballard *et al.*[75] similarly argued that participants choose not to operate at the maximum capacity of short-term memory but instead seek to minimize its use by frequently accessing the sensory input via eye movements. The reluctance to use short-term memory can be explained if such memory is expensive to use with respect to the often smaller cost of just looking again, since the scene itself tends to serve as an "outside memory."[76]

Memory for searched objects in scenes

Although there is some dispute on how much episodic memory generated from distractor fixations is used to guide search, there is strong

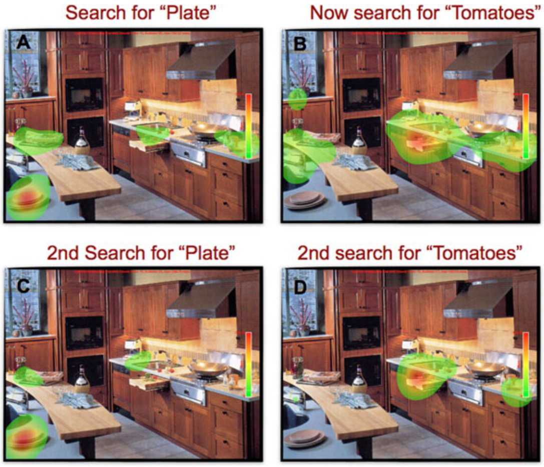

Figure 2. Fixation heatmaps that indicate (A) search space when looking for a plate in the kitchen for the first time and (B) subsequently looking for the tomatoes that had been looked at previously during the search for the plate, and (C) after several hundred trials looking for the plates again, (D) as well as for the tomatoes. Data replotted from Ref. 59.

agreement that searching for the same target object a second time is marked by great reductions of search time and search space, even with hundreds of trials intervening.[38,63,64,73] This suggests that different memory representations are generated during scene viewing by the act of looking at an object compared to looking for that object[63] (Fig. 2).

Within a search task, participants show substantially better memory for targets than distractors.[68,70,77] Memory for distractors increases when they share a feature with the target, and longer fixation durations tend to be related to better memory consolidation of the distractor items.[77] How does memory for targets compare to memory for objects that were intentionally memorized? Tatler and Tatler[78] tested memory performance as a function of encoding instructions in a real-world setting. Eye movements were recorded with a portable eye tracker, while participants performed one of three different tasks: (1) free viewing of a room; (2) intentional memorization of the whole room; and (3) intentional memorization of only tea-related objects. Results showed that performance was above chance in the free-viewing task, despite the lack of intentional encoding, and was presumably based on incidental encoding (see also Refs. 67, 68, 77, and 79). As expected, memory for objects was much better in the intentional memory conditions and best for relevant (i.e., tea-related) objects. These results suggest that intentional encoding should outperform incidental encoding.

Although great reductions in reaction times from the first to second searches of the same object have been attributed to strong target memory representations after search, memory for those targets was never explicitly tested. Draschkow *et al.*[80] directly compared incidental encoding of search targets in naturalistic scenes with intentional memorization of the same scenes. They found that recall memory performance—assessed by asking participants to draw the scenes—was actually markedly better for searched objects than for objects they had intentionally tried to memorize, even though participants in the search condition were not explicitly asked to memorize objects and did not know there would be a memory test. This effect was robust despite comparable gaze durations on the critical objects across tasks. Interestingly, the mere act of finding an object does not seem to be sufficient to create this memory benefit for searched items, since the effect disappeared when random object displays were used rather than naturalistic scenes. Thus, scene semantics may not only help search for objects in real scenes, but also create scene representations that boost memory for objects that have been sought to find.

Further important issues

Although the fundamental investigations on search guidance in artificial displays are still heavily relied on, there is a progressively greater need to test whether specific insights from that research still hold when searching in more complex, realistic environments. Initial progress has been made, and this paper aimed at discussing the differences in search guidance that arise when moving from simple T among L style displays to images of naturalistic scenes. However, study of the search for objects in 2D scenes on computer monitors is still far away from search in the real 3D world where the searchers themselves may move through the volume of space. We therefore strongly believe that further efforts are needed to test which of the important insights gained so far still hold truth when actually moving around in the real world.

Finally, the cognitive abilities that are taken for granted are usually the ones least well understood. We are not born with the rich set of knowledge that is effortlessly used to guide search in our environments; this knowledge has to be learned. Exactly how and when these knowledge structures develop over the life span raises another interesting but unanswered set of questions.

Conclusions

The ease with which we continuously search through our environment often makes us forget how many cognitive processes are involved. Abstract knowledge and rules stored in long-term memory, specific episodic memory representations generated online from recent search activities, and current task requirements all constantly interact to promote efficiently guided attention and action. Memory in its many forms plays a prominent role in the daily endeavor of searching the visual world, but not every memory that could assist a search does assist that search. A search engine has many sources of assistance and will use whichever sources allow it to reach the target most quickly.

Acknowledgments

The authors thank the ZiF research group "Competition and priority control in mind and brain: new perspectives from task-driven vision," as well as Emilie Josephs, Dejan Draschkow, and Sage Boettcher for helpful discussions on the topics reviewed here. This work was supported by Office of Naval Research (ONR) Grant N000141010278, National Institutes of Health (NIH) Grant EY017001, and NSF CELEST Science of Learning Center to J.M. Wolfe, and by NIH Grant F32EY022558, as well as Deutsche Forschungsgemeinschaft (DFG) Grant VO 1683/2-1 to M.L-H.V. M.L-H.V. drafted the manuscript, and J.M.W. provided critical comments for revisions. Both authors approved the final version of the manuscript for submission.

Conflicts of interest

The authors declare no conflicts of interest.

References

1. Enoch, J.M. 1959. Effect of the size of a complex display upon visual search. *J. Opt. Soc. Am.* **49:** 280–286.
2. Kingsley, H.L. 1932. An experimental study of "Search." *Am. J. Psychol.* **44:** 314–318. doi:10.2307/1414831.
3. Wolfe, J.M. 1998. Visual search. In *Attention.* H. Pashler, Ed.: 13–74. Hove, East Sussex, UK: Psychology Press Ltd.
4. Wolfe, J.M. 2014. Theoretical and behavioral aspects of selective attention. In *The Cognitive Neurosciences,* fifth edition. M.S. Gazzaniga & G.R. Mangun, Eds. Cambridge, MA: MIT Press.
5. Wolfe, J.M. & T.S. Horowitz. 2004. What attributes guide the deployment of visual attention and how do they do it? *Nat. Rev. Neurosci.* **5:** 495–501. doi:10.1038/nrn1411.
6. Wolfe, J.M. 1994. Guided Search 2.0 A revised model of visual search. *Psychon. Bull. Rev.* **1:** 202–238. doi:10.3758/BF03200774.
7. Wolfe, J.M. 2007. Guided Search 4.0: Current Progress with a model of visual search. In *Integrated Models of Cognitive Systems.* W. Gray, Ed.: 99–119. New York: Oxford University Press.
8. Wolfe, J.M., K.R. Cave & S.L. Franzel. 1989. Guided search: an alternative to the feature integration model for visual search. *J. Exp. Psychol. Hum. Percept. Perform.* **15:** 419–433. doi:10.1037/0096-1523.15.3.419.
9. Malcolm, G.L. & J.M. Henderson. 2010. Combining top-down processes to guide eye movements during real-world scene search. *J. Vision* **10:** 4.
10. Wolfe, J.M. *et al.* 2004. How fast can you change your mind? The speed of top-down guidance in visual search. *Vis. Res.* **44:** 1411–1426. doi:10.1016/j.visres.2003.11.024.
11. Kristjánsson, A. 2006. Simultaneous priming along multiple feature dimensions in a visual search task. *Vis. Res.* **46:** 2554–2570. doi:10.1016/j.visres.2006.01.015.
12. Posner, M.I. & Y. Cohen. 1984. Components of visual orienting. In *Attention and Performance X.* H. Bouma & D. Bouwhuis, Eds.: 531–556. London: Erlbaum.
13. Klein, R. 1988. Inhibitory tagging system facilitates visual-search. *Nature* **334:** 430–431.
14. Klein, R.M. & W.J. MacInnes. 1999. Inhibition of return is a foraging facilitator in visual search. *Psychol. Sci.* **10:** 346.
15. Wolfe, J.M. & C.W. Pokorny. 1990. Inhibitory tagging in visual-search—a failure to replicate. *Percept. Psychophys.* **48:** 357–362.
16. Klein, R.M. & T.L. Taylor. 1994. Categories of cognitive inhibition with reference to attention. In *Inhibitory processes in attention, memory, and language.* D. Dagenbach & T. H. Carr, Eds: 113–150. San Diego, CA: Academic Press.
17. Takeda, Y. & A. Yagi. 2000. Inhibitory tagging in visual search can be found if search stimuli remain visible. *Atten. Percept. Psychophys.* **62:** 927–934.
18. Horowitz, T.S. & J.M. Wolfe. 1998. Visual search has no memory. *Nature* **394:** 575–577.
19. Wang, Z. & R.M. Klein. 2010. Searching for inhibition of return in visual search: a review. *Vis. Res.* **50:** 220–228. doi:10.1016/j.visres.2009.11.013.
20. Gilchrist, I.D. & M. Harvey. 2000. Refixation frequency and memory mechanisms in visual search. *Curr. Biol.* **10:** 1209–1212.
21. Hooge, I.T.C. & M.A. Frens. 2000. Inhibition of saccade return (ISR): spatio-temporal properties of saccade programming. *Vis. Res.* **40:** 3415–3426. doi:10.1016/S0042-6989(00)00184-X.

22. Snyder, J.J. & A. Kingstone. 2000. Inhibition of return and visual search: how many separate loci are inhibited? *Percept. Psychophys.* **62:** 452–458. doi:10.3758/BF03212097.
23. Itti, L. & C. Koch. 2000. A saliency-based search mechanism for overt and covert shifts of visual attention. *Vis. Res.* **40:** 1489–1506.
24. Zelinsky, G.J. 2008. A theory of eye movements during target acquisition. *Psychol. Rev.* **115:** 787–835. doi:10.1037/a0013118.
25. Luke, S.G., J. Schmidt & J.M. Henderson. 2013. Temporal oculomotor inhibition of return and spatial facilitation of return in a visual encoding task. *Front. Psychol.* doi:10.3389/fpsyg.2013.00400/abstract.
26. Luke, S.G., T.J. Smith, J. Schmidt & J.M. Henderson. 2014. Dissociating temporal inhibition of return and saccadic momentum across multiple eye movement tasks. *J. Vision* **14:** doi:10.1167/14.10.202.
27. Smith, T.J. & J.M. Henderson. 2009. Facilitation of return during scene viewing. *Visual Cogn.* **17:** 1083–1108. doi:10.1080/13506280802678557.
28. Kim, G., J.A. Lewis-Peacock, K.A. Norman & N.B. Turk-Browne. 2014. Pruning of memories by context-based prediction error. *Proc. Natl. Acad. Sci.* **111:** 8997–9002. doi:10.1073/pnas.1319438111.
29. Chun, M.M. & Y. Jiang. 1998. Contextual cueing: implicit learning and memory of visual context guides spatial attention. *Cogn. Psychol.* **36:** 28–71.
30. Chun, M.M. & N.B. Turk-Browne. 2008. Associative learning mechanisms in vision. In *Visual Memory*. S.J. Luck and A.R. Hollingworth, Eds. 1–1. Oxford University.
31. Kunar, M.A., S. Flusberg, T.S. Horowitz & J.M. Wolfe. 2007. Does contextual cuing guide the deployment of attention? *J. Exp. Psychol. Hum. Percept. Perform.* **33:** 816–828. doi:10.1037/0096-1523.33.4.816.
32. Wolfe, J.M., N. Klempen & K. Dahlen. 2000. Postattentive vision. *J. Exp. Psychol. Hum. Percept. Perform.* **26:** 693–716.
33. Kunar, M.A., S. Flusberg & J.M. Wolfe. 2008. The role of memory and restricted context in repeated visual search. *Percept. Psychophys.* **70:** 314–328. doi:10.3758/PP.70.2.314.
34. Solman, G.J.F. & D. Smilek. 2012. Memory benefits during visual search depend on difficulty. *J. Cogn. Psychol.* **24:** 689–702. doi:10.1086/651260?ref=no-x-route:ca7f60a6dd9f21508f333f3b308f899f.
35. Hout, M.C. & S.D. Goldinger. 2012. Incidental learning speeds visual search by lowering response thresholds, not by improving efficiency: evidence from eye movements. *J. Exp. Psychol.* **38:** 90–112.
36. Körner, C. & I.D. Gilchrist. 2008. Memory processes in multiple-target visual search. *Psychol. Res.* **72:** 99–105. doi:10.1007/s00426-006-0075-1.
37. Peterson, M.S., M.R. Beck & M. Vomela. 2007. Visual search is guided by prospective and retrospective memory. *Percept. Psychophys.* **69:** 123–135.
38. Wolfe, J.M., G.A. Alvarez, R. Rosenholtz, *et al.* 2011. Visual search for arbitrary objects in real scenes. *Atten. Percept. Psychophys.* **73:** 1650–1671. doi:10.3758/s13414-011-0153-3.
39. Chun, M.M. & N.B. Turk-Browne. 2007. Interactions between attention and memory. *Curr. Opin. Neurobiol.* **17:** 177–184.
40. Brockmole, J.R. & J.M. Henderson. 2006. Using real-world scenes as contextual cues for search. *Visual Cogn.* **13:** 99–108. doi:10.1080/13506280500165188.
41. Brockmole, J.R., D.Z. Hambrick, D.J. Windisch & J.M. Henderson. 2008. The role of meaning in contextual cueing: evidence from chess expertise. *Q. J. Expt. Psych.* **61:** 1886–1896. doi:10.1080/17470210701781155.
42. Henderson, J.M. 2007. Regarding Scenes. *Curr. Direct. Psychol. Sci.* **16:** 219–222. doi:10.2307/20183200?ref=no-x-route:fc2118610c84af682e5748454a31f300.
43. Henderson, J.M., G.L. Malcolm, C. Schandl. 2009. Searching in the dark: cognitive relevance drives attention in real-world scenes. *Psychon. Bull. Rev.* **16:** 850–856. doi:10.3758/PBR.16.5.850.
44. Biederman. 1976. On processing information from a glance at a scene: some implications for a syntax and semantics of visual processing. *UODICS'76*. 75–88. doi:10.1145/1024273.1024283.
45. Biederman, I., R.J. Mezzanotte & J.C. Rabinowitz. 1982. Scene perception: detecting and judging objects undergoing relational violations. *Cogn. Psychol.* **14:** 143–177.
46. Hollingworth, A. & J. Henderson. 1998. Does consistent scene context facilitate object perception? *J. Exp. Psychol. Gen.* **127:** 398–415.
47. Võ, M.L.-H. & J.M. Henderson. 2009. Does gravity matter? Effects of semantic and syntactic inconsistencies on the allocation of attention during scene perception. *J. Vision* **9:** 24–24. doi:10.1167/9.3.24.
48. Võ, M.L.-H. & J.M. Henderson. 2011. Object–scene inconsistencies do not capture gaze: evidence from the flash-preview moving-window paradigm. *Atten. Percept. Psychophys.* **73:** 1742–1753. doi:10.3758/s13414-011-0150-6.
49. Võ, M.L.-H. & J.M. Wolfe. 2013. Differential electrophysiological signatures of semantic and syntactic scene processing. *Psychol. Sci.* **24:** 1816–1823. doi:10.1177/0956797613476955.
50. Patel, A.D. 2003. Language, music, syntax and the brain. *Nat. Neurosci.* **6:** 674–681.
51. Neider, M. & G. Zelinsky. 2008. Exploring set size effects in scenes: identifying the objects of search. *Visual Cogn.* **16:** 1–10. doi:10.1080/13506280701381691.
52. Torralba, A., A. Oliva, M.S. Castelhano & J.M. Henderson. 2006. Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search. *Psychol. Rev.* **113:** 766–786. doi:10.1037/0033–295X.113.4.766.
53. Oliva, A. & P.G. Schyns. 1997. Coarse blobs or fine edges? Evidence that information diagnosticity changes the perception of complex visual stimuli. *Cogn. Psychol.* **34:** 72–107.
54. Potter, M.C. 1975. Meaning in visual search. *Science* **187:** 965–966.
55. Thorpe, S., D. Fize & C. Marlot, 1996. Speed of processing in the human visual system. *Nature* **381:** 520–522.
56. Greene, M.R. & A. Oliva. 2009. Cognitive psychology. *Cogn. Psychol.* **58:** 137–176. doi:10.1016/j.cogpsych.2008.06.001.
57. Oliva, A. 2004. Gist of the scene. *Neurobiol. Atten.* 1–7.
58. Wolfe, J.M., M.L.-H. Võ, .K. Evans & M.R. Greene. 2011. Visual search in scenes involves selective and nonselective pathways. *Trends Cogn. Sci. (Regul. Ed).* **15:** 77–84. doi:10.1016/j.tics.2010.12.001.

59. Castelhano, M.S. & J.M. Henderson. 2007. Initial scene representations facilitate eye movement guidance in visual search. *J. Exp. Psychol. Hum. Percept. Perform.* **33:** 753–763. doi:10.1037/0096-1523.33.4.753.

60. Võ, M.L.-H. & J.M. Henderson. 2010. The time course of initial scene processing for eye movement guidance in natural scene search. *J. Vision* **10:** 14. 1–13. doi:10.1167/10.3.14.

61. Võ, M.L.-H. & W.X. Schneider. 2010. A glimpse is not a glimpse: differential processing of flashed scene previews leads to differential target search benefits. *Visual Cogn.* **18:** 171–200. doi:10.1080/13506280802547901.

62. Hillstrom, A.P., H. Scholey, S.P. Liversedge & V. Benson. 2012. The effect of the first glimpse at a scene on eye movements during search. *Psychon. Bull. Rev.* **19:** 204–210. doi:10.3758/s13423-011-0205-7.

63. Võ, M.L.-H. & J.M. Wolfe. 2012. When does repeated search in scenes involve memory? Looking at versus looking for objects in scenes. *J. Exp. Psychol. Hum. Percept. Perform.* **38:** 23–41. doi:10.1037/a0024147.

64. Hollingworth, A. 2012. Task specificity and the influence of memory on visual search: comment on Võ and Wolfe (2012). *J. Exp. Psychol. Hum. Percept. Perform.* **38:** 1596–1603. doi:10.1037/a0030237.

65. Brady, T.F., T. Konkle, G.A. Alvarez & A. Oliva. 2008. Visual long-term memory has a massive storage capacity for object details. *Proc. Natl. Acad. Sci.* **105:** 14325–14329. doi:10.1073/pnas.0803390105.

66. Konkle, T., T.F. Brady, G.A. Alvarez & A. Oliva. 2010. Scene memory is more detailed than you think: the role of categories in visual long-term memory. *Psychol. Sci.* **21:** 1551–1556. doi:10.1177/0956797610385359.

67. Castelhano, M. & J. Henderson. 2005. Incidental visual memory for objects in scenes. *Visual Cogn.* **12:** 1017–1040. doi:10.1080/13506280444000634.

68. Võ, M.-H., W.X. Schneider & E. Matthias. 2008. Transsaccadic scene memory revisited: a "Theory of Visual Attention (TVA)" based approach to recognition memory and confidence for objects in naturalistic scenes. *J. Eye-Movement Res.* **2:** 7:1–13.

69. Williams, C.C., J.M. Henderson & F. Zacks. 2005. Incidental visual memory for targets and distractors in visual search. *Percept. Psychophys.* **67:** 816–827. doi:10.3758/BF03193535.

70. Hollingworth, A. 2006. Visual memory for natural scenes: evidence from change detection and visual search. *Visual Cogn.* **14:** 781–807. doi:10.1080/13506280500193818.

71. Oliva, A., J.M. Wolfe & H.C. Arsenio. 2004. Panoramic search: the interaction of memory and vision in search through a familiar scene. *J. Exp. Psychol. Hum. Percept. Perform.* **30:** 1132–1146. doi:10.1037/0096-1523.30.6.-1132.

72. Kit, D., L. Katz, B. Sullivan, *et al.* 2014. Eye movements, visual search and scene memory, in an immersive virtual environment. *PLoS One* **9:** e94362.

73. Võ, M.L.-H. & J.M. Wolfe. 2013. The interplay of episodic and semantic memory in guiding repeated search in scenes. *Cognition* **126:** 198–212. doi:10.1016/j.cognition.2012.09.017.

74. Solman, G.J.F. & A. Kingstone. 2014. Balancing energetic and cognitive resources: memory use during search depends on the orienting effector. *Cognition* **132:** 443–454. doi:10.1016/j.cognition.2014.05.005.

75. Ballard, D.H., M.M. Hayhoe & J.B. Pelz. 1995. Memory representations in natural tasks. *J. Cogn. Neurosci.* **7:** 66–80.

76. O'Regan, J.K. 1992. Solving the "real" mysteries of visual perception: the world as an outside memory. *Can. J. Psychol.* **46:** 461–488.

77. Olejarczyk, J.H., S.G. Luke & J.M. Henderson. 2014. Incidental memory for parts of scenes from eye movements. *Visual Cogn.* **22:** 975–995. doi:10.1080/13506285.2014.941433.

78. Tatler, B.W. & S.L. Tatler. 2013. The influence of instructions on object memory in a real-world setting. *J. Vision* **13:** 5–5. doi:10.1167/13.2.5.

79. Hout, M.C. & S.D. Goldinger. 2010. Learning in repeated visual search. *Atten. Percept. Psychophys.* **72:** 1267–1282. doi:10.3758/APP.72.5.1267.

80. Draschkow, D., J.M. Wolfe & M.L.-H. Võ. 2014. Seek and you shall remember: scene semantics interact with visual search to build better memories. *J. Vision* **14:** 10–10. doi:10.1167/14.8.10.