

# Guidance of Visual Search by Memory and Knowledge

Andrew Hollingworth

**Abstract** To behave intelligently in the world, humans must be able to find objects efficiently within the complex environments they inhabit. A growing proportion of the literature on visual search is devoted to understanding this type of natural search. In the present chapter, I review the literature on visual search through natural scenes, focusing on the role of memory and knowledge in guiding attention to task-relevant objects.

**Keywords** Visual search · Memory · Goal-directed vision · Natural scenes

The laboratory study of visual search began as a means to an end. Early research used visual search as an experimental paradigm for probing perceptual efficiency in identification and selection (e.g., Schneider and Shiffrin 1977; Treisman and Gelade 1980). The purpose was not necessarily to understand how we find objects in natural scenes (where on earth did I leave my keys?) but rather to isolate the mechanisms that enable humans to select goal-relevant items in the presence of perceptual competition. The visual search paradigm has become indispensable to the study of attention and has been applied in research ranging from the single unit study of neurons in the superior colliculus (e.g., McPeck and Keller 2002) to the study of individual differences in psychopathology (see Weierich et al. 2008). Over the last 10–15 years, however, a large and growing proportion of the research on visual search has been devoted to understanding how we find goal-relevant objects within the types of complex environments that comprise daily life. Several of the chapters in the present volume are examples of this trend.

How we find objects in scenes is one of the central topics a science of cognition must address, because finding objects efficiently is essential for everyday behavior. Consider the activity of making tea discussed by Land et al. (1999). Making tea requires the sequential selection and use of several different objects, and they must be selected in a particular sequence defined by the task. The tea kettle must be found to fill it with water. A cup must be located, and then the tea bags are needed to place one in the cup. Milk is required later in the task, as is sugar, and a spoon

---

A. Hollingworth (✉)

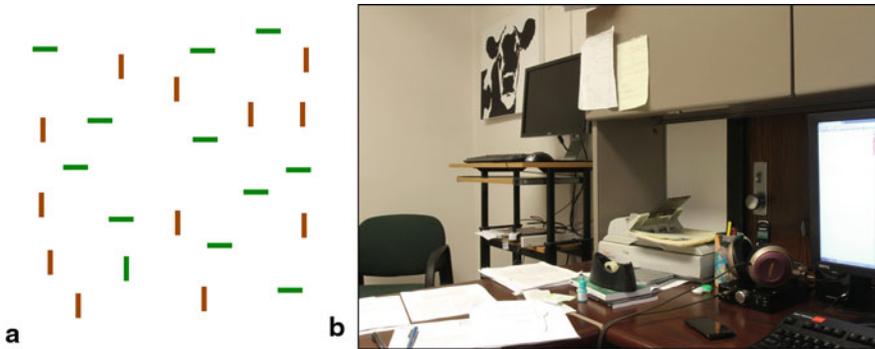
Department of Psychology, University of Iowa, Iowa City, IA 52242, USA  
e-mail: andrew-hollingworth@uiowa.edu

must be found to measure the sugar and stir it into the liquid. The objects in a scene cannot be perceived simultaneously, so each of these component operations requires visual search for a particular target object. Moreover, the goal of search and the representation of the relevant target must evolve as the sub-goals of the task change. For example, the template specifying the relevant target object must switch at some point from representing features of the sugar bowl to features of a spoon. The efficiency by which we find each object controls the overall efficiency of the task, as anyone who has attempted to make tea in an unfamiliar kitchen can attest. Similar sequential search demands are imposed by many everyday tasks, from brushing one's teeth to retrieving a letter from the mailbox. Thus, efficient human behavior is dependent on numerous visual searches that unfold, one after another, as goals change.

To what extent can research using the traditional visual search paradigm be applied to understanding object search in scenes? Many researchers have identified a direct relationship between search arrays and natural environments (both consist of numerous objects), and many have identified a relationship between the task in visual search (find the "T") and real-world tasks (find my keys). Hundreds of papers and talks using traditional visual search paradigms have started with a real-world example of search, indicating that the authors saw their results as relevant to understanding how we find objects in natural environments. However, the traditional visual search task is, in many key respects, ill-suited to this endeavor. Below, I discuss the properties of real-world scenes that are important for understanding natural search, most of which are poorly captured by traditional search paradigms (see Fig. 1). This discussion is not a criticism of existing approaches to visual search. In fact, the properties that make the visual search paradigm ill-suited to understanding natural search are often precisely those properties that have given researchers sufficient control to draw inferences about basic mechanisms of visual perception and attention.

*Scenes and the Objects within them are Meaningful* Visual scenes are meaningful environments, such as a kitchen or office (Henderson and Hollingworth 1999). We know a great deal about scene categories that could be brought to bear on search operations. For example, mixers tend to be found in kitchens, whereas staplers tend to be found in offices. One's search for a stapler will proceed much differently if one is searching for the stapler within an office versus within a kitchen (Henderson et al. 1999; Neider and Zelinsky 2006; Torralba et al. 2006).

*Objects are Complex, and their Identities and Visual Forms are Highly Variable* Most traditional search arrays are composed of highly similar (often identical) items. In contrast, natural environments are composed of objects that vary widely in their identities and visual properties. In addition, objects typically have complex shapes composed of multiple parts and contain several values on any given feature dimension (e.g., multiple colors). It is not clear that targets in a natural scenes can be segregated efficiently from distractors on the basis of simple visual features, such as those studied in the traditional search literature (see Wolfe et al. 2011).



**Fig. 1** **a** Typical stimulus in the traditional visual search paradigm (modeled after Treisman and Gelade 1980). The array consists of simple, meaningless stimuli arranged randomly. Across trials, locations are generated anew, precluding any possibility of spatial learning. Inter-item similarity is very high, including multiple item repetitions. Objects can be discriminated from each other on the basis of a small set of well-defined feature values. Object identification does not necessarily require movements of the eyes to fixate the elements of the array. **b** Typical real-world scene (the author's office). The scene is clearly identifiable as belonging to a particular category, as are the individual objects that comprise the scene. Objects are constrained to appear on surfaces. In addition, their locations are related to their functions (the keyboard appears near the monitor) and are often consistent across multiple searches (the monitor tends to stay put). Inter-object similarity is very low compared with the traditional search array. Objects are complex, composed of multiple values on several feature dimensions. It would be difficult to discriminate some of the objects from distractors on the basis of a single (or even several) feature values. The photograph is static, but under dynamic conditions characteristic of natural vision, changes in illumination, occlusion, perspective, viewing distance, and so forth would alter the appearance of the objects significantly. Within such a large, complex, crowded environment, object identification depends on movements of the eyes to obtain foveal information from individual items

*Search Through Scenes Depends on the Retrieval of Target Features from Memory* When searching for a stapler in an office, one is not provided with a visual image of the stapler immediately before commencing search (If one saw the stapler before searching, there would be no need to search for it). Therefore, the representation of the target object (i.e., the target template) must be retrieved from memory. If one knows something about the features of the particular stapler, then the template could be specific to that object. Alternatively, one could draw upon categorical knowledge to establish a search template that is general to the class of stapler (Yang and Zelinsky 2009).

*Scenes have Coherent Spatial Structure* Unlike standard search arrays, the locations of objects in scenes are not randomly determined. Spatial constraints arise from several sources. The most obvious is that objects are constrained by gravity to be located on surfaces (or to be supported in some other way, such as hanging from the wall) (see, e.g., Biederman et al. 1982). If one seeks an object in a scene, one can generally limit the search to surfaces that plausibly support the object. In addition, object locations are related to object function. Staplers often appear on desks because

that location provides efficient access in the course of stapling. Similar functional considerations apply across exemplars of a particular scene type, enabling categorical inferences about the probable locations of objects.

*Scenes are Experienced Repeatedly* The vast majority of environments we inhabit are familiar to us, and searches are repeated many times for common tasks. This raises the possibility that search through scenes is strongly influenced by learning (Chun 2000).

*Selection in Scenes Depends on Overt Shifts of Gaze* Traditional work on visual search has often minimized the contribution of eye movements. Yet, the size and complexity of natural scenes requires eye movements (and head and body movements) to obtain high-resolution, foveal information from objects. Moreover, gaze fixation is used to link objects to the motor programs that operate over them (Ballard et al. 1997; Hayhoe 2000; Land and Hayhoe 2001; Land et al. 1999). These two roles mesh seamlessly. In the course of search, the eyes are directed sequentially to possible targets until the required object is fixated. Once fixated, gaze supports motor interaction with that object so that it can be used in the current task. (For discussion of the importance of studying eye movements in search, see Eckstein 2011; Findlay and Gilchrist 2003; Henderson 2003; Zelinsky 1996, 2008; Zelinsky et al. 1997).

Weaving through this analysis of the properties of scenes and search within them is the idea that natural search is likely to be strongly guided by knowledge and memory. We have extended experience with scene exemplars and scene types, we search repeatedly through particular environments, and the nature of the search operation changes dynamically as goals change. In fact, we construct and organize our environments with the intention of maximizing the efficient use of memory and knowledge in visual search. The remote control is kept in a little basket on the side table so that one need only inspect a single location to find it. After using a pen, one tends to place it back in the same location on the desk so that memory for its location can guide attention back to it. Progress toward understanding natural search will therefore depend on understanding how memory, knowledge, and goals guide the search process and how these mechanisms interact with the visual processing of the scene.

Before discussing the roles of memory and knowledge in natural search, let's first consider the possibility that search is instead dominated by the *visual* properties of the scene. In an influential model, Itti and Koch (2000) developed a "saliency map" approach to explain search through scenes (and saccade target selection in scenes more generally). The model computes the relative saliency of local scene regions as a function of various low-level visual properties (such as luminance, contrast, and contour density). These regions are then ranked within a saccade "priority map" and fixated sequentially, with initial fixations on the scene directed to the most salient regions. The approach is attractive in that it does not demand any role at all for memory, knowledge, or goals, is relatively straightforward to implement in computer systems, and incorporates properties of neural systems involved in visual selection.

However, such a model would appear to have difficulty accounting for natural search behavior. Returning to the example of making tea, consider the possibility that as each object becomes relevant to the task (kettle, then cup, then tea bag), gaze is directed to the most visually conspicuous object in the scene. This would obviously be a complete disaster; the tea would not get made. The visual salience of an object in a scene is not closely related to its functional relevance. In addition, the visual salience of objects does not change systematically as goals change, providing no means to adapt search to changing task demands.

A large body of research supports this analysis. Low-level visual salience accounts for only a small proportion of the variance in selection during goal-directed search through natural scenes. Selection is much more strongly influenced by memory and knowledge (Einhäuser et al. 2008a, b; Foulsham et al. 2011; Foulsham and Underwood 2007, 2008; Henderson et al. 2007; Henderson et al. 2009; Itti 2005; Tatler et al. 2011; Torralba et al. 2006; Underwood et al. 2009; Zelinsky et al. 2006). When given the task to search for a target object in a novel scene, participants tend to direct their gaze to regions of the scene where an object of that type is likely to be found (Eckstein et al. 2006; Ehinger et al. 2009; Henderson et al. 1999; Neider and Zelinsky 2006; Torralba et al. 2006), independently of the visual salience of those regions (Henderson et al. 2009). In addition, attention is guided toward regions of the scene that match the features of the search target, regardless of whether the target itself is physically salient (Hwang et al. 2009; Kanan et al. 2009; Malcolm and Henderson 2009, 2010; Pomplun 2006; Zelinsky 2008). Although the visual properties of the scene over which search occurs are obviously central to the search operation, the selective mechanisms that determine where attention and gaze will be allocated within the scene are driven primarily by the observer's goals combined with knowledge about the scene context and knowledge of the visual properties of the object the observer is trying to find.

### **Memory in Traditional Visual Search Tasks**

There is no question that memory and knowledge play a central role in visual search through natural scenes. However, the role of memory in traditional search tasks has been more controversial. Horowitz and Wolfe (1998) argued that search through random arrays of symbols has no memory. In their experiments, the positions of search items were static or were scrambled every 111 ms. If search depends on using memory to avoid previously attended locations (e.g., Klein 1988) then scrambling should impair search efficiency. Yet, the search slopes did not differ between the two conditions. Converging evidence came from a repeated search paradigm, in which multiple searches over the same static array did not generate a major benefit in performance (Wolfe et al. 2000). Wolfe (1999) proposed that after attention is withdrawn from an object, the visual object representation dissolves into its elementary features, leaving no memory that could facilitate search either within a trial or across trials.

This claim has been subject to significant debate (e.g., Shore and Klein 2000), and subsequent research has established a central role for memory in traditional visual search paradigms. If targets in the scrambled condition of the Horowitz and Wolfe (1998) paradigm are placed in previous distractor locations, search efficiency is impaired relative to the static condition (Kristjánsson 2000). Larger set sizes also generate slope differences between static and scrambled conditions (Kristjánsson 2000) as does search for multiple targets (Gibson et al. 2000). von Mühlenen et al. (2003) showed that the original Horowitz and Wolfe finding was probably caused by an idiosyncratic strategy: Participants kept their attention in a single location and waited for the target to appear nearby. Other experiments demonstrated memory in several components of the search operation. Refixations during search were observed less frequently than predicted by a no-memory model (McCarley et al. 2003; Peterson et al. 2001), and previous demonstrations of inhibition of return during search (Klein 1988) were confirmed (Müller and Von Mühlenen 2000; Takeda and Yagi 2000). Memory for distractors during search has been found to facilitate subsequent searches when the distractor becomes a target (Körner and Gilchrist 2007), and repeated search across multiple trials generates reliable savings (e.g., Chun and Jiang 1998). Thus, the balance of evidence indicates that memory processes plays multiple important roles in traditional visual search tasks.

## Guidance by Categorical Knowledge of Scene Types

Our extensive experience with scene types allows us to predict where particular objects are likely to be found. Milk tends to be found in the refrigerator and not in the pantry. Staplers tend to be found on top of desks rather than beneath them. Indeed, search for objects whose locations are predictable within a scene is more efficient than search for objects whose locations are not (Eckstein et al. 2006; Henderson et al. 1999; Mack and Eckstein 2011; Neider and Zelinsky 2006). This type of knowledge can be applied rapidly within a search operation to guide attention and gaze toward regions of the scene that have a high probability of containing the target. For example, Torralba et al. (2006) found that eye movements during search for pedestrians were rapidly directed to sidewalks and other regions of the scene where pedestrians were likely to be found. Torralba et al. developed a *contextual guidance model* to account for these effects of scene and object knowledge on visual search. The model learns the associations between large-scale image features within a scene and the locations of particular object types. During a search operation, global image features and local salience are computed in parallel. These two sources of guidance are combined within a priority map that governs the order of scene regions fixated. The inclusion of contextual guidance allows the model outperform a model based solely on visual

saliency. In addition, if one adds a second top-down guidance component to the model that uses knowledge of the target appearance to prioritize candidate regions, the combined model can account for an impressive proportion of the variance in human eye movements during visual search (Ehinger et al. 2009).

## Guidance from the Learning of Scene Exemplars and Repeated Search

We experience scenes repeatedly. We conduct multiple searches through the same scene for different objects and search for the same object in a scene multiple times. We conduct all kinds of non-search activities through those same scenes. Humans have ample opportunity to encode information from particular scene exemplars, such as the spatial structure of the environment and the locations of individual objects. And humans have the *capability* to remember these features of environments: Long-term memory (LTM) for objects and scenes can be quite precise (Hollingworth and Henderson 2002) and has a remarkably large capacity (Brady et al. 2008; Hollingworth 2004, 2005; Konkle et al. 2010; Standing 1973; Standing et al. 1970). The role of scene memory in search has been addressed in studies using scene previews and repeated search and has been focused on several key questions. Does scene memory facilitate visual search? If so, what types of scene information are functional? How much experience is necessary? Is learning task-general or specific to visual search?

A preview of the scene facilitates later search (Becker and Rasmussen 2008; Castelhana and Heaven 2011; Castelhana and Henderson 2007; Hillstrom et al. 2012; Hollingworth 2009, *in press*; Vö and Henderson 2010) and repeated search leads to reliable savings (Brockmole et al. 2006; Brockmole and Henderson 2006b; Brooks et al. 2010; Ehinger and Brockmole 2008; Vö and Wolfe 2012). Thus, there is no doubt that memory for a scene exemplar is indeed employed to facilitate visual search. Using a scene preview design, (Hollingworth 2009; see also Hollingworth 2006) examined two forms of scene memory that are likely to control the allocation of attention and gaze: memory for object locations (which could guide attention directly to the target location) and memory for the spatial structure of a scene (which could guide attention to the locations where the target object is likely to be found). Participants viewed a scene preview for 10 s. Then, a single search target object was presented. Next, participants searched through the scene and reported whether the target in the scene had the same or different left-right orientation. Two preview conditions were compared with a no-preview control condition. A preview that was identical to the search scene except for the presence of the target improved search, as measured both by reaction time and the elapsed time to the first fixation on the target. Thus, memory for the spatial structure of the scene guides attention during search. A preview that contained the target led to further savings, demonstrating that participants had encoded the locations of specific objects from the preview and had used this to guide attention to the remembered location.

Guidance from scene memory can be observed even with an extremely brief glimpse of a scene before search. Castelhana and Henderson (2007) presented a preview of a scene for 250 ms, followed by a label indicating the search target, followed by search through the scene. [In this and related experiments, Castelhana, Henderson and colleagues used a moving window technique to limit visible information during search to a small region around the current fixation point; however, this aspect of the design does not appear to be necessary, as similar effects are found for search through fully visible scenes (Hillstrom et al. 2012)]. A brief glimpse facilitated search relative to a no-preview baseline, and subsequent work has found that a masked preview of only 50 ms is sufficient to generate a search benefit (Võ and Henderson 2010). Thus, scene information can be extracted very rapidly from an image in a manner that can guide attention and gaze to plausible target locations. The effect is not due to simply identifying the *type* of scene over which search will occur, because a preview using a different exemplar from the same scene category provided no benefit at all (Castelhana and Henderson 2007), and preview effects are observed even for target objects that are semantically inconsistent with the scene in which they appear (Castelhana and Heaven 2011). Instead, the preview allows participants to encode the spatial structure of the scene exemplar and locate the plausible surfaces on which an object could appear. Note that this spatial representation is likely to be quite abstract; a change in the absolute size of the scene from preview to search had no effect on the magnitude of the preview benefit (Castelhana and Henderson 2007).

In preview studies, participants do not conduct explicit search during the preview itself, but the preview is embedded within the context of a search experiment, and participants know that they will later search through the previewed scene. To what extent is the scene learning that facilitates search specific to the context of a visual search task? Evidence from Võ and Wolfe (2012) suggested that learning is highly task-specific. They included an initial preview session in which each scene was viewed for 30 s. Participants did not know, during the preview session, that they would later search through the scenes. The preview task was either closely related to the informational demands of search (memorize the locations of the objects in the scenes) or not closely related (decide whether the room is inhabited by a man or a woman). Relative to a no-preview control, there was no evidence at all that a scene preview facilitated search, even in the case that participants were instructed to memorize object locations. However, repeated *searches* generated significant savings. Thus, the transfer of memory to visual search appeared to be limited to memory representations formed during previous searches for a particular object. Võ and Wolfe reasoned that in order to transfer, learning must occur within the context of visual search. This finding is broadly consistent with theoretical approaches holding that visual memory encoding and application is strongly constrained by task (Ballard et al. 1995; Droll et al. 2005).

The interpretation of the Võ and Wolfe (2012) results is limited by the fact that their preview and no-preview trials were assigned to different participant groups, and they had relatively little power to detect an effect of preview. Hollingworth (*in press*) replicated the method but in a more sensitive within-subjects design. Participants viewed half the scenes items in a preview session followed by a search session

containing all scene items. A preview task of memorizing object locations led to a robust benefit in later search. Further, a preview task that did not involve remembering object locations (decide which object is the least likely to be present in a scene of this type) also generated a robust search benefit, and the magnitude of the effect was similar to that found for the location-memorization preview. Contrary to the evidence of Võ and Wolfe, our results establish that the transfer of learning is not strongly constrained by task. Object and scene information acquired outside of the context of visual search transfers reliably to the task of finding objects in scenes. Scene memory appears to be applied flexibly.

Consistent with this conclusion is the finding that when distractors are fixated during search (i.e., when they are not directly task relevant), a memory trace is established (Castelano and Henderson 2005; Williams et al. 2005) that facilitates later search when that distractor becomes a target (Howard et al. 2011; Võ and Wolfe 2012). In three recent studies (Hollingworth *in press*; Howard et al. 2011; Võ and Wolfe 2012), participants searched for a series of different objects within the same scene image. For example, in a static bedroom scene a participant would first search for an alarm clock, then for a cell phone, then for a candle, and so on. As multiple searches progressed, participants became faster as they searched for objects that had been distractors during previous searches (Hollingworth *in press*; Võ and Wolfe 2012) and this benefit was related to whether the distractor had been fixated during a previous search (Howard et al. 2011). Again, scene memory appears to be quite flexible. Information acquired incidentally from an object when it was not task relevant is stored and used later to facilitate search for that object.

Further insights into the nature of scene learning across repeated searches comes from the literature on contextual cuing (for a review, see Chun and Turk-Browne 2008; Chap. 6). Initial studies used relatively simple arrays of randomly organized symbols. In the basic task, participants search for a target symbol. The spatial configuration of a subset of search arrays is repeated throughout the experiment. Search is facilitated as participants learn the spatial relationship between repeated search contexts and target locations. In the standard contextual cuing paradigm, the learning of target locations is incremental, implicit, and local. Arrays require several repetitions before an advantage emerges over the novel condition, and RTs for repeated arrays do not reach asymptote until at least 10–15 repetitions. Once established, the magnitude of the contextual cuing effect is modest (typically 40–70 ms). At the end of an experiment, participants are at chance when asked to discriminate between repeated arrays and novel arrays, indicating that the memory representation responsible for facilitating search is not consciously accessible. Finally, the learning in contextual cuing studies appears to be highly local, with the effect dependent on the objects closest to the target (Brady and Chun 2007; Olson and Chun 2002). For example, Brady and Chun (2007) found that repetition limited to the two distractors closest to the target was sufficient to produce contextual cuing equivalent to that produced by repetition of the entire array of 11 distractors.

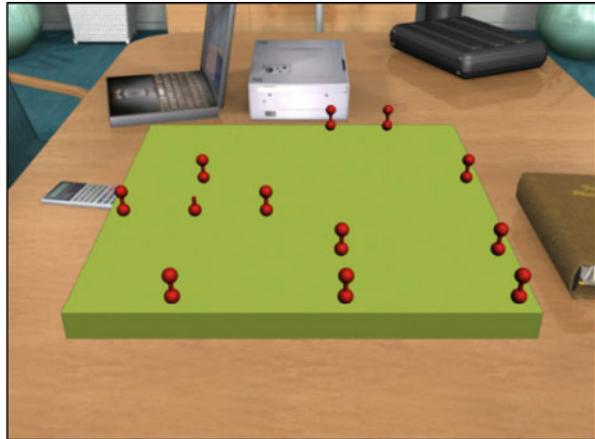
But the story is, in fact, much more complicated. Studies of repeated search through *naturalistic* scenes have found learning to be neither incremental, implicit, nor local; instead, the learning of target location is extremely rapid, is explicitly available, and is strongly dependent on global properties of the scene context (Brockmole

et al. 2006; Brockmole and Henderson 2006a, b; Brooks et al. 2010). In Brockmole and Henderson (2006b) participants searched for small letter targets embedded within images of natural scenes. The very first repetition of a particular scene context led to search savings, and search reached floor-level RTs following only 3–4 repetitions. In addition, the magnitude of the repetition benefit was extremely large. After a few repetitions, search through repeated scenes was approximately 2 s faster than search through novel scenes.

Differences between paradigms are not limited to the rate of learning or to the magnitude of the effect; the nature of the learning process itself appears to differ. At the end of the session in Brockmole and Henderson (2006b), participants reliably discriminated repeated scenes from novel scenes, indicating that the learning of target locations in scenes depends on memory processes that are available to conscious report. More importantly, object location was coded relative to the global structure and/or identity of the scene, rather than locally. Brockmole et al. (2006) included a transfer session after the main search session. For a repeated scene, either the local context around the target was disrupted (e.g., replacing the table on which the target appeared with a different table) or the global context was disrupted (e.g., replacing all of the rest of the scene except the table on which the target appeared). The former did not have any observable influence on the magnitude of the repetition effect, whereas the latter eliminated it completely. For real-world scenes, then, it is clear that target learning and/or the expression of learning is strongly dependent on global processing of the scene. In the experiments of Brockmole, Henderson, and colleagues, memory for the location of the target was likely to have been associated with a representation of the identity of the scene. Brockmole and Henderson (2006a) included a transfer session in which they mirror-reversed repeated scene items. Mirror reversal did not radically impair scene identification, but it disrupted the spatial structure of the scene and the absolute location of the target. Initial saccades during search were very frequently directed to the original *screen* location of the target rather than to the original *scene* location of the target. Participants appeared to associate an absolute target location directly with a scene identity, independently—at least to some extent—of the internal spatial structure of the scene.

The preceding discussion depends on a distinction between global and local properties of search contexts, but these are difficult to define for natural scenes. The functional context involved in search for a desk in an office (the office) will be different from the context involved in the search for a stapler on the desk (the desk). Brooks et al. (2010) proposed that memory for scenes is organized hierarchically and is accessed hierarchically in search. It would highly inefficient to code the location of the stapler relative to the coordinate system established by the large-scale features of the office (or office building). The locations of local objects are more likely to be coded relative to the immediate context for search, such as the desk on which the stapler usually appears. This localist coding assumption is consistent with evidence from the traditional contextual cuing literature, reviewed above. However, Brooks et al. proposed that that retrieval of the appropriate search context will depend on scene recognition. That is, retrieval of the desk context will be strongly dependent on

**Fig. 2** Example of a hybrid scene stimulus in which a search array is embedded within a scene background. (From Brooks et al. 2010)



identifying the office in which it appears, consistent with the results of Brockmole et al. (2006).

To test this proposed resolution of the conflicting theoretical accounts of contextual cuing, Brooks et al. (2010) embedded search arrays within images of natural scenes (see Fig. 2). During several search blocks, a particular target location was consistently paired with a particular array configuration and with a particular scene background. Participants showed reliable contextual cuing. In a transfer session, we disrupted either the search array (by scrambling the distractor locations) or the scene background (by replacing it with a different background). A purely localist view predicts transfer in the latter condition but not the former. A purely globalist view predicts the reverse. The hierarchical hypothesis predicts transfer in neither: disrupting the array eliminates the functional context within which the target location was learned; disrupting the background blocks the identification of the scene and the retrieval of the appropriate local context. The data confirmed the hierarchical prediction. Evidence of local learning and effects of scene identity can therefore be integrated within a larger view holding that learning is relative to a particular functional search context, but that the retrieval of this contextual information depends on scene recognition. Brooks et al. found that search arrays could be learned in the absence of a consistent background, suggesting that the search context and scene identity are integrated only when there is a consistent pairing between the two.

## Guidance from a Target Template

Thus far, we have discussed the guidance of attention and gaze by categorical knowledge and by exemplar memory. Guidance also derives from knowledge of the visual properties of the target object; searches will be more efficient when we know something specific about the appearance of what we are looking for (e.g., Wolfe et al.

2004). Most theories of attentional selection depend on a mechanism that can specify and keep active the features of the to-be-selected item (Bundesen 1990; Desimone and Duncan 1995; Duncan and Humphreys 1989; Wolfe 1994), providing a set of parameters for the guidance of selection. The mechanism of guidance has traditionally been conceived as a *search template* specifying the relevant features that distinguish the target from distractors (e.g., Wolfe 1994). The search template properties must be maintained during the course of search, which can extend over multiple shifts of attention and gaze to individual objects. Clearly, such extended maintenance requires a relatively robust form of memory. VWM is the natural candidate, and several theories of attention assume that VWM is indeed the substrate of the search template (Bundesen et al. 2005; Desimone and Duncan 1995). VWM also provides the type of flexibility needed to configure search in the course of real-world tasks. As goals change, new target information can be rapidly loaded into VWM (Vogel et al. 2006), providing a means to re-set the parameters of search efficiently (e.g., from those specifying the features of the sugar bowl to those specifying the features of the spoon).

Recent research has provided strong confirmatory evidence that VWM plays precisely this role. First, a concurrent VWM load interferes with search efficiency (Woodman et al. 2007), except when the target features repeat over many trials and do not themselves require active maintenance in VWM (Hollingworth and Maxcey-Richard *in press*; Woodman and Luck 2004). Second, the capability to store multiple objects in VWM is reflected in the ability to establish multiple templates that guide attention simultaneously to different sets of objects (Beck et al. 2012). Third, an electrophysiological correlate of VWM maintenance, the contralateral delay activity (CDA) event-related potential (ERP) component, is sustained across the course of visual search, and the amplitude of the CDA predicts search success (Woodman and Arita 2011).

Finally, and most directly, VWM can automatically bias selection during visual search. Several studies have found that attention is oriented to distractors that match the content of VWM (Han and Kim 2009; Hollingworth et al. *in press*; Hollingworth and Luck 2009; Olivers 2009; Olivers et al. 2006; Soto et al. 2005; Soto et al. 2006; Soto et al. 2010). In these experiments, participants maintain a VWM load (typically a color) during a visual search task. The search array is made up of colored items, but the target feature is orthogonal to color. On a subset of trials, a distractor is rendered in the remembered color. Participants know that this object will never contain the target value, yet its presence attracts gaze and slows search. There are some circumstances in which participants can avoid or even reverse this *memory-driven capture*, but these appear to be the exception rather than the rule, occurring when the search task is extremely difficult and when participants have sufficient time to configure a template that excludes the remembered feature value (Downing and Dodds 2004; Han and Kim 2009; Woodman and Luck 2007). The basic memory-driven capture effect suggests that VWM and attentional selection are tightly coupled. VWM modulates the competition among multiple objects for selection (Desimone and Duncan 1995), with attention biased toward items in a scene that match VWM content. In capture studies, this bias impairs performance, but under normative conditions, when the

content of VWM is a set of features that define the desired target, the same bias would guide attention toward relevant objects.

VWM guidance plays a central role in a particular form of visual search generated by the simple act of making saccades (Hollingworth et al. 2008). Saccades are notoriously inaccurate. In cluttered, natural environments, when the eyes fail to land on the intended saccade target, there may be multiple objects visible near the landing position. This creates a classic visual search problem: The original saccade target must be located among distractors so as to generate an appropriate corrective saccade. We make tens of thousands of saccades each day, and a significant proportion of these fail to land on the intended object. Search for the original saccade target is almost certainly the most common form of search behavior in which humans engage, and the success of correction will determine the efficiency by which the eyes are ultimately directed to each task-relevant object.

Hollingworth et al. (2008) had participants execute a saccade to one target object in a circular array of colored disks. During the saccade, the array was rotated so that the eyes landed between the target and an equidistant distractor. Gaze correction to the original target required VWM for the target color, because the rotation during the saccade was not directly visible. Gaze correction was accurate, rapid, and automatic, suggesting a fundamental, and potentially low-level, interaction between the content of VWM and saccadic orienting. In addition, gaze correction was impaired by loading VWM with task-irrelevant information, just as found in more traditional search tasks (Woodman et al. 2007). Moreover, if the distractor object matched the content of a secondary VWM load, correction was also impaired (Hollingworth and Luck 2009), consistent with the finding of memory-driven capture in standard search tasks (Olivers et al. 2006; Soto et al. 2005). Thus, a similar guidance mechanism appears to be functional at multiple levels of the search process. VWM guides attention and the eyes to the ultimate target of the search. VWM also guides gaze correction for the individual eye movements that are embedded within the larger search task.

What is the locus of the interaction between VWM and attentional orienting? Recent studies indicate that one form of interaction occurs during the initial sensory processing of objects. This is plausible given evidence that VWM maintenance generates feature-specific activation in early sensory regions of the brain (Harrison and Tong 2009; Serences et al. 2009). Two studies have demonstrated that VWM directly influences perceptual experience: Holding a particular motion direction in VWM modifies motion perception (Kang et al. 2011); holding a particular orientation in VWM modulates the competition between stimuli in binocular rivalry, biasing perception toward the matching orientation (Pearson et al. 2008). In addition, attentional selection on the basis of color (i.e., feature-based selection, which is likely to depend on a VWM template) modulates the P1 ERP component at latencies of ~100 ms post stimulus (Zhang and Luck 2009), indicating an effect on the initial feed-forward sweep of sensory processing. An early sensory interaction is also consistent with the fact that VWM modulates capture by abrupt onsets during search (Mannan et al. 2010).

The effects of VWM on perceptual processing and orienting are observed even in the absence of stimulus competition. In Hollingworth et al. (in press), orienting

saccades to single onset targets were influenced by whether the object did or did not match the content of a task-irrelevant VWM load: Saccades to matching targets were generated more rapidly and were more accurate (i.e., landed closer to the center of the target). These effects were observed on saccades with latencies near the limit of human capabilities, between 100 and 150 ms post stimulus (a result that contrasts with theories claiming that rapid saccades are influenced solely by physical salience, Ludwig and Gilchrist 2002; van Zoest et al. 2004). Further, VWM modulated the competition between the target and a task-irrelevant distractor, with the landing position of the saccade strongly biased toward whichever object matched VWM. Thus, even the most elementary operations of the oculomotor system are influenced by the interaction between VWM and perceptual processing, providing a plausible mechanism by which VWM can guide selection during visual search.

These recent findings provide a means to revive theories that explain oculomotor selection during natural search in terms of low-level visual salience (Itti and Koch 2000, 2001; Parkhurst et al. 2002). Visual salience has been modeled on the basis of the physical properties of local scene regions. Physical salience does not correlate strongly with task relevance, nor does it change systematically as goals change. If, as suggested above, visual salience is instead a joint property of an object's physical properties and their match with perceptual features maintained in VWM, then visual salience would be directly related to the features of task-relevant objects (maintained in VWM), and visual salience would vary systematically as goals change (and the content of VWM is modified to reflect the properties of newly relevant objects). This possibility is broadly consistent with models of search in which salience is modulated by the match between local scene regions and the target features maintained in a search template (Kanan et al. 2009). Of course, a modulating effect of VWM on visual salience would not necessarily provide a comprehensive explanation of guidance, as categorical knowledge and scene memory will nonetheless place strong constraints on selection during search through scenes.

Is template-based guidance plausible when searching for objects in natural environments? Natural search introduces significant hurdles to the generation and implementation of a search template. In the traditional visual search literature, the trial usually begins with the presentation of an image of the target, which simplifies the problem of guidance significantly: the target image typically matches the target exactly, containing all of the features of the target, and the perceptual representation of the target image can be loaded efficiently into VWM. But when we are looking for a particular object in a scene, rarely do we get a chance to see what we are looking for before we start. Instead, target features must be retrieved from LTM, which may limit both their precision and the extent to which those features are represented robustly in VWM. In addition, the visual features of an actual object in a scene will vary significantly as a result of idiosyncrasies in orientation, lighting, occlusion, distance, and so forth. It would be an extraordinary coincidence to obtain an exact visual match between the internal representation of the target (retrieved from memory) and any particular object in the scene. Moreover, if one is searching broadly for any object matching a particular category, such as "sugar bowl", variation in the perceptual properties of the exemplars that make up that category will also limit the precision of

the search template (Yang and Zelinsky 2009). Wolfe et al. (2011) identified several additional constraints on template use in natural search: (1) natural target objects have multiple feature values on multiple dimensions, and it is not clear that guidance can be implemented simultaneously for a complex set of features; (2) these values are often shared by many other objects in the scene, limiting their discriminative power; and (3) effects of template guidance may be quite small relative to guidance from scene knowledge (e.g., by knowledge of the where the object is likely to be found in the scene).

The literature on guidance in scenes suggests that, despite these limitations, template-based guidance is possible, is not necessarily overshadowed by the effects of scene knowledge, and occurs even when the template must be retrieved from LTM (Bravo and Farid 2009, 2012; Hollingworth and Matsukura 2011; Hwang et al. 2009; Malcolm and Henderson 2009, 2010; Pomplun 2006; Wolfe et al. 2011; Zelinsky 2008; Zelinsky et al. 1997). Malcolm and Henderson (2009) had participants search for objects in photographs of natural scenes, such as a kettle in a kitchen. They were provided either a verbal description of the target or a picture of the target immediately before scene onset (see Schmidt and Zelinsky 2009; Vickery et al. 2005; Wolfe et al. 2004). The availability of a precise visual template in the picture condition reduced the time taken to scan the scene for the target and reduced the time needed to verify a target match. The effect of visual guidance was observed above and beyond any guidance attributable to scene knowledge, as knowledge of the plausible location of a kettle in a kitchen was available in both conditions. Indeed, manipulation of both template precision and contextual constraint (i.e., the extent to which the target location was predictable on the basis of scene knowledge) indicated that the two sources of guidance combine additively (Malcolm and Henderson 2010).

Template-based guidance during natural tasks depends on the retrieval of features from LTM. Bravo and Farid (2009) trained participants to associate five photographs of tropical fish with five different species labels. They then had participants determine whether scenes of coral reefs did or did not contain a fish. Before each search, participants received a specific verbal cue (the label of one of the five species) or a non-specific cue (“fish”). In addition, the actual fish in the scene was either identical to a studied fish, a transformed version of the studied fish (rescaled, rotated, and mirror reversed), or a different exemplar from the same species. The use of verbal labels meant that participants had to retrieve visual details of the studied images from LTM. Relative to a non-specific cue, the specific cue facilitated search equally in the identical and transformed conditions but did not facilitate search at all in the different exemplar condition. That a specific verbal cue facilitated search at all demonstrates guidance from a visual template that was retrieved from LTM. The absence of facilitation in the different exemplar condition suggests that the retrieved template was specific to the particular fish image that had been studied. Because different exemplars of the same species had similar visual properties, the template must have been quite precise, retaining particular visual details of the studied image. However, the equivalence of facilitation in the identical and transformed conditions indicates that template guidance was not necessarily based on low-level image matching. The template generalized over metric differences in size and orientation. As discussed by

Bravo and Farid, such generalization makes a good deal of sense in dynamic, natural search, as the viewer's distance from the target and orientation with respect to the target are subject to change both within a search episode and across search episodes.

Converging evidence that LTM supports visual template guidance comes from Wolfe et al. (2011). In a repeated search paradigm, participants located natural objects in scenes on the basis of a verbal label. The second search for a particular object was much faster than the first search. Much of this improvement was attributable to learning the specific visual features of the target object. On the first search, participants had only a verbal label and did not know anything specific about the visual form of the object exemplar in the scene. Once the target had been found the first time, however, participants were able to encode its visual form into LTM. On the second search, retrieval of the target's features from LTM provided substantial guidance, improving search efficiency. This effect was observed above that attributable to learning the locations of the target objects.

Does retrieval from LTM lead to active maintenance of the search template in VWM? If so, then we can establish a bridge between the literature on VWM guidance in traditional search tasks and template guidance while searching for natural objects in scenes. In the memory-driven capture literature, capture is observed even if the memory item is specified by a label, such as "red" (Mannan et al. 2010; Soto and Humphreys 2007). In this case, activation in VWM is presumably driven by retrieval from LTM. Similarly, Olivers (2011) found that long-term perceptual associations for an item maintained in VWM influence search. Participants were given a verbal description of a canonically colored item (such as "stop sign"). They then searched for a grayscale version this item within an object array. Significantly greater distraction was generated by a colored distractor that matched the canonical color of the target (in this case, red) versus a colored distractor that did not, despite the fact that the colored distractor was never the target. Again, the feature value associated with the target must have been retrieved from LTM, and the capture effect indicates an interaction between the sustained activation of that feature in VWM and perceptual processing of the search display. Thus, it seems quite plausible that template features retrieved from LTM in the course of natural search would be represented in VWM and would interact with selection to increase the perceptual salience of scene regions matching VWM content.

Formal theories of search through naturalistic environments differ in their dependence on a search template. The original salience map models (Itti and Koch 2000; Parkhurst et al. 2002) contain no target representation and thus did not implement guidance on the basis of target features (but see Navalpakkam and Itti 2005). The contextual guidance model (Torralba et al. 2006) adds guidance by knowledge of plausible target locations to the salience-map account, but it likewise does not implement guidance by a search template (but see Ehinger et al. 2009). In contrast, several models of natural search depend centrally on a search template without introducing constraints from scene knowledge (Hwang et al. 2009; Zelinsky 2008). For example, in Zelinsky's target acquisition model, the search template, maintained in memory, is a multidimensional feature vector derived from an exact image representation of

the region of the scene containing the target. During search, this feature vector is correlated with feature vectors computed at each pixel location within the search image (the scene vectors are strongly influenced by fixation position within the scene and thus vary with each eye movement). The “salience” of each location in the search image is then a function of the extent to which features present in the template are present at that location and features absent from the template are absent from that location. Saccades are directed to regions of high correlation.

The target acquisition model (Zelinsky 2008) has been highly adept at capturing key features of human eye movements during search through scenes. Its assumption that a memory representation of the target features is combined with perceptual information to bias selection toward matching regions of the scene is consistent with the general claims of VWM-based guidance discussed above. However, the type of template specified in the target acquisition model and in other image-based models (Hwang et al. 2009; Pomplun 2006) would be difficult to integrate with this view. VWM representations are not imagistic; rather, they are abstracted away from the precise, metric structure of early vision (Irwin 1991; Phillips 1974). The only visual memory system capable of maintaining the type of low-level image representation specified in these models is sensory persistence (Averbach and Coriell 1961; Coltheart 1980; Sperling 1960), but sensory persistence decays within less than 500 ms of a perceptual event, so it could not plausibly support template maintenance during natural search. In addition, templates will rarely be derived from precise image representations, as they will need to be retrieved from LTM, and LTM representations, like VWM representations, are not imagistic (for a review, see Hollingworth 2008). Instead, template-guidance is likely to depend on visual representations that are abstracted away from the precise metric properties of early vision, allowing generalization over image-level variation within and across searches (e.g., Bravo and Farid 2009).

## Online Memory for the Search: Inhibition of Return

The efficiency of search for objects in scenes would be improved by a mechanism that prevents attention and gaze from returning to previously fixated objects. Posner and Cohen (1984) discovered that target detection at a peripherally cued location is impaired starting approximately 300 ms after a neutral cue. This inhibition of return (IOR) effect has been taken to suggest that attention is biased against returning to a previously attended location (but see Berlucchi 2006; Posner and Cohen 1984 for alternative accounts). Salience models of visual search make this assumption explicit, with an IOR mechanism that deprioritizes previously fixated regions so that gaze does not cycle continuously between the most visually salient regions of the scene (Itti and Koch 2001). More generally, IOR has been proposed to act as a foraging facilitator in complex environments, ensuring that attention is consistently oriented to new locations that might contain the target (for a review, see Wang and Klein 2010).

The most direct evidence that IOR plays a functional role in the task of finding objects in scenes comes from a study by Klein and MacInnes (1999). Participants completed a very difficult visual search task: Finding well-camouflaged Waldos in “Where’s Waldo?” images. At a pre-specified point during search, a dot onset was introduced, and participants executed a saccade to the onset dot as rapidly as possible. Oculomotor IOR was observed as increased saccade latency to targets that appeared in the reverse direction from the preceding saccade (i.e., back to a previously fixated location) relative to targets that appeared in the same direction as the preceding saccade (i.e., toward a new fixation location). This basic finding has been replicated several times (Dodd et al. 2009; Smith and Henderson 2011a, b). In addition, the duration of a fixation before a saccade to a new location tends to be shorter than the duration preceding a saccade that returns gaze to a just-fixated location (Hooge et al. 2005).

A functional account of IOR is bolstered by two additional lines of evidence. First, oculomotor IOR is strongly tied to the particular scene context in which the search occurs. If the scene background is removed immediately before the onset dot, the latency deficit for saccades to previously fixated locations is eliminated (Klein and MacInnes 1999; see also Müller and Von Mühlenen 2000; Takeda and Yagi 2000). Second, the presence of IOR is strongly tied to the demands of the task. Dodd et al. (2009) modified the Klein and MacInnes task with a manipulation of viewing instructions. Participants either searched for a small letter target within scenes, memorized the scenes, assessed the aesthetics of the scenes, or viewed freely. Orienting latency to onset targets was measured at previously fixated locations and new locations. Only the search task placed a premium on inhibiting orienting to previously fixated locations, and only in the search task was IOR observed. Saccades to old locations were actually facilitated under the other task conditions, a phenomenon of *facilitation of return*. Thus, IOR appears to be tied directly to the demands of visual search.

The phenomenon of oculomotor IOR—increased saccadic latency to targets at previously fixated locations—is well established, but its functional relevance to natural search depends on whether this effect reduces the probability that a scene location will be refixated under free viewing conditions (Hooge et al. 2005). That is, the latency deficit must be indicative of a spatial bias against directing saccades back to previously fixated regions (i.e., making refixations). Such biases certainly exist. Refixations during visual search through scenes are generally rare, much lower than would be expected by chance (Gilchrist et al. 2001; for similar results using abstract arrays, see Peterson et al. 2001). Moreover, refixations are less frequent during visual search than during non-search tasks (Dodd et al. 2009). However, it has not been demonstrated that these large-scale biases against refixation during search are caused by the same IOR mechanism that produces latency differences, and several recent studies suggest that they may not be.

At question is whether the increased latency of return saccades (that reverse a saccade just executed) is accompanied by a reduction in the probability that a return saccade will be executed. Several studies have now found that return saccades are generated more often than would be expected by chance, indicating a bias to

direct gaze *back to*, rather than *away from*, the most recently fixated item, and researchers have interpreted this finding as evidence against the idea that IOR interacts significantly with the selection of saccade locations to facilitate foraging (Hooge et al. 2005; Smith and Henderson 2009, 2011a, b).<sup>1</sup> The relatively high frequency of return saccades eliminates the possibility that IOR dominates other sources of constraint to ensure that the eyes are *always* directed to new objects. It is, however, possible that an immediate return saccade is a special case. Return saccades are quite likely to reflect incomplete processing of the target during the initial fixation. Gaze is returned relatively rapidly (Hooge et al. 2005) to the object so as to compete the acquisition of information. This does not violate the general goal that saccades should be biased away from previously examined objects, because the immediate return can be considered as completion of the process of initial examination. Diagnostic evidence would come from objects fixated two or more fixations earlier, when the processing of the fixated object has clearly been completed. Smith and Henderson (2011b) found that the probability of return to a 2-back location was not reduced relative to other locations within the scene, and thus there is currently no evidence that IOR facilitates foraging by reducing the probability of refixation.

It is still the case that over the entire course of search, refixations are relatively rare. There must be *some* memory mechanism that implements this behavioral bias. Although IOR might contribute, it is unlikely to be the only mechanism. Oculomotor IOR is observed for, at most, four preceding fixations (Dodd et al. 2009) and appears to depend on the VWM system (Castel et al. 2003), which has a severely limited capacity. If IOR is found, ultimately, to cause reduced refixation probability, that bias could be operational for only a few of objects over a couple of seconds. Yet, real-world searches often unfold over many seconds of exploration. A bias to avoid refixation across extended search cannot plausibly depend on IOR alone. Instead, it is likely that there are longer-term learning mechanisms that serve this purpose. The memory mechanisms supporting longer-term biases during search through scenes have received no systematic study whatsoever.

In summary, it still remains possible that IOR plays a functional role in the task of finding objects in scenes. But there is currently no positive evidence that IOR is associated with a reduced probability of refixation, and tests of the immediate return of gaze suggest that return is actually more likely than expected by chance. Without evidence of a bias against refixations, there is currently no direct link between the IOR phenomenon and the facilitation of search within scenes, and thus no direct support for the foraging facilitator hypothesis. Moreover, any influence of IOR on the selection of saccade targets must be limited, as it does not appear to prevent immediate returns, and the temporal range of IOR is quite brief (four fixations at most). Biases against refixation during extended search events must therefore depend on longer-term memory mechanisms.

---

<sup>1</sup> Smith and Henderson (2009) use the term “facilitation of return” to refer to the finding that return saccades are observed more frequently than expected by chance. Dodd et al. (2009) use “facilitation of return” to refer to the finding that, in non-search tasks, saccades are generated more quickly to abrupt onsets at previously fixated locations than at new locations.

## Conclusions

Visual search is one of the most common behaviors in which humans engage. When looking for objects in scenes, multiple sources of guidance are used to direct attention and gaze to the target location efficiently, including categorical knowledge of plausible object positions in scenes, memory for the particular environment in which search occurs, and knowledge of the perceptual features that characterize the desired object. In addition, online memory for inspected locations is likely to improve efficiency by biasing attention and gaze to new locations. These various sources of guidance appear to be complementary (Castelhano and Heaven 2010; Ehinger et al. 2009; Malcolm and Henderson 2010; Wolfe et al. 2011), and future work will need to account for the manner in which they jointly contribute to search efficiency.

## References

- Averbach, E., & Coriell, A. S. (1961). Short-term memory in vision. *Bell System Technical Journal*, 40(1), 309–328.
- Ballard, D. H., Hayhoe, M. M., & Pelz, J. B. (1995). Memory representations in natural tasks. *Journal of Cognitive Neuroscience*, 7(1), 66–80. doi: 10.1162/jocn.1995.7.1.66.
- Ballard, D. H., Hayhoe, M. M., Pook, P. K., & Rao, R. P. (1997). Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences*, 20(4), 723–767.
- Beck, V. M., Hollingworth, A., & Luck, S. J. (2012). Simultaneous control of attention by multiple working memory representations. *Psychological Science*, 23(8), 887–898. doi: 10.1177/0956797612439068
- Becker, M. W., & Rasmussen, I. P. (2008). Guidance of attention to objects and locations by long-term memory of natural scenes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(6), 1325–1338. doi: 10.1037/a0013650.
- Berlucchi, G. (2006). Inhibition of return: A phenomenon in search of a mechanism and a better name. *Cognitive Neuropsychology*, 23(7), 1065–1074. doi: 10.1080/02643290600588426.
- Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: Detecting and judging objects undergoing relational violations. *Cognitive Psychology*, 14(2), 143–177. doi: 10.1016/0010-0285(82)90007-X.
- Brady, T. F., & Chun, M. M. (2007). Spatial constraints on learning in visual search: Modeling contextual cuing. *Journal of Experimental Psychology: Human Perception and Performance*, 33(4), 798–815. doi: 10.1037/0096-1523.33.4.798.
- Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences of the United States of America*, 105(38), 14325–14329. doi: 10.1073/pnas.0803390105.
- Bravo, M. J., & Farid, H. (2009). The specificity of the search template. *Journal of Vision*, 9(1), 34, 31–39. doi: 10.1167/9.1.34.
- Bravo, M. J., & Farid, H. (2012). Task demands determine the specificity of the search template. *Attention, Perception, & Psychophysics*, 74(1), 124–131. doi: 10.3758/s13414-011-0224-5.
- Brockmole, J. R., Castelhano, M. S., & Henderson, J. M. (2006). Contextual cueing in naturalistic scenes: Global and local contexts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32(4), 699–706. doi: 10.1037/0278-7393.32.4.699.
- Brockmole, J. R., & Henderson, J. M. (2006a). Recognition and attention guidance during contextual cueing in real-world scenes: Evidence from eye movements. *Quarterly Journal of Experimental Psychology*, 59(7), 1177–1187. doi: 10.1080/17470210600665996.

- Brockmole, J. R., & Henderson, J. M. (2006b). Using real-world scenes as contextual cues for search. *Visual Cognition*, *13*(1), 99–108. doi: 10.1080/13506280500165188.
- Brooks, D. I., Rasmussen, I. P., & Hollingworth, A. (2010). The nesting of search contexts within natural scenes: Evidence from contextual cuing. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(6), 1406–1418. doi: 10.1037/a0019257.
- Bundesen, C. (1990). A theory of visual attention. *Psychological Review*, *97*(4), 523–547. doi: 10.1037/0033-295X.97.4.523.
- Bundesen, C., Habekost, T., & Kyllingsbaek, S. (2005). A neural theory of visual attention: Bridging cognition and neurophysiology. *Psychological Review*, *112*(2), 291–328. doi: 10.1037/0033-295X.112.2.291.
- Castel, A. D., Pratt, J., & Craik, F. I. M. (2003). The role of spatial working memory in inhibition of return: Evidence from divided attention tasks. *Perception & Psychophysics*, *65*(6), 970–981. doi: 10.3758/BF03194827.
- Castelhano, M. S., & Heaven, C. (2010). The relative contribution of scene context and target features to visual search in scenes. *Attention, Perception, & Psychophysics*, *72*(5), 1283–1297. doi: 10.3758/app.72.5.1283.
- Castelhano, M. S., & Heaven, C. (2011). Scene context influences without scene gist: Eye movements guided by spatial associations in visual search. *Psychonomic Bulletin & Review*, *18*(5), 890–896. doi: 10.3758/s13423-011-0107-8.
- Castelhano, M. S., & Henderson, J. M. (2005). Incidental visual memory for objects in scenes. *Visual Cognition*, *12*(6), 1017–1040. doi: 10.1080/13506280444000634.
- Castelhano, M. S., & Henderson, J. M. (2007). Initial scene representations facilitate eye movement guidance in visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(4), 753–763. doi: 10.1037/0096-1523.33.4.753.
- Chun, M. M. (2000). Contextual cueing of visual attention. *Trends in Cognitive Sciences*, *4*(5), 170–178. doi: 10.1016/S1364-6613(00)01476-5.
- Chun, M. M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology*, *36*(1), 28–71. doi: 10.1006/cogp.1998.0681.
- Chun, M. M., & Turk-Browne, N. B. (2008). Associative learning mechanisms in vision. In S. J. Luck, & A. Hollingworth (Eds.), *Visual memory* (pp. 209–246). New York: Oxford University Press.
- Coltheart, M. (1980). The persistences of vision. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *290*(1038), 269–294 (1980). doi: 10.1098/rstb.1980.0082.
- Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*, 193–222. doi: 10.1146/annurev.ne.18.030195.001205.
- Dodd, M. D., Van der Stigchel, S., & Hollingworth, A. (2009). Novelty is not always the best policy: Inhibition of return and facilitation of return as a function of visual task. *Psychological Science*, *20*(3), 333–339. doi: 10.1111/j.1467-9280.2009.02294.x.
- Downing, P. E., & Dodds, C. M. (2004). Competition in visual working memory for control of search. *Visual Cognition*, *11*(6), 689–703. doi: 10.1080/13506280344000446.
- Droll, J. A., Hayhoe, M. M., Triesch, J., & Sullivan, B. T. (2005). Task demands control acquisition and storage of visual information. *Journal of Experimental Psychology: Human Perception and Performance*, *31*(6), 1416–1438. doi: 10.1037/0096-1523.31.6.1416.
- Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, *96*(3), 433–458. doi: 10.1037//0033-295X.96.3.433.
- Eckstein, M. P. (2011). Visual search: A retrospective. *Journal of Vision*, *11*(5), 14: 11–36. doi: 10.1167/11.5.14.
- Eckstein, M. P., Drescher, B. A., & Shimozaki, S. S. (2006). Attentional cues in real scenes, saccadic targeting, and Bayesian priors. *Psychological Science*, *17*(11), 973–980. doi: 10.1111/j.1467-9280.2006.01815.x.
- Ehinger, K. A., & Brockmole, J. R. (2008). The role of color in visual search in real-world scenes: Evidence from contextual cuing. *Perception & Psychophysics*, *70*(7), 1366–1378. doi: 10.3758/pp.70.7.1366.

- Ehinger, K. A., Hidalgo-Sotelo, B., Torralba, A., & Oliva, A. (2009). Modelling search for people in 900 scenes: A combined source model of eye guidance. *Visual Cognition*, *17*(6–7), 945–978. doi: 10.1080/13506280902834720.
- Einhauser, W., Rutishauser, U., & Koch, C. (2008a). Task-demands can immediately reverse the effects of sensory-driven saliency in complex visual stimuli. *Journal of Vision*, *8*(2), 2: 1–19. doi: 10.1167/8.2.2.
- Einhauser, W., Spain, M., & Perona, P. (2008b). Objects predict fixations better than early saliency. *Journal of Vision*, *8*(14), 18: 11–26. doi: 10.1167/8.14.18.
- Findlay, J. M., & Gilchrist, I. D. (2003). *Active vision: The psychology of looking and seeing*. Oxford: Oxford University Press.
- Foulsham, T., Alan, R., & Kingstone, A. (2011). Scrambled eyes? Disrupting scene structure impedes focal processing and increases bottom-up guidance. *Attention, Perception, & Psychophysics*, *73*(7), 2008–2025. doi: 10.3758/s13414-011-0158-y.
- Foulsham, T., & Underwood, G. (2007). How does the purpose of inspection influence the potency of visual salience in scene perception? *Perception*, *36*(8), 1123–1138. doi: 10.1068/p5659.
- Foulsham, T., & Underwood, G. (2008). What can saliency models predict about eye movements? Spatial and sequential aspects of fixations during encoding and recognition. *Journal of Vision*, *8*(2), 6: 1–17. doi: 10.1167/8.2.6.
- Gibson, B. S., Li, L., Skow, E., Brown, K., & Cooke, L. (2000). Searching for one versus two identical targets: When visual search has a memory. *Psychological Science*, *11*(4), 324–327. doi: 10.1111/1467-9280.00264.
- Gilchrist, I. D., North, A., & Hood, B. (2001). Is visual search really like foraging? *Perception*, *30*(12), 1459–1464. doi: 10.1068/p3249.
- Han, S. W., & Kim, M. S. (2009). Do the contents of working memory capture attention? Yes, but cognitive control matters. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(5), 1292–1302. doi: 10.1037/a0016452.
- Harrison, S. A., & Tong, F. (2009). Decoding reveals the contents of visual working memory in early visual areas. *Nature*, *458*(7238), 632–635. doi: 10.1038/nature07832.
- Hayhoe, M. (2000). Vision using routines: A functional account of vision. *Visual Cognition*, *7*(1–3), 43–64. doi: 1080/135062800394676.
- Henderson, J. M. (2003). Human gaze control during real-world scene perception. *Trends in Cognitive Sciences*, *7*(11), 498–504. doi: 10.1016/j.tics.2003.09.006.
- Henderson, J. M., Brockmole, J. R., Castelano, M. S., & Mack, M. (2007). Visual saliency does not account for eye movements during search in real-world scenes. In R. van Gompel, M. Fischer, W. Murray, & R. Hill (Eds.), *Eye movements: A window on mind and brain* (pp. 537–562). Oxford: Elsevier.
- Henderson, J. M., & Hollingworth, A. (1999). High-level scene perception. *Annual Review of Psychology*, *50*, 243–271. doi: 10.1146/annurev.psych.50.1.243.
- Henderson, J. M., Malcolm, G. L., & Schandl, C. (2009). Searching in the dark: Cognitive relevance drives attention in real-world scenes. *Psychonomic Bulletin & Review*, *16*(5), 850–856. doi: 10.3758/pbr.16.5.850.
- Henderson, J. M., Weeks, P. A., & Hollingworth, A. (1999). The effects of semantic consistency on eye movements during complex scene viewing. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(1), 210–228. doi: 10.1037/0096-1523.25.1.210.
- Hillstrom, A. P., Scholey, H., Liversedge, S. P., & Benson, V. (2012). The effect of the first glimpse at a scene on eye movements during search. *Psychonomic Bulletin & Review*, *19*(2), 204–210. doi: 10.3758/s13423-011-0205-7.
- Hollingworth, A. (2004). Constructing visual representations of natural scenes: The roles of short- and long-term visual memory. *Journal of Experimental Psychology: Human Perception and Performance*, *30*(3), 519–537. doi: 10.1037/0096-1523.30.3.519.
- Hollingworth, A. (2005). The relationship between online visual representation of a scene and long-term scene memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(3), 396–411. doi: 10.1037/0278-7393.31.3.396.

- Hollingworth, A. (2006). Visual memory for natural scenes: Evidence from change detection and visual search. *Visual Cognition, 14*(4–8), 781–807. doi: 10.1080/13506280500193818.
- Hollingworth, A. (2008). Visual memory for natural scenes. In S. J. Luck, & A. Hollingworth (Eds.), *Visual memory* (pp. 123–162). New York: Oxford University Press.
- Hollingworth, A. (2009). 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, 17*(1–2), 273–291. doi: 10.1080/13506280802193367.
- Hollingworth, A. (in press). Task specificity and the influence of memory on visual search: Commentary on Vö and Wolfe (2012). *Journal of Experimental Psychology: Human Perception and Performance*.
- Hollingworth, A., & Henderson, J. M. (2002). Accurate visual memory for previously attended objects in natural scenes. *Journal of Experimental Psychology: Human Perception and Performance, 28*(1), 113–136. doi: 10.1037//0096-1523.28.1.113.
- Hollingworth, A., & Luck, S. J. (2009). The role of visual working memory (VWM) in the control of gaze during visual search. *Attention, Perception, & Psychophysics, 71*(4), 936–949. doi: 10.3758/APP.71.4.936.
- Hollingworth, A., & Matsukura, M. (2011). *Visual working memory content biases the allocation of gaze during search through natural scenes*. Paper presented at the Annual Meeting of the Psychonomic Society, Seattle, WA.
- Hollingworth, A., Matsukura, M., & Luck, S. J. (in press). Visual working memory modulates rapid eye movements to simple onset targets. *Psychological Science*.
- Hollingworth, A., & Maxcey-Richard, A. M. (in press). Selective maintenance in visual working memory does not require sustained visual attention. *Journal of Experimental Psychology: Human Perception and Performance*.
- Hollingworth, A., Richard, A. M., & Luck, S. J. (2008). Understanding the function of visual short-term memory: Transsaccadic memory, object correspondence, and gaze correction. *Journal of Experimental Psychology: General, 137*(1), 163–181. doi: 10.1037/0096-3445.137.1.163.
- Hooge, I. T. C., Over, E. A. B., van Wezel, R. J. A., & Frens, M. A. (2005). Inhibition of return is not a foraging facilitator in saccadic search and free viewing. *Vision Research, 45*(14), 1901–1908. doi: 10.1016/j.visres.2005.01.030.
- Horowitz, T. S., & Wolfe, J. M. (1998). Visual search has no memory. *Nature, 394*(6693), 575–577. doi: 10.1038/29068.
- Howard, C. J., Pharaon, R. G., Körner, C., Smith, A. D., & Gilchrist, I. D. (2011). Visual search in the real world: Evidence for the formation of distractor representations. *Perception, 40*(10), 1143–1153. doi: 10.1068/p7088.
- Hwang, A. D., Higgins, E. C., & Pomplun, M. (2009). A model of top-down attentional control during visual search in complex scenes. *Journal of Vision, 9*(5), 25: 21–18. doi: 10.1167/9.5.25.
- Irwin, D. E. (1991). Information integration across saccadic eye movements. *Cognitive Psychology, 23*(3), 420–456. doi: 10.1016/0010-0285(91)90015-G.
- Itti, L. (2005). Quantifying the contribution of low-level saliency to human eye movements in dynamic scenes. *Visual Cognition, 12*(6), 1093–1123. doi: 10.1080/13506280444000661.
- Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research, 40*(10–12), 1489–1506. doi: 10.1016/S0042-6989(99)00163-7.
- Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience, 2*(3), 194–203. doi: 10.1038/35058500.
- Kanan, C., Tong, M. H., Zhang, L., & Cottrell, G. W. (2009). SUN: Top-down saliency using natural statistics. *Visual Cognition, 17*(6–7), 979–1003. doi: 10.1080/13506280902771138.
- Kang, M. S., Hong, S. W., Blake, R., & Woodman, G. F. (2011). Visual working memory contaminates perception. *Psychonomic Bulletin & Review, 18*(5), 860–869. doi: 10.3758/s13423-011-0126-5.
- Klein, R. M. (1988). Inhibitory tagging system facilitates visual search. *Nature, 334*(6181), 430–431. doi: 10.1038/334430a0.
- Klein, R. M., & MacInnes, W. J. (1999). Inhibition of return is a foraging facilitator in visual search. *Psychological Science, 10*(4), 346–352. doi: 10.1111/1467-9280.00166.

- Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010). Scene memory is more detailed than you think: The role of categories in visual long-term memory. *Psychological Science, 21*(11), 1551–1556. doi: 10.1177/0956797610385359.
- Körner, C., & Gilchrist, I. D. (2007). Finding a new target in an old display: Evidence for a memory recency effect in visual search. *Psychonomic Bulletin & Review, 14*(5), 846–851. doi: 10.3758/bf03194110.
- Kristjánsson, A. (2000). In search of remembrance: Evidence for memory in visual search. *Psychological Science, 11*(4), 328–332. doi: 10.1111/1467-9280.00265.
- Land, M. F., & Hayhoe, M. (2001). In what ways do eye movements contribute to everyday activities? *Vision Research, 41*(25–26), 3559–3565. doi: 10.1016/S0042-6989(01)00102-X.
- Land, M. F., Mennie, N., & Rusted, J. (1999). The roles of vision and eye movements in the control of activities of daily living. *Perception, 28*(11), 1311–1328. doi: 10.1068/p2935.
- Ludwig, C. J. H., & Gilchrist, I. D. (2002). Stimulus-driven and goal-driven control over visual selection. *Journal of Experimental Psychology: Human Perception and Performance, 28*(4), 902–912. doi: 10.1037//0096-1523.28.4.902.
- Mack, S. C., & Eckstein, M. P. (2011). Object co-occurrence serves as a contextual cue to guide and facilitate visual search in a natural viewing environment. *Journal of Vision, 11*(9), 9: 1–16. doi: 10.1167/11.9.9.
- Malcolm, G. L., & Henderson, J. M. (2009). The effects of target template specificity on visual search in real-world scenes: Evidence from eye movements. *Journal of Vision, 9*(11), 8: 1–13. doi: 10.1167/9.11.8.
- Malcolm, G. L., & Henderson, J. M. (2010). Combining top-down processes to guide eye movements during real-world scene search. *Journal of Vision, 10*(2), 4: 1–11. doi: 10.1167/10.2.4.
- Mannan, S. K., Kennard, C., Potter, D., Pan, Y., & Soto, D. (2010). Early oculomotor capture by new onsets driven by the contents of working memory. *Vision Research, 50*(16), 1590–1597. doi: 10.1016/j.visres.2010.05.015.
- McCarley, J. S., Wang, R. X. F., Kramer, A. F., Irwin, D. E., & Peterson, M. S. (2003). How much memory does oculomotor search have? *Psychological Science, 14*(5), 422–426. doi: 10.1111/1467-9280.01457.
- McPeck, R. M., & Keller, E. L. (2002). Superior colliculus activity related to concurrent processing of saccade goals in a visual search task. *Journal of Neurophysiology, 87*(4), 1805–1815. doi: 10.1152/jn.00501.2001.
- Müller, H. J., & Von Mühlénen, A. (2000). Probing distractor inhibition in visual search: Inhibition of return. *Journal of Experimental Psychology: Human Perception and Performance, 26*(5), 1591–1605. doi: 10.1037/0096-1523.26.5.1591.
- Navalpakkam, V., & Itti, L. (2005). Modeling the influence of task on attention. *Vision Research, 45*(2), 205–231. doi: 10.1016/j.visres.2004.07.042.
- Neider, M. B., & Zelinsky, G. J. (2006). Scene context guides eye movements during visual search. *Vision Research, 46*(5), 614–621. doi: 10.1016/j.visres.2005.08.025.
- Olivers, C. N. L. (2009). What drives memory-driven attentional capture? The effects of memory type, display type, and search type. *Journal of Experimental Psychology: Human Perception and Performance, 35*(5), 1275–1291. doi: 10.1037/a0013896.
- Olivers, C. N. L. (2011). Long-term visual associations affect attentional guidance. *Acta Psychologica, 137*(2), 243–247. doi: 10.1016/j.actpsy.2010.07.001.
- Olivers, C. N. L., Meijer, F., & Theeuwes, J. (2006). Feature-based memory-driven attentional capture: Visual working memory content affects visual attention. *Journal of Experimental Psychology: Human Perception and Performance, 32*(5), 1243–1265. doi: 10.1037/0096-1523.32.5.1243.
- Olson, I. R., & Chun, M. M. (2002). Perceptual constraints on implicit learning of spatial context. *Visual Cognition, 9*(3), 273–302. doi: 10.1080/13506280042000162.
- Parkhurst, D., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision Research, 42*(1), 107–123. doi: 10.1016/S0042-6989(01)00250-4.

- Pearson, J., Clifford, C. W. G., & Tong, F. (2008). The functional impact of mental imagery on conscious perception. *Current Biology*, *18*(13), 982–986. doi: 10.1016/j.cub.2008.05.048.
- Peterson, M. S., Kramer, A. F., Wang, R. F., Irwin, D. E., & McCarley, J. S. (2001). Visual search has memory. *Psychological Science*, *12*(4), 287–292. doi: 10.1111/1467-9280.00353.
- Phillips, W. A. (1974). On the distinction between sensory storage and short-term visual memory. *Perception & Psychophysics*, *16*(2), 283–290. doi: 10.3758/BF03203943.
- Pomplun, M. (2006). Saccadic selectivity in complex visual search displays. *Vision Research*, *46*(12), 1886–1900. doi: 10.1016/j.visres.2005.12.003.
- Posner, M. I., & Cohen, Y. (1984). Components of visual orienting. In H. Bouma, & D. G. Bouwhuis (Eds.), *Attention and performance X* (pp. 531–556). Hillsdale: Erlbaum.
- Schmidt, J., & Zelinsky, G. J. (2009). Search guidance is proportional to the categorical specificity of a target cue. *Quarterly Journal of Experimental Psychology*, *62*(10), 1904–1914. doi: 10.1080/17470210902853530.
- Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information-processing: I. Detection, search, and attention. *Psychological Review*, *84*(1), 1–66. doi: 10.1037/0033-295x.84.1.1.
- Serences, J. T., Ester, E. F., Vogel, E. K., & Awh, E. (2009). Stimulus-specific delay activity in human primary visual cortex. *Psychological Science*, *20*(2), 207–214. doi: 10.1111/j.1467-9280.2009.02276.x.
- Shore, D. I., & Klein, R. M. (2000). On the manifestations of memory in visual search. *Spatial Vision*, *14*(1), 59–75. doi: 10.1163/156856801741369.
- Smith, T. J., & Henderson, J. M. (2009). Facilitation of return during scene viewing. *Visual Cognition*, *17*(6–7), 1083–1108. doi: 10.1080/13506280802678557.
- Smith, T. J., & Henderson, J. M. (2011a). Does oculomotor inhibition of return influence fixation probability during scene search? *Attention, Perception, & Psychophysics*, *73*(8), 2384–2398 (2011a). doi: 10.3758/s13414-011-0191-x.
- Smith, T. J., & Henderson, J. M. (2011b). Looking back at Waldo: Oculomotor inhibition of return does not prevent return fixations. *Journal of Vision*, *11*(1), 3: 1–11 (2011b). doi: 10.1167/11.1.3.
- Soto, D., Heinke, D., Humphreys, G. W., & Blanco, M. J. (2005). Early, involuntary top-down guidance of attention from working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *31*(2), 248–261. doi: 10.1037/0096-1523.31.2.248.
- Soto, D., & Humphreys, G. W. (2007). Automatic guidance of visual attention from verbal working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(3), 730–737. doi: 10.1037/0096-1523.33.3.730.
- Soto, D., Humphreys, G. W., & Heinke, D. (2006). Working memory can guide pop-out search. *Vision Research*, *46*(6–7), 1010–1018. doi: 10.1016/j.visres.2005.09.008.
- Soto, D., Wriglesworth, A., Bahrami-Balani, A., & Humphreys, G. W. (2010). Working memory enhances visual perception: Evidence from signal detection analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(2), 441–456. doi: 10.1037/a0018686.
- Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs*, *74*, (11, Whole no. 498).
- Standing, L. (1973). Learning 10,000 pictures. *Quarterly Journal of Experimental Psychology*, *25*(2), 207–222. doi: 10.1080/14640747308400340.
- Standing, L., Conezio, J., & Haber, R. N. (1970). Perception and memory for pictures: Single-trial learning of 2500 visual stimuli. *Psychonomic Science*, *19*(2), 73–74.
- Takeda, Y., & Yagi, A. (2000). Inhibitory tagging in visual search can be found if search stimuli remain visible. *Perception & Psychophysics*, *62*(5), 927–934 (2000). doi: 10.3758/BF03212078.
- Tatler, B. W., Hayhoe, M. M., Land, M. F., & Ballard, D. H. (2011). Eye guidance in natural vision: Reinterpreting salience. *Journal of Vision*, *11*(5), 5: 1–23. doi: 10.1167/11.5.5.
- Torralba, A., Oliva, A., Castelhano, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. *Psychological Review*, *113*(4), 766–786. doi: 10.1037/0033-295X.113.4.766.

- Treisman, A., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*(1), 97–136. doi: 10.1016/0010-0285(80)90005-5.
- Underwood, G., Foulsham, T., & Humphrey, K. (2009). Saliency and scan patterns in the inspection of real-world scenes: Eye movements during encoding and recognition. *Visual Cognition*, *17*(6–7), 812–834. doi: 10.1080/13506280902771278.
- van Zoest, W., Donk, M., & Theeuwes, J. (2004). The role of stimulus-driven and goal-driven control in saccadic visual selection. *Journal of Experimental Psychology: Human Perception and Performance*, *30*(4), 746–759. doi: 10.1037/0096-1523.30.4.746.
- Vickery, T. J., King, L. W., & Jiang, Y. (2005). Setting up the target template in visual search. *Journal of Vision*, *5*(1), 81–92. doi: 10.1167/5.1.8.
- Võ, M. L. H., & Henderson, J. M. (2010). The time course of initial scene processing for eye movement guidance in natural scene search. *Journal of Vision*, *10*(3), 14: 11–13. doi: 10.1167/10.3.14.
- Võ, M. L. H., & Wolfe, J. M. (2012). When Does Repeated Search in Scenes Involve Memory? Looking At Versus Looking For Objects in Scenes. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(1), 23–41. doi: 10.1037/a0024147.
- Vogel, E. K., Woodman, G. F., & Luck, S. J. (2006). The time course of consolidation in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(6), 1436–1451. doi: 10.1037/0096-1523.32.6.1436.
- von Mühlenen, A., Müller, H. J., & Müller, D. (2003). Sit-and-wait strategies in dynamic visual search. *Psychological Science*, *14*(4), 309–314. doi: 10.1111/1467-9280.14441.
- Wang, Z., & Klein, R. M. (2010). Searching for inhibition of return in visual search: A review. *Vision Research*, *50*(2), 220–228. doi: 10.1016/j.visres.2009.11.013.
- Weierich, M. R., Treat, T. A., & Hollingworth, A. (2008). Theories and measurement of visual attentional processing in anxiety. *Cognition & Emotion*, *22*(6), 985–1018. doi: 10.1080/02699930701597601.
- Williams, C. C., Henderson, J. M., & Zacks, R. T. (2005). Incidental visual memory for targets and distractors in visual search. *Perception & Psychophysics*, *67*(5), 816–827. doi: 10.3758/BF03193535.
- Wolfe, J. M. (1994). Guided Search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review*, *1*(2), 202–238. doi: 10.3758/bf03200774.
- Wolfe, J. M. (1999). Inattentive amnesia. In V. Coltheart (Ed.), *Fleeting memories* (pp. 71–94). Cambridge: MIT Press.
- Wolfe, J. M., Alvarez, G. A., Rosenholtz, R., Kuzmova, Y. I., & Sherman, A. M. (2011). Visual search for arbitrary objects in real scenes. *Attention, Perception, & Psychophysics*, *73*(6), 1650–1671 (2011). doi: 10.3758/s13414-011-0153-3.
- Wolfe, J. M., Horowitz, T. S., Kenner, N., Hyle, M., & Vasan, N. (2004). How fast can you change your mind? The speed of top-down guidance in visual search. *Vision Research*, *44*(12), 1411–1426. doi: 10.1016/j.visres.2003.11.024.
- Wolfe, J. M., Klempe, N., & Dahlen, K. (2000). Postattentive vision. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(2), 693–716. doi: 10.1037/0096-1523.26.2.693.
- Wolfe, J. M., Võ, M. L., Evans, K. K., & Greene, M. R. (2011). Visual search in scenes involves selective and nonselective pathways. *Trends in Cognitive Sciences*, *15*(2), 77–84. doi: 10.1016/j.tics.2010.12.001.
- Woodman, G. F., & Arita, J. T. (2011). Direct electrophysiological measurement of attentional templates in visual working memory. *Psychological Science*, *22*(2), 212–215. doi: 10.1177/0956797610395395.
- Woodman, G. F., & Luck, S. J. (2004). Visual search is slowed when visuospatial working memory is occupied. *Psychonomic Bulletin & Review*, *11*(2), 269–274. doi: 10.3758/BF03196569.
- Woodman, G. F., & Luck, S. J. (2007). Do the contents of visual working memory automatically influence attentional selection during visual search? *Journal of Experimental Psychology: Human Perception and Performance*, *33*(2), 363–377. doi: 10.1037/0096-1523.33.2.363.

- Woodman, G. F., Luck, S. J., & Schall, J. D. (2007). The role of working memory representations in the control of attention. *Cerebral Cortex, 17*, 118–124. doi: 10.1093/cercor/bhm065.
- Yang, H., & Zelinsky, G. J. (2009). Visual search is guided to categorically-defined targets. *Vision Research, 49*(16), 2095–2103. doi: 10.1016/j.visres.2009.05.017.
- Zelinsky, G. J. (1996). Using eye saccades to assess the selectivity of search movements. *Vision Research, 36*(14), 2177–2187. doi: 10.1016/0042-6989(95)00300-2.
- Zelinsky, G. J. (2008). A theory of eye movements during target acquisition. *Psychological Review, 115*(4), 787–835. doi: 10.1037/a0013118.
- Zelinsky, G. J., Rao, R. P. N., Hayhoe, M. M., & Ballard, D. H. (1997). Eye movements reveal the spatiotemporal dynamics of visual search. *Psychological Science, 8*(6), 448–453. doi: 10.1111/j.1467-9280.1997.tb00459.x.
- Zelinsky, G. J., Zhang, W., Yu, B., Chen, X., & Samaras, D. (2006). The role of top-down and bottom-up processes in guiding eye movements during visual search. In Y. Weiss, B. Scholkopf, & J. Platt (Eds.), *Advances in neural information processing systems* (Vol. 18, pp. 1569–1576). Cambridge: MIT Press.
- Zhang, W., & Luck, S. J. (2009). Feature-based attention modulates feedforward visual processing. *Nature Neuroscience, 12*(1), 24–25 (2009). doi: 10.1038/nn.2223.