Neuromethods (2020) 151: 23–35
DOI 10.1007/7657_2019_30
© Springer Science+Business Media, LLC 2019
Published online: 2 October 2019

Eye Tracking in Visual Search Experiments

Andrew Hollingworth and Brett Bahle

Abstract

Over the last 30 years, eye tracking has grown in popularity as a method to understand attention during visual search, principally because it provides a means to characterize the spatiotemporal properties of selective operations across a trial. In the present chapter, we review the motivations, methods, and measures for using eye tracking in visual search experiments. This includes a discussion of the advantages (and some disadvantages) of eye tracking data as a measure spatial attention, compared with more traditional reaction time paradigms. In addition, we discuss stimulus and design considerations for implementing experiments of this type. Finally, we will discuss the major measures that can be extracted from an eye tracking record and discuss the inferences that each allow. In the course of this discussion, we address both experiments using abstract arrays and experiments using real-world scene stimuli.

Keywords Eye movements, Eye tracking, Visual search, Attention, Scene perception

1 Why Track Gaze Position in Visual Search Experiments?

The vast majority of visual search experiments have used manual, end-of-trial reaction time (RT) as the dependent measure. RT provides a single data point per trial with which to draw inferences about the component operations involved in finding a particular target object. Great progress has been made using end-of-trial RT by using sophisticated experimental designs to isolate component operations and to assess them with sufficient precision (e.g., [1]). However, RT approaches provide limited insight into how search evolves over the course of the trial; that is, they provide little direct evidence about which objects in a search display were attended, for how long, and in what sequence. This is the primary benefit of eye tracking: it provides a continuous window on the allocation of attention over a display in a manner that can characterize the spatiotemporal evolution of the search process on individual trials. Thus, instead of assessing the selectivity of guidance from differences in RT as a function of set size, one can directly observe the probability that fixated objects either match or do not match a particular cued feature [2–4]. Instead of inferring attention capture from small increases in RT when a particular distractor value is present, one can directly observe the probability that distractors with that feature value are fixated [4–9]. Direct observation of the phenomena of interest leads to increased sensitivity, and it allows one to precisely characterize how a particular manipulation leads to changes in behavior. A reliable increase of, say, 50 ms in mean RT on distractor-present trials of a capture paradigm tells us only that the distractor likely interfered with search in some manner and on some proportion of trials. In contrast, eye tracking allows us to measure the probability that the distractor was fixated, how early during the course of search it was fixated, for how long, whether gaze returned to that object later in the trial, and the effect of these events on the time required to direct gaze to the target object.

As a concrete example, consider a recent study by Beck, Luck, and Hollingworth [8]. The research question concerned whether search templates in VWM could be configured to deprioritize particular feature values for selection (i.e., a *negative template* or *template for rejection*). Previous studies using end-of-trial RT as the dependent measure had produced conflicting results. On each trial of the Beck et al. study, participants saw a color cue indicating a color to be avoided: objects with this color in the search array were never targets. Gaze was monitored as participants searched through arrays of colored, circular objects for a target with an extremely small gap on the top or bottom (target feature) among distractors with gaps on the left or right. Gap discrimination required foveation. The search arrays consisted of four objects in four different colors (a total of 16) that were randomly arrayed. By monitoring gaze during search, Beck et al. were able to examine the evolution of selectivity across the trial. At the very onset of search, participants were *more likely* than expected by chance to fixate objects in the cued color (i.e., capture), but by approximately the third object fixated in the array, this pattern reversed, with cued-color objects *less likely* to be fixated than expected by chance (i.e., successful avoidance). This pattern was not observable on end-of-trial RT. Relative to a neutral-cue condition, capture early in the trial (increasing overall RT) and successful avoidance later in the trial (decreasing overall RT) largely cancelled. Thus, the ability to examine selection across the entire trial was key to understanding the mechanisms involved in the use of a negative template, and a purely RT-based design would likely have led to an erroneous conclusion (i.e., that there was no capture by matching items nor any benefit from the cue information).

The eye tracking method allowed several further analyses critical to understanding the underlying mechanisms in Beck, Luck, and Hollingworth [8]. First, fixating a cue-matching object led to a substantial increase in overall search time, reflected on both the elapsed time to the first fixation on the target object and the manual RT to report gap location (these two measures will be strongly correlated in a design such as this one). Second, there was no observable relationship between fixation of a cue-matching object early in the trial and later avoidance, indicating that fixation of the cued color was not necessary for producing later selectivity. Finally, the latency of the very first saccade on the array was longer on negative-cue trials compared with neutral and positive-cue trials, potentially indicating that it took additional time to set up the guidance operation for a negative template or that participants had to exert control to avoid oculomotor capture by matching distractors before they could initiate a saccade to a relevant array item. In sum, eye tracking produces data about the allocation of attention across time and space that can support a rich understanding of the underlying attentional mechanisms.

In addition to the increasing use of eye tracking within traditional visual search tasks, the use of eye tracking in visual search has grown more prominent as interest in real-world scene perception has increased (for a review, see [10]). Given the size and complexity of natural environments, movements of the eyes (and head and body) are typically required to obtain information from task-relevant objects. Coincident with this interest, the field of visual search has been transitioning gradually from using search paradigms as a tool for understanding basic properties of visual perception and attention (typically using abstract arrays) to investigating visual search as an important behavior to be understood in its own right: how people are able to find relevant objects within complex displays (often naturalistic scenes). The field has progressed so that current, prominent general theories of visual search seek to explain how gaze is oriented sequentially to objects within natural scenes (e.g., [11]). The same inferential advantages for eye tracking using abstract displays apply to real-world scene studies. In addition, real-world scenes have visual, spatial, conceptual, and episodic structure that are not typically found in abstract arrays and can be critical to understanding the search operation. For example, kitchens tend to contain blenders, these tend to appear on the counter rather than on the floor, and the blender will also tend to appear in the same place as it was observed previously. Thus, scenes allow additional forms of guidance that can be observed in the sequence of eye movements during search. For example, one can ask whether and how early during visual search gaze is directed to regions of the scene where a target object is likely to be found [12–19] or to locations where the target has been observed previously [20–24].

Before discussing in more detail the methods and measures involved in implementing eye tracking studies using abstract arrays and complex scenes, it is important to discuss the potential limitations in using eye tracking as means to infer the properties of *covert* attention. A relatively consistent literature (for a review, see [25]) demonstrates that saccade execution is necessarily preceded by a covert shift of spatial attention to the saccade target location [26–28]. However, attention can be shifted covertly in the absence of saccade preparation [29–34]. Thus, if the goal of the study is, specifically, to understand covert attentional processes independently of saccade preparation, then eye movements will need to be eliminated or controlled.[1] However, if a saccade is observed, one can be quite certain that, immediately preceding the saccade, there was a corresponding shift of covert attention. Thus, the sequence of eye movements provides strong evidence regarding the spatial allocation of covert attention across time, but this record may not be exhaustive; that is, there may be additional shifts of covert attention that were not associated with saccade preparation or did not ultimately lead to a saccade generation (e.g., through competition or direct inhibition). Despite this limitation, the ability of eye movement data to deliver a close approximation of the spatiotemporal properties of covert selection is a major strength of the method.

2 Methods

In this section we discuss methods for conducting experiments on eye movements in visual search, using both abstract arrays and natural scene stimuli. Then, we turn to eye tracking measures and their interpretation. Figure 1 illustrates typical designs and methods.

Abstract Arrays: Methods for conducting eye tracking studies using abstract search arrays vary only modestly from traditional RT experiments, and thus it is 'quite simple to delineate the necessary modifications, some of which are quite commonsensical (for an example of a fully implemented method, see [8]).

First, search arrays need to be appropriately constructed to minimize confusion about which object is fixated. This is mostly a simple matter of making sure that there is adequate spacing between objects. During search, if objects are well distinguished from the background, a very large majority of saccades will be directed to an object rather than to the spaces between objects. Some care should be taken in selection of the areas of interest (AOI, i.e., the spatial region around each object that is used to classify a fixation as "on" a particular object) so that they are large enough to tolerate some noise in tracking and calibration (extending beyond the physical boundaries of the objects) but not so large that they

[1] Note that instructions are rarely sufficient to ensure that participants do not make eye movements. Thus, even if the goal is to eliminate eye movements, gaze still needs to be monitored. Ideally, an eye tracker can be used, but there is another option. In covert attentions studies, we often use a simple video camera to display a large image of one of the eyes, and the experimenter monitors this image throughout the experiment (a human eye tracker). Movements of the eyes are quite easy to observe, and the experimenter both notes trials with eye movements and reminds the participant, when an eye movement is observed, to keep gaze focused on the relevant reference point. With appropriate, well-timed feedback of this sort, most participants quickly learn how to keep gaze focused centrally and rarely make eye movements after an initial practice session.

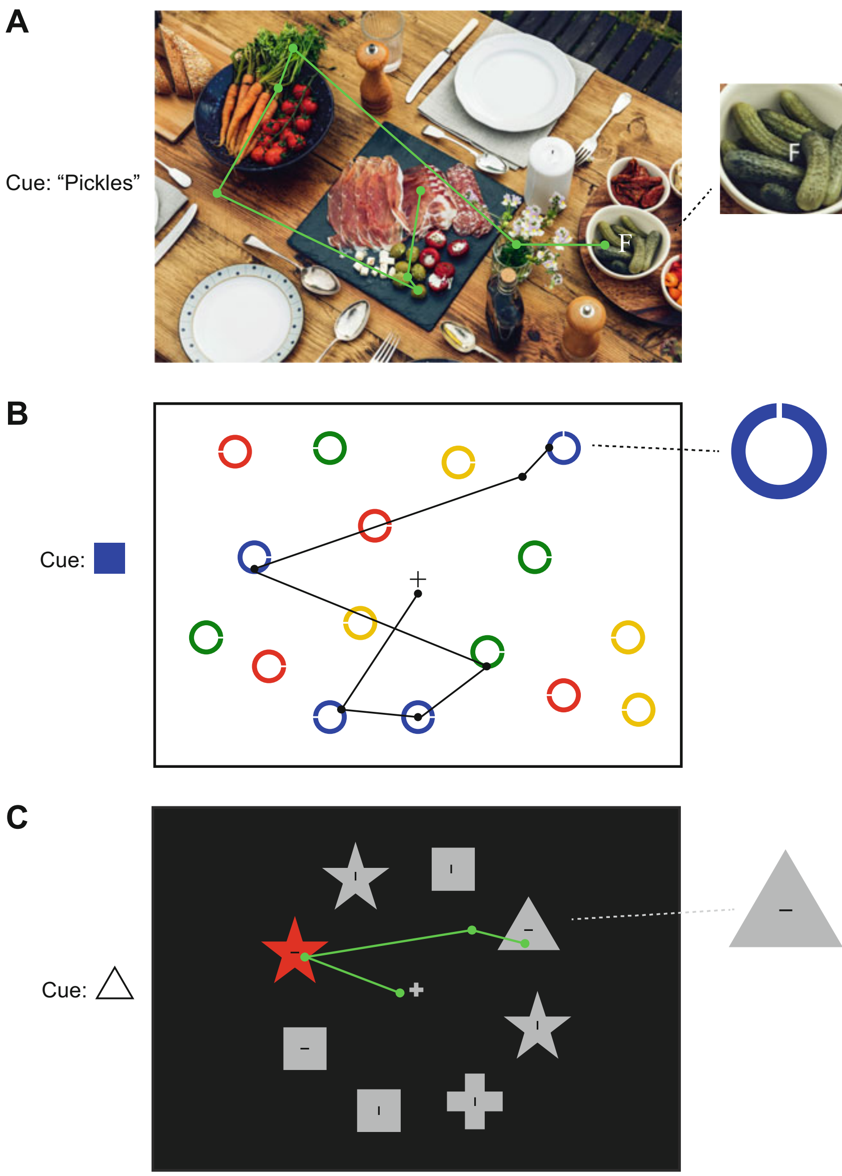

Fig. 1 Three common eye movement paradigms used to assess visual search processes. Lines indicate saccades and dots fixations. (**a**) A paradigm that probes search efficiency. The participant searches for a single cued object within a scene or array, reporting a secondary feature superimposed over it (here, the orientation of an “F”). The primary measure typically would be the elapsed time to the first fixation on the target, but, as discussed in the text, this could be subdivided into *initiation*, *scanning*, and *verification* times. (**b**) A paradigm that probes selectivity. In this example, the cued feature (blue) indicates the color of the target object, with multiple possible target objects present in the display, only one of which has the to-be-reported feature (gap on top or bottom). The primary measure typically would be the probability that a particular fixated object matches the cued value. (**c**) A paradigm that probes attention capture. Here, the participant searches for a cued shape in the presence of a physically salient distractor (uniquely colored item). The typical-dependent measure would be the probability of critical distractor fixation, either limited to the first saccade on the array or at any point before fixation of the target

miscategorize fixations that are clearly directed to the background or to other objects.

Second, and this may seem an obvious point, but search stimuli should be excluded from the immediate region around the position where the participant will be fixating when the array appears (typically the center). All stimuli should be at least a few degrees of visual angle away from this starting position so that a saccade is required for fixation of any array object.

Third, and most importantly, the search task itself should be configured so that participants must translate covert selection into overt shifts of gaze. That is, the search task should *require* fixation of individual objects.[2] This could be implemented in several ways. One could simply make the task to fixate the target. However, we typically do not want to have subjects to reflect too deeply on their gaze patterns, as this could cause them to consciously control gaze. A good alternative is to have participants report a small, secondary feature of the target object that requires foveation for discrimination [2], such as the orientation of a small bar superimposed over the object, the location of a small gap in the object, or the identity of a small letter (see Fig. 1). There is no need in this sort of method to even mention eye movements in the instructions (except that they will be monitored). One must be careful to ensure that the secondary feature is not so salient that participants can search for this feature rather than for the cued object. If concerned about this possibility, the appearance of the secondary feature can be made contingent on fixation of the target region.

Natural Scenes: Methods for constructing experiments using eye tracking on scene stimuli introduce a much greater set of challenges, most of which revolve around how to construct naturalistic images while maintaining appropriate experimental control. The suggestions outlined below apply not only to visual search experiments but to other types of experiments using natural scenes stimuli. To make the design challenges and solutions concrete, let's consider an experiment in which one asks whether objects that appear in plausible scene locations are found faster than objects that appear in implausible locations, testing the hypothesis that a key aspect of guidance in visual search through scenes is knowledge of typical object locations (e.g., that blenders tend to appear on kitchen counters).

The first consideration is how we will construct our scene stimuli to present objects in both plausible and implausible locations within images of natural scenes and how we will retain a

[2] Note that a present/absent design is not always ideal for an eye tracking study, as the mere presence or absence of the target can often be determined without foveation.

strong degree of stimulus control across this manipulation. The choice of stimulus production method will be key. First, we could use line drawings of scenes [17]. This would simplify the process of moving objects to different scene locations, as we would not need to worry about differences in illumination, and so on. But line drawings do not get us very far toward the goal of realism. Second, we could use real environments and take a photograph with a particular object in a typical location, then move the object to an atypical location, and take another photograph. However, a physically moved object will often lead to substantial differences in object appearance caused by changes in illumination, distance from the camera, perspective, and so on. Moreover, such an approach is time-consuming, and it is brittle, in that one cannot alter the manipulation post hoc: it would be practically impossible to re-create the original photographic context if, say, you later decided to add a third object location. Third, we could use a 3D modeling and rendering program that can produce near-photo-realistic images. We would then simply render the scene with the critical object in one location, move the object within the model in a manner that minimized changes in visual appearance, and then re-render the image [23, 35]. This method allows for a great deal of control over the scene stimuli and has the advantage of being robust to the addition of future manipulations. Moreover, some manipulations are possible in 3D that are virtually impossible in photographic stimuli, such as changes of object in-depth orientation, viewer perspective, illumination, and so on. Limitations to the 3D modeling approach primarily revolve around the investment of time necessary to develop expertise in 3D graphics, the availability (and perhaps cost) of 3D scene models, the computing resources necessary to render the images (especially if one uses methods such as raytracing and highly detailed models to produce near-photo-realistic images), and the time required for rendering. However, many of these limitations can be ameliorated by using programs that have been optimized for efficient rendering, such as home design programs or game engines (e.g., [20, 36]). Finally, we could simply obtain existing photographic images (from a web search or from one of the several research databases of scene images) and implement our manipulation of object position using a 2D graphics program such as Adobe Photoshop. With some expertise, it is possible to add, move, or otherwise modify objects in scene photographs in a manner that results in relatively seamless integration of the changes.[3]

[3] Note, however, that in this example experiment, if we were to manipulate photographic images, we would need to use an object image from a different source and then paste and integrate it into *both* the plausible and implausible locations within the experimental scene; this would control our two conditions for artifacts generated by the process of adding an object to a particular scene location.

In general, the choice of stimulus production method will depend on the needs of the experiment. For example, in a relatively recent study [23], we needed to have participants search for 12 different objects in each scene. This was most efficiently implemented by using a 3D modeling program to carefully place the 12 different target objects in discrete locations within each scene, so that there was sufficient object-to-object spacing. It would have been very difficult to find photographs that met these requirements (but see [22]). In another study, we had to manipulate the color of target and distractor objects in scenes [37]. This was implemented most efficiently by using Photoshop to change object color within scene photographs, which is quite easily done in a manner that looks realistic.

After ensuring control over visual properties of objects across manipulations, we also need to exert control over scene locations across manipulations. In our example experiment, we would want to ensure that, due to random variation, the locations chosen for the plausible and implausible conditions did not differ in the ease of search (independently of plausibility). To minimize such effects, we could simply make sure that, across conditions, mean eccentricity of targets was similar, but this would not control for factors such as location variability, spatial biases, or the properties of the local context at plausible an implausible positions (e.g., targets in the plausible condition might tend to be located in sparser regions of the scenes, making them easier to find). To solve this problem, we could manipulate *two* different objects within each scene [38]. For example, in an office scene, a framed picture is likely to appear on a desk and a wastebasket on the floor. Thus, we create *four* versions of the scene: picture on desk, picture on floor, wastebasket on floor, and wastebasket on desk. Now, each scene location and each target object are used in both conditions. When creating a design like this, where there are multiple versions of each scene but only one version can be shown to each subject, then scene-item-to-condition assignments will need to be counterbalanced across subjects (this will also require a relatively large number of scene items). Of course, other experimental designs will differ in the type of implementation necessary to ensure control over locations across conditions.

For the search task itself, we again need to make fixation of the target necessary for trial completion, but not the primary goal of the task. As described above, a good means to this end is to superimpose a small discrimination target over the object [20, 37]. For example, in Bahle, Matsukura, and Hollingworth [37], a small letter "F" was superimposed over the target, and this could either be normally oriented or mirror reversed. The search task was to find the cued object and then report the orientation of the superimposed "F." On the vast majority of trials, participants fixated the target immediately before reporting the orientation of the "F."

3 Measures

In this final section, we discuss four major types of eye movement measures relevant to visual search tasks: measures of search efficiency, measures of selectivity, measures of capture, and measures of memory. Before discussing these measures, we note that each depends on defining areas of interest around the array objects (and, perhaps, other relevant locations, such as the screen center). The eye tracking record is typically analyzed with respect to the fixations that occur within AOIs. In addition, it is common to eliminate trials if the starting gaze position deviates substantially from the reference point (e.g., array center)[4] or if there is no entry into the target AOI during the trial.

Measures of Search Efficiency: The most direct measure of the time taken to find a single target object in an array or natural scene is the elapsed time from the onset of the search stimulus to the beginning of the first fixation with in the target AOI (for the entry that immediately precedes the manual response).[5] We will refer to this measure at *elapsed time to target fixation* (see Fig. 1a). There are several other measures that will almost always be highly correlated with elapsed time to target fixation. *Number of saccades to target fixation* produces essentially the same data as elapsed time but with coarser grain, typically adding little to the overall analysis. *Manual RT* for discrimination of the secondary target feature (e.g., the embedded line orientation) also produces essentially the same data as elapsed time to target fixation but with increased variability due to the addition of discrimination and response processes. Finally, *Path ratio* provides a spatial, rather than temporal, measure of search efficiency. Path ratio is the sum of the amplitudes of all of the saccades before the first entry into the target region *divided by* the distance between the fixation position at search onset (e.g., the center of the scene or array) and the location of the target object. Thus, a value of 1 indicates maximally efficient search, with values increasingly greater than 1 indicating increasingly less efficient search.

A recent approach to understanding the efficiency of component processes during the search task has been to divide the elapsed

[4] An alternative is to make trial initiation contingent on central fixation.

[5] It is possible that, on a small proportion of trials, a participant fixates the target, fails to recognize it at such, leaves the target region to fixate other objects, and returns only later during search, leading to the manual response. Thus, elapsed time to target fixation should be the time until the entry that immediately precedes the response and not necessarily the elapsed time to the very first entry.

time to manual response into three different epochs [16, 38, 39]: (1) time to initiate the first saccade during search (*initiation time*), (2) time from the onset of the fist saccade to the first fixation in the target region (*scanning time*), and (3) time from the first fixation in the target region to the manual response (*verification time*). The former two constitute elapsed time to target fixation. Researchers have considered initiation time to reflect processes related to establishing the target template and to selecting the first item for scrutiny. Scanning time is considered a relatively pure measure of the search process. Finally, verification time is considered a measure of the time taken to decide that the fixated object is in, in fact, the target object (this is relevant primarily to present/absent paradigms where there are no decision processes related to a secondary target feature). Although this division can provide a mapping of eye tracking epochs to particular component processes, it is important to note that, as with any eye tracking measure, the mapping must be treated with considerable care and caution. For example, the very first saccade on an array or scene can quite often be poorly guided, driven by the sensory transient caused by the onset of the search stimulus, and thus initiation time would not necessarily reflect the time required to establish goal-directed guidance. Second, verification time assumes that target identification began only after target fixation, which does not take into account the strong possibility that verification began before the saccade that took the eyes to the target region. That is, it is likely that the target object was selected for fixation, at least in part, because it was identified as the target object in the periphery.

Measures of Selectivity: It is often of interest to attention researchers to characterize the selectivity of visual search: that is, the extent to which attention is limited to cued or otherwise goal-relevant items (Fig. 1b). For example, in a classic study, Williams [2] probed whether selective guidance was implemented more or less efficiently for different feature dimensions (color, size, and shape). This involved constructing displays in which there were multiple objects matching the cued feature value (e.g., multiple red items among items of different colors), with only one of the cue-matching objects containing a secondary target feature. The measure of selectivity was simply the probability that, for any given fixation on an array object, the fixated object matched the cued value.[6] Zelinsky [3] used a related method to show that factors influencing RT as a function of set size had corresponding effects on oculomotor selectivity. And, as noted above, it is possible to examine how selectivity changes over the course of a trial [8] by computing the probability of cued-item fixation for each ordinal fixation number (i.e., following the first saccade during search, the second, the third,

[6] An alternative would be to consider each entry and exit from an object as a single event, collapsing across multiple fixations between entry and exit.

and so on) or using some other means to temporally divide the trial into quantiles.

Measures of Capture: Similar to measuring selectivity as the probability that a cue-matching object is fixated, one can measure the probability that a *distractor* with some additional or unique feature property is fixated (relative to control items that do not have the critical attribute). That attribute could be anything from physical salience (Fig. 1c) to emotional relevance. In one of the first studies of this type, Theeuwes, Kramer, Hahn, Irwin, and Zelinsky [5] examined whether a singleton item in a search display (e.g., uniquely colored) would attract gaze preferentially in a manner consistent with RT results. Relative to non-salient distractors, singletons had a higher probability of being fixated following the first saccade on the array.[7] As in this study, researchers sometimes limit their capture analysis to the very first saccade on the array (especially if the goal is to probe processes that as hypothesized to be based on low-level stimulus properties). However, we have found that, in related tasks, differences in fixation probability can extend multiple saccades into the search process, and thus we have tended to define distractor fixation probability as an object entry at *any point* between the onset of the search array and the first fixation on the target immediately preceding the response [7, 37].

Measures of Memory: During visual search tasks, participants often have to keep track of the items that have been previously scrutinized and rejected, so that gaze can be selectively directed to possible target items [40, 41]. One direct way to measure these processes is to examine the probability of distractor refixation during search. Here, *refixation* refers to the situation in which gaze is directed to a particular region, exits that region for some period of time, and then returns. To make this type of design concrete, consider the task in Fig. 1b, but with no color cue, so that participants have to search every item in the display until the target is found. This would require keeping track of the locations of multiple, previously fixated objects, and the probability of refixation would then serve as a measure of memory-based avoidance. Using this type of method, Peterson, Kramer, Wang, Irwin, and McCarley [42] tested the hypothesis that visual search is *memoryless* in the sense that each selective operation is amnesic with respect to previous events on that trial [43]. Falsifying this hypothesis, Peterson et al. confirmed a major role for memory in search efficiency by showing that distractor refixation probability was much lower than would have been predicted by an amnesic search operation (see also [44]).

[7] It can also be useful to examine saccade latency in this context. For trials *without* oculomotor capture, several studies have observed that saccades directed to the target were delayed when a critical distractor was present versus when it was not, indicating that the programming of the saccade required additional time to resolve the competition between the salient distractor and the target.

References

1. Wolfe JM (2007) Guided search 4.0: current progress with a model of visual search. In: Gray W (ed) Integrated models of cognitive systems. Oxford University Press, Oxford, pp 99–119
2. Williams LG (1967) The effects of target specification on objects fixated during visual search. Acta Psychol 27:355–360. https://doi.org/10.1016/0001-6918(67)90080-7
3. Zelinsky GJ (1996) Using eye saccades to assess the selectivity of search movements. Vis Res 36:2177–2187. https://doi.org/10.1016/0042-6989(95)00300-2
4. Beck VM, Hollingworth A, Luck SJ (2012) Simultaneous control of attention by multiple working memory representations. Psychol Sci 23:887–898. https://doi.org/10.1177/0956797612439068
5. Theeuwes J, Kramer AF, Hahn S, Irwin DE, Zelinsky GJ (1999) Influence of attentional capture on oculomotor control. J Exp Psychol Hum Percept Perform 25:1595–1608. https://doi.org/10.1037/0096-1523.25.6.1595
6. Gaspelin N, Leonard CJ, Luck SJ (2015) Direct evidence for active suppression of salient-but-irrelevant sensory inputs. Psychol Sci 26:1740–1750. https://doi.org/10.1177/0956797615597913
7. Bahle B, Beck VM, Hollingworth A (2018) The architecture of interaction between visual working memory and visual attention. J Exp Psychol Hum Percept Perform 44:992–1011. https://doi.org/10.1037/xhp0000509
8. Beck VM, Luck SJ, Hollingworth A (2018) Whatever you do, don't look at the...: evaluating guidance by an exclusionary attentional template. J Exp Psychol Hum Percept Perform 44:645–662. https://doi.org/10.1037/xhp0000485
9. Le Pelley ME, Pearson D, Griffiths O, Beesley T (2015) When goals conflict with values: counterproductive attentional and oculomotor capture by reward-related stimuli. J Exp Psychol Gen 144:158–171. https://doi.org/10.1037/xge0000037
10. Henderson JM (2003) Human gaze control during real-world scene perception. Trends Cogn Sci 7:498–504. https://doi.org/10.1016/j.tics.2003.09.006
11. Itti L, Koch C (2000) A saliency-based search mechanism for overt and covert shifts of visual attention. Vis Res 40:1489–1506. https://doi.org/10.1016/S0042-6989(99)00163-7
12. Castelhano MS, Henderson JM (2007) Initial scene representations facilitate eye movement guidance in visual search. J Exp Psychol Hum Percept Perform 33:753–763. https://doi.org/10.1037/0096-1523.33.4.753
13. Castelhano MS, Heaven C (2011) Scene context influences without scene gist: eye movements guided by spatial associations in visual search. Psychon Bull Rev 18:890–896. https://doi.org/10.3758/s13423-011-0107-8
14. Eckstein MP, Drescher BA, Shimozaki SS (2006) Attentional cues in real scenes, saccadic targeting, and Bayesian priors. Psychol Sci 17:973–980. https://doi.org/10.1111/j.1467-9280.2006.01815.x
15. Henderson JM, Malcolm GL, Schandl C (2009) Searching in the dark: cognitive relevance drives attention in real-world scenes. Psychon Bull Rev 16:850–856. https://doi.org/10.3758/pbr.16.5.850
16. Malcolm GL, Henderson JM (2009) The effects of target template specificity on visual search in real-world scenes: evidence from eye movements. J Vis 9(8):1–13. https://doi.org/10.1167/9.11.8
17. Henderson JM, Weeks PA, Hollingworth A (1999) The effects of semantic consistency on eye movements during complex scene viewing. J Exp Psychol Hum Percept Perform 25:210–228. https://doi.org/10.1037//0096-1523.25.1.210
18. Võ MLH, Henderson JM (2010) The time course of initial scene processing for eye movement guidance in natural scene search. J Vis 10:14. https://doi.org/10.1167/10.3.14
19. Torralba A, Oliva A, Castelhano MS, Henderson JM (2006) Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search. Psychol Rev 113:766–786. https://doi.org/10.1037/0033-295X.113.4.766
20. Brockmole JR, Castelhano MS, Henderson JM (2006) Contextual cueing in naturalistic scenes: global and local contexts. J Exp Psychol Learn Mem Cogn 32:699–706. https://doi.org/10.1037/0278-7393.32.4.699
21. Brockmole JR, Henderson JM (2006) Using real-world scenes as contextual cues for search. Vis Cogn 13:99–108. https://doi.org/10.1080/13506280500165188
22. Võ MLH, Wolfe JM (2012) When does repeated search in scenes involve memory? Looking at versus looking for objects in scenes. J Exp Psychol Hum Percept Perform

38:23–41. https://doi.org/10.1037/a0024147
23. Hollingworth A (2012) Task specificity and the influence of memory on visual search: comment on Võ and Wolfe (2012). J Exp Psychol Hum Percept Perform 38:1596–1603. https://doi.org/10.1037/A0030237
24. Bahle B, Hollingworth A (2019) Contrasting episodic and template-based guidance during search through natural scenes. J Exp Psychol Hum Percept Perform 45:523–536. https://doi.org/10.1037/xhp0000624
25. Hunt AR, Reuther J, Hilchey MD, Klein R (2019) The relationship between spatial attention and eye movements. Curr Top Behav Neurosci. https://doi.org/10.1007/7854_2019_95
26. Hoffman JE, Subramaniam B (1995) The role of visual attention in saccadic eye movements. Percept Psychophys 57:787–795. https://doi.org/10.3758/BF03206794
27. Kowler E, Anderson E, Dosher B, Blaser E (1995) The role of attention in the programming of saccades. Vis Res 35:1897–1916. https://doi.org/10.1016/0042-6989(94)00279-U
28. Deubel H, Schneider WX (1996) Saccade target selection and object recognition: evidence for a common attentional mechanism. Vis Res 36:1827–1837. https://doi.org/10.1016/0042-6989(95)00294-4
29. Hunt AR, Kingstone A (2003) Covert and overt voluntary attention: linked or independent? Cogn Brain Res 18:102–105. https://doi.org/10.1016/j.cogbrainres.2003.08.006
30. Klein RM (1980) Does oculomotor readiness mediate cognitive control of visual attention? In: Nickerson RS (ed) Attention and performance VIII. Erlbaum, Hillsdale, pp 259–276
31. Klein RM, Pontefract A (1994) Does oculomotor readiness mediate cognitive control of visual attention? Revisited! In: Umilta C, Moscovitch M (eds) Attention and performance XV—conscious and nonconscious information processing. MIT Press, Cambridge, pp 333–350
32. Juan CH, Shorter-Jacobi SM, Schall JD (2004) Dissociation of spatial attention and saccade preparation. Proc Natl Acad Sci U S A 101:15541–15544. https://doi.org/10.1073/pnas.0403507101
33. Schafer RJ, Moore T (2011) Selective attention from voluntary control of neurons in prefrontal cortex. Science 332:1568–1571. https://doi.org/10.1126/science.1199892
34. Thompson KG, Biscoe KL, Sato TR (2005) Neuronal basis of covert spatial attention in the frontal eye field. J Neurosci 25:9479–9487. https://doi.org/10.1523/jneurosci.0741-05.2005
35. Hollingworth A, Henderson JM (2002) Accurate visual memory for previously attended objects in natural scenes. J Exp Psychol Hum Percept Perform 28:113–136. https://doi.org/10.1037//0096-1523.28.1.113
36. Koehler K, Eckstein MP (2017) Beyond scene gist: objects guide search more than scene background. J Exp Psychol Hum Percept Perform 43:1177–1193. https://doi.org/10.1037/xhp0000363
37. Bahle B, Matsukura M, Hollingworth A (2018) Contrasting gist-based and template-based guidance during real-world visual search. J Exp Psychol Hum Percept Perform 44:367–386. https://doi.org/10.1037/xhp0000468
38. Malcolm GL, Henderson JM (2010) Combining top-down processes to guide eye movements during real-world scene search. J Vis 10(4):1–11. https://doi.org/10.1167/10.2.4
39. Castelhano MS, Pollatsek A, Cave KR (2008) Typicality aids search for an unspecified target, but only in identification and not in attentional guidance. Psychon Bull Rev 15:795–801. https://doi.org/10.3758/pbr.15.4.795
40. Smith TJ, Henderson JM (2011) Does oculomotor inhibition of return influence fixation probability during scene search? Atten Percept Psychophys 73:2384–2398. https://doi.org/10.3758/s13414-011-0191-x
41. Klein RM, MacInnes WJ (1999) Inhibition of return is a foraging facilitator in visual search. Psychol Sci 10:346–352. https://doi.org/10.1111/1467-9280.00166
42. Peterson MS, Kramer AF, Wang RF, Irwin DE, McCarley JS (2001) Visual search has memory. Psychol Sci 12:287–292. https://doi.org/10.1111/1467-9280.00353
43. Horowitz TS, Wolfe JM (1998) Visual search has no memory. Nature 394:575–577. https://doi.org/10.1038/29068
44. Gilchrist ID, Harvey M (2000) Refixation frequency and memory mechanisms in visual search. Curr Biol 10:1209–1212. https://doi.org/10.1016/s0960-9822(00)00729-6