The effects of target template specificity on visual search in real-world scenes: Evidence from eye movements

George L. Malcolm   University of Edinburgh, UK

John M. Henderson   University of Edinburgh, UK

We can locate an object more quickly in a real-world scene when a specific target template is held in visual working memory, but it is not known exactly how a target template's specificity affects real-world search. In the present study, we compared word and picture cues in real-world scene search. Using an eye-tracker, we segmented search time into three behaviorally defined epochs: search initiation time, scanning time, and verification time. Results from three experiments indicated that target template specificity affects scanning and verification time. Within the scanning epoch, target template specificity affected the number of scene regions visited and the mean fixation duration. Changes to SOA did not affect this pattern of results. Similarly, the pattern of results did not change when participants were familiarized with target images prior to testing, suggesting that an immediately preceding picture provides a more useful search template than one stored in long-term memory. The results suggest that the specificity of the target cue affects both the activation map representing potential target locations and the process that matches a fixated object to an internal representation of the target.

Keywords: active vision, eye movement, search

Citation: 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, http://journalofvision.org/9/11/8/, doi:10.1167/9.11.8.

Introduction

We typically search for specific objects in our environment by moving our eyes (Findlay & Gilchrist, 2003; Land & Hayhoe, 2001). Eye movements are generally characterized by short bursts of movements (saccades) interleaved by pauses (fixations). The purpose of saccadic movements is to direct the region of highest resolution on the retina (the fovea) toward an informative part of the external environment. A central question therefore is what determines where we direct our eyes in a scene.

Eye movement guidance is typically thought to involve two general types of information: image features and cognitive knowledge (Henderson, 2007). Recent computational models have used image features to identify visually salient regions of the scene (areas of local contrast) that could attract eye movements (Itti & Koch, 2000; Itti, Koch, & Niebur, 1998; Parkhurst, Law, & Niebur, 2002). With this view, attention in a search task is assumed to be deployed to the most visually salient scene region, and if the region does not contain the target attention is then directed to the next most salient region. Models of this type have had some success in predicting human scan patterns, but they tend to fail in visual search tasks involving real-world scenes in which context can play a role (Einhäuser, Spain, & Perona, 2008; Foulsham & Underwood, 2007; Henderson, Brockmole, Castelhano, & Mack, 2007; Zelinsky, Zhang, Yu, Chen, & Samaras, 2006).

Knowledge about the nature of the target object and its relationship to the scene can also be used to guide search in a top–down manner. A person searching for a particular target in a novel scene can orient visual attention based on the scene constraints associated with their semantic knowledge of the world, often within the very first eye movement (Castelhano & Henderson, 2007; Eckstein, Drescher, & Shimozaki, 2006; Neider & Zelinsky, 2006; Torralba, Oliva, Castelhano, & Henderson, 2006). For example, if an observer is looking for pedestrians in an urban scene, they can use scene context to direct their eyes to the sidewalk (Neider & Zelinsky, 2006; Torralba et al., 2006; Zelinsky & Schmidt, 2009). However, scene context is less useful for locating a target among distracter objects within the same global region (parking meters, fire hydrants, etc.).

Another source of top–down information used to guide search is the target template. A target template's visual features—the observer's mental representation of the target—can guide eye movements during search tasks (Beutter, Eckstein, & Stone, 2003; Eckstein, Beutter, Pham, Shimozaki, & Stone, 2007; Findlay, 1997; Luria & Strauss, 1975; Motter & Belky, 1998; Rajashekar, Bovik, & Cormack, 2006; Tavassoli, van der Linde, Bovik, & Cormack, 2007, 2009; Williams, 1967; Zelinsky, Rao, Hayhoe, & Ballard, 1997) and facilitate perceptual decision tasks (Burgess, 1985; Greenhouse & Cohn, 1978; Judy, Kijewski, Fu, & Swensson, 1995) by increasing the weight of signals from target similar features in the scene percept while de-weighting signals from target dissimilar features. This has led to the recent generation of target template guided eye movement models (Rao, Zelinsky, Hayhoe, & Ballard, 2002; Zelinsky, 2008).

The specificity of the target template has also been recently shown to affect response time (RT) performance in search tasks (Bravo & Farid, 2009; Vickery, King, & Jiang, 2005; Wolfe, Horowitz, Kenner, Hyle, & Vasan, 2004). If a specific target template is stored in visual working memory (VWM) prior to a search task (generated from a picture cue of the target), search time will be reduced. Conversely, if an abstract target template is stored in VWM (generated from a word cue), search will be longer. Two key features can be taken from the research on the effect of target template specificity on visual search RT. First, visually specific cues facilitate search because they provide for a more specific target template. Second, in the studies that manipulated stimulus onset asynchrony (SOA) between target cue presentation and search display, SOA was found to interact with cue specificity: An abstract target template was found to benefit search more as SOA increased (Vickery et al., 2005; Wolfe et al., 2004). Conversely, the benefit provided by a specific cue was slightly reduced as SOA increased. This interaction has been taken to suggest that a target template created from an abstract cue takes longer to set up before becoming fully useful, whereas a precise target template created from a picture cue can be established more quickly but decays over time (Vickery et al., 2005; Wolfe et al., 2004).

Collectively, these results demonstrate that a target template's specificity affects search performance. However, it is not understood precisely how a target template's specificity affects search, nor the eye movements involved. The dependent variable in most previous target template specificity studies has been RT, a unitary measurement that makes it difficult to draw inferences about the subprocesses affecting visual search (but see Schmidt & Zelinsky, 2009). This leaves open several questions about the role of target template specificity on eye movements during the search process. For instance, does the availability of a more specific target template affect the activation map[1] used to select probable target regions for fixation? Or does a more specific template allow for faster comparison to each item fixated during search, reducing the time needed to reject distracters and accept the target? Or does a more specific template simply allow search to begin faster?

To begin to address these questions, we used eye-tracking to investigate target template-guided visual search. We divided search into three behaviorally defined epochs. Each of these epochs reflected a separate hypothesized underlying search process. This method of using eye-tracking to differentiate epochs of search to examine particular subprocesses more closely has recently been used to good effect (Castelhano, Pollatsek, & Cave, 2008). Castelhano et al. (2008) divided search into two epochs based on eye movement data: target latency or the time to search for the target, and target verification or the time to accept the target once found. A primary result was

that picture cues lead to shorter target latency and verification epochs.

In the current study, we divided search through real-world scenes into three behaviorally independent epochs based on the eye movement record. First, we examined search initiation time, defined as the time from appearance of the search scene until the first saccade away from the initial fixation point (i.e., initial saccade latency). We assume search initiation time reflects processes related to time needed to establish the search template plus time needed to select a first search target candidate in the scene. Second, we measured scanning time, defined as the elapsed time between the first saccade (the end of the search initiation epoch) to the first fixation on the target object. This epoch was taken to represent the actual search process (cf. Castelhano et al., 2008). Third, verification time was defined as the participant's gaze duration on the target object[2] and was taken to reflect the time needed to decide that the fixated object was in fact the target. Total trial duration, the measure typically reported in visual search studies, equals the sum of these three epochs (Figure 1). Segmenting total trial duration into these three epochs helps elucidate the effect that target template specificity has on the search process, particularly for correct trials. Therefore, trials in which the participant fixated the target and then continued searching were removed from analyses (see below) as these cases would tend to distort the verification measures.

It should be noted that both the scanning and verification epochs contain fixations that reflect at least some similar processes such as deciding whether the target is present in the fixated region. However, there is enough of a functional difference to separate these two measures. Scanning epoch fixations involve two processes: a reject decision—which only needs a single fixated feature to mismatch with a target template—and the selection of the next fixation location. The verification epoch contains an accept decision which would likely be based on a much more complete analysis of the fixated item to ensure that it is a match to the sought target. In addition, in the verification epoch minimal emphasis need be placed on deciding where in the scene to fixate next.

Present study

In the present study, we asked participants to search for targets in photographs of real-world scenes while their eye movements were recorded. We then divided total trial durations into the three independent epochs for analysis. The two manipulated properties in the experiments were the type of cue shown (word or picture) and the SOA between cue presentation and onset of the search scene (long or short).

If cuing a participant with a specific (picture) cue of the target rather than a more abstract (word) cue benefits
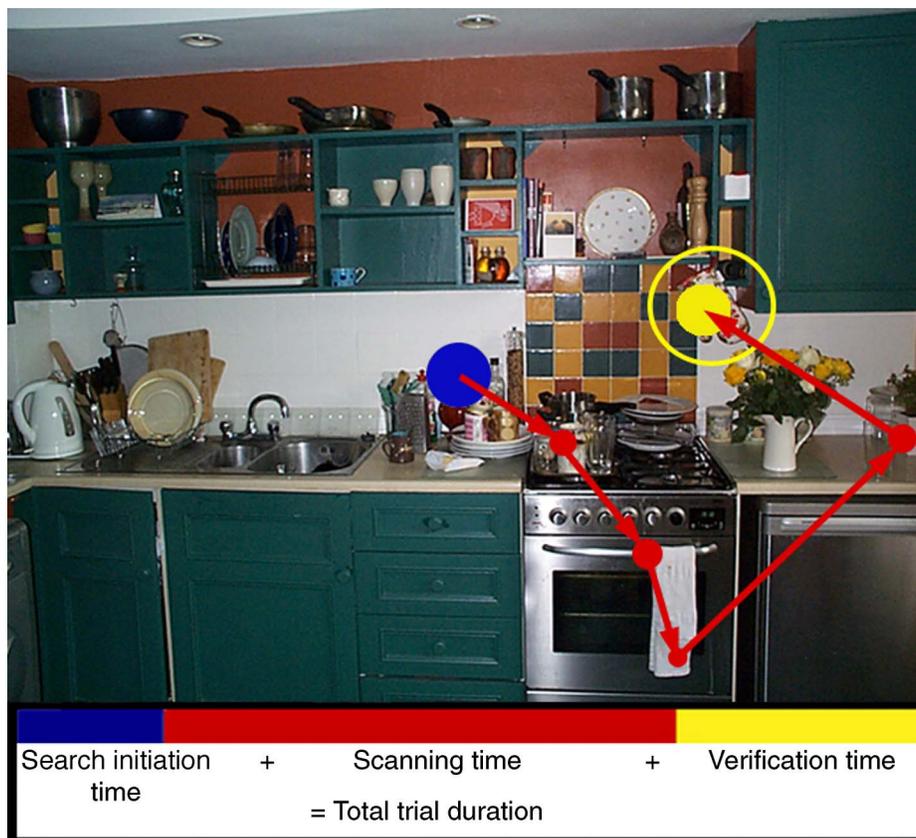
Figure 1. Dividing up visual search time. Blue represents initial saccade latency; red, scanning time; and yellow, verification time. When summed, they yield the total trial duration. Lines represent saccades; circles represent fixations. The thin yellow line outlines the target (the oven mitt).

visual search, then total trial duration should be reduced. If this benefit is due to a picture cue facilitating faster target template set up in VWM prior to search, we should replicate the search time interaction found in previous studies (Vickery et al., 2005; Wolfe et al., 2004). Thus, word cues should benefit from longer SOAs since more time is available to establish a search template prior to appearance of the scene. Picture cues on the other hand should suffer with longer SOAs because visually specific templates created from picture cues should decay with the extra time. However, total trial duration is a relatively coarse measure for determining target template set up time. Another more sensitive method for analyzing target template set up time is provided by search initiation time. If cue specificity does affect target template set up time, search initiation times for word cue trials should be longer when the SOA is shorter. This result is expected because more time would be needed to finish establishing the template than a short SOA would allow, extending into the initial fixation once the search scene appears. Long SOAs would allow the visual system to set up the template prior to the scene's appearance. Conversely, with short SOAs, picture cues should provide a quickly available and precise template that would allow search to begin as soon as the search display appeared. Therefore, the prediction

is that search initiation time should be shorter for picture than word cues with short SOAs, but this difference should be reduced or reverse as SOA increases.

If increased template specificity benefits the search process after initiation, then search time should be faster following a picture cue. Such a benefit could be revealed in the scanning and/or verification epoch (Castelhano et al., 2008). Furthermore, within the scanning epoch, cue specificity may affect further subprocesses differently. For example, at each fixation during the scanning epoch there are two processes occurring sequentially (van Diepen, Wampers, & d'Ydewalle, 1998): the visual system must first process the fixated object and then, if that object is not accepted as the target, must decide where to fixate next. In order to gauge whether cue type affects either of these scanning processes, we compared the number of scene regions visited as well as the mean scanning fixation duration across conditions.

The number of scene regions visited during the scanning epoch provides an indication of whether the target cue affects the activation map. To count the number of regions visited during scanning, each scene was divided into 48 square regions of 100×100 pixels. The number of regions visited in each trial by each participant across all conditions was then calculated. As we were examining

how many regions were visited, but not how often, regions fixated more than once were still scored as one. Previous models suggest that the visual system exploits the precise visual properties of a target template to improve an activation map's selection of target-probable regions (Rao et al., 2002; Zelinsky, 2008). If this result generalizes to complex scene search, then fewer regions of the scene should be visited during the scanning epoch following a picture cue.

The mean scanning fixation duration provides an indication of whether the target cue affects the process of matching a fixated object to an internal representation of the target. Faster processing of the item at fixation reduces the fixation duration (Henderson & Ferreira, 1990). If a more specific template facilitates faster processing and so faster rejecting of a distracter, picture cue trials would have shorter scanning fixation durations than word cue trials.

We conducted three experiments to address these issues. Each experiment cued the search target immediately prior to the search scene, either in the form of an exactly matching picture (e.g., a picture of a coffee mug exactly as it appeared in the scene) or using a word that described the target (e.g., the words "coffee mug"). The search scene then followed 125 to 1000 ms later. Participants had to respond via a button press as soon as they found the target. Experiments 1 and 2 manipulated cue type and SOA. In Experiment 3, we investigated whether target familiarity affected search by familiarizing participants with all the targets prior to testing.

Experiment 1

Methods

Participants

Twelve participants gave informed consent in accordance with the institutional review board of University of Edinburgh. All participants were naïve about the purpose of the study.

Stimulus materials

Sixty photographs of real-world scenes from a variety of categories (indoor and outdoor, natural and man-made) were used as stimuli. Search targets were chosen such that they occurred only once in the scene, did not appear at the scene center, were not occluded, were large enough to be easily recognized yet smaller than a maximum of 3° in diameter, and were easily identifiable when presented alone (as determined by initial pilot testing).

Once the scenes were selected, they were scaled to 800 × 600 pixel resolution. To create the picture cues, the target objects were copied and pasted into a new blank background using Adobe Photoshop CS (Adobe, San Jose, CA). Picture cues were edited so that they did not contain any of the surrounding scene context and then were placed at the center of an 800 × 600 pixel gray background. A further 60 corresponding word cues were created that contained only the names of the target objects, presented in a 30-point font subtending 0.89 degrees in height centered within the same gray background.

Apparatus

Eye movements were recorded using an EyeLink 1000 eye-tracker sampling at 1000 Hz. Viewing was binocular but only the right eye was tracked. Experiments were programmed in Experiment Builder. Initial data reduction was accomplished with DataViewer (SR-Research, Mississauga, ON). Stimuli were shown on a 21-in. ViewSonic G225f cathode ray tube monitor (ViewSonic, London, UK) positioned 90 cm away from the participant, taking up an 18.72° × 24.28° field of view, with a refresh rate of 140 Hz.

Procedure

Prior to the experiment, each participant underwent the EyeLink calibration procedure: Eye positions were recorded as participants fixated a series of nine dots arranged in a square grid extending to 19.25° eccentricity. Calibration was then validated against a second set of nine dots.

For the experiment, the trial structure was as follows. Each trial began with eye-tracking drift assessment and correction. Participants then pressed the spacebar to start the trial. A central fixation cross appeared for 400 ms, followed by a cue identifying the search target for 200 ms. The cue was either a word identifying the target or an exactly matching picture of the target. The cue was followed by a central fixation point lasting either 100 or 800 ms, creating two SOA conditions of 300 and 1000 ms. These SOA times were selected because they replicated the longest SOA duration and one of the shortest SOA durations used in previous studies (Vickery et al., 2005; Wolfe et al., 2004). The experiment was thus a 2 × 2 design with cue type (word vs. picture) and SOA (300 vs. 1000 ms) as the variables. Once the delay was over, the corresponding real-world scene appeared. Participants were asked to locate the target as quickly and as accurately as possible, and to press a response key as soon as the target was found. Participants were given eight practice trials prior to the experiment.

Results

Trials with errors were removed from analysis; these included trials in which participants incorrectly identified

the target; participants fixated the target, moved away, and returned before correctly identifying it; or the total trial duration exceeded 5500 ms. Overall, 7% of trials were removed by these criteria. If a participant fixated the target once, moved fixation off the target for one fixation, and then immediately returned to the target in the next fixation, this was accepted as a correct trial. On such trials, a single fixation deviating away from the target was considered to be the result of a pre-programmed oculo-motor command and not due to a decision to attend to a different possible target. This fixation sequence occurred on 4.3% of the correct trials.

Analysis of total trial duration and the three scanning epochs

Repeated-measures ANOVAs with cue type (word vs. picture) and SOA (300 vs. 1000 ms) as factors were conducted on total trial duration, search initiation time, scanning time, and verification time (Table 1).

For total trial duration, there was a significant main effect of cue type, $F(1, 11) = 19.35$, $MS_E = 13215.64$, $p < 0.005$, with faster response times for picture cues than word cues, indicating that search was facilitated by the ability to establish a more precise target template. SOA, however, did not produce a significant main effect, $F(1, 11) = 1.57$, $MS_E = 17252.93$, $p = 0.236$, nor was there a significant interaction between cue type and SOA, $F(1, 11) = 1.51$, $MS_E = 21816.95$, $p = 0.246$. Thus, SOA failed to influence total trial duration.

For search initiation time, there was no main effect of cue type, $F(1, 11) = 0.302$, $MS_E = 553.90$, $p = 0.59$; participants began their search equivalently given a picture or a word cue. There was a significant main effect of SOA, $F(1, 11) = 30.64$, $MS_E = 469.24$, $p < 0.001$, with a longer SOA producing quicker search initiation. Cue type and SOA did not interact, $F(1, 11) = 2.240$, $MS_E = 213.03$, $p = 0.163$. Search initiation was therefore only affected by SOA and not the specificity of the template.

In contrast to search initiation time, cue type produced a significant main effect on both the scanning and verification epochs. Scanning and verification times were shorter for picture than word cues ($F(1, 11) = 12.64$, $MS_E = 7123.55$, $p < 0.01$, and $F(1, 11) = 7.62$, $MS_E = 5079.60$, $p < 0.05$, respectively), demonstrating an advantage for a more precise target template. There was no effect of SOA in either the scanning or verification epochs ($F(1, 11) = 0.43$, $MS_E = 14031.42$, $p = 0.524$, and $F(1, 11) = 0.86$, $MS_E = 1854.04$, $p = 0.373$, respectively) and no interaction of cue type and SOA in either epoch ($F(1, 11) = 0.96$, $MS_E = 19258.60$, $p = 0.348$, and $F(1, 11) = 0.32$, $MS_E = 1160.88$, $p = 0.582$, respectively). Thus, two epochs of visual search, scanning for the target and verifying that the target had been found, were both facilitated by the more precise target cue, supporting the hypothesis that search in real-world scenes is facilitated by a more precise target template.

The underlying behavior affecting scanning time

Since picture cues resulted in shorter scanning epochs than word cues, further analyses were conducted to specify how the cue affected the scanning processes. Specifically, we examined the number of scene regions visited and fixation durations during the scanning epoch (Table 1).

Participants visited fewer regions during scanning when they had been shown a picture cue, $F(1, 11) = 14.43$, $MS_E = 0.150$, $p < 0.005$, consistent with the hypothesis that a more precise target template led to a more selective placement of fixations. There was no main effect of SOA, $F(1, 11) = 0.013$, $MS_E = 0.131$, $p = 0.910$, and no interaction between cue type and SOA, $F(1, 11) = 1.475$, $MS_E = 0.153$, $p = 0.250$.

Scanning fixation durations were marginally shorter following picture cues than word cues, $F(1, 11) = 3.778$, $MS_E = 3777.348$, $p = 0.078$. There was no main effect of SOA, $F(1, 11) = 0.004$, $MS_E = 842.327$, $p = 0.953$, nor was there an interaction between cue type and SOA, $F(1, 11) = 2.292$, $MS_E = 367.474$, $p = 0.158$. These data suggest faster rejection of non-targets in each fixation given a more specific target template.

	Word Cue, 1000 ms SOA		Word Cue, 300 ms SOA		Picture Cue, 1000 ms SOA		Picture Cue, 300 ms SOA	
	Mean	*SE*	Mean	*SE*	Mean	*SE*	Mean	*SE*
Total trial duration	1245.45	(42.42)	1240.65	(35.46)	1047.14	(42.63)	1146.95	(80.73)
Search initiation time	242.76	(10.57)	271.07	(9.53)	232.73	(6.40)	273.65	(7.36)
Scanning time	500.43	(34.73)	483.65	(35.72)	374.53	(30.46)	436.32	(65.15)
Regions visited	1.95	(0.12)	2.08	(0.15)	1.66	(0.15)	1.52	(0.13)
Fixation durations	165.49	(6.60)	157.41	(6.66)	146.22	(5.63)	154.89	(7.22)
Verification time	502.26	(27.82)	485.13	(27.62)	439.88	(34.75)	433.92	(32.62)

Table 1. Results from Experiment 1. All means are in millisecond units, except for the item Regions visited, which is measured in the number of regions visited on the display screen (maximum 48).

Discussion

The results of Experiment 1 clearly indicate that visual search in real-world scenes is facilitated by a specific target template. Total trial duration was reduced given a picture cue than a word cue. A more precise target template affected both scanning time and verification time, with picture cues yielding shorter scanning and verification epochs. Closer analysis of the scanning epoch revealed that picture cues allowed for fewer regions to be visited and a tendency for fixations to be shorter in duration during search. The results suggest that knowledge of a target's appearance prior to search can benefit scanning in two ways: by facilitating selection of potential target locations beyond the current fixation location and by shortening the time needed to reject fixated distracters before moving on to the next potential target.

Interestingly, we did not observe an interaction between cue specificity and SOA, either in total trial duration or in search initiation time. Therefore, in our paradigm we did not see evidence either for lengthened template set up time given a word cue or for decay of visual specificity given a picture cue. The finding that search initiation time was faster overall given a longer SOA can be accommodated by the common finding that responses are faster given more pre-trial warning time to prepare.

Before we accept the conclusion that there is no effect of cue specificity on search initiation time, we must consider a possible alternative explanation for the null results. It has been reported that specific cues reach close to their full advantage with SOAs around 200 ms (Vickery et al., 2005; Wolfe et al., 2004). Our shorter SOA (300 ms) may have been too long to reveal an effect of cue type. We therefore replicated Experiment 1 with SOAs of 200 and 800 ms.

Experiment 2

Methods

Participants

Thirteen participants gave informed consent in accordance with the institutional review board of University of Edinburgh. All participants were naïve about the purpose of the study and none of them participated in Experiment 1.

Stimulus materials

The stimuli were the same as Experiment 1, except the word cues' font was increased to 72 point (2.14°).

Procedure

The procedure was the same as Experiment 1 with the following exceptions. First, cues were shown for 150 ms, followed by a fixation cross for 50 or 750 ms, producing SOAs of 200 or 800 ms. Second, participants now responded via a response pad (SR Research, Mississauga, ON). Third, the experimenter initiated each trial after the participant fixated a central drift-correction dot.

Results

Data from 12 participants were accepted; data from a 13th participant were eliminated due to an unusually high error rate (18 errors in 60 trials vs. 93% mean accuracy rate among the accepted 12 participants). Trials in which participants fixated, moved off, and returned to the target in the next fixation occurred on 3.8% of the correct trials. All correct trials were subject to four repeated-measures ANOVAs with cue type (word vs. picture) and SOA (200 ms vs. 800 ms) as factors and total trial duration, search initiation, scanning time, and verification time as dependent measures. As in Experiment 1, fixation durations and number of regions visited during scanning were also analyzed (Table 2).

Total trial duration was faster for picture than word cues, $F(1, 11) = 33.08$, $MS_E = 16957.615$, $p < 0.001$, again demonstrating a search advantage when a more precise search template could be established. However, the reduced SOA still failed to produce a main effect, $F(1, 11) = 0.001$, $MS_E = 39273.990$, $p = 0.974$, and it did not interact with cue type, $F(1, 11) = 0.957$, $MS_E = 13890.135$, $p = 0.349$.

	Word Cue, 800 ms SOA		Word Cue, 200 ms SOA		Picture Cue, 800 ms SOA		Picture Cue, 200 ms SOA	
	Mean	*SE*	Mean	*SE*	Mean	*SE*	Mean	*SE*
Total trial duration	1246.50	(49.45)	1215.13	(53.83)	997.01	(37.80)	1032.19	(36.95)
Search initiation time	266.59	(7.34)	312.30	(10.51)	261.41	(8.22)	310.70	(6.00)
Scanning time	584.95	(46.06)	516.29	(69.10)	387.63	(45.70)	382.18	(45.43)
Regions visited	1.90	(0.17)	1.71	(0.17)	1.31	(0.12)	1.24	(0.12)
Fixation durations	180.13	(6.36)	170.65	(6.70)	161.52	(10.41)	157.53	(6.42)
Verification time	450.01	(30.43)	437.93	(33.27)	384.05	(20.99)	393.25	(37.19)

Table 2. Results from Experiment 2. All means are in millisecond units, except for the item Regions visited, which is measured in the number of regions visited on the display screen (maximum 48).

As in Experiment 1, search initiations were faster for longer SOAs, $F(1, 11) = 82.28$, $MS_E = 329.105$, $p < 0.001$, but there was no effect of cue type, $F(1, 11) = 0.402$, $MS_E = 343.357$, $p = 0.539$, and no interaction between cue type and SOA, $F(1, 11) = 0.064$, $MS_E = 604.580$, $p = 0.805$. These results are consistent with a general warning benefit from the longer SOA.

For both the scanning and verification epochs, picture cues produced shorter times than word cues ($F(1, 11) = 19.21$, $MS_E = 17151.66$, $p < 0.001$, and $F(1, 11) = 15.10$, $MS_E = 2432.13$, $p < 0.005$, for scanning and verification time, respectively). There was again no effect of SOA for either the scanning or verification epochs ($F(1, 11) = 0.62$, $MS_E = 26419.31$, $p = 0.446$, and $F(1, 11) = 0.01$, $MS_E = 5200.40$, $p = 0.946$, respectively) nor an interaction between cue type and SOA in either of the epochs ($F(1, 11) = 1.21$, $MS_E = 9910.41$, $p = 0.295$, and $F(1, 11) = 0.42$, $MS_E = 3215.25$, $p = 0.529$).

As with Experiment 1, number of regions visited and fixation durations during the scanning epoch were examined to specify more precisely how the scanning epoch was influenced by the two variables (Table 2). In the regions-visited analysis, fewer regions were visited following picture than word cues, $F(1, 11) = 14.38$, $MS_E = 0.24$, $p < 0.005$. Again, there was no main effect of SOA, $F(1, 11) = 1.23$, $MS_E = 0.16$, $p = 0.292$, and no interaction between cue type and SOA, $F(1, 11) = 0.39$, $MS_E = 0.11$, $p = 0.546$.

Picture cues yielded shorter mean fixation durations than word cues, $F(1, 11) = 5.92$, $MS_E = 510.73$, $p < 0.05$, but there was no main effect of SOA, $F(1, 11) = 0.58$, $MS_E = 945.19$, $p = 0.464$, and no interaction between cue type and SOA, $F(1, 11) = 0.25$, $MS_E = 359.29$, $p = 0.626$.

In summary, the pattern of results remained identical to those of Experiment 1. A more specific cue facilitated search, primarily due to faster reject decisions and better targeting during the scanning epoch, along with faster acceptance of the target once it was fixated. These results demonstrate that a more precise target cue, leading to a more precise search template, facilitates search in real-world scenes.

Discussion

The results of Experiment 2 were almost identical to those of Experiment 1, with total trial duration, scanning time, and verification time all facilitated by a specific picture cue. The shorter scanning epoch again resulted from significantly faster reject decisions at each fixation and a better targeting of the next fixation. Search initiation was faster following longer SOAs. However, we did not observe an interaction of cue type and SOA on either total trial duration or search initiation time.

Prior studies reported an interaction of cue type and SOA on RT, hypothesized to be partially a result of more time needed to establish a target template following a word cue than a picture cue (Vickery et al., 2005; Wolfe et al., 2004). Why have we not replicated this interaction, either in the total trial duration or search initiation measure? One possible explanation is that search initiation time is independent of cue specificity and that a longer SOA may simply provide more general warning that the trial will begin. However, this does not explain the lack of an interaction of factors on total trial duration. Another possible explanation for the null effect in our current results, not addressed in the first two experiments, is related to participants' overall familiarity with the targets. In Vickery et al. (2005), participants were familiarized with the appearance and name of all targets prior to testing. Similarly, in both that study and the study by Wolfe et al. (2004), each target was presented several times during the experiment. It is possible that in those studies, participants used word cues to retrieve from visual long term memory (VLTM) an image of the target learned during the experiment, either before initiating search or during the course of search. This might be a time-consuming process that could be completed prior to the trial given a longer SOA, but that might be less likely to be completed until after the trial had started given a shorter SOA. Picture cues, in contrast, would not require any retrieval from VLTM and could be used to establish a specific template in the duration allowed by a short SOA.

Experiment 3 investigated this possibility by familiarizing participants with targets prior to the experiment. If the failure to find evidence for varying target template set up time in Experiments 1 and 2 was due to lack of familiarity with the pictorial properties of the targets in the word cue condition, then we should observe an interaction between cue specificity and SOA in Experiment 3. This would be demonstrated in either the total trial duration or search initiation time measures.

It is also possible that 200 ms is still not a short enough SOA to reveal an interaction of SOA and cue type. Previous reports suggest that most of the advantage of real-world picture cues can be gained by an SOA of around 200 ms (Vickery et al., 2005; Wolfe et al., 2004), so in Experiment 3 we decreased the shorter SOA to 125 ms: a duration short enough that it should probe a period of target template set up and long enough that the word cues are still identifiable.

Finally, one might argue that search in the first two experiments was too easy, leading to a ceiling effect that might mask an effect of SOA on scanning. Participants found the targets relatively quickly with total trial durations means in the range of 1000–1250 ms. Participants in previous research took up to 1700 ms (Vickery et al., 2005). Therefore, all scenes with mean

total trial durations less than 950 ms in Experiments 1 and 2 were replaced with more difficult search scenes in Experiment 3.

Experiment 3

Methods

Participants

Fifteen participants gave informed consent in accordance with the institutional review board of University of Edinburgh. All participants were naïve about the purpose of the study and none of them participated in Experiments 1 and 2.

Procedure

Experiment 3 followed the same procedure as Experiments 1 and 2 with the following exceptions. First, participants were shown all possible target pictures together with their associated words four times each prior to the experiment. Target picture-word pairs were shown in random order and were self-paced. Participants were told to pay attention to this learning session because target cues during the experiment would be presented briefly and it would benefit them to know a target's appearance in advance. Second, the target cue was displayed for 75 ms followed by a fixation cross for 50 ms in the short SOA condition and 725 ms in the long SOA condition. This resulted in SOAs of 125 and 800 ms.

Stimulus materials

All scene images from Experiment 2 with mean total trial duration across participants below 950 ms or that contained human faces were replaced in Experiment 3. This resulted in replacement of 24 scene images and their corresponding picture and name templates. Selection of new scenes followed the same criteria as before.

Results

Three participants were removed from analysis, one due to poor calibration and two for failing to follow instructions. Data from 12 participants were analyzed. There was a 92% mean accuracy rate among these 12 participants. Trials where participants fixated, moved off, and returned to the target within one fixation occurred on 3.5% of the correct trials. Data analyses mirrored Experiments 1 and 2 (Table 3).

As in Experiments 1 and 2, total trial duration was faster following picture cues, $F(1,11) = 40.78$, $MS_E = 22593.65$, $p < 0.001$. In addition, total trial duration was also shorter following long SOAs, $F(1,11) = 5.46$, $MS_E = 19585.81$, $p < 0.05$. However, there was still no interaction between the two variables, $F(1,11) < 0.01$, $MS_E = 29752.11$, $p = 0.985$.

Search initiation time was faster following longer SOAs, $F(1,11) = 44.72$, $MS_E = 1136.03$, $p < 0.001$, just as in Experiments 1 and 2, but again there was no significant main effect of cue type, $F(1,11) = 5.01$, $MS_E = 805.49$, $p = 0.939$, and no significant interaction between cue type and SOA, $F(1,11) = 1.41$, $MS_E = 680.54$, $p = 0.260$. Thus, even when participants were familiarized with target appearances prior to the experiment, SOA only affected general rather than cue-specific preparatory processes.

Both scanning and verification epochs were shorter for picture- than word-cued trials ($F(1,11) = 44.15$, $MS_E = 14410.98$, $p < 0.001$, and $F(1,11) = 33.41$, $MS_E = 766.19$, $p < 0.001$, respectively). Even with the shorter SOA of 125 ms, there was still no effect of SOA on either the scanning or verification epochs ($F(1,11) = 0.58$, $MS_E = 15691.90$, $p = 0.462$, and $F(1,11) = 0.05$, $MS_E = 704.70$, $p = 0.823$, respectively). There was also no interaction between cue type and SOA in the scanning epoch ($F(1,11) = 0.55$, $MS_E = 31980.30$, $p = 0.473$) and only a marginal trend towards an effect in the verification epoch, $F(1,11) = 3.29$, $MS_E = 2948.79$, $p = 0.097$.

Fixation durations and regions visited during the scanning epoch were examined as a function of cue type and SOA (Table 3). Fewer regions of a scene were visited following picture cues, $F(1, 11) = 5.64$, $MS_E = 0.15$, $p <$

	Word Cue, 800 ms SOA		Word Cue, 125 ms SOA		Picture Cue, 800 ms SOA		Picture Cue, 125 ms SOA	
	Mean	*SE*	Mean	*SE*	Mean	*SE*	Mean	*SE*
Total trial duration	1411.76	(36.13)	1507.12	(74.19)	1135.62	(56.77)	1229.04	(58.59)
Search initiation time	264.52	(17.11)	320.64	(11.03)	254.92	(11.81)	328.94	(16.65)
Scanning time	772.15	(38.13)	838.08	(63.49)	580.24	(51.83)	569.44	(36.68)
Regions visited	2.59	(0.20)	2.41	(0.17)	2.15	(0.19)	2.32	(0.15)
Fixation durations	179.78	(5.17)	182.02	(7.14)	165.51	(6.61)	171.31	(6.72)
Verification time	375.09	(17.71)	348.39	(22.57)	300.45	(18.04)	330.65	(19.32)

Table 3. Results from Experiment 3. All means are in millisecond units, except for the item Regions visited, which is measured in the number of regions visited on the display screen (maximum 48).

0.05. There was no main effect of SOA, $F(1, 11) < 0.01$, $MS_E = 0.34$, $p = 0.957$, nor a significant interaction between cue type and SOA, $F(1, 11) = 2.24$, $MS_E = 0.16$, $p = 0.163$. Fixation durations were shorter for picture than word cued trials, $F(1, 11) = 30.54$, $MS_E = 139.16$, $p < 0.001$, and for longer than shorter SOAs $F(1, 11) = 5.37$, $MS_E = 75.37$, $p < 0.05$. However, cue type and SOA did not interact, $F(1, 11) = 0.01$, $MS_E = 182.20$, $p = 0.927$. These results replicate the results from Experiments 1 and 2.

Discussion

By shortening the short SOA condition to 125 ms, the SOA variable now produced an effect on total trial duration. However, despite this effect and despite the fact that participants were familiarized with the search targets prior to the experiment, we still did not observe an interaction between cue type and SOA on either total trial duration or on search initiation time. These results appear most consistent with the hypothesis that search initiation time reflects a general preparatory process rather than the time needed to set up a target template.

At the same time, cue specificity affected total trial duration, as in Experiments 1 and 2. This effect was seen in the scanning and verification epochs, with picture cues producing shorter times. Within the scanning epoch, picture cues were again found to reduce the number of regions visited and to shorten the fixation durations. Even though participants were familiarized with the visual forms of the search targets prior to the experiment, a more specific pictorial cue presented just before the search scene facilitated search. The facilitation was primarily due to faster reject decisions and better targeting during the scanning epoch, and faster target verification after it was fixated.

We again failed to replicate the cue type × SOA interaction found in Vickery et al. (2005) and Wolfe et al. (2004). Since we familiarized participants with targets prior to the experiment rather than repeating targets several times over the course of the experiment, it is possible that participants in our study did not remember targets as well as participants in the above studies. However, we note that all participants reported post hoc that they clearly remembered target appearances, even after one repetition of the familiarization phase. Secondly, these results are consistent with visual search studies reporting that participants prefer searching for a target rather than recalling it (Oliva, Wolfe, & Arsenio, 2004; Wolfe et al., 2004).

General discussion

Humans constantly search for task relevant objects to assist them in their daily lives. This search behavior is achieved by directing the eyes toward specific regions in the external environment. An important question therefore is what guides eye movements during search. Typically, research on gaze control during scene viewing has tended to focus on two sources of information: image features operating in a bottom–up manner (Itti & Koch, 2000; Itti et al., 1998; Parkhurst et al., 2002) and cognitive knowledge structures, particularly scene context, operating in a top–down manner (Castelhano & Henderson, 2007; Eckstein et al., 2006; Neider & Zelinsky, 2006; Torralba et al., 2006). Another top–down factor is the target template, which has been shown to guide eye movements during search tasks (Beutter et al., 2003; Eckstein et al., 2007; Findlay, 1997; Luria & Strauss, 1975; Motter & Belky, 1998; Rajashekar et al., 2006; Tavassoli et al., 2007, 2009; Williams, 1967). Knowledge about a target's exact properties allows the visual system to increase topographical activity associated with target similar features from the incoming scene percept and ignore (bias down) noisy activity from target irrelevant features.

The specificity of the template has also been shown to affect RT performance in search tasks (Bravo & Farid, 2009; Vickery et al., 2005; Wolfe et al., 2004). However, most previous studies have not addressed exactly how the specificity of a target template affects search and particularly how it affects eye movements related to search (but see Schmidt & Zelinsky, 2009). Schmidt and Zelinsky (2009) investigated how increased text label specificity facilitated eye movement behaviors during search. In the present study, we investigated how target template specificity affected the search process by analyzing three independent, behaviorally defined epochs of search, as well as their respective eye movement behaviors.

Three experiments confirmed that picture cues reduce total trial duration, replicating with real-world scenes previous studies that used object arrays (Castelhano et al., 2008; Schmidt & Zelinsky, 2009; Vickery et al., 2005; Wolfe et al., 2004). Furthermore, by using eye movement measures to divide the search process into functional epochs, we found that the shorter total trial duration in the picture cue condition was due to facilitated scanning and verification times (for similar results in object arrays, see also Castelhano et al., 2008) but not the time needed to set up a target template prior to search. Within the scanning epoch, a specific target template reduced the number of scene regions the participant visited and also the mean fixation duration.

The current results support previous research indicating that a target template can guide visual attention during real-world search (Zelinsky, 2008; Zelinsky et al., 1997, 2006). It has been known for some time that target features can guide visual search in a top–down manner (Wolfe, 1994; Wolfe, Cave, & Franzel, 1989). For example, Wolfe's *guided search model* posited that a stored representation of the search target modulates low-level feature maps to enhance particular channels; for example, if the target is green, a target template can be used to enhance the activation of all green items in the

color feature map. When the feature maps are summed they form an activation map that highlights regions in the visual display with target-similar features. The more definite the target template the more selective the feature maps, and thereby the activation map, can be. However, the guided search model does not explain how attention is sequentially distributed during a search task, an essential concern when considering eye movements in information-rich real-world scenes. Recent models predict that the visual system exploits precise visual properties of a target template to improve selection of peripheral locations on the activation map, with highly activated regions drawing attention, and eye movements (Rao et al., 2002; Zelinsky, 2008).

The present data extend these models in two particular ways. First, it shows that target template specificity affects how well eye movements are distributed during search. These models indicate that an exact target template affects attention distribution but do not speculate as to the effects of manipulating the specificity of the template (but see Zelinsky, 2008). Secondly, these models predict that a target template will benefit search by modulating the activation map to improve the selection of fixation locations but make no predictions about the time required for processing potential targets once fixated. Here, we find that the processing of fixated objects, whether they are the target or a distracter, is faster when a specific template is available for comparison.

The current results also extend the results from the Castelhano et al. (2008) study. That study similarly demonstrated that a picture cue facilitates the scanning epoch but did not address why this would be apart from participants needing fewer fixations to reach the target. Fixation count, however, is an ambiguous measure: fewer fixations could mean that an exact target template improves selectivity in the activation map so that more probable targets are fixated and less probable ones are ignored, thus reducing scanning time. Alternatively, fewer fixations could mean that an exact template allows attended distracters to be processed faster once fixated. Faster processing at fixation reduces the chances of a re-fixation occurring (Henderson & Ferreira, 1990), and fewer fixations would reduce scanning time. We found evidence for both possibilities: with an exact template participants visited fewer scene regions during scanning, indicating a better ability to select probable target regions; and scanning fixations were shorter, indicating that distracters were matched with the template and rejected quicker.

Another key feature of the results was the failure to replicate the interaction between cue type and SOA, found in both Vickery et al. (2005) and Wolfe et al. (2004). In these studies, trials cued with a word led to longer RTs than trials cued with a picture. Yet at longer SOAs, RTs for trials cued with a word were significantly reduced, while those cued with pictures grew slightly longer. The interpretation was that a target template created from an abstract cue takes longer to set up before becoming fully useful, whereas a precise target template created from a picture cue can be established more quickly but decays over time. Each of the previous studies, and the current one, used a similar type of categorical cueing in the word cue condition (e.g., the word cue *apple* as opposed to *fruit*), suggesting that our failure to replicate the previously found interaction was unlikely to be due to different text labels. An obvious difference between the studies of Vickery et al. (2005) and Wolfe et al. (2004) and the current one is that both previous studies used objects arrays—sets of objects with no established spatial relations between each other—whereas the present study used semantically rich real-world scenes. In object arrays the only top–down information available to guide search is knowledge of the target's appearance, so RTs might be sensitive to any changes to the specificity of the target template. In real-world scenes other information is available to guide search (e.g., scene context: Castelhano & Henderson, 2007; Eckstein et al., 2006; Torralba et al., 2006). A potential reason for not finding an interaction between cue type and SOA in the present study is that when a participant knows that other information will be available in the scene to guide search, they may devote fewer cognitive resources to creating a detailed target template. Participants may simply store a coarser version of the template—either extracting and storing fewer features from a picture cue or forming a less specific template from a word cue—leaving the template less susceptible to changes in SOA.

The present results also appear to contrast with those of Foulsham and Underwood (2007), who found no difference in search time or fixation count between categorical- and instance-cued search. The two studies differed, however, in that Foulsham and Underwood (2007) used a single word cue prior to a block of trials, either indicating a categorical target type (in their case fruit, meaning that several different types of fruit could be the target over the block of trials) or an instance cue (for example, apple, meaning that the target for every trial in the block would be an apple). Participants were never given an exact picture of a cue, and since the angle, size, luminance, color, and other features of the target naturally changed from trial to trial, it would be impossible for their participants to generate a specific target template. Even if participants could store a few features from trial to trial in the instance cue block (e.g., it was a red apple), our results (Experiment 3) and those of previous experiments (Wolfe et al., 2004) indicate that seeing a picture cue of the target is more effective than recalling it.

Toward an integrated real-world search model

Recent research has identified several sources of information that help guide search in novel, real-world scenes. However, most research to date has focused on how the visual system processes these forms of information

individually. For instance, the processing of image properties has tended to be studied in isolation in the saliency model (Itti & Koch, 2000; Parkhurst et al., 2002). Similarly, the effect of scene context has tended to be studied in isolation (Castelhano & Henderson, 2007; Eckstein et al., 2006; Neider & Zelinsky, 2006; Zelinsky & Schmidt, 2009). And in the present study, we studied the effect of target template specificity in isolation. In the real world, however, when all these forms of information are available, a more efficient method of guiding visual attention would be to integrate two or more information sources during a search task (see Ehinger, Hidalgo-Sotelo, Torralba, & Oliva, 2009).

The benefit of integrating different processes in real-world search has been demonstrated in the *contextual guidance model* (Torralba et al., 2006). In this model saliency at a local spatial scale is constrained by scene context at a global spatial scale. Areas of high salience within a selected global region are given higher weights on an activation map than those that fall outside the selected global region. The model accurately predicted participant's first few eye movements in a counting search task. This suggests that the visual system benefits from integrating multiple sources of information. Recent empirical evidence, however, indicates that the visual system relies less on low-level saliency than the contextual guidance model suggests (Einhäuser et al., 2008; Foulsham & Underwood, 2007; Henderson et al., 2007; Zelinsky et al., 2006). A future integrated real-world search model may benefit from substituting a form of top–down information, such as a target template, in place of saliency at the local spatial scale. Instead of selecting where within a selected global scene region to fixate based on saliency, a model could compare the target template stored in VWM with the incoming scene percept to weight the activation map accordingly. Local regions with high correlations to the target template are given higher weights on the activation map, particularly if they fall within the selected global region. This way the visual system is looking for target-similar features within a target-probable region of the scene rather than highly salient features which may not have any correlation with the target object.

In addition, a complete model of real-world search must consider fixation durations. When sites of fixations are also weighted by their durations, the distribution of attention changes dramatically (Henderson, 2007). The current data demonstrate that when there is a more specific target template available, scanning fixations are shorter as the visual system will have a more definite representation of the target to compare with a fixated region. Since fixation durations vary with the internal representation of the target while the display image and task stay constant, the current data support the hypothesis that fixation durations are, at least in part, under direct control during real-world tasks (Henderson & Pierce, 2008; Henderson & Smith, 2009).

Conclusion

The present study indicates that a target template can guide attention during real-world visual search. A more specific target template facilitates scanning for and verification of the target. Search initiation time appears to be an automatic process affected only by SOA, with longer SOAs producing shorter initial saccade latencies. Familiarizing participants with the target prior to testing did not affect the target template guidance process.

The current study also demonstrates that eye-tracking allows us insight into the processes underlying real-world visual search. By separating traditional unitary RT measures into three behaviorally defined epochs that probe the time taken to initiate search, the time to locate the target, and the time to accept an object as the target, we were able to examine the underlying processes that are affected by target template specificity and so increase our knowledge of search processes.

Acknowledgments

Commercial relationships: none.
Corresponding author: George L. Malcolm.
Email: g.l.malcolm@sms.ed.ac.uk.
Address: Psychology Department, 7 George Square, Edinburgh, EH8 9JZ, Scotland, UK.

Footnotes

[1]We chose the term activation map as it reflects the dynamic nature of the visual attention topographical map. There are several synonyms for this term, most notably saliency map. However, we avoided this term as it could cause confusion with bottom–up-driven saliency models.

[2]Target objects had invisible boundaries drawn around them, extending a mean of 0.47° outside the target's edge. Any fixations within this boundary were counted as on the target object.

References

Beutter, B. R., Eckstein, M. P., & Stone, L. S. (2003). Saccadic and perceptual performance in visual search

tasks: I. Contrast detection and discrimination. *Journal of the Optical Society of America A, Optics, Image Science, and Vision, 20,* 1341–1355. [PubMed]

Bravo, M. J., & Farid, H. (2009). The specificity of the search template. *Journal of Vision, 9*(1):34, 1–9, http://journalofvision.org/9/1/34/, doi:10.1167/9.1.34. [PubMed] [Article]

Burgess, A. (1985). Visual signal detection: III. On Bayesian use of prior knowledge and cross correlation. *Journal of the Optical Society of America A, Optics and Image Science, 2,* 1498–1507. [PubMed]

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,* 753–763. [PubMed]

Castelhano, M. S., Pollatsek, A., & Cave, K. R. (2008). Typicality aids search for an unspecified target, but only in identification and not in attentional guidance. *Psychonomic Bulletin & Review, 15,* 795–801. [PubMed] [Article]

Eckstein, M. P., Beutter, B. R., Pham, B. T., Shimozaki, S. S., & Stone, L. S. (2007). Similar neural representations of the target for saccades and perception during search. *Neuron, 27,* 1266–1270. [PubMed] [Article]

Eckstein, M. P., Drescher, B. A., & Shimozaki, S. S. (2006). Attentional cues in real scenes, saccadic targeting, and Bayesian priors. *Psychological Science, 17,* 973–980. [PubMed]

Ehinger, K. A., Hidalgo-Sotelo, B., Torralba, A., & Oliva, A. (2009). Modeling search for people in 900 scenes: A combined source model of eye guidance. *Visual Cognition, 17,* 945–978.

Einhäuser, W., Spain, M., & Perona, P. (2008). Objects predict fixations better than early saliency. *Journal of Vision, 8*(14):18, 1–26, http://journalofvision.org/8/14/18/, doi:10.1167/8.14.18. [PubMed] [Article]

Findlay, J. M. (1997). Saccade target selection during visual search. *Vision Research, 37,* 617–631. [PubMed]

Findlay, J. M., & Gilchrist, I. D. (2003). *Active vision: The psychology of looking and seeing.* King's Lynn: Oxford University Press.

Foulsham, T., & Underwood, G. (2007). How does the purpose of inspection influence the potency of visual saliency in scene perception? *Perception, 36,* 1123–1138. [PubMed]

Greenhouse, D. S., & Cohn, T. E. (1978). Effect of chromatic uncertainty on detectability of a visual stimulus. *Journal of the Optical Society of America, 68,* 266–267. [PubMed]

Henderson, J. M. (2007). Regarding scenes. *Current Directions in Psychological Science, 16,* 219–222.

Henderson, J. M., Brockmole, J. R., Castelhano, M. S., & Mack, M. (2007). Visual saliency does not account for eye movements during visual search in real world scenes. In R. P. G. van Gompel, M. H. Fischer, W. S. Murray, & R. L. Hill (Eds.), *Eye movements: A window on mind and brain* (pp. 537–562). Oxford: Elsevier Ltd.

Henderson, J. M., & Ferreira, F. (1990). Effects of foveal processing difficulty on the perceptual span in reading: Implications for attention and eye movement control. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 417–429. [PubMed]

Henderson, J. M., & Pierce, G. L. (2008). Eye movements during scene viewing: Evidence for mixed control of fixation durations. *Psychonomic Bulletin & Review, 15,* 566–573. [PubMed] [Article]

Henderson, J. M., & Smith, T. J. (2009). How are eye fixation durations controlled during scene viewing? Further evidence from a scene onset delay paradigm. *Visual Cognition, 17,* 1055–1072.

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research, 40,* 1489–1506. [PubMed]

Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 20,* 1254–1259.

Judy, P. F., Kijewski, M. F., Fu, X., & Swensson, R. G. (1995). Observer detection efficiency with target size uncertainty. *SPIE, 2436,* 10–17.

Land, M. F., & Hayhoe, M. M. (2001). In what ways do eye movements contribute to everyday activities? *Vision Research, 41,* 3559–3565. [PubMed]

Luria, S. M., & Strauss, M. S. (1975). Eye movements during search for coded and uncoded targets. *Perception & Psychophysics, 17,* 303–208.

Motter, B. C., & Belky, E. J. (1998). The guidance of eye movements during active visual search. *Vision Research, 38,* 1805–1815. [PubMed]

Neider, M. B., & Zelinsky, G. J. (2006). Scene context guides eye movements during visual search. *Vision Research, 46,* 614–621. [PubMed]

Oliva, A., Wolfe, J. M., & Arsenio, H. C. (2004). Panoramic search: The integration of memory and vision in search through a familiar scene. *Journal of Experimental Psychology: Human Perception and Performance, 30,* 1132–1146. [PubMed]

Parkhurst, D., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision Research, 42,* 107–123. [PubMed]

Rajashekar, U., Bovik, A. C., & Cormack, L. K. (2006). Visual search in noise: Revealing the influence of structural cues by gaze-contingent classification image analysis. *Journal of Vision, 6*(4):7, 379–386, http://journalofvision.org/6/4/7/, doi:10.1167/6.4.7. [PubMed] [Article]

Rao, R. P. N., Zelinsky, G. J., Hayhoe, M. M., & Ballard, D. H. (2002). Eye movements in iconic visual search. *Vision Research, 42,* 1447–1463. [PubMed]

Schmidt, J., & Zelinsky, G. J. (2009). Search guidance is proportional to the categorical specificity of a target cue. *Quarterly Journal of Experimental Psychology, 62,* 1904–1914. [PubMed]

Tavassoli, A., van der Linde, I., Bovik, A. C., & Cormack, L. K. (2007). An efficient technique for revealing visual search strategies with classification images. *Perception & Psychophysics, 69,* 103–112. [PubMed] [Article]

Tavassoli, A., van der Linde, I., Bovik, A. C., & Cormack, L. K. (2009). Eye movements selective for spatial frequency and orientation during active visual search. *Vision Research, 49,* 173–181. [PubMed]

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,* 766–786. [PubMed]

van Diepen, P. M. J., Wampers, M., & d'Ydewalle, G. (1998). Functional division of the visual field: Moving masks and moving windows. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 337–355). Oxford: Elsevier.

Vickery, T. J., King, L. W., & Jiang, Y. H. (2005). Setting up the target template in visual search. *Journal of Vision, 5*(1):8, 81–92, http://journalofvision.org/5/1/8/, doi:10.1167/5.1.8. [PubMed] [Article]

Williams, L. G. (1967). The effects of target specification on objects fixated during visual search. *Acta Psychologica, 27,* 355–360. [PubMed]

Wolfe, J. M. (1994). Guided Search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review, 1,* 202–238.

Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided search: An alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 419–433. [PubMed]

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,* 1411–1426. [PubMed]

Zelinsky, G. J. (2008). A theory of eye movements during target acquisition. *Psychological Review, 115,* 787–835. [PubMed]

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,* 448–453.

Zelinsky, G. J., & Schmidt, J. (2009). An effect of referential scene constraint on search implies scene segmentation. *Visual Cognition, 17,* 1004–1028.

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, MA: MIT Press.