

Contrasting Gist-Based and Template-Based Guidance During Real-World Visual Search

Brett Bahle, Michi Matsukura, and Andrew Hollingworth
The University of Iowa

Visual search through real-world scenes is guided both by a representation of target features and by knowledge of the semantic properties of the scene (derived from scene gist recognition). In 3 experiments, we compared the relative roles of these 2 sources of guidance. Participants searched for a target object in the presence of a critical distractor object. The color of the critical distractor either matched or mismatched (a) the color of an item maintained in visual working memory for a secondary task (Experiment 1), or (b) the color of the target, cued by a picture before search commenced (Experiments 2 and 3). Capture of gaze by a matching distractor served as an index of template guidance. There were 4 main findings: (a) The distractor match effect was observed from the first saccade on the scene, (b) it was independent of the availability of scene-level gist-based guidance, (c) it was independent of whether the distractor appeared in a plausible location for the target, and (d) it was preserved even when gist-based guidance was available before scene onset. Moreover, gist-based, semantic guidance of gaze to target-plausible regions of the scene was delayed relative to template-based guidance. These results suggest that feature-based template guidance is not limited to plausible scene regions after an initial, scene-level analysis.

Public Significance Statement

Many tasks in our modern world are visual search tasks, such as baggage screeners looking for dangerous items or radiologists looking for tumors. The present study aids in understanding how real-world searches of this sort are performed and can be optimized. In real-world contexts, searchers use the visual properties of target objects and knowledge of where these objects are likely to be located. The findings from the present study expand our understanding of the relationship between these 2 sources of guidance.

Keywords: visual search, attentional guidance, eye movements, visual attention

Real-world tasks require that attention is guided efficiently to goal-relevant objects. For example, making a sandwich for lunch depends on a series of visual searches in which attention is directed to different objects as each becomes relevant (Land & Hayhoe, 2001). How is such strategic guidance implemented? Early theories of visual search emphasized the importance of guidance by a *target template* to explain search performance. This classic approach is exemplified in Wolfe's (1994) guided search model. Bottom-up saliency is determined by the physical properties of the constituent objects, their similarity to neighboring stimuli, and their spatial proximity. Top-down guidance is implemented as a selective increase in the activity of feature values matching the

target template. Attention is therefore biased toward items in the visual field that share target features. This type of model, and similar approaches (e.g., Desimone & Duncan, 1995; Duncan & Humphreys, 1989), capture key processes of visual search in the domain for which they were developed: simple visual stimuli arrayed in randomly determined locations.

In contrast to traditional search displays, real-world scenes have conceptual structure (objects tend to appear in similar locations across different scenes from a category) and episodic consistency (objects tend to appear in the same places as they were previously observed). Thus, scenes allow additional forms of strategic guidance beyond that provided by a template. Indeed, a large set of studies has demonstrated that, during search through real-world scenes, gaze is directed to regions of the scene where an object of that type is likely to be found (Castelhano & Heaven, 2011; Castelhano & Henderson, 2007; Eckstein, Drescher, & Shimozaki, 2006; Henderson, Malcolm, & Schandl, 2009; Henderson, Weeks, & Hollingworth, 1999; Malcolm & Henderson, 2009; Neider & Zelinsky, 2008; Vö & Henderson, 2010) and to locations where the target has been observed previously (Brockmole & Henderson, 2006a, 2006b; Brooks, Rasmussen, & Hollingworth, 2010; Hollingworth, 2012; Vö & Wolfe, 2012). The allocation of attention during search through scenes is also influenced by the physical

This article was published Online First August 10, 2017.

Brett Bahle, Michi Matsukura, and Andrew Hollingworth, Department of Psychological and Brain Sciences, The University of Iowa.

This research was supported by a National Institute of Health grant (R01EY017356) to Andrew Hollingworth.

Correspondence concerning this article should be addressed to Brett Bahle, Department of Psychological and Brain Sciences, University of Iowa, W311 Seashore Hall, Iowa City, IA 52242. E-mail: brett-bahle@uiowa.edu

salience of particular objects (Itti & Koch, 2001), but salience is typically unrelated to task relevance and does not adapt to changing task demands. Thus, salience appears to play a relatively minor role in the guidance of goal-directed search (Henderson et al., 2009).

In sum, search through real-world scenes is guided by several forms of knowledge: knowledge of the visual properties of the target object (template guidance), knowledge of where the object is likely to be found in a scene of that type (semantic guidance),¹ and knowledge of where the object previously appeared in that particular scene exemplar (episodic guidance). However, the precise interactions between these sources of guidance are not well established. The contributions of template and semantic guidance are central to most current theories of search through real-world scenes, and here we will focus on the relative contributions of these two sources and the time course of their implementation.

Similar to classical approaches, some theories of scene-based search stress a primary role for a target template (e.g., Hwang, Higgins, & Pomplun, 2009; Navalpakkam & Itti, 2005; Zelinsky, 2008), and empirical work demonstrates the importance of template information in search through scenes (e.g., Malcolm & Henderson, 2009). However, other theories have suggested that guidance in natural scenes is dominated by scene-level information, particularly at early stages of the search process (Ehinger, Hidalgo-Sotelo, Torralba, & Oliva, 2009; Torralba, Oliva, Castelano, & Henderson, 2006; Wolfe, Võ, Evans, & Greene, 2011). Specifically, rapid recognition of the scene category and layout (i.e., gist) allows search to be limited, in large part, to regions of the scene where the target is likely to be found.

For example, in the contextual guidance model of Torralba et al. (2006), search is guided by semantic information (learned associations between scene types and object locations) and saliency. These are integrated prior to the first saccade on a scene. Thus, areas selected for fixation tend to be contextually probable locations for the target, and the effects of salience are generally limited to these regions. This contextual guidance model was later updated to include a template representation (Ehinger et al., 2009), but the revised model also stresses a strong role of initial guidance by semantic knowledge. In a similar vein, the two-pathway architecture of Wolfe et al. (2011) consists of a nonselective pathway driven by scene gist recognition and a selective pathway driven by guidance from a target template. The former pathway is nonselective in that it depends on a rapid, global analysis of scene information, leading to gist recognition, and the parsing of the scene into regions where the target is and is not likely to be found. The latter pathway employs template guidance and is primarily restricted to those regions identified as plausible for the target from the global analysis.

The primacy of a scene-based pathway is possible given evidence for highly efficient gist recognition. A scene's basic or superordinate level category can be recognized from presentations as brief as 40 ms (Greene & Oliva, 2009; Oliva & Schyns, 2000; Potter & Levy, 1969; Rousselet, Joubert, & Fabre-Thorpe, 2005; Sanocki, 2003).² Moreover, very brief presentations of a scene allow guidance to be implemented based on perception of the scene category and its broad spatial layout (Castelhano & Henderson, 2007; Võ & Henderson, 2010). For example, Castelano and Henderson (2007) used a flash-preview-moving-window paradigm to investigate the role of initial scene representations in facilitating

visual search. Participants were first shown a 250-ms preview of an upcoming search scene, followed by a mask, and then a category label for a search target. Finally, a scene was displayed in which participants searched for the target through a 2° circular window dynamically centered at fixation. Search times decreased when a preview of the search scene was displayed compared with control conditions displaying either a different scene or a pattern mask, suggesting that a brief glance of a scene is sufficient to significantly improve search efficiency.

In the present study, we tested whether gist-based, semantic guidance dominates the early stages of visual search, with the working hypothesis that classic, template-based guidance may play a larger role in early search than has been proposed in recent work (e.g., Wolfe et al., 2011). Why might this be the case? First, gist-based guidance requires not only recognizing the scene type or exemplar but also perceiving the spatial layout of the scene, perceiving the identities of particular surfaces (e.g., kitchen counter vs. kitchen table), and mapping stored knowledge of typical object locations (that blenders tend to be found on kitchen counters but not on kitchen tables) to the parsed structural representation of the scene. Thus, although gist can be extracted with a 40-ms presentation, this does not necessarily mean that gist-based guidance can be *applied* to the search process so quickly, as application requires several additional processes. How long might it take to implement gist-based guidance? Võ and Henderson (2010, Experiment 3) used the flash-preview-moving-window paradigm and varied the stimulus onset asynchrony (SOA) between the brief scene preview and the onset of the search scene. Longer SOAs led to a larger search benefit, suggesting that gist-based guidance of search may take as long as 3,000 ms to configure optimally. This leaves open the possibility that other forms of guidance (e.g., a template specifying target features) may influence the early stages of search. That is, template guidance may not be limited to a secondary role applied only to semantically plausible regions of the scene.

Second, recent studies have shown that the content of visual working memory (VWM), the standard substrate for template-based guidance (Desimone & Duncan, 1995; Duncan & Humphreys, 1989; Woodman, Carlisle, & Reinhart, 2013), has a strong influence on attention and saccade target selection even when there are strong competing biases from knowledge of relevant locations (Hollingworth, Matsukura, & Luck, 2013; Schneegans, Spencer, Schöner, Hwang, & Hollingworth, 2014). These studies had participants execute a saccade to a target that always appeared on the horizontal midline. When a distractor on the vertical midline matched an incidental color held in VWM, saccades were directed to the distractor on a large proportion of trials. That is, even with the possibility of extremely strong guidance from knowledge of the relevant target region, initial eye movements were often

¹ We treat gist-based, *semantic guidance* as encompassing all structural effects of scene categorical knowledge on attention during search. These can include the relationship between scene category and object presence and location, the relationship between scene surface structure and typical object location (Castelhano & Heaven, 2011), and the relationship between prominent landmarks and typical object location (Koehler & Eckstein, 2017).

² Potter, Wyble, Hagmann, and McCourt (2014) reported gist detection at 13 ms per scene image, but this may have been due to inadequate masking (Maguire & Howe, 2016).

controlled based on incidental VWM content. Thus, when the content of VWM is a template specifying the features of the target object, it is quite possible that the template would override guidance based on knowledge of plausible spatial locations, such as that derived from recognition of the scene gist.

Although several studies have manipulated template and gist-based guidance (e.g., [Malcolm & Henderson, 2010](#)), neither their relative priority nor time course has been examined independently because both template and gist-based guidance would have driven attention to the same scene region. In three experiments, we dissociated guidance from these two sources of information using a version of a VWM-based capture technique ([Hollingworth & Luck, 2009](#); [Olivers, 2009](#); [Soto, Heinke, Humphreys, & Blanco, 2005](#)). Eye movements were monitored as participants searched for target objects in real-world scenes. On some trials, a distractor was present that either matched a color held in VWM for a secondary task (Experiment 1) or matched the color of the search target (Experiments 2 and 3). In addition, this distractor could either appear in a semantically plausible region for the target object or a semantically implausible region, allowing us to test whether template-based guidance was limited to regions of the scene that were likely to contain the target and to examine the time course of the template-based guidance effect.

To preview the results, in Experiments 1 and 2, a VWM- or template-matching distractor was at least twice as likely to be fixated than a nonmatching distractor. This difference was observed from the very first saccade on the scene and was independent of the relevance of the distractor location for finding the target. Moreover, in Experiment 1, the *gist-based* guidance of attention to target-plausible scene regions was delayed relative to

template-based guidance to the critical distractor. In Experiment 3, a VWM-matching distractor likewise captured gaze, even when participants were shown a scene preview that should have allowed them to prepare a gist representation for search guidance at scene onset. The data indicate a key role for template-based guidance in search through scenes, a role that is not strongly limited, either temporally or spatially, by the application of scene-level knowledge.

Experiment 1

In Experiment 1, we investigated the relationship between semantic and template guidance by placing gist-based, semantic guidance in conflict with guidance from a secondary color maintained in VWM. The procedure is illustrated in [Figure 1](#). Participants searched for a letter superimposed on a target object (indicated by green, right-hand rectangle in [Figure 1](#)) in a natural scene, receiving no cue, a category label cue, or an exact picture cue of the target object. Simultaneously, they maintained a color in VWM for a later memory test. A distractor (indicated by red, left-hand rectangle in [Figure 1](#)) in the scene either matched (*match* trials) or did not match (*mismatch* trials) the VWM color. Note that the color in VWM was not functioning as a template per se, as it did not indicate the color of the target. However, the incidental content of VWM guides attention in a manner similar to the explicit template of search ([Olivers, 2009](#); [Olivers, Meijer, & Theeuwes, 2006](#); [Soto et al., 2005](#)). By separating the target representation (as specified by the target cue) and the VWM color, we could manipulate the specificity of the target information independently of the VWM color. This allowed us to introduce different levels of

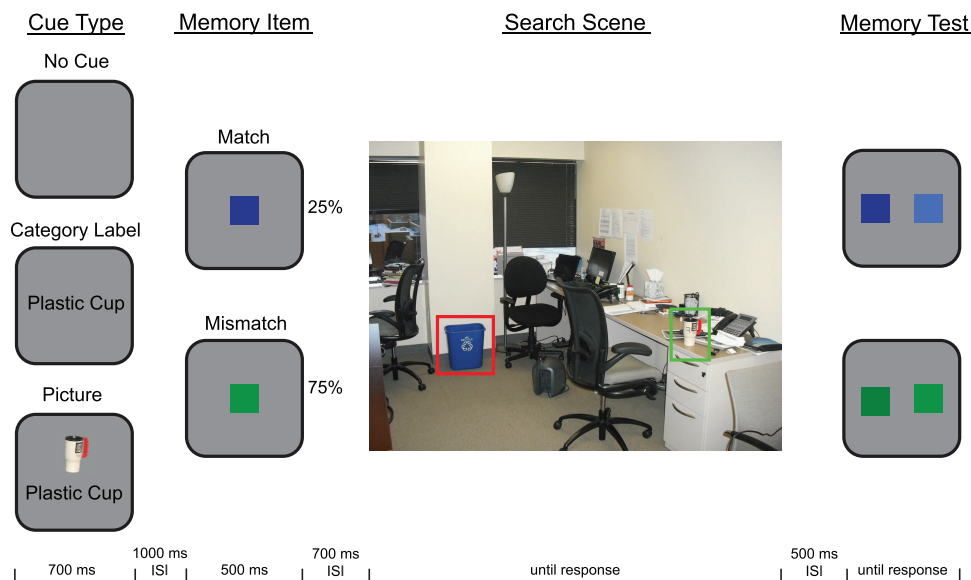


Figure 1. Trial event sequence for Experiment 1. Each trial began with a cue presentation (700 ms). After a 1,000-ms delay, a memory color was presented for 500 ms for a postsearch memory test. After another 700-ms delay, a search scene was presented, which always contained the target (marked by green, right-hand square, not present in experimental image). The memory color could either match or mismatch a critical distractor object in the scene (marked by red, left-hand square, not present in experimental image). After responding to the orientation of an “F” (Arial font) on their target object, there was a 500-ms delay before subjects completed a two-alternative forced choice color memory test. See the online article for the color version of this figure.

gist-based, semantic guidance while keeping the potential interaction between VWM and the colors of the objects in the scene relatively constant. The *no-cue* condition provided no information about the target object, and thus semantic information (plausible target locations) could not be used to guide search. The *category-label-cue* condition and *picture-cue* condition both provided information about the identity of the object that could be combined with scene gist and layout to guide attention to plausible scene locations. That is, knowing that the target was a blender, participants could use gist recognition of a kitchen scene to limit search to plausible locations, such as the kitchen counter.

The primary measure was the probability of fixating the critical distractor across the course of search as a function of three variables: (a) the match between the distractor color and the VWM color; (b) the availability of gist-based guidance, as implemented by the cue manipulation; and (c) the location of the distractor either in a plausible or implausible location for the target. Moreover, we investigated the gist-based guidance of attention by examining the probability of fixating scene regions in which the target could plausibly appear. This allowed us to compare the relative contributions of template-based and gist-based guidance within the search process and the time course of each.

If feature-based template guidance plays a more substantial role than claimed under recent theories that stress gist-based guidance, the probability of fixating the critical distractor early during search should be significantly higher on match compared with mismatch trials. In addition, if feature-based template guidance is applied across the scene and is not limited to plausible regions, then the effect of VWM match should be observed for distractors that appear both in plausible and implausible locations for the target. Finally, the time course of feature-based guidance need not be delayed relative to the time course of semantic guidance. In contrast, theories that stress the primacy of gist-based, semantic guidance predict that gaze should be directed rapidly to plausible scene regions, and if an effect of VWM match is present, it should be limited, in large part, to later stages of search and to regions of the scene where the target is likely to appear.

In addition to the central goals of the experiment, described above, Experiment 1 was designed to contribute to an important debate regarding the architecture of interaction between VWM and attentional selection. Under one view (Olivers, Peters, Houtkamp, & Roelfsema, 2011), only a single representation in VWM is maintained in a state that guides selection. In contrast, we have argued that, because multiple VWM items can be maintained as active representations in sensory cortex (Emrich, Riggall, Larocque, & Postle, 2013), there need not be any such architectural constraint (Beck & Hollingworth, 2017; Beck, Hollingworth, & Luck, 2012; Hollingworth & Beck, 2016). The primary behavioral evidence supporting the single-item-template hypothesis comes from studies in which participants maintained both an immediately relevant search target in VWM and a secondary memory item for a later test (Downing & Dodds, 2004; Houtkamp & Roelfsema, 2006). However, the results from these studies have been ambiguous. Some experiments have demonstrated no capture by a distractor matching the secondary memory item (Downing & Dodds, 2004; Houtkamp & Roelfsema, 2006, Experiments 1A, 2A, and 3), consistent with the single-item-template hypothesis, and others have demonstrated capture (Hollingworth & Beck, 2016; Houtkamp & Roelfsema, 2006, Experiments 1B and 2B), consistent

with the multiple-item-template hypothesis. In Experiment 1, we included the picture-cue condition to create a test similar to previous studies. A representation of the target picture should be maintained as a template in VWM. If VWM-based guidance is limited to a single item, the memory color should not interact with oculomotor selection, producing no effect of distractor match. However, if both the target representation and the memory color interact with selection, we should observe both relatively efficient guidance toward the search target and capture by an object matching the memory color.

Method

Participants. Forty-four (34 female) participants from the University of Iowa community completed the experiment for course credit. Each was between the ages of 18 and 30 and reported normal or corrected-to-normal vision (we excluded participants who needed contact lenses to achieve normal vision). Cue type was manipulated between subjects. Analysis of a pilot experiment ($N = 20$) similar to the category-label-cue condition indicated that for the effect of VWM match on distractor fixation probability, eight participants would be required to obtain 80% power (Cohen, 1988). As a conservative approach, in Experiment 1, we ran 12 participants in the no-cue condition and 12 in the category-label-cue condition. Twenty participants were allocated to the picture-cue condition, with the expectation that effective size might be reduced when both the search target and memory color were maintained in VWM. However, the results from the picture-cue condition were functionally equivalent in the full analysis and in an analysis limited to the first 12 participants run. We report data from the full set of 20 participants.

Stimuli. For the search task, 96 photographs of real-world scenes were used, each subtending $26.32^\circ \times 19.53^\circ$ visual angle at a resolution of 1280×960 pixels. The scenes were mostly indoor locations (e.g., an office or a bedroom) collected online using several different search engines. Target objects were chosen as semantically consistent with the scene and always appeared in a typical location (e.g., a stapler on an office desk). The Appendix lists the category of each scene and the identity of each target object. Note that each scene and each target were unique (there was no repetition of scene backgrounds or targets). Targets subtended between $1.63^\circ \times 1.59^\circ$ and $9.30^\circ \times 7.56^\circ$ visual angle, with a mean of $3.27^\circ \times 3.34^\circ$. None of the targets appeared at the center of the scene, so that participants had to execute at least one saccade to complete the search task. The mean eccentricity of targets (scene center to object center) was 7.62° . Superimposed on the search target object in each scene was a left- or right-facing "F" in Arial font, subtending $0.25^\circ \times 0.41^\circ$. The "F" was black, white, or gray, chosen to ensure visibility when superimposed over each target.

Half ($n = 48$) of the scenes contained a critical distractor object that could either match or mismatch a color maintained in VWM. The other half of the scenes consisted of filler scenes. However, for the filler scenes, the color held in VWM was chosen so that it was never a close match for any object in the scene; thus, these filler scenes functioned as additional mismatch scenes, limiting the match trials to only 25% of all trials in the experiment. For the other 48 scenes containing a critical distractor object, one fairly salient object was chosen as the distractor. In addition, it was

chosen to have a relatively uniform color across its surface. Each distractor was the only object with that particular color in the scene, so that any capture effect would be limited to the distractor object. To do this, in some scenes, we used photo-editing software to change the color of the distractor so that it was unique within the scene. Distractor colors varied across the set of scene items. The distractors ranged from subtending $1.63^\circ \times 1.59^\circ$ to $9.40^\circ \times 8.70^\circ$, with a mean of $3.34^\circ \times 3.86^\circ$. Note that the appearance of the scene and distractor was identical in the match and mismatch conditions; VWM match was manipulated by changing the remembered color, not the scene. Thus, the physical saliency of the distractor objects was controlled across match and mismatch conditions.

The memory color square subtended $1.64^\circ \times 1.64^\circ$, presented centrally. On match trials, the memory color was calculated as the average RGB color value across all pixels of the critical distractor object. Note that in this design, there was rarely any major part of the object that was an exact match with the remembered color. Thus, given the subtle color difference in the memory test stimuli (described below), it is unlikely that participants would have attended to the distractor in order to improve memory performance. On mismatch trials, the memory color was chosen from a different color category than on match trials (e.g., if the memory color was a red hue on match trials, the memory color could be a green hue on mismatch trials). The particular memory colors associated with a scene item (match and mismatch) remained the same across participants. In the memory test at the end of the trial, two colored squares were presented to the left and right of central fixation. One of the colored squares was an exact match to the color presented for memorization. The other colored square varied from the exact match square by ± 20 on each of the three RGB channels, with the \pm direction determined randomly for each channel. The location of the matching color (and thus the correct memory test response) was also determined randomly.

There were three target-cue conditions: no-cue, category-label-cue, and picture-cue. In the category-label-cue condition, the cue was a basic- or subordinate-level label for the object (e.g., “blender” or “coffee mug”) presented in black Arial font. In the picture-cue condition, the cue was an image of the target object extracted from the scene itself, presented centrally. To avoid any ambiguity in the identity of the pictured target, the category label was also presented beneath the picture.

Apparatus. The apparatus was the same for all experiments. The stimuli were presented on an LCD monitor with a refresh rate of 100 Hz at a viewing distance of 77 cm maintained by a chin and forehead rest. The position of the right eye was monitored using an SR Research Eyelink 1000 eye tracker, sampling at 1,000 Hz. Manual responses to both the search and memory test were collected with a response pad. The experiment was controlled by E-Prime software.

Procedure. After arriving for the experiment session, participants provided informed consent and received both oral and written instructions. They were instructed that the primary task was to find the letter “F” in each scene based on the cue (if present) indicating the target object, and to report whether the “F” was normally oriented or mirror-reversed. They were also instructed that there would sometimes be an object in the scene with a color similar to the color of the remembered square, but this object would never contain the “F.” The eye tracker was calibrated at the

beginning of the session and was recalibrated if the estimate of gaze position deviated from the central fixation point by more than approximately 0.75° .

The experimenter initiated each trial as the participant maintained central fixation. In the category-label and picture-cue conditions, a fixation cross appeared for 400 ms, followed by the target cue for 700 ms, followed by another fixation cross for 1,000 ms. In the no-cue condition, there was just a 1,000-ms fixation cross following the trial onset. Next, in all conditions, the colored memory square appeared for 500 ms, to be maintained in VWM for the later memory test. A final fixation cross followed the square for 700 ms, and the search scene was presented until response. Participants pressed the right button on a button box to indicate a normal “F” and the left button to indicate a mirror-reversed “F.” After this response, there was a 500-ms delay, followed by the color memory test display until response. Participants used the same buttons to indicate the color square (left or right) that was the exact match for the memory color. No feedback was given for either response.

Participants first completed a practice session of six trials. Then, they completed an experiment session of 96 trials: 24 match trials, 24 mismatch trials, and 48 filler trials, with trial order randomly determined. Participants saw each scene item once. Across an experiment, each scene item appeared in each condition an equal number of times, with the assignment of scenes to the match conditions counterbalanced across pairs of participants. For the first participant in the pair, the assignment of scene items to match conditions was determined randomly. For the second, the assignment was reversed. The entire session lasted approximately 30 min.

Data analysis. Saccades were defined by a combined velocity ($30^\circ/\text{s}$) and acceleration ($8000^\circ/\text{s}^2$) threshold. Eye-tracking data was analyzed with respect to two regions of interest (ROIs): the *target region* and the *critical distractor region*, which never overlapped. Both regions were rectangular and extended approximately 0.3° beyond the edges of the target and critical distractor objects, respectively. Trials were eliminated from the analysis if the target object was not fixated before the response, if the response to the letter orientation was incorrect, if the first fixation on the scene was not within a $1.54^\circ \times 1.54^\circ$ central square region, or if a given search time was greater than 2.5 standard deviations from a subject’s condition mean. This process resulted in the removal of 19.3% of trials in the no-cue condition, 16.5% of trials in the category-label-cue condition, and 17.1% of trials in the picture-cue condition. The pattern of results was not affected by trial removal.

Results and Discussion

We begin by presenting summary statistics for the memory and search tasks. We then present the critical results concerning measures of distractor fixation.

Memory accuracy. Memory and search accuracy data are reported in Table 1. Overall, memory accuracy was 67%. The data were entered into a 3 (cue condition: no-cue, category-label-cue, picture-cue) \times 2 (distractor color match condition: match, mismatch) mixed-design analysis of variance (ANOVA). There was no main effect of cue condition, $F(2, 41) = 1.35$, $p = .271$, $\eta_p^2 = .062$, no main effect of distractor-match conditions, $F(1, 41) =$

Table 1
Memory and Search Accuracy From Experiment 1

Condition	Memory accuracy		Search accuracy	
	Match	Mismatch	Match	Mismatch
No-cue	64.9%	71.2%	92.4%	95.1%
Category-label-cue	70.1%	68.1%	99.0%	98.6%
Picture-cue	64.2%	64.9%	97.2%	95.3%

0.512, $p = .478$, $\eta_p^2 = .012$, and no interaction, $F(2, 41) = 1.01$, $p = .373$, $\eta_p^2 = .047$.

Search accuracy. Overall search accuracy was very high (96%). The data were entered into a 3 (cue condition) \times 2 (distractor color match condition) mixed-design ANOVA. As with memory accuracy, there was no main effect of cue condition, $F(2, 41) = 1.67$, $p = .201$, $\eta_p^2 = .075$, and no main effect of distractor match condition, $F(1, 41) = .086$, $p = .771$, $\eta_p^2 = .002$. There was a significant interaction, $F(2, 41) = 4.73$, $p = .014$, $\eta_p^2 = .187$, driven by slightly lower performance in the no-cue condition on match (92.4%) versus mismatch trials (95.1%), consistent with the capture effects reported below.

Search time. Search time was defined as the elapsed time from the onset of the search scene to the beginning of the first fixation on the target object (Figure 2A). Reaction time for the manual response produced an equivalent pattern of results. First, there was a main effect of cue type, $F(2, 41) = 36.57$, $p < .001$, $\eta_p^2 = .641$, with progressively faster searches as the cue became more specific. Second, there was a main effect of distractor color match, with faster search on mismatch (960 ms) compared with match (1,070 ms) trials, $F(1, 41) = 7.15$, $p = .011$, $\eta_p^2 = .148$. The interaction did not reach significance, $F(2, 41) = .472$, $p = .627$, $\eta_p^2 = .023$. Examining the pairwise differences between cue conditions, there was a reliable advantage for the category-label-cue condition (825 ms) over the no-cue condition (1,663 ms), $F(1, 22) = 26.84$, $p < .001$, $\eta_p^2 = .550$, and a numerical trend toward an advantage for the picture-cue (740 ms) condition over the category-label-cue condition, $F(1, 30) = 2.94$, $p = .097$, $\eta_p^2 = .089$.

Measures of distractor fixation (template guidance). Next, we examined the probability of fixating the critical distractor object as a function of VWM color match (Figure 2B). Participants were more likely to fixate the distractor on match (.48) compared with mismatch (.25) trials, $F(1, 41) = 68.5$, $p < .001$, $\eta_p^2 = .626$, demonstrating guidance from the incidental content of VWM. This effect was seen for all three cue conditions: for the no-cue condition, $F(1, 11) = 9.42$, $p = .011$, $\eta_p^2 = .461$ (.62 vs. .40), for the category-label-cue condition, $F(1, 11) = 68.23$, $p < .001$, $\eta_p^2 = .861$ (.46 vs. .21), and for the picture-cue condition, $F(1, 19) = 35.05$, $p < .001$, $\eta_p^2 = .648$ (.40 vs. .18).³ There was also a main effect of cue condition, with decreasing probability of distractor fixation as the cue became more specific, $F(2, 41) = 23.59$, $p < .001$, $\eta_p^2 = .535$, consistent with more efficient guidance with more specific cues. Critically, there was no interaction between cue condition and distractor match, $F(2, 41) = 0.14$, $p = .872$, $\eta_p^2 = .007$. In particular, the capture effect (i.e., the difference in fixation probability on match compared with mismatch) was numerically at least as large when participants were provided with a cue (.25 for

the category-label cue, .22 for the picture-cue), and thus could have used gist-based guidance, as when they were given no information about the target (.22 for the no-cue).

To examine the time course of the distractor match effect, we calculated the probability of fixating the distractor object for each ordinal fixation index during search. In Figure 2C, the first participant-controlled fixation (after the first saccade on the scene) is denoted as Fixation 1. Figure 2C shows each probability for the match and mismatch trials in the three cue conditions. A trial was removed from the analysis upon target fixation. Given that different trials ended at different ordinal fixation indices, as ordinal fixation increased, fewer and fewer trials contributed to the analysis. A participant's data were included only if the participant contributed at least eight trials to each condition. In addition, a value is plotted in the figure only if nine of the 12 participants' data were available for the no-cue and category-label-cue conditions, and 15 of 20 for the picture-cue condition. All 12 participants were included in all ordinal fixations bins for the no-cue condition; all 12 participants were included in all bins in the category-label-cue condition; all 20 subjects in the picture-cue condition were included for Bins 1 through 5, and 15 contributed to Fixation Bin 6. Thus, fewer data points are plotted for the cue conditions in which search was more efficient. The key result, as evident in Figure 2C, was that the distractor color-match effect was largest for Ordinal Fixations 1 and 2. Maintaining a particular color in VWM had a substantial influence on saccade target selection from the very first saccade on the scene.

Figure 2D displays the Euclidean distance (in degrees of visual angle) of the current fixation point from the center of the target

³ It is important to eliminate the possibility that the effect of match by the remembered color was caused by low-level priming from the mere presentation of the memory color before the search. Thus, we ran a control experiment ($N = 12$; 6 female) that was identical to the category-label-cue condition of Experiment 1, except there was no posttrial color memory test. Although participants saw the memory color, there was no demand to actively maintain the color in VWM during search. Participants were told the colored square signaled that the trial was about to begin. Trials were removed according to the same criteria as in Experiment 1 (12.8%). In contrast to the results reported above, color match did not reliably influence search time, $F(1, 11) = 1.00$, $p = .337$, $\eta_p^2 = .083$ (811 ms for match vs. 872 ms for mismatch). In addition, there was no reliable difference in the probability of fixating the distractor object between match (0.24) and mismatch (0.21) trials, $F(1, 11) = 1.19$, $p = .297$, $\eta_p^2 = .097$. The control experiment had sufficient power to detect a mean fixation probability difference of .08 between match conditions. Thus, it had enough power to detect an effect substantially smaller than the effect observed in the main experiment (.23 in the corresponding category-label-cue condition of the main experiment). In sum, the distractor color-match effect in Experiment 1 was unlikely to have been caused by low-level priming.

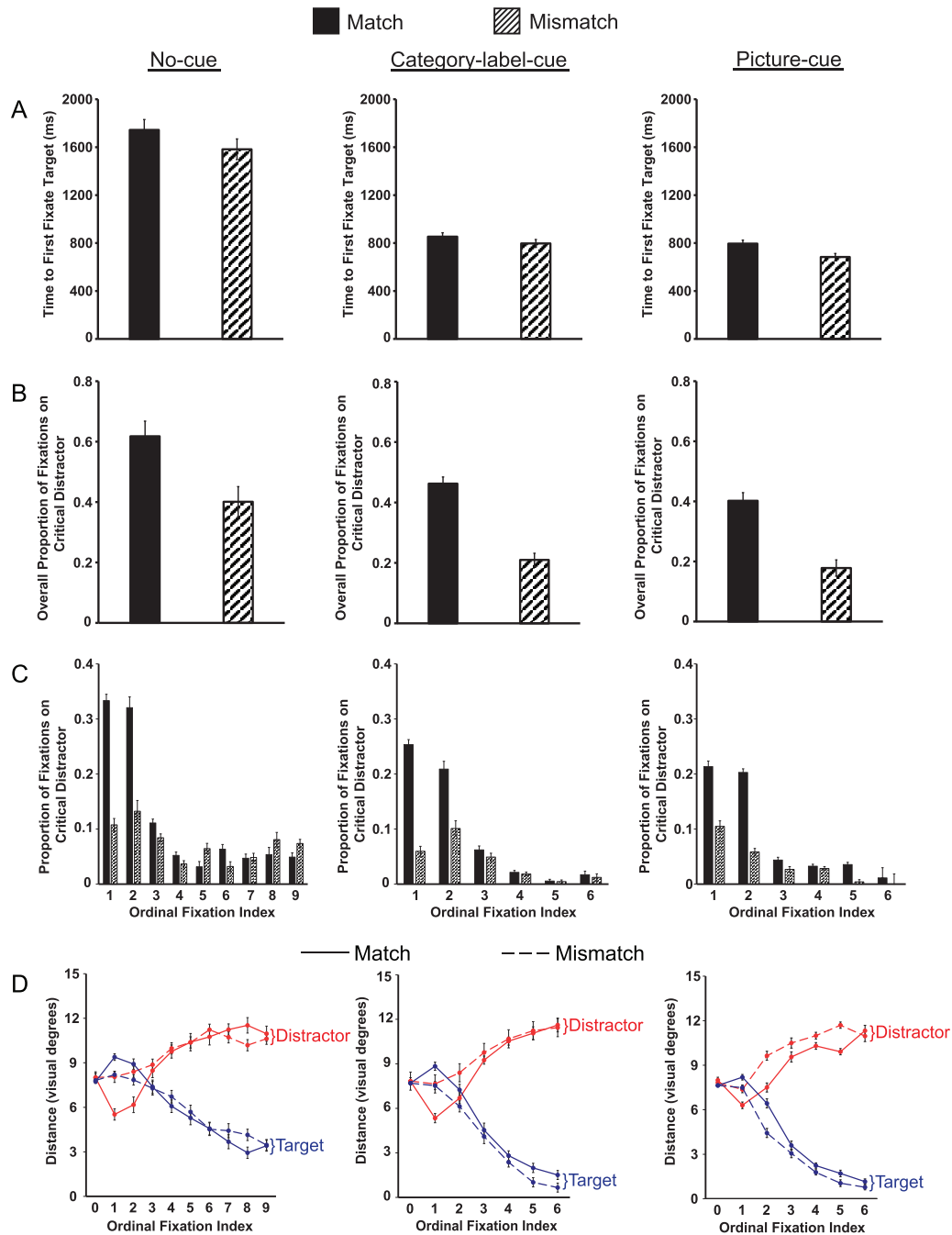


Figure 2. Data figures for Experiment 1. From left to right, data are plotted for the no-cue, category-label-cue, and picture-cue conditions. (A) Time (ms) to first fixate the target as a function of distractor color match. (B) Overall probability of fixating the critical distractor as a function of color match. (C) Proportions of fixations landing on the critical distractor object as a function of ordinal fixation index and color match. (D) The distance, in visual degrees, of the current fixation to the target and critical distractor object as a function of ordinal fixation index/number. Error bars indicate within-subjects 95% confidence intervals (Morey, 2008). See the online article for the color version of this figure.

object (blue lines) and from the center of the distractor object (red lines), as a function of ordinal fixation index. Note that Fixation 0 (the first fixation on the scene, controlled by the experimenter) was included to show that there was minimal variability in the distance

from the two objects (i.e., target, distractor) at the start of each trial. For this analysis, data from ordinal fixations after the target had been fixated were included (with a value of zero) to reflect the overall progress of guidance toward the target. On match trials, the

result of the first few saccades indicates that the eyes were taken away from the target and directed toward the distractor, on average. On mismatch trials, the first saccade did not reduce the distance to the target, but subsequent saccades brought the eyes systematically closer to the target. Thus, on match trials, the incidental distractor match interfered with the process of guiding the eyes toward the target.

Next, we considered whether the distractor color-match effect was modulated by the semantic plausibility of the distractor location as a location for the target. Such an effect would be expected if the application of VWM/template information was limited to target-plausible regions of the scene. A new group of eight (five female) undergraduates from the University of Iowa rated the plausibility of the 48 critical distractor locations. Specifically, they saw each scene for a self-determined amount of time, with the target marked by a green circle and the distractor marked by a white square. They reported whether the target could plausibly occur in the location occupied by the distractor, on a scale of 1 to 7. The interclass correlation showed the across-subjects ratings to be reliably similar, $R = .864$, $F(47, 329) = 7.33$, $p < .001$. We collapsed the rated values across subjects and performed a median split of the scene items (see Experiment 2 for an illustration of a scene with a moderately high rating of plausibility and a scene with a moderately low rating of plausibility). The data were entered into a 2 (plausibility) \times 2 (distractor color match) ANOVA for both the category-label and picture-cue conditions. For the category-label-cue condition, there no main effect of plausibility, $F(1, 11) = 0.782$, $p = .396$, $\eta_p^2 = .066$ (.35 vs. .31), and no interaction with match condition, $F(1, 11) = 3.91$, $p = .074$, $\eta_p^2 = .262$, though there was a numerical trend toward a larger match effect for plausible (.32) than for implausible (.18) distractors. Similarly, in the picture-cue condition, there was no main effect of plausibility, $F(1, 19) = 2.54$, $p = .128$, $\eta_p^2 = .118$, (.31 vs. .27), and no interaction, $F(1, 19) = 2.46$, $p = .134$, $\eta_p^2 = .115$.

though, again, there was a numerical trend toward a larger match effect for plausible (.28) compared with implausible (.17) distractors. In sum, there remained a robust effect of distractor color match even when the matching distractor appeared in a location that was relatively implausible for the target object. However, this analysis is limited by the post hoc separation of scene items, resulting in variability in the number of scene items in each plausibility condition for each participant. In Experiment 2, location plausibility was manipulated experimentally.

Measures of semantic guidance. Finally, we examined the gist-based, semantic guidance of attention to target-plausible regions for the scene. A new group of four observers (one an author, Brett Bahle), outlined the regions in each of the 48 critical scene items where the target object for that scene was likely to be located. Identified regions were highly consistent across raters. For example, in a kitchen scene with a mixer target, all raters outlined the two regions of kitchen counter as plausible. For the few inconsistencies, we chose the maximally inclusive set of regions. These were then converted into software ROIs for eye movement analysis. To assess guidance, we compared the elapsed time until the first fixation in one of the target-plausible regions for the no-cue condition, in which there could not have been gist-based guidance, and the category-label-cue condition, in which gaze could have been preferentially allocated to plausible regions for the cued object type. The analysis was limited to mismatch trials so that effects would not be contaminated by guidance from VWM. Moreover, to minimize the potential effect of guidance by a template representation of the target itself (potentially derived from the verbal label), we excluded the target region (conservatively, 25% larger than the target region for the main analysis) from the set of plausible-region ROIs.

The elapsed time to the first fixation on a target-plausible region was reliably shorter in the category-label-cue condition (404 ms) compared with the no-cue condition (712 ms), $F(1, 22) = 22.66$, $p < .001$, $\eta_p^2 = .507$ (Figure 3A). We also examined the mean

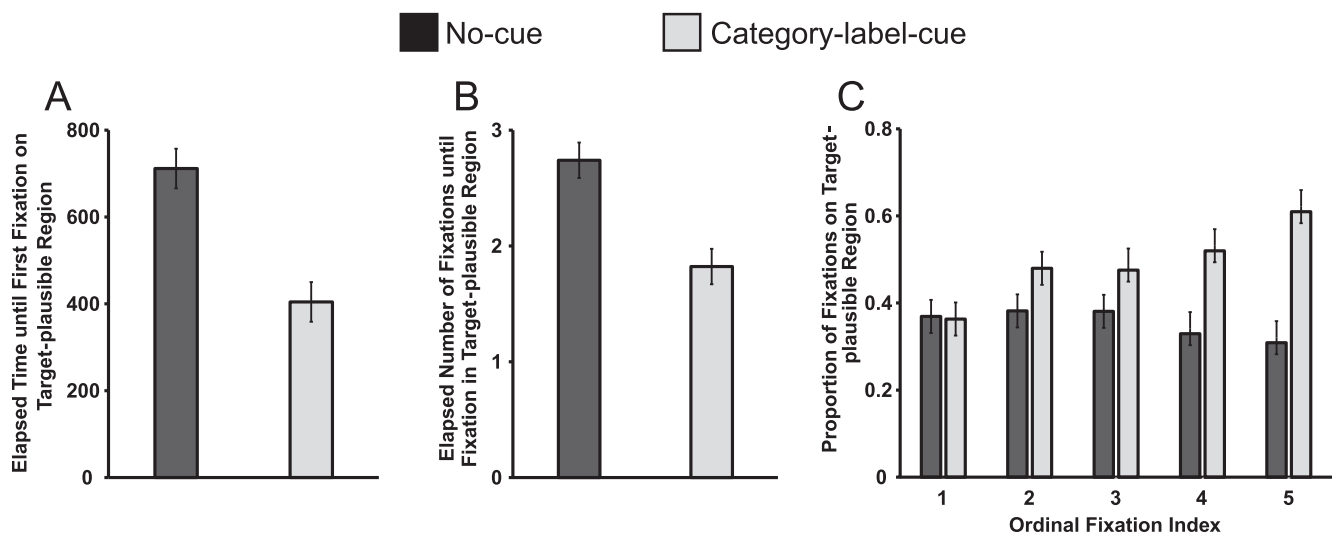


Figure 3. Data figures for the analysis of semantic guidance for Experiment 1. (A) Elapsed time to first fixation on a target-plausible region as a function of cue type. (B) Elapsed number of fixations until first fixation on a target-plausible region as a function of cue type. (C) Proportion of fixations on target-plausible regions as a function of cue type and ordinal fixation index. Error bars are standard errors of the mean.

elapsed number of fixations to the first fixation on a plausible region for the target. Participants required fewer fixations to fixate a plausible region in the category-label-cue condition (1.82) compared with the no-cue condition (2.74), $F(1, 22) = 17.95$, $p < .001$, $\eta_p^2 = .449$ (Figure 3B). Equivalent results were observed for a comparison between the no-cue and picture-cue conditions. Thus, we can be confident that participants used the category information in the label and picture cues to implement gist-based, semantic guidance during search.

Further, we conducted a time-course analysis of semantic guidance similar to the analysis for the time course of the distractor match effect. We calculated the probability of fixating a target-plausible region for each ordinal fixation index during search. As is evident in Figure 3C, there was no effect of semantic guidance for the very first saccade on the scene, and the guidance effect increased gradually over the course of the first five fixations on the scene (note that the no-cue data are plotted only out to Fixation 5, as this was the limit for the category-label-cue data). Consistent with this pattern, there was a reliable interaction between cue condition and ordinal fixation, $F(4, 88) = 4.79$, $p = .002$, $\eta_p^2 = .179$, indicating that the effect of gist-based guidance increased during the course of search. Pairwise comparisons demonstrated that there was no reliable difference in plausible-region fixation probability for Fixation 1, $F(1, 22) = 0.03$, $p = .862$, $\eta_p^2 = .001$. However, this probability differed between the two cue conditions for Fixations 2 to 5 (except for Fixation 3): for Fixation 2, $F(1, 22) = 4.90$, $p = .037$, $\eta_p^2 = .182$; for Fixation 3, $F(1, 22) = 2.35$, $p = .139$, $\eta_p^2 = .097$; for Fixation 4, $F(1, 22) = 10.57$, $p = .004$, $\eta_p^2 = .325$; and for Fixation 5, $F(1, 22) = 11.81$, $p = .002$, $\eta_p^2 = .349$. Thus, the effect of semantic guidance was delayed relative to the effect of template-based guidance, with the latter observed from the very first saccade on the scene (compare Figures 2C and 3C).

Summary

The match between a distractor object and a color held in VWM captured attention from the first saccade on the scene during visual search. Because VWM acts as the primary substrate of feature-based attention templates, this result suggests that template-based guidance can dominate the initial stages of visual search through scenes, at least under the present conditions. In addition, the effect of VWM was independent of the availability of semantic guidance within the scene: The effect was just as large when participants knew the identity of the target object (and could thus use scene gist to limit attention to plausible scene regions) as when they had no information about target identity (providing no opportunity for gist-based guidance). The effect of VWM match also remained when the analysis was limited to scenes with distractors that appeared in an implausible location for the target, suggesting that VWM/template guidance was implemented broadly across the scene and was not limited to target-plausible regions based on an earlier gist-driven parsing. Moreover, analysis of the time course of gist-based guidance indicated that it was delayed relative to the implementation of template-based guidance, the reverse of the prediction generated from the hypothesis that gist-based, semantic guidance dominates the early stages of search (Torralba et al., 2006; Wolfe et al., 2011). This observed delay is consistent with our hypothesis that gist-based guidance may take substantially

longer to configure than estimated from the speed of gist recognition.

Note that the VWM-match results are particularly strong, because the VWM color was not the explicit target of search and, in the category-label-cue and picture-cue conditions, it had to compete with template-based guidance from a representation of the target object. Moreover, in the picture-cue condition, the target template should have been maintained itself in VWM. Thus, the results inform our secondary question about whether VWM guidance is limited to a single item or can span multiple items: The significant capture from a secondary VWM color in the picture-cue condition supports the latter view.

Experiment 2

In Experiment 1, the test of early template guidance depended on capture from a secondary color maintained in VWM, which allowed us to independently manipulate the availability of semantic guidance via the target-cue condition. In Experiment 2, we developed a complementary test to gather converging evidence with Experiment 1. The memory task was eliminated, and the cuing condition was limited to the picture cue. On match trials, the critical distractor object shared the color of the cued target object. On mismatch trials, it did not. This was implemented by manipulating the color of the target object in a particular scene so that it either did or did not match the color of the critical distractor, as illustrated in Figure 4. Additionally, the semantic plausibility of the distractor location (as a location for the target) was experimentally manipulated. If the application of feature-based template guidance is constrained by an initial gist-based parsing of the scene into target-plausible and target-implausible regions, then the effect of distractor color match should be limited to the plausible condition. In contrast, if feature-based template guidance is implemented broadly from the very beginning of search (i.e., before the application of semantic guidance), as suggested by the results from Experiment 1, then the effect of distractor match should be largely independent of location plausibility.

Method

Participants. Twelve (nine female) new participants from the University of Iowa community completed the experiment for course credit. Each was between the ages of 18 and 30 and reported normal or corrected-to-normal vision.

Stimuli. The stimuli were 96 photographs of real-world scenes (48 experimental scenes and 48 filler scenes). Many of the experimental scenes were the same as in Experiment 1. However, some targets from Experiment 1 did not have an appropriate color for a distractor match (e.g., a black stapler, in a scene full of black objects), and these scenes were replaced with new scenes for Experiment 2. Target object size ranged from $1.40^\circ \times 1.56^\circ$ to $10.55^\circ \times 9.11^\circ$ visual angle, with a mean of $3.80^\circ \times 3.34^\circ$, and distractor objects' size ranged from subtending $1.56^\circ \times 1.55^\circ$ to $9.53^\circ \times 9.64^\circ$ of visual angle, with a mean of $4.03^\circ \times 3.54^\circ$. To create the match and mismatch stimuli, for each scene, we used photo-editing software to manipulate the colors of the target and critical distractor (see Figure 4). In the match condition, the two objects had the same color. This was achieved by first taking the

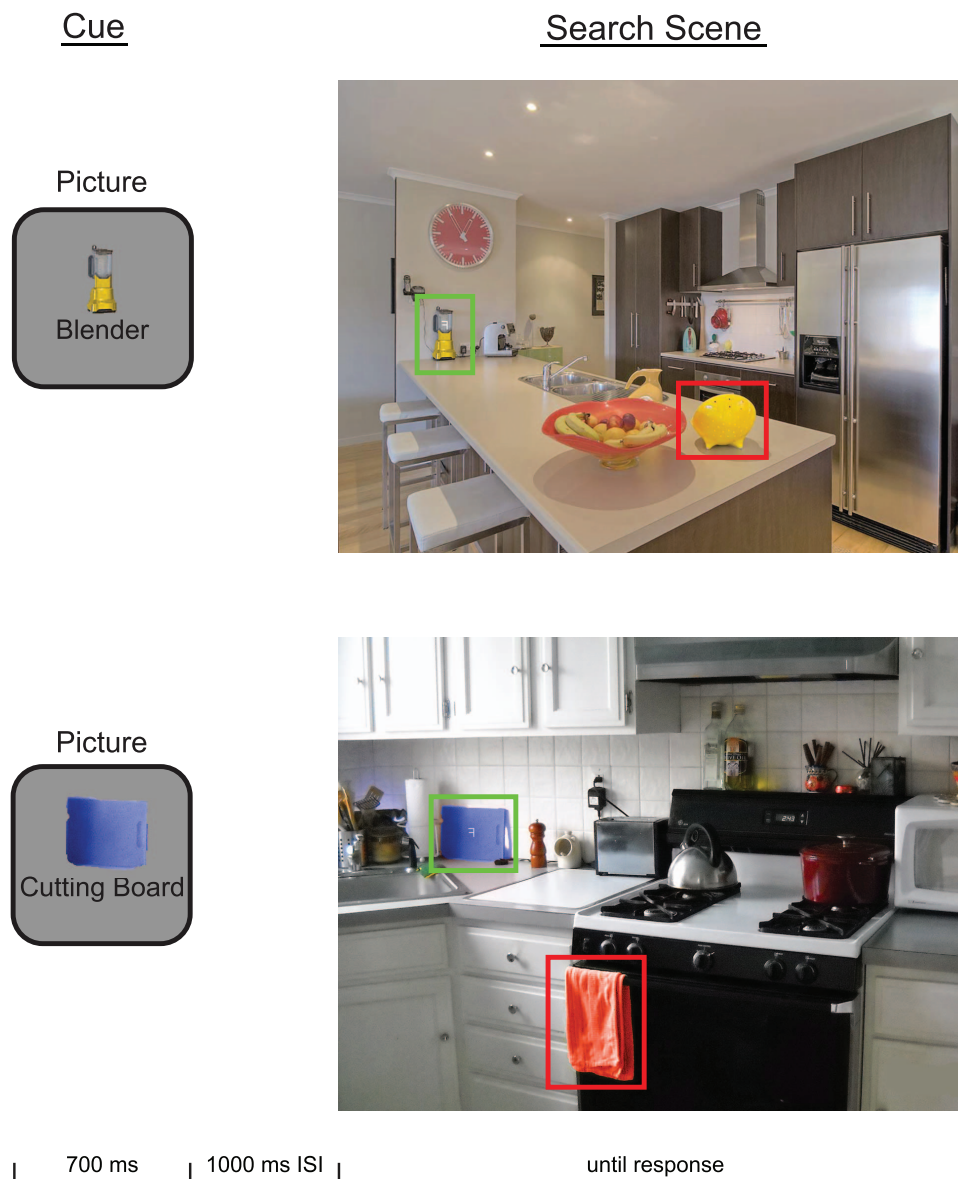


Figure 4. Example stimuli from Experiment 2. The top row depicts a match-plausible scene. The target (blender, outlined in green, left-hand rectangle) matches the color of the critical distractor (outlined in red, right-hand rectangle). Additionally, the critical distractor is in a plausible location for the blender. The bottom row depicts a mismatch-implausible scene. The target (cutting-board, outlined in green, left-hand rectangle) mismatches the color of the critical distractor object (outlined in red, right-hand rectangle). The critical distractor is in an implausible location for the cutting-board. See the online article for the color version of this figure.

average color of the target object, then changing the color of the critical distractor object to match this average color. In the mismatch condition, the target color was different from the critical distractor color. Thus, the color of the critical distractor did not vary across match conditions, controlling for low-level differences that might influence distractor fixation probability. The target picture cue that appeared before search was specific to the color of the target in that scene. For example, in a match trial using the scene depicted in Figure 4, the participant saw a yellow blender as the picture cue and searched for that object in the presence of a

yellow critical distractor. In a mismatch trial using the same scene, the participant saw a blue blender as the picture cue and searched for that object in the presence of the same yellow distractor.

We experimentally manipulated whether the critical distractor appeared in a plausible or implausible location for the target. To do this, we originally selected the 48 scenes, so that 24 had a distractor in a plausible location for the target and 24 had a distractor in an implausible location for the target. We then had a new group of eight (three female) undergraduates from the University of Iowa rate the plausibility of the 48 critical distractor locations for the 48

experimental scenes, using the same method as described in Experiment 1. The interclass correlation showed the across-subjects ratings to be reliably similar, $R = .839$, $F(47, 329) = 7.33$, $p < .001$. We performed a median split, assigning the scenes with the lower half of scores to the implausible location condition and scenes with the higher half of scores to the plausible location condition. The final assignment of items to plausibility conditions was largely consistent with the choices made by the experimenters: 42 of the 48 scenes fell into the plausibility category for which they were originally selected. The analyses, reported below, were equivalent when including all 48 scene items (reported) and when limited to the set of 42 for which the independent ratings agreed with the initial experimenter selection.

Procedure. The procedure was the same as in Experiment 1's picture-cue condition, except there was no secondary color to remember and, thus, no postsearch color memory test. As in Experiment 1, each participant saw each scene item once. The assignment of experimental scene items to the match and plausibility conditions was counterbalanced across sets of four participants, with each participant shown 12 items in each of the four conditions (match-plausible, match-implausible, mismatch-plausible, mismatch-implausible).

Results and Discussion

Trials were removed based on the same criteria as in Experiment 1 (15.1%). The pattern of results was not affected by trial removal.

Search accuracy. Search accuracy was overall very high (99%; see Table 2). There was no difference in accuracy between match (99.3%) and mismatch (98.6%) trials, $F(1, 11) = 0.646$, $p = .438$, $\eta_p^2 = .056$.

Search time. The time to first fixation of the target data (Figure 5A) were entered in a 2 (distractor color match condition: match, mismatch) \times 2 (distractor location plausibility: plausible, implausible) ANOVA. There was a main effect of distractor color match, with longer mean search time on match (430 ms) compared with mismatch (378 ms) trials, $F(1, 11) = 4.91$, $p = .049$, $\eta_p^2 = .309$. There was also a main effect of plausibility: Targets were found more quickly when the critical distractor was in implausible location for the target (370 ms vs. 438 ms), $F(1, 11) = 21.18$, $p = .001$, $\eta_p^2 = .658$. Moreover, there was a numerical trend toward an interaction between these two factors, $F(1, 11) = 2.82$, $p = .121$, $\eta_p^2 = .204$, driven by the fact that the distractor color-match effect was larger for implausible distractor scenes than for the plausible distractor scenes. The most likely explanation for this interaction (and for the main effect difference between plausibility conditions) derives from idiosyncratic differences in the scene items used for the plausible and implausible conditions. For the plausible condition, the scene items producing the highest rates of distractor

fixation also tended to have a distractor that was relatively close to the target location (the 12 scene items with the highest distractor fixation rate had a mean distance of 10.2° between target and distractor), whereas in the implausible condition, the scene items producing the highest rates of distractor fixation tended to have a distractor that was far from the target location (the 12 scene items with the highest distractor fixation rate had a mean distance of 16.4° between target and distractor). Thus, capture was more disruptive in the implausible condition. Note that this is unlikely to have been caused by semantic guidance of attention to the target region in the plausible condition, as the overall rates of distractor fixation did not differ between plausible and implausible distractor scenes, as reported next.

Measures of distractor fixation. The data for probability of fixating the critical distractor (Figure 5B) were entered into a 2 (distractor color match condition) \times 2 (distractor location plausibility) ANOVA. There was a main effect of distractor color match, with a higher probability of distractor fixation on match (.30) compared with mismatch (.04) trials, $F(1, 11) = 35.66$, $p < .001$, $\eta_p^2 = .764$. There was no main effect of plausibility, suggesting that distractors were fixated as often whether they were in plausible (.20) or implausible (.15) locations, $F(1, 11) = 1.64$, $p = .226$, $\eta_p^2 = .130$. Critically, there was no interaction between these two factors, $F(1, 11) = 0.020$, $p = .890$, $\eta_p^2 = .002$, indicating that the effect of distractor color match was not modulated by distractor location plausibility. Planned contrasts in each of the plausibility conditions indicated a reliable distractor match effect when the critical distractor appeared both in a plausible location for the target, $F(1, 11) = 11.33$, $p = .006$, $\eta_p^2 = .507$, and in an implausible location for the target, $F(1, 11) = 22.51$, $p < .001$, $\eta_p^2 = .672$. Comparison of the absolute size of the capture effect across plausibility conditions is limited by the fact that, by necessity, the two conditions used different sets of scene items. However, the critical point is that the distractor match effect was observed robustly in the implausible condition, indicating that template guidance was not limited to target-plausible regions of the scene.

As in Experiment 1, we again considered the time course of the distractor color-match effect by computing the probability of fixating the distractor at each ordinal fixation index as a function of color match (Figure 5C). A participant's data were included only if the participant contributed at least eight trials to each condition, and a value is plotted only if nine of the 12 participants' data were available. All 12 participants were included for all three fixation bins for the plausible condition; for the implausible condition, all 12 participants were included for Fixation Bins 1 and 2, and 10 were included in Fixation Bin 3. Because search was quite efficient in Experiment 2, data points only extend out to the third fixation. As in Experiment 1, the distractor color-match effect was evident from the beginning of search in both the plausible and implausible conditions, suggesting that template-based guidance was deployed broadly across the scene, rather than being limited to target-plausible regions.

We also report the Euclidean distance analysis for Experiment 2, using the same method as in Experiment 1 (Figure 5D). Note that for Fixation 0 (the first fixation on the scene, controlled by the experimenter), gaze tended to be slightly closer to the target than to the critical distractor, on average, reflecting the fact that in this modified set of scenes, targets were slightly closer to the scene center than were distractors. The first participant-directed fixations

Table 2
Search Accuracy From Experiment 2

Condition	Search accuracy	
	Match	Mismatch
Plausible	100.0%	99.3%
Implausible	98.6%	97.9%

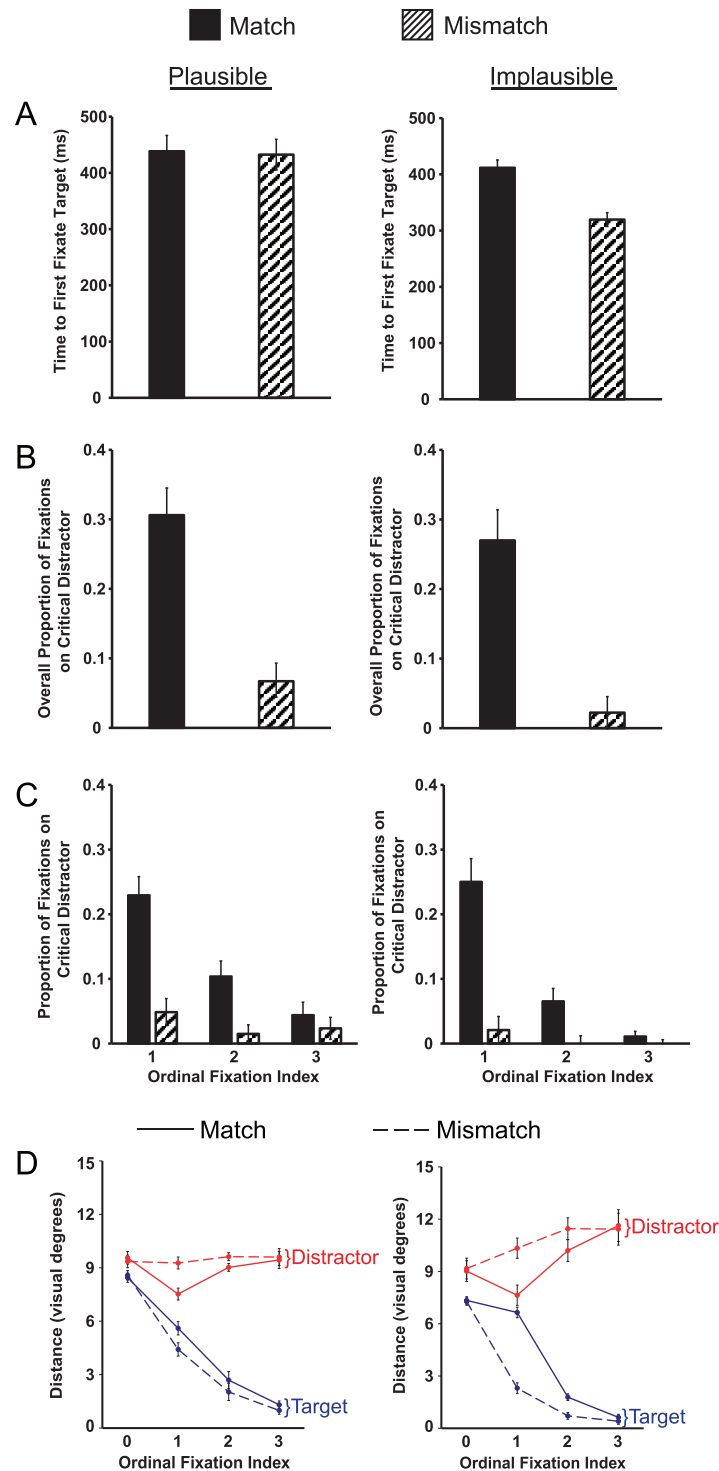


Figure 5. Data figures for Experiment 2. From left to right, data are plotted for the *plausible* and *implausible* conditions. (A) Time (ms) to first fixate the target as a function of distractor color match. (B) Overall probability of fixating the critical distractor as a function of color match. (C) Proportions of fixations landing on the critical distractor object as a function of ordinal fixation index and color match. (D) The distance, in visual degrees, of the current fixation to the target and critical distractor object as a function of ordinal fixation index. Error bars indicate within-subjects 95% confidence intervals (Morey, 2008). See the online article for the color version of this figure.

on match trials was biased toward the distractor compared with mismatch trials. This effect was observed for both plausible and implausible distractors.

In sum, there was a substantial effect of template color match on the probability of distractor fixation for the first two saccades on the scene. This pattern was observed both for distractors that appeared in locations where the target would typically appear and, critically, for distractors that appeared in locations where the target would not typically appear. Thus, feature-based template guidance was implemented from the beginning of search and was not restricted to target-plausible locations.

Experiment 3

In Experiments 1 and 2, we demonstrated that a feature-based template representation can guide selection early during search in scenes. However, one might argue that template-based guidance and gist-based, semantic guidance were not given an equal opportunity to influence the early stages of search. That is, participants were shown the target cue before the presentation of the scene, allowing them to establish a template representation before scene onset. In contrast, semantic guidance could not have been configured before the scene onset (except to the limited extent that the identity of the target object allowed a participant to predict the category of the scene). Thus, the early distractor color-match effect may have arisen only because semantic guidance was not available from the beginning of search.

In Experiment 3, we used the basic method of Experiment 2, but on half of the trials, we added a scene preview before search to familiarize participants with the identity and general layout of the scene. The preview was a low-pass filtered grayscale image of the scene, so that gist could be extracted but there would be limited opportunity to extract information about the specific objects in the scene (such as the target and its location). Previous experiments have demonstrated that gist can be extracted under these viewing conditions for the particular type of scenes chosen (Castelano & Henderson, 2008; Oliva & Schyns, 2000).

Note that this method is a particularly strong test of early feature-based template guidance. The scene preview allowed participants to establish a representation of scene gist and layout and combine this with the cue indicating the target of search, so as to localize plausible regions of the scene. It is therefore possible that participants could have directed attention covertly to a plausible screen location even before the onset of the search scene. In addition, although the filtering manipulation eliminated high-frequency information above 2 cycles/degree, it did not eliminate all of the information from local objects, so it is also possible that sufficient target information could have been extracted from the preview to allow participants to localize the target before search commenced, at least on some trials. Thus, we anticipated that the availability of a scene preview would reduce the effect of distractor color match. However, if the distractor color-match effect were still observed, even with reduced magnitude, then such a result would indicate that feature-based template information can guide early search despite strong spatial constraints that were available before the onset of search.

Method

Participants. Twenty-four (16 female) new participants from the University of Iowa community completed the experiment for course credit. We tripled the N suggested by the power analysis of the pilot experiment (eight participants), with the expectation that the preview might diminish the effect of distractor match. Each participant was between the ages of 18 and 30 and reported normal or corrected-to-normal vision.

Stimuli. The stimuli were the same as used in Experiment 2, with the exception of the scene preview images. The scene previews were generated by first creating a monochrome version of each scene. This was accomplished by converting each scene from RGB to L^*a^*b color space, then removing the a^*b components. Each scene was then low-pass filtered at 2 cycles/degree of visual angle, which maintained most mid- and low-level frequency information. To ensure that participants could indeed extract the gist of each scene from the filtered preview, we performed a control experiment. Twelve new subjects (eight female) from the University of Iowa community saw the filtered version of each of the 48 experimental scenes for 250 ms, followed by a 50-ms mask and a two-alternative forced choice test consisting of two scene category labels. For example, after viewing the filtered version of a kitchen scene, the participant chose between the labels "kitchen" and "dining room." Foil labels were chosen so that indoor scenes always had indoor foils and outdoor scenes outdoor foils. Mean discrimination accuracy was 90.7%, indicating that the previews contained sufficient information to reliably extract the gist of the scene.

The mask that appeared after the scene preview was a random noise mask: Each pixel in the mask had a randomly determined luminance value, with the a^*b components in L^*a^*b color space equal to zero.

Procedure. The procedure was the similar to that in Experiment 2. For half of the scene items (both experimental scenes and filler scenes), a preview scene was displayed before the target cue (see Figure 6). The preview appeared for 250 ms, followed by a 50-ms mask. The other half of trials had no preview, and the sequence of events was the same as in Experiment 2. The assignment of scene items to preview and no-preview conditions was counterbalanced in the same manner as in Experiment 1. That is, for the first participant in a pair, the assignment of scene items to preview conditions was determined randomly. For the second, the assignment was reversed. There was no experimental manipulation of distractor location plausibility.

Results and Discussion

Trials were removed based on the same criteria as in Experiment 1 (14.4%). The pattern of results was not affected by trial removal.

Search accuracy. As in previous experiments, search accuracy was very high (98%, see Table 3). Accuracy data were entered in a 2 (distractor color match condition: match, mismatch) \times 2 (preview condition: preview, no-preview) ANOVA. There was no main effect of distractor color match, $F(1, 23) = 0.264$, $p = .612$, $\eta_p^2 = .011$, no main effect of preview, $F(1, 23) = 0.057$, $p = .814$, $\eta_p^2 = .002$, and no interaction, $F(1, 23) = 0.046$, $p = .833$, $\eta_p^2 = .002$.

Search time. The time until the first fixation on the target data (Figure 7A) was entered in a 2 (distractor color match condi-

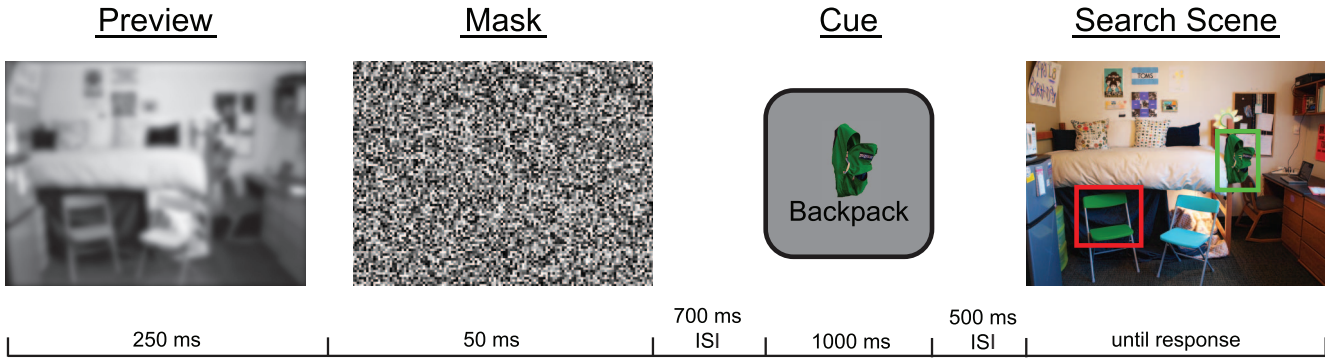


Figure 6. Trial event sequence for Experiment 3. Each trial began with a 250-ms low-pass filtered preview of the upcoming search scene, allowing participants to extract the scene gist. The preview was followed by a 50-ms mask. After a 700-ms ISI, the picture cue of the search target was displayed for 1,000 ms followed by a 500-ms ISI. Finally, the search scene appeared. On a match trial, there was a color match between the target object (backpack, outlined in green) and the critical distractor object (chair, outlined in red). ISI = Interstimulus interval. See the online article for the color version of this figure.

tion) \times 2 (preview condition) ANOVA. There was no main effect of distractor color match, $F(1, 23) = 2.70, p = .114, \eta_p^2 = .105$, but a nonsignificant trend toward a main effect of preview, $F(1, 23) = 3.13, p = .090, \eta_p^2 = .120$, with numerically faster search in the preview condition ($M = 383$ ms) than in the no-preview condition ($M = 405$ ms).⁴ There was no reliable interaction between these two conditions, $F(1, 23) = .214, p = .648, \eta_p^2 = .009$. Thus, the search time data provide some evidence that the scene preview aided search above the type of gist-based guidance available after the search scene appeared.

Measures of distractor fixation. The data for probability of fixating the distractor (Figure 7B) were entered into a 2 (distractor color match condition) \times 2 (preview condition) ANOVA. There was a main effect of distractor color match, with a higher probability of distractor fixation on match (.24) compared with mismatch (.05) trials, $F(1, 23) = 102.49, p < .001, \eta_p^2 = .817$. There was a main effect of preview, suggesting a higher probability of distractor fixation on no-preview (.17) compared with preview (.12) trials, $F(1, 23) = 4.98, p = .036, \eta_p^2 = .178$. There was a trend toward a reliable interaction between these conditions, $F(1, 23) = 3.56, p = .072, \eta_p^2 = .134$, suggesting the distractor color-match effect may have been smaller on preview trials compared with no-preview trials. Critically, however, the distractor color-match effect was observed robustly for both preview and no-preview trials: for preview, $F(1, 23) = 26.52, p < .001, \eta_p^2 = .536$, with a match effect of .14, and for no-preview, $F(1, 23) = 57.61, p < .001, \eta_p^2 = .715$, with a match effect of .23. Thus, although the preview manipulation may have reduced the match effect, it did not eliminate it.

As with both previous experiments, we investigated the time course of the distractor color-match effect by plotting the probability of distractor fixation at each ordinal fixation index as a function of color match for both the preview and no-preview conditions (Figure 7C). All 24 participant's data are included in each ordinal fixation bin for both preview and no-preview trials. The effect of distractor color match was evident for the earliest fixations on the scene in both preview conditions.

The data for the distance of gaze from both the target and distractor are depicted in Figure 7D. As in previous experiments, the first participant-directed fixations on match trials, compared with mismatch trials, tended to reduce the distance to the distractor. This effect was true for both preview and no-preview trials, though it was diminished in the former condition.

In sum, although participants were shown a preview of the scene before they searched, giving them ample opportunity to both extract the gist and layout of the scene and prepare for the application of gist-based information to the search process, their eyes were still directed selectively to distractor objects that matched the color of the target template. As was found in all three experiments, this distractor color-match effect was evident from the first saccade on the scene.

General Discussion

In the present experiments, we investigated visual search through real-world scenes, comparing the contributions of gist-based, semantic guidance (from knowledge of where a target is likely to be found in a scene) and template-based guidance (from a representation of the visual form of the target). A critical dis-

Table 3
Search Accuracy From Experiment 3

Condition	Search accuracy	
	Match	Mismatch
Preview	97.6%	97.6%
No-preview	97.9%	97.9%

⁴ We also attempted to examine the elapsed time until fixation on a plausible region for the target, as in Experiment 1, which could indicate that participants used the preview to help guide attention to plausible scene regions. However, because searches were so efficient with a picture cue, there was a large number of trials in which participants directed gaze almost immediately to the target. That is, there were no fixations on target-plausible regions that were not also fixations on the target object, leaving too little data to support analysis (e.g., three subjects had no available data in the preview condition).

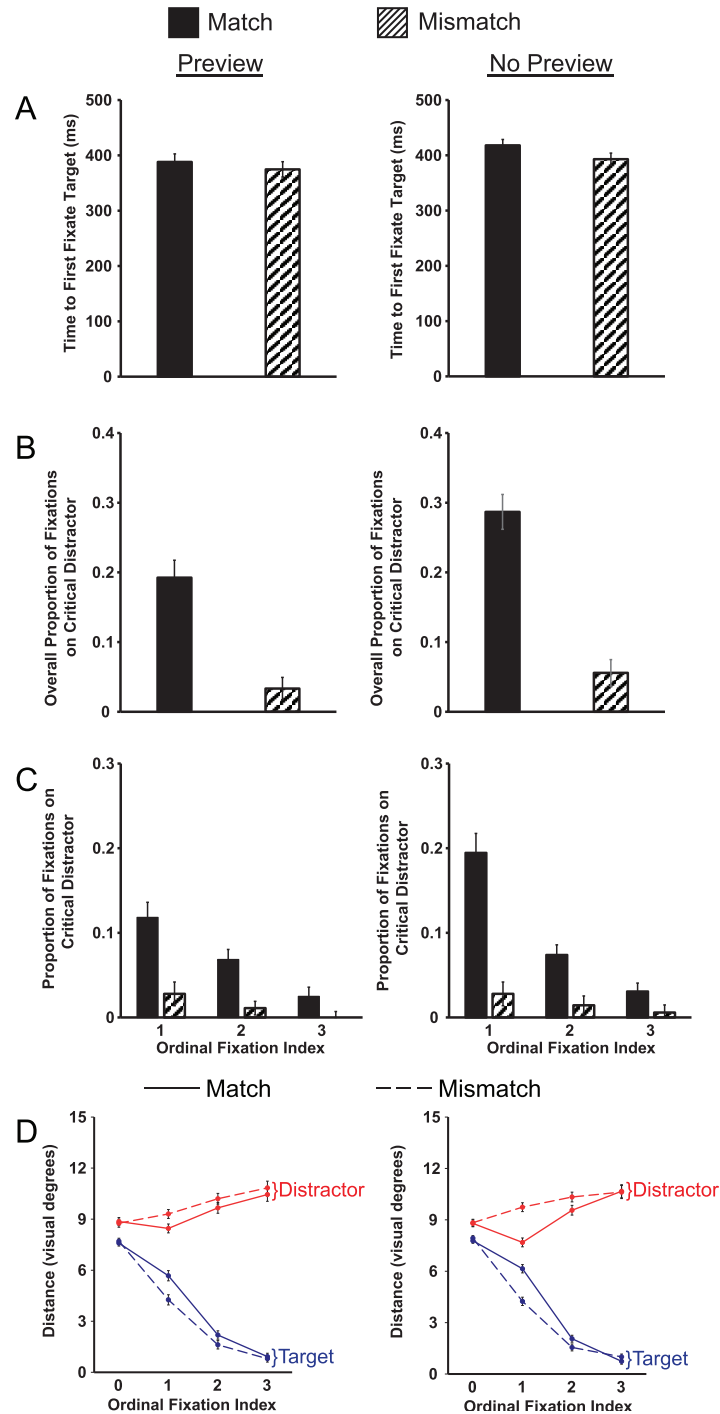


Figure 7. Data figures for Experiment 3. From left to right, data are plotted for the *preview* and *no-preview* conditions. (A) Time (ms) to first fixate the target as a function of distractor color match. (B) Overall probability of fixating the critical distractor as a function of color match. (C) Proportions of fixations landing on the critical distractor object as a function of ordinal fixation index and color match. (D) The distance, in visual degrees, of the current fixation to the target and critical distractor object. Error bars indicate within-subjects 95% confidence intervals (Morey, 2008). See the online article for the color version of this figure.

tractor in the scene either matched or mismatched a color held in VWM (Experiment 1) or the color of the target object (Experiments 2 and 3). Significant oculomotor capture by the matching color distractor was observed in all experiments, an index of template-based guidance. There were six further findings critical for understanding the relationship between template-based and gist-based guidance. First, the distractor color-match effect was observed from the very first saccade on the scene, suggesting strong initial guidance by a template representation. Second, the implementation of gist-based, semantic guidance (Experiment 1) was actually delayed relative to the implementation of template guidance. Third, the magnitude of the color-match effect was independent of the availability of semantic guidance from knowledge of target identity. Fourth, the color-match effect was independent of the plausibility of the distractor location as a location for the target object. Fifth, the color-match effect, though reduced, remained even when participants were given a preview of their search scene and could potentially establish a scene-level contextual representation before search commenced. Finally, early template-based guidance was observed across search tasks that ranged from relatively inefficient (Experiment 1, no-cue condition: mean time to target fixation = ~1600 ms) to relatively efficient (Experiments 2 and 3: mean time to target fixation <500 ms). Together, these data indicate that template guidance plays a key role in search through natural scenes, a role that is not necessarily limited to semantically plausible regions of the scene after an initial, gist-based parsing.

The results inform current theories of visual search through natural scenes. Prominent accounts have come to stress the role of gist-based, semantic guidance in modulating the application of a feature-based template representation (Torralba et al., 2006; Wolfe et al., 2011). Although semantic guidance clearly plays a key role in search for common objects, the present results indicate that the relationship between semantic and template guidance should not be considered as strictly ordered (gist guidance before template guidance) or strictly hierarchical (template application limited to semantically plausible scene regions). In particular, the present results indicate that gist and template information are not necessarily integrated prior to the onset of search, constraining the spatial application of the template (Torralba et al., 2006). The very early stages of search in the present experiments were dominated by broad, feature-based guidance from the template.

As discussed in the introduction, the application of gist-based information to search requires more than simple gist recognition. Although recognition can be achieved from an extremely brief exposure, gist-based guidance of search requires retrieval of stored information about typical object locations and the parsing of the scene into plausible regions based on the identity of the target. Such demands may have been minimized in experiments that have found dominance by semantic guidance (e.g., Torralba et al., 2006). Specifically, in the Torralba et al. (2006) experiments, the task involved searching, on each trial, for the same object type (e.g., people) in a series of semantically similar scene scenes (e.g., street scenes). Such a design may have minimized the demands for configuring gist-based guidance, as knowledge about the plausible locations for the target type could be maintained consistently across trials. The present method used a different target object and a different scene on each trial, requiring the participant to retrieve and configure new semantic information for every search. These

conditions are characteristic of typical real-world search behavior (only in relatively rare, repetitive tasks do the target and scene remain constant), and it appears that the feature-based representation of the target object can exert considerable influence before gist-based guidance becomes operational later in the search process. Thus, template-based approaches to search guidance (e.g., Wolfe, 1994; Zelinsky, 2008) may have greater explanatory relevance to real-world search behavior than has been claimed recently (Wolfe et al., 2011).

Converging evidence is provided by another recent study that attempted to dissociate feature-based and semantic guidance (Spotorno, Malcolm, & Tatler, 2014). These authors manipulated the specificity of the target template (picture vs. word) and the plausibility of the target location (plausible vs. implausible) orthogonally. Although there have been similar manipulations (e.g., Malcolm & Henderson, 2010), Spotorno et al.'s (2014) study used clearly defined plausible and implausible regions so as to examine semantic guidance independently of template guidance. Specifically, their scenes were divided by the horizon. Half of the targets were objects that tend to be found on the ground, and the other half were objects that tend to be found in the sky. The authors interpreted their results (additive effects of template specificity and position plausibility on overall search time) as evidence that the target's appearance and its likely location were integrated before search commenced. However, close examination of their results revealed that early template guidance was not strongly modulated by position plausibility. Upon the presentation of a picture cue, there was no difference in the probability that the first saccade was directed to the target when the target appeared in a plausible location and when it appeared in an implausible location. Thus, we consider Spotorno et al.'s results as converging evidence that early template guidance need not be strongly constrained by gist-based, semantic guidance.

The effect of VWM load (Experiment 1) and the effect of color match to a feature of the target template (Experiments 2 and 3) were very similar, showing the same time course of preferential fixation for matching distractors. The results are therefore consistent with the orthodox position that the active maintenance of features in VWM operates as an attentional template (Desimone & Duncan, 1995; Woodman et al., 2013). Moreover, the results show that VWM-based attention capture (Olivers, 2009; Soto et al., 2005), which has been studied using simple search arrays with geometric shapes and colors without spatial structure, generalizes to search for objects in real-world scenes. The finding indicates that theories developed to explain search under highly controlled settings can explain more naturalistic behavior, potentially opening up research in several domains in which VWM-based distraction could have important consequences (e.g., driving) and/or in which disordered attentional biases (e.g., spider phobia, posttraumatic stress disorder) may be grounded in preservative focus in VWM on particular object types.

Although the time course of the color-match effect was similar between Experiments 1 and 2, the effect was numerically slightly larger when the matching feature was a property of the search target in Experiment 2 than when it was the property of a secondary memory item in the picture cue condition of Experiment 1. However, this is to be expected given that Experiment 1 placed greater demands on VWM, requiring participants to maintain two items (the target image and the secondary memory color) rather

than just one (the target image) in Experiment 2. The fact that there was color-based capture *at all* in the picture-cue condition of Experiment 1 has major implications for theories of the architecture of interaction between VWM and attention. One prominent theory holds that only a single item in VWM can interact with attention to guide selection (Olivers et al., 2011). Any other items in VWM are maintained in an *accessory* state that is inert with respect to attentional guidance. In this view, a secondary memory color should not interact with selection if the template slot is occupied by the search target, as should have been the case in the picture-cue condition of Experiment 1, particularly as the search target changed on every trial (Woodman et al., 2013). Thus, capture by the secondary memory color is inconsistent with the single-item-template hypothesis. The results support, instead, the claim that multiple items in VWM can guide attention simultaneously (Beck & Hollingworth, 2017; Beck et al., 2012; Hollingworth & Beck, 2016). There are now several lines of evidence supporting this multiple-item-template hypothesis, from the capture of attention by multiple items in VWM (Experiment 1; Hollingworth & Beck, 2016) to the efficient guidance of attention to items matching either of two target values (Beck et al., 2012; Roper & Vecera, 2012), to increased competition for selection when both items in a two-item display match a feature in VWM (Beck & Hollingworth, 2017). Although the present results do not necessarily contradict the broad distinction between items in VWM that are maintained in an active versus accessory state (Hollingworth & Hwang, 2013; van Moorselaar, Theeuwes, & Olivers, 2014), they suggest that more than one item can be maintained in a state that interacts with selection.

References

- Beck, V. M., & Hollingworth, A. (2017). Competition in saccade target selection reveals attentional guidance by simultaneously active working memory representations. *Journal of Experimental Psychology: Human Perception and Performance*, 43, 225–230. <http://dx.doi.org/10.1037/xhp0000306>
- Beck, V. M., Hollingworth, A., & Luck, S. J. (2012). Simultaneous control of attention by multiple working memory representations. *Psychological Science*, 23, 887–898. <http://dx.doi.org/10.1177/0956797612439068>
- Brockmole, J. R., & Henderson, J. M. (2006a). Recognition and attention guidance during contextual cuing in real-world scenes: Evidence from eye movements. *The Quarterly Journal of Experimental Psychology*, 59, 1177–1187. <http://dx.doi.org/10.1080/17470210600665996>
- Brockmole, J. R., & Henderson, J. M. (2006b). Using real-world scenes as contextual cues for search. *Visual Cognition*, 13, 99–108. <http://dx.doi.org/10.1080/13506280500165188>
- Brooks, D. I., Rasmussen, I. P., & Hollingworth, A. (2010). The nesting of search contexts within natural scenes: Evidence from contextual cuing. *Journal of Experimental Psychology: Human Perception and Performance*, 36, 1406–1418. <http://dx.doi.org/10.1037/a0019257>
- Castelhano, M. S., & Heaven, C. (2011). Scene context influences without scene gist: Eye movements guided by spatial associations in visual search. *Psychonomic Bulletin & Review*, 18, 890–896. <http://dx.doi.org/10.3758/s13423-011-0107-8>
- 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. <http://dx.doi.org/10.1037/0096-1523.33.4.753>
- Castelhano, M. S., & Henderson, J. M. (2008). The influence of color on the perception of scene gist. *Journal of Experimental Psychology: Human Perception and Performance*, 34, 660–675. <http://dx.doi.org/10.1037/0096-1523.34.3.660>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, 18, 193–222. <http://dx.doi.org/10.1146/annurev.ne.18.030195.001205>
- Downing, P. E., & Dodds, C. M. (2004). Competition in visual working memory for control of search. *Visual Cognition*, 11, 689–703. <http://dx.doi.org/10.1080/13506280344000446>
- Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, 96, 433–458. <http://dx.doi.org/10.1037/0033-295X.96.3.433>
- 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. <http://dx.doi.org/10.1111/j.1467-9280.2006.01815.x>
- 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(6–7), 945–978. <http://dx.doi.org/10.1080/13506280902834720>
- Emrich, S. M., Riggall, A. C., Larocque, J. J., & Postle, B. R. (2013). Distributed patterns of activity in sensory cortex reflect the precision of multiple items maintained in visual short-term memory. *The Journal of Neuroscience*, 33, 6516–6523. <http://dx.doi.org/10.1523/JNEUROSCI.5732-12.2013>
- Greene, M. R., & Oliva, A. (2009). Recognition of natural scenes from global properties: Seeing the forest without representing the trees. *Cognitive Psychology*, 58, 137–176. <http://dx.doi.org/10.1016/j.cogpsych.2008.06.001>
- Henderson, J. M., Malcolm, G. L., & Schandl, C. (2009). Searching in the dark: Cognitive relevance drives attention in real-world scenes. *Psychonomic Bulletin & Review*, 16, 850–856. <http://dx.doi.org/10.3758/PBR.16.5.850>
- Henderson, J. M., Weeks, P. A., & Hollingworth, A. (1999). The effects of semantic consistency on eye movements during complex scene viewing. *Journal of Experimental Psychology: Human Perception and Performance*, 25, 210–228. <http://dx.doi.org/10.1037/0096-1523.25.1.210>
- Hollingworth, A. (2012). Task specificity and the influence of memory on visual search: Comment on Vö and Wolfe (2012). *Journal of Experimental Psychology: Human Perception and Performance*, 38, 1596–1603. <http://dx.doi.org/10.1037/a0030237>
- Hollingworth, A., & Beck, V. M. (2016). Memory-based attention capture when multiple items are maintained in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, 42, 911–917. <http://dx.doi.org/10.1037/xhp0000230>
- Hollingworth, A., & Hwang, S. (2013). The relationship between visual working memory and attention: Retention of precise colour information in the absence of effects on perceptual selection. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, 368, 20130061. <http://dx.doi.org/10.1098/rstb.2013.0061>
- Hollingworth, A., & Luck, S. J. (2009). The role of visual working memory (VWM) in the control of gaze during visual search. *Attention, Perception, & Psychophysics*, 71, 936–949. <http://dx.doi.org/10.3758/APP.71.4.936>
- Hollingworth, A., Matsukura, M., & Luck, S. J. (2013). Visual working memory modulates rapid eye movements to simple onset targets. *Psychological Science*, 24, 790–796. <http://dx.doi.org/10.1177/0956797612459767>
- Houtkamp, R., & Roelfsema, P. R. (2006). The effect of items in working memory on the deployment of attention and the eyes during visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 32, 423–442. <http://dx.doi.org/10.1037/0096-1523.32.4.423>

- Hwang, A. D., Higgins, E. C., & Pomplun, M. (2009). A model of top-down attentional control during visual search in complex scenes. *Journal of Vision*, 9, 25. <http://dx.doi.org/10.1167/9.5.25>
- Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience*, 2, 194–203. <http://dx.doi.org/10.1038/35058500>
- Koehler, K., & Eckstein, M. P. (2017). Beyond scene gist: Objects guide search more than scene background. *Journal of Experimental Psychology: Human Perception and Performance*, 43, 1177–1193. <http://dx.doi.org/10.1037/xhp0000363>
- Land, M. F., & Hayhoe, M. (2001). In what ways do eye movements contribute to everyday activities? *Vision Research*, 41, 3559–3565. [http://dx.doi.org/10.1016/S0042-6989\(01\)00102-X](http://dx.doi.org/10.1016/S0042-6989(01)00102-X)
- Maguire, J. F., & Howe, P. D. (2016). Failure to detect meaning in RSVP at 27 ms per picture. *Attention, Perception, & Psychophysics*, 78, 1405–1413. <http://dx.doi.org/10.3758/s13414-016-1096-5>
- 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, 8. <http://dx.doi.org/10.1167/9.11.8>
- Malcolm, G. L., & Henderson, J. M. (2010). Combining top-down processes to guide eye movements during real-world scene search. *Journal of Vision*, 10, 4. <http://dx.doi.org/10.1167/10.2.4>
- Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorials in Quantitative Methods for Psychology*, 4, 61–64. <http://dx.doi.org/10.20982/tqmp.04.2.p061>
- Navalpakkam, V., & Itti, L. (2005). Modeling the influence of task on attention. *Vision Research*, 45, 205–231. <http://dx.doi.org/10.1016/j.visres.2004.07.042>
- Neider, M. B., & Zelinsky, G. J. (2008). Exploring set size effects in scenes: Identifying the objects of search. *Visual Cognition*, 16, 1–10. <http://dx.doi.org/10.1080/13506280701381691>
- Oliva, A., & Schyns, P. G. (2000). Diagnostic colors mediate scene recognition. *Cognitive Psychology*, 41, 176–210. <http://dx.doi.org/10.1006/cogp.1999.0728>
- Olivers, C. N. L. (2009). What drives memory-driven attentional capture? The effects of memory type, display type, and search type. *Journal of Experimental Psychology: Human Perception and Performance*, 35, 1275–1291. <http://dx.doi.org/10.1037/a0013896>
- Olivers, C. N. L., Meijer, F., & Theeuwes, J. (2006). Feature-based memory-driven attentional capture: Visual working memory content affects visual attention. *Journal of Experimental Psychology: Human Perception and Performance*, 32, 1243–1265. <http://dx.doi.org/10.1037/0096-1523.32.5.1243>
- Olivers, C. N. L., Peters, J., Houtkamp, R., & Roelfsema, P. R. (2011). Different states in visual working memory: When it guides attention and when it does not. *Trends in Cognitive Sciences*, 15, 327–334. <http://dx.doi.org/10.1016/j.tics.2011.05.004>
- Potter, M. C., & Levy, E. I. (1969). Recognition memory for a rapid sequence of pictures. *Journal of Experimental Psychology*, 81, 10–15. <http://dx.doi.org/10.1037/h0027470>
- Potter, M. C., Wyble, B., Hagmann, C. E., & McCourt, E. S. (2014). Detecting meaning in RSVP at 13 ms per picture. *Attention, Perception, & Psychophysics*, 76, 270–279. <http://dx.doi.org/10.3758/s13414-013-0605-z>
- Roper, Z. J. J., & Vecera, S. P. (2012). Searching for two things at once: Establishment of multiple attentional control settings on a trial-by-trial basis. *Psychonomic Bulletin & Review*, 19, 1114–1121. <http://dx.doi.org/10.3758/s13423-012-0297-8>
- Rousselet, G. A., Joubert, O. R., & Fabre-Thorpe, M. (2005). How long to get to the “gist” of real-world natural scenes? *Visual Cognition*, 12, 852–877. <http://dx.doi.org/10.1080/13506280444000553>
- Sanocki, T. (2003). Representation and perception of scenic layout. *Cognitive Psychology*, 47, 43–86. [http://dx.doi.org/10.1016/S0010-0285\(03\)00002-1](http://dx.doi.org/10.1016/S0010-0285(03)00002-1)
- Schneegans, S., Spencer, J. P., Schöner, G., Hwang, S., & Hollingworth, A. (2014). Dynamic interactions between visual working memory and saccade target selection. *Journal of Vision*, 14, 9. <http://dx.doi.org/10.1167/14.11.9>
- Soto, D., Heinke, D., Humphreys, G. W., & Blanco, M. J. (2005). Early, involuntary top-down guidance of attention from working memory. *Journal of Experimental Psychology: Human Perception and Performance*, 31, 248–261. <http://dx.doi.org/10.1037/0096-1523.31.2.248>
- Spotorno, S., Malcolm, G. L., & Tatler, B. W. (2014). How context information and target information guide the eyes from the first epoch of search in real-world scenes. *Journal of Vision*, 14, 7. <http://dx.doi.org/10.1167/14.2.7>
- 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. <http://dx.doi.org/10.1037/0033-295X.113.4.766>
- van Moorselaar, D., Theeuwes, J., & Olivers, C. N. (2014). In competition for the attentional template: Can multiple items within visual working memory guide attention? *Journal of Experimental Psychology: Human Perception and Performance*, 40, 1450–1464. <http://dx.doi.org/10.1037/a0036229>
- Võ, M. L. H., & Henderson, J. M. (2010). The time course of initial scene processing for eye movement guidance in natural scene search. *Journal of Vision*, 10, 14. <http://dx.doi.org/10.1167/10.3.14>
- Võ, M. L. H., & Wolfe, J. M. (2012). When does repeated search in scenes involve memory? Looking at versus looking for objects in scenes. *Journal of Experimental Psychology: Human Perception and Performance*, 38, 23–41. <http://dx.doi.org/10.1037/a0024147>
- Wolfe, J. M. (1994). Guided Search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review*, 1, 202–238. <http://dx.doi.org/10.3758/BF03200774>
- Wolfe, J. M., Võ, M. L., Evans, K. K., & Greene, M. R. (2011). Visual search in scenes involves selective and nonselective pathways. *Trends in Cognitive Sciences*, 15, 77–84. <http://dx.doi.org/10.1016/j.tics.2010.12.001>
- Woodman, G. F., Carlisle, N. B., & Reinhart, R. M. G. (2013). Where do we store the memory representations that guide attention? *Journal of Vision*, 13, 1. <http://dx.doi.org/10.1167/13.3.1>
- Zelinsky, G. J. (2008). A theory of eye movements during target acquisition. *Psychological Review*, 115, 787–835. <http://dx.doi.org/10.1037/a0013118>

(Appendix follows)

Appendix A

Scene and Critical Object Exemplar Types

Table A1
Experiment 1

Experimental scenes			Filler scenes	
Scene	Target object	Critical distractor	Scene	Target object
Bathroom 1	Deodorant	Lotion	Bathroom 5	Floss
Bathroom 2	Electric razor	Towel	Bathroom 6	Hair brush
Bathroom 3	Mouthwash	Lotion	Bathroom 7	Hair dryer
Bathroom 4	Shampoo	Mouthwash	Bathroom 8	Toilet paper
Bedroom 1	Alarm clock	Purse	Bedroom 3	Banjo
Bedroom 2	Whiskey bottle	Lamp	Bedroom 4	Eye glasses
Den 1	Candle	Cushion	Bedroom 5	Fan
Den 2	Cushion	Vase	Bedroom 6	Game controller
Den 3	Ottoman	Cushion	Bedroom 7	Guitar case
Den 4	Television	Chair	Bedroom 8	Pet bed
Den 5	Vase	Storage container	Bedroom 9	Shoes
Kitchen 1	Blender	Strainer	Bedroom 10	Skateboard
Kitchen 2	Paper towels	Carrots	Bedroom 11	Stereo
Kitchen 3	Pepper grinder	Tomato	Bedroom 12	Teddy bear
Kitchen 4	Skillet	Pot	Den 6	Broom
Kitchen 5	Teapot	Towel	Den 7	Christmas tree
Kitchen 6	Toaster	Dish soap	Den 8	Laptop
Kitchen 7	Trash bin	Teapot	Den 9	Painting
Kitchen 8	Bowl	Oranges	Den 10	Wicker chair
Kitchen 9	Cookbooks	Lemons	Kitchen 16	Apples
Kitchen 10	Dish soap	Lemonade	Kitchen 17	Bananas
Kitchen 11	Egg	Lemon	Kitchen 18	Bar stool
Kitchen 12	Garlic	Green pepper	Kitchen 19	Red bell pepper
Kitchen 13	Microwave	Candle	Kitchen 20	Bread loaf
Kitchen 14	Mixer	Plate	Kitchen 21	Coffee pot
Kitchen 15	Oven mitt	Coffee container	Kitchen 22	Cooler
Office 1	Cell phone	Folder	Kitchen 23	Cutting board
Office 2	Speaker	Soda bottle	Kitchen 24	Drain plug
Office 3	Stapler	Paper	Kitchen 25	Drinking glass
Office 4	Tape dispenser	Coffee mug	Kitchen 26	Dry erase board
Office 5	Telephone	Folder	Kitchen 27	Flour container
Office 6	Tissue box	Paper	Kitchen 28	Food scale
Office 7	Clock	Flowers	Kitchen 29	Fork
Office 8	Coffee mug	Folder	Kitchen 30	Grater
Office 9	Fax machine	File cabinet	Kitchen 31	Knife set
Office 10	Headphones	Book	Kitchen 32	Pot
Office 11	Lamp	Folder	Kitchen 33	Rolling pin
Office 12	Plastic cup	Recycling bin	Kitchen 34	Sponge
Office 13	Printer	Paper	Kitchen 35	Stove top
Office 14	Soda can	Folder	Kitchen 36	Toaster oven
Outdoor 1	Church tower	Building	Kitchen 37	Wine bottle
Outdoor 2	Mailbox	Sign	Office 15	Backpack
Porch 1	Bench	Sandbox	Office 16	Globe
Porch 2	Deck chair	Pumpkin	Office 17	Hand sanitizer
Porch 3	Flower pot	Chair	Office 18	Keys
Porch 4	Hose	Flowers	Office 19	Pencil cup
Porch 5	Lounge chair	Tile	Office 20	Sticky note
Porch 6	Wreath	Pumpkin	Porch 7	Water bottle

(Appendix continues)

Table A2
Experiments 2 and 3

Experimental scenes			Filler scenes	
Scene	Target object	Critical distractor	Scene	Target object
Bathroom 1	Deodorant	Lotion	Bathroom 5	Floss
Bathroom 2	Electric razor	Towel	Bathroom 6	Hair brush
Bathroom 3	Mouthwash	Lotion	Bathroom 7	Hair dryer
Bathroom 4	Shampoo	Mouthwash	Bathroom 8	Toilet paper
Bedroom 1	Backpack	Chair	Bedroom 6	Banjo
Bedroom 2	Chair	Lamp	Bedroom 7	Eye glasses
Bedroom 3	Coat	Ottoman	Bedroom 8	Fan
Bedroom 4	Purse	Bag	Bedroom 9	Game controller
Bedroom 5	Whiskey bottle	Light	Bedroom 10	Guitar case
Den 1	Blanket	Purse	Bedroom 11	Shoes
Den 2	Candle	Cushion	Bedroom 12	Skateboard
Den 3	Coaster	Cushion	Bedroom 13	Stereo
Den 4	Cushion	Couch	Bedroom 14	Teddy bear
Den 5	Flowers	Lamp	Den 8	Broom
Den 6	Ottoman	Cushion	Den 9	Christmas tree
Den 7	Rug	Cushion	Den 10	Painting
Garage 1	Gas can	Chair	Den 11	Television
Kitchen 1	Apples	Flower pot	Den 12	Vase
Kitchen 10	Oven mitt	Place mat	Kitchen 17	Bananas
Kitchen 11	Pitcher	Plate	Kitchen 18	Bar stool
Kitchen 12	Pot	Bowl	Kitchen 19	Red bell pepper
Kitchen 13	Teapot	Green bell pepper	Kitchen 20	Bread loaf
Kitchen 14	Toaster	Bowl	Kitchen 21	Coffee pot
Kitchen 15	Towel	Lamp	Kitchen 22	Cookbooks
Kitchen 16	Trash bin	Bird	Kitchen 23	Cooler
Kitchen 2	Blender	Strainer	Kitchen 24	Dish soap
Kitchen 3	Bowl	Container	Kitchen 25	Drain plug
Kitchen 4	Cutting board	Towel	Kitchen 26	Dry erase board
Kitchen 5	Drinking glass	Towel	Kitchen 27	Food scale
Kitchen 6	Flower pot	Towel	Kitchen 28	Fork
Kitchen 7	Hand soap	Towel	Kitchen 29	Grater
Kitchen 8	Kitchen cleaner	Coffee filter	Kitchen 30	Knife set
Kitchen 9	Mixer	Skillet	Kitchen 31	Microwave
Office 1	Book	Candle	Kitchen 32	Paper towels
Office 10	Pen	Folder	Kitchen 33	Sponge
Office 11	Pencil cup	Computer mouse	Kitchen 34	Toaster oven
Office 12	Plastic cup	Recycling bin	Kitchen 35	Wine bottle
Office 13	Tape dispenser	Coat	Outdoor 1	Mailbox
Office 14	Water bottle	Paper weight	Office 15	Glob
Office 2	Clock	Tool container	Office 16	Hand sanitizer
Office 3	Coffee mug	Lamp	Office 17	Headphones
Office 4	Fax machine	File cabinet	Office 18	Keys
Office 5	Folder	Mousepad	Office 19	Printer
Office 6	Lamp	Folder	Office 20	Soda can
Office 7	Laptop	Chair	Office 21	Speaker
Office 8	Mousepad	Tissue box	Office 22	Stapler
Office 9	Notepad	Cushion	Office 23	Telephone
Porch 1	Lounge chair	Tile	Porch 2	Bench

Received January 13, 2017
Revision received April 10, 2017
Accepted May 23, 2017 ■