How context information and target information guide the eyes from the first epoch of search in real-world scenes

Sara Spotorno — School of Psychology, University of Dundee, Dundee, Scotland, UK; Institut de Neurosciences de la Timone (INT), CNRS & Aix-Marseille University, Marseille, France

George L. Malcolm — Department of Psychology, The George Washington University, Washington, DC, USA

Benjamin W. Tatler — School of Psychology, University of Dundee, Dundee, Scotland, UK

This study investigated how the visual system utilizes context and task information during the different phases of a visual search task. The specificity of the target template (the picture or the name of the target) and the plausibility of target position in real-world scenes were manipulated orthogonally. Our findings showed that both target template information and guidance of spatial context are utilized to guide eye movements from the beginning of scene inspection. In both search initiation and subsequent scene scanning, the availability of a specific visual template was particularly useful when the spatial context of the scene was misleading and the availability of a reliable scene context facilitated search mainly when the template was abstract. Target verification was affected principally by the level of detail of target template, and was quicker in the case of a picture cue. The results indicate that the visual system can utilize target template guidance and context guidance flexibly from the beginning of scene inspection, depending upon the amount and the quality of the available information supplied by either of these high-level sources. This allows for optimization of oculomotor behavior throughout the different phases of search within a real-world scene.

Introduction

Most of our activities first of all require that we locate a target for action from among other objects. Visual search studies have provided key understanding about the decisions that underlie when and where we move the eyes in one of the most frequent and important tasks in our everyday life (see Wolfe & Reynolds, 2008). Search involves both low- and high-level information in scenes, with the two key sources of guidance (see Tatler, Hayhoe, Land, & Ballard, 2011) coming from what we expect the target to look like (Kanan, Tong, Zhang, & Cottrell, 2009) and where we expect to find it (Ehinger, Hidalgo-Sotelo, Torralba, & Oliva, 2009; Torralba, Henderson, Oliva, & Castelhano, 2006). What is less well understood is how these two sources of information are used together to guide search and whether their relative uses vary over the course of search. The present work considers the relative contribution of these two sources of information in guiding search.

Expectations about target appearance

Prior information about the target allows observers to form a representation (i.e., a template) in visual working memory, which can be compared with the attributes of the current percept. The more detailed this representation, the more efficient the ensuing search. Response times are indeed faster when the target is cued by its picture than when it is described by a text label (e.g., Vickery, King, & Jiang, 2005; Wolfe, Horowitz, Kenner, Hyle, & Vasan, 2004). This facilitation also holds true for oculomotor behavior. Objects having highly matching properties with the template are likely to be selected by the eyes for further processing (Findlay, 1997; Rao, Zelinsky, Hayhoe, &

Citation: 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*(2):7, 1–21, http://www.journalofvision.org/content/14/2/7, doi:10.1167/14.2.7.

doi: 10.1167/14.2.7 Received June 27, 2013; published February 11, 2014 ISSN 1534-7362

Ballard, 2002; Scialfa & Joffe, 1998; Williams & Reingold, 2001; see Zelinsky, 2008, for a more detailed theoretical account of these results). When searching object arrays (Castelhano, Pollatsek, & Cave, 2008; Schmidt & Zelinsky, 2011; Yang & Zelinsky, 2009) or real-world scenes (Castelhano & Heaven, 2010; Malcolm & Henderson, 2009, 2010) using a picture cue prior to search results in faster search than cuing with a verbal label describing the target. These picture/word differences are likely to reflect the level of detail that they permit in forming the representation of the search target rather than reflecting the nature of the processing (visual versus verbal) required by the type of the prior information. This is demonstrated by the fact that picture cues that differ from the target in scale and orientation are less effective than an exactly matching pictures (Bravo & Farid, 2009; Vickery et al., 2005). Moreover, the benefit for search provided by a verbal cue increases as the description of the target becomes more precise. Maxfield and Zelinsky (2012) showed that objects cued with subordinate category labels (e.g., "taxi") were found faster than those cued by basic-level category labels (e.g., "car"), which were in turn found faster than objects cued by superordinate category labels (e.g., "vehicle"). Schmidt and Zelinsky (2009) had shown the same effect of the narrowing of the category level, comparing basic with superordinate labels. These authors had also reported a similar facilitating effect on search obtained by adding some information about target features (e.g., the color) to either of these two types of verbal cues.

Expectations about target location

Observers can access knowledge about the overall gist and spatial structure of a scene within 100 ms or less (e.g., Biederman, 1981; Potter, 1976; Greene & Oliva, 2009). This knowledge can assist subsequent search. Eye guidance is improved and response times are shortened, for instance, with a scene preview, even brief, compared to situations without a scene preview (e.g., Castelhano & Heaven, 2010; Castelhano & Henderson, 2007; Hillstrom, Scholey, Liversedge, & Benson, 2012; Hollingworth, 2009; Võ & Henderson, 2010) or when the preview is just a jumbled mosaic of scene parts (Castelhano & Henderson, 2007; Võ & Schneider, 2010). However, neither a preview of another scene from the same basic-level category (Castelhano & Henderson, 2007) nor cueing the searching scene with its basic category verbal label (Castelhano & Heaven, 2010) seem to facilitate search. What appears crucial, indeed, is the guidance supplied by the physical background context of the scene: Previewing the component objects without background is not beneficial (Võ & Schneider, 2010). This is in line with the fact that searching for arbitrary objects is far more efficient when they are embedded in scenes with consistent background than when they are arranged in arrays on a blank background: While the estimated search slope in a consistent scene is about 15 ms/item, it increases to about 40 ms/item in the absence of any scene context (Wolfe, Alvarez, Rosenholtz, & Kuzmova, 2011). In visual search, knowledge about the spatial structure of scenes enables rapid selection of plausible target locations, biasing search to a subset of regions in the scene. This has been mainly shown with images of everyday scenes presented on a computer screen (Eckstein, Drescher, & Shimozaki, 2006; Henderson, Weeks, & Hollingworth, 1999; Malcolm & Henderson, 2010; Neider & Zelinsky, 2006; Torralba et al., 2006; Võ & Wolfe, 2013; Zelinsky & Schmidt, 2009), but there is evidence that placing the target in an expected location facilitates search also in real-world environments. On this point, Mack and Eckstein (2011) used a search task in which the target object was on a cluttered table in a real room, placed next to objects usually co-occurring with it or among unrelated objects: Fewer fixations were necessary to find it and search times were shorter in the first case.

Combining expectations about target appearance and target location

It seems reasonable that expectations about target appearance and target location should be integrated during search to guide the eyes optimally. Indeed a dual-pathway architecture underlying visual search in scenes has been proposed (Wolfe, Võ, Evans, & Greene, 2011) in which global scene structure and local scene details are used to guide search. Similarly, Ehinger et al. (2009) evaluated a model of scene viewing that combines low-level salience (e.g., Itti & Koch, 2000), expected target appearance, and expected target placement in the scene. This model was able to account for a high proportion of human fixations during search. A similar approach was employed by Kanan et al. (2009) to show that a model containing low-level salience, expected appearance, and expected object placement outperformed models containing only a subset of these factors. While both studies suggest that all three components contribute to attention guidance in search, they drew different conclusions about the relative importance of appearance and expected location: Kanan et al. (2009) suggested a more prominent role for appearance than expected location; Ehinger et al. (2009) suggested the opposite.

One way to study the relative contribution of expectations about target appearance and target placement in scenes is to manipulate the reliability and availability of each source of information (Castelhano

& Heaven, 2010; Malcolm & Henderson, 2010). While both of these previous studies varied target template information by comparing verbal versus pictorial target cues, they differed in the way of manipulating the importance of expected target placement in the scene: Castelhano and Heaven (2010) took into account the specificity of prior information about the scene (scene preview vs. word cue indicating scene's gist); Malcolm and Henderson (2010) investigated the effect of consistency in the placement of targets in the scenes. Both studies found contributions of target appearance and spatial context in search: Usefulness of target information and usefulness of context information shortened additively the time needed to first fixate the target and enhanced spatial selectivity, in terms of number and spatial distribution of fixations.

Differential reliance on appearance and placement during search

Using knowledge about likely target placement and appearance to guide search requires extraction of global and local information, respectively. It is unclear whether both sources of information are available simultaneously or sequentially in search. In particular, they may be differentially available at the initiation of search. Overall scene structure may be available very early to guide search to regions where the target might be expected (Greene & Oliva, 2009; Neider & Zelinsky, 2006; Nijboer, Kanai, de Haan, & van der Smagt, 2008; Oliva & Torralba, 2001) and this may be prior to processing a scene's local components. Alternatively, some local properties may also be available in a single glance at a scene (Joubert, Rousselet, Fize, & Fabre-Thorpe, 2007; Mack & Palmeri, 2010; Quattoni & Torralba, 2009) and may affect attentional allocation from the outset of search (Torralba et al., 2006). Empirical evidence about the relative contribution of these types of information to search initiation is inconclusive. Malcolm and Henderson (2010) divided search into three phases (initiation, scanning, and verification) and suggested that neither target template information nor target placement influenced search initiation, but both impacted on later phases of search. In contrast, Schmidt and Zelinsky (2009, 2011) demonstrated that the percentage of initial saccades directed toward the target object within an array of distractors was higher when the target template was specific.

Võ and Henderson (2010) found that prior exposure to scene structure via a preview resulted in shorter latency to initiate search and larger amplitudes of initial saccades (but see Hillstrom et al., 2012, for a study that did not find such an early effect on amplitude). Neider and Zelinsky (2006) found that initial saccades were more likely to be directed toward a target plausible region than toward a target implausible region. Eckstein et al. (2006) showed that landing points of initial saccades were closer to the target when it was plausibly placed, independently of its detectability and eccentricity. Importantly, landing points of initial saccades were closer to plausible than implausible locations even when the target was absent. This was true for human observers and also for initial saccades generated by a Bayesian model of differential weighting of scene locations, based on visual evidence for the presence of target-relevant features and on expectations associated to each location.

The present study

In the present study we manipulated the precision of target template information (visual or verbal) and the utility of spatial expectations (by placing targets in expected or unexpected scene regions). We considered the impact of these two types of information on search initiation, scanning, and verification. By manipulating the expected appearance and placement of objects in this way and dividing search into three phases we were able to consider the relative contributions of expectations about appearance and location to each phase of search. We focused particularly on search initiation because it remains largely unanswered whether (and how) the first saccade is guided by both knowledge about target appearance and expectations about target location.

The rare previous studies manipulating both target template and spatial context information (e.g., Castelhano & Heaven, 2010; Malcolm & Henderson, 2010) used scenes that were complex in nature, with high-probability and low-probability regions not easily divisible, making it hard to understand precisely what source of high-level information was being utilized. For example, a saccade could move in the direction of a highly probable region, but then land on a low-probability region, masking the goal of the saccade (e.g., when searching for a plate in a restaurant, the observer might saccade downward toward a table in the foreground, but land on the floor in a gap between to table). Issues like this, therefore, made it hard to determine which particular region a participant was intending to fixate. We used scenes that included two clearly differentiated regions, separated by easily defined boundaries, and clearly having a low or high probability to include the target object (see Method and Figure 1). In this way, the direction of the first saccade can be a critical and reliable measure of the selectivity of early eye guidance.

A key condition to understand how target template and spatial expectation guide search is to put them in

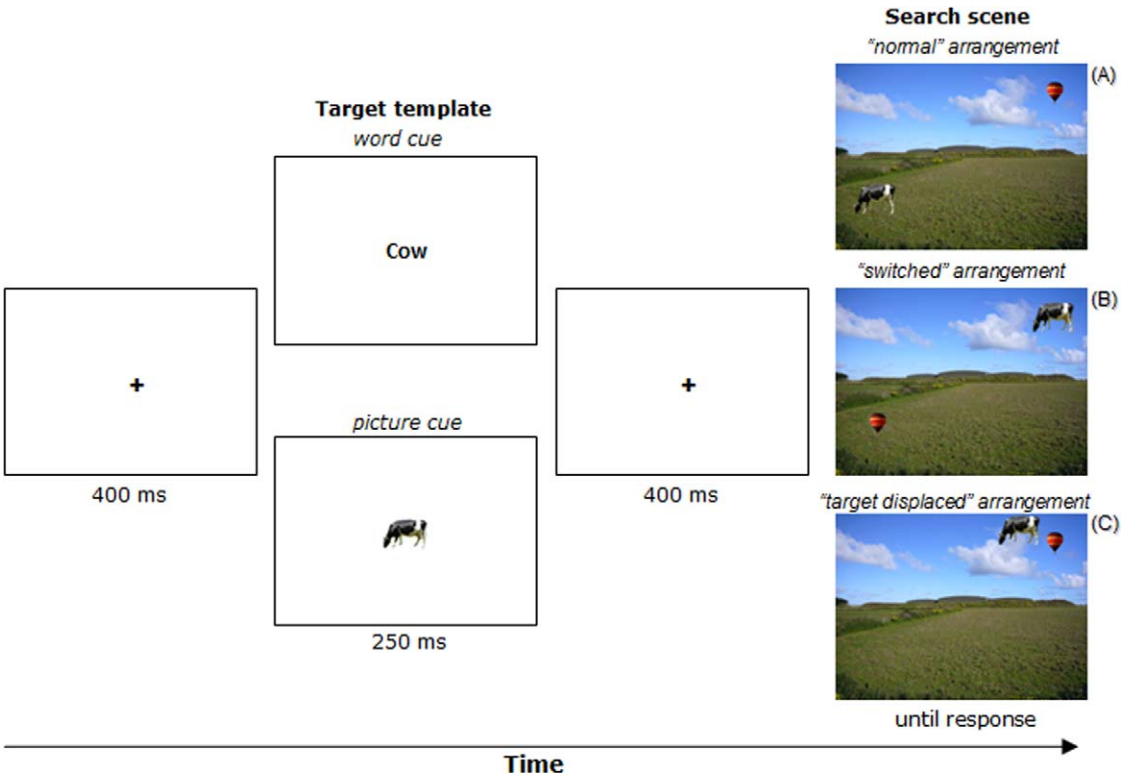

Figure 1. Example of screen shots of trials. This example shows the two types of target template and the three scene arrangements. The types of scene arrangement were made by placing in different positions the two objects (i.e., the target, here the cow, and the distractor, here the hot-air balloon) added in two regions (here, the field and the sky) of each scene. These objects were inserted in their respective high-probability region (A), one in the region plausible for the other (B), or both in the plausible region for the distractor (C). Please note that each trial started with a drift check screen (here not depicted).

conflict, placing the target in an unexpected (i.e., inconsistent) location, and comparing this situation to when the target is plausibly located within the scene. In doing that, however, it is necessary to control for a potential confounding effect. We need to distinguish the impacts of target template and expectations concerning target position from that of an attentional prioritization due to spatial inconsistency per se (e.g., Biederman, Mezzanotte, & Rabinowitz, 1982), which would result in earlier (Underwood, Templeman, Lamming, & Foulsham, 2008) or longer ocular exploration (Võ & Henderson, 2009, 2011). This effect cannot be teased apart from previous findings in search, as the spatially inconsistent target always had as counterparts only spatially consistent objects (Castelhano & Heaven, 2011; Võ & Henderson, 2009, 2011) or a variety of spatially consistent and inconsistent objects (Võ & Wolfe, 2013). In the present study, a target in an unexpected location was paired, in half of the trials, with another spatially inconsistent object by placing a distractor in the target high-probability region, equating in this way every attentional effect of inconsistency.

The role of spatial expectations concerning the target might be to guide the eyes either toward a plausible (even empty) location or toward objects that are potential target candidates and are placed in the region where the target was expected. The first possibility would be coherent with the space-based account of attentional allocation (e.g., Eriksen & Yeh, 1985), whereas the latter possibility would give support to the object-based account (e.g., Egly, Driver, & Rafal, 1994). Previous investigations of search in scenes cannot differentiate these two possibilities because when the target was in an unexpected location, other objects were always placed in the target plausible region. In this study, we left the target plausible region empty (i.e., without any foreground object) in one third of the trials. Therefore, our findings are relevant to this ongoing debate in the literature, giving direct support to one of these competing accounts.

Our study allowed us to consider a number of alternative hypotheses about how the precision of prior information about the target object and the plausibility of target position within the scene influence search. If only the target template is used, we would expect that the target object is saccaded to equally well irrespective of the scene region in which it is placed, and also that the time to match (verify) the target object to its template would not be influenced by target position. If, conversely, only expectations about location guide

search, we would expect that saccades would be directed initially toward the target plausible region, independent of the type of target template and the actual target position in the scene. No effects on saccade latency should be reported as well. If the visual system needs the presence of an object in the target plausible region to direct the eyes to that region, in line with an object-based account of attention, we should find that the plausible target region is not saccaded to when it is empty, but only when it contains an object (either the target or another distractor object). It is obviously very unlikely that the visual system utilizes only one of these sources of information even to plan the first saccade in visual search (e.g., Ehinger et al., 2009); rather both sources of guidance are likely to be utilized from the outset of search. If this is the case, a particularly interesting situation for understanding how expectations about the target's appearance and its position guide search is when these two sources of information are in conflict. If the contribution of these two sources of information is comparable, we would expect a similar percentage (close to 50%) of initial saccades directed either toward the target object or the target plausible region, with similar latencies. Any greater contribution of one type of guidance over the other should result in a change in this proportion of initial saccades and/or their latency. While the initiation of search is particularly diagnostic for the early use of information in guiding the eyes, differential reliance on target template and spatial expectations may persist in later epochs of search. We therefore consider scanning and target verification phases of search.

Method

Participants

Twenty-four native English-speaking students (16 females), aged 18–21 ($M = 19.54$, $SD = 1.19$) participated for course credit and gave informed consent in accordance with the institutional review board of the University of Dundee. All participants were naïve about the purpose of the study and reported normal or corrected-to-normal vision.

Apparatus

Eye movements were recorded using an EyeLink 1000 at a sampling rate of 1000 Hz (SR Research, Canada). Viewing was binocular, but only the dominant eye was tracked. Experimental sessions were carried out on a Dell Optiplex 755 computer running OS Windows XP. Stimuli were shown on a ViewSonic G90f-4 19-in. CRT monitor, with a resolution of 800 × 600 pixels, and a refresh rate of 100 Hz. A chin rest stabilized the eyes 60 cm away from the display. Manual responses were made on a response pad. Stimulus presentation and response recording was controlled by Experiment Builder (SR Research, Canada).

Materials

Forty-eight full-color photographs (800 × 600 pixels, 31.8° × 23.8°) of real-world scenes from a variety of categories (outdoor and indoor, natural and man-made) were used as experimental scenes. Each of them had two distinct regions (e.g., field and sky). Two objects taken from Hemera Images database (Hemera Technologies, Gatineau, Canada) or Google Images were modified and placed into each scene with Adobe Photoshop CS (Adobe, San Jose, CA). One of the two inserted objects was designated as the target on the search task, while the other had the function of distractor. The designated target object in each scene was counterbalanced across participants.

In order to manipulate the arrangement of the objects within scene context, four versions of each experimental scene were made by inserting the two objects in different positions. This created three types of scene arrangement (see Figure 1). In the first scene version, the target and the distractor were added in their respective high-probability regions ("normal" scene arrangement; e.g., a cow in the field and a hot-air balloon in the sky). In the second version, these objects were switched, so that so that they were both in low-probability locations ("switched" scene arrangement: e.g., the hot-air balloon in the field and the cow in the sky). In the third and fourth versions, finally, no objects were in the target probable region, as both objects were placed in the other region ("target displaced" arrangement: the cow and the hot-air balloon both in the field, if the target was the hot-air balloon, or both in the sky, if the target was the cow). All the experimental scenes, with the four versions, are available online at http://www.activevisionlab.org.

In order to manipulate the template of the target, picture and word cues were created. To create the picture cues, each object was pasted in the middle of a white background, appearing exactly as it would in the scene regarding size, color, etc. To create the word cues, 48 verbal labels (up to three words) of the objects (font: Courier, color: black, font size: 72 point), subtending 2.14° in height, were centered on a white background.

Seventy-eight further scenes were added to the experiment, four for practice and the others as fillers,

using an existing object in the scene as the target. Thirty-nine picture cues and 39 word cues were created for these scenes.

Evaluation of the experimental scenes

In an evaluation study, the normal arrangement and switched arrangement versions of the experimental scenes were evaluated by 10 participants (aged 22–35, mean age = 30.3, $SD = 4.41$). None of them had seen the images before and none took part subsequently in the search experiment. They were divided into two groups of five in order to counterbalance across participants the versions of the images presented. Each group evaluated half of the scenes with the normal arrangement and the other half with the switched arrangement. A participant, therefore, never saw the same object at two different locations within the scene. For each experimental scene (plus two images as practice) several aspects were rated on Likert scales (from one, "low," to six, "high"): the degree of matching between the verbal label and the picture of the object, the quality of object insertion (i.e., how much it seemed to belong in the scene in terms of visual features, independent of the plausibility of its location), the plausibility of the object's position in the scene, the object's perceptual salience (in terms of brightness, color, size, etc.) and the object's semantic relevance for the global meaning (i.e., the gist) of the scene. Finally, they rated on the same six-point scale the complexity of the whole image, defined with regard to the number of objects, their organization, and image textures. Before starting the rating experiment, participants were given a written definition of each aspect to rate, immediately followed by an example. After practice, the experimental scenes were presented in random order, while the series of judgments respected always the above described sequence. For each scene, each judge scored the two inserted objects, whose order of presentation was counterbalanced across participants. First of all, the picture of the first object was presented in the center of the screen, followed by its name; once the participant had rated the degree of name-picture matching, the scene was presented and remained visible on the screen for all the required evaluations. The same sequence was then repeated for the second object. Finally, the complexity of the image was rated.

Results showed that, overall, the scenes were rated as having medium complexity ($M = 3.40$, $SD = 0.98$). The chosen verbal label matched their corresponding objects well ($M = 5.87$, $SD = 0.27$) and object insertions were of good quality ($M = 4.30$, $SD = 0.59$), without a significant difference depending on scene version, $t(95) < 1$, $p = 0.510$. Scores of objects meant to be in high- and low-probability regions confirmed the plausibility ($M = 5.33$, $SD = 0.76$) or implausibility ($M = 1.58$, $SD = 0.75$), respectively, of the chosen locations. The difference between these two groups of scores was significant, $t(95) = 39.04$, $p < 0.001$. Objects in both location conditions were rated, on average, as rather salient ($M = 4.33$, $SD = 0.84$) and relevant ($M = 4.05$, $SD = 0.82$).

Procedure

Prior to the experiment each participant underwent a randomized nine-point calibration procedure, that was validated in order to ensure that the average error was less than 0.5° and the maximum error in one of the calibration points was less than 1°. Recalibrations were performed during the task if necessary. Before each trial sequence, a drift check was applied as the participant fixated a dot in the center of the screen. When the drift check was deemed successfully (drift error less than 1°), the experimenter initiated the trial. A central fixation cross appeared for 400 ms followed by a 250-ms cue indicating the search target. The cue was either the name of the target or an exactly matching picture of the target. This was followed by a central fixation point lasting another 400 ms, making a stimulus onset asynchrony of 650 ms. The scene then appeared and participants searched for the target object, responding with a button press as soon as it was located.

The experiment had a 2 (Template Type) × 3 (Scene Arrangement) design. Half of the scenes were cued with the picture of the target object, the other half of the scenes were cued with the name of the target object (see Figure 1). The picture and word cues were fixed for the filler scenes and counterbalanced across participants for the experimental scenes.

Each scene was displayed only once during the experiment. Each participant saw one third of the 48 experimental scenes having a normal scene arrangement, one third of them with the switched scene arrangement, and one third of them with the target displaced arrangement. The three manipulations of object position were rotated through scenes across participants in a Latin Square design. Targets in filler scenes were positioned in high-probability locations, meaning that 75% of all the scenes viewed by participants had target objects in high-probability regions. This percentage ensured that participants would recognize scene context as a potential source of guidance throughout the experiment. Test scenes and filler scenes were intermixed and presented in a random order for each participant. The eye movements from the filler trials were not analyzed. The experiment lasted for about 30 min.

ROIs definition and data analyses

The regions of interest (ROIs) for scoring eye movements were defined as the smallest fitting rectangle that encompassed both the target and the distractor when placed in the same scene region. Two ROIs (i.e., "target high-probability region" and "distractor high-probability region") in each scene were defined with this criterion. Thus, the two scoring regions per image were the same for all the conditions to allow for better comparisons. A saccade was considered as being directed toward a specific ROI if its angular direction was within 22.5° of the angular direction to the center of the ROI.

In complementary analyses, we defined an alternative set of ROIs, named here as "extensive scene regions," that encompassed (a) the entire region of the scene that was plausible for the target object and (b) the entire region of the scene that was plausible for the distractor object. This allowed us to check for saccades that targeted the scene region but not the location in which an object was inserted.

Data from two participants were eliminated due to an imbalance in the experimental design in the condition where the target object and the distractor object were in the same scene region and the expected target location was empty. Raw data were parsed into saccades and fixations using the SR Research algorithm. Subsequent analyses of fixations, saccades and individual samples were conducted using routines written in Matlab 7.12.0. We discarded from analyses trials for which the target was meant to be in an unexpected location but the rates of plausibility of positions (rated in the evaluation study) were not sufficiently low (1.82%). Trials in which participants were not maintaining the central fixation when the scene appeared (1.82%) and trials with errors (2.95%) were also removed. Responses were considered correct if the participant looked at the target when pressing the buttons or during the immediately preceding fixation. Trials with first saccade latency shorter than 50 ms (3.47%) or with RTs greater than two standard deviations from the mean for each condition (3.13%) were excluded as outliers. Overall, 13.19% of trails were removed by these criteria.

Repeated-measures analyses of variance (ANOVAs) with template type (word vs. picture) and scene arrangement (normal vs. switched vs. target displaced) as factors were conducted on total trial duration and on oculomotor behavior considering separately three phases (see Malcolm & Henderson, 2009): search initiation (planning and execution of the first saccade), image scanning (from the end of the first saccade until the target is first fixated), and target verification (i.e., the acceptance of the currently inspected object as being the target). Partial η^2 is reported as measure of effect size, considering an effect as being small when the partial η^2 value is less than 0.06, medium when it is more or equal to 0.06 but less than 0.14, and large when it is more or equal to 0.14 (see Cohen, 1988). The measure and, in particular, these conventional benchmarks, should be considered carefully (see also Fritz, Morris, & Richler, 2012), but they offer a way to assess the practical significance of an effect beyond its statistical significance. In cases in which the assumption of sphericity was violated, Mauchly's W value and degrees of freedom adjusted with the Greenhouse-Geisser correction are reported. Differences between means of conditions were analyzed with Bonferroni corrected paired-sample t tests (all two-tailed); the reported adjusted p values where obtained by multiplying the unadjusted p value for the number of the comparisons made.

In order to understand the manner in which scene context and target information interact, it is important to explore and differentiate thoroughly effects arising from covert (indexed by saccade latency) and overt (indexed by saccade direction) selection of the target object and those arising from covert or overt selection of the scene region in which the target is expected to appear. As a result we ran separate analyses for targeting with respect to each of the target object and the expected scene region. By conducting separate analyses in this way we were able to better describe the manner in which targeting decisions are influenced by target information and scene context. Of course, these two approaches to analysis are related but in order to differentiate effects of each type of guidance information separate analyses were required. Note that the data for the normal scene arrangement (where the target object is in its plausible location) were the same in the analyses conducted for targeting with respect to the target object and expected scene region; however, to allow meaningful comparisons for each measure, these data were included in both analysis.

Our proposed hypotheses for the roles of scene context and target information are distinguished by the patterns of differences between template types for each scene arrangement (and vice versa). As such, we were a priori interested in the comparisons of these conditions and we report these comparisons in the sections that follow even in cases where the interaction is not significant. The exception to this is that we do not break down interactions if the F ratio is less than one.

Results

Figure 2 depicts fixation density distributions across all participants for an example scene in each of the

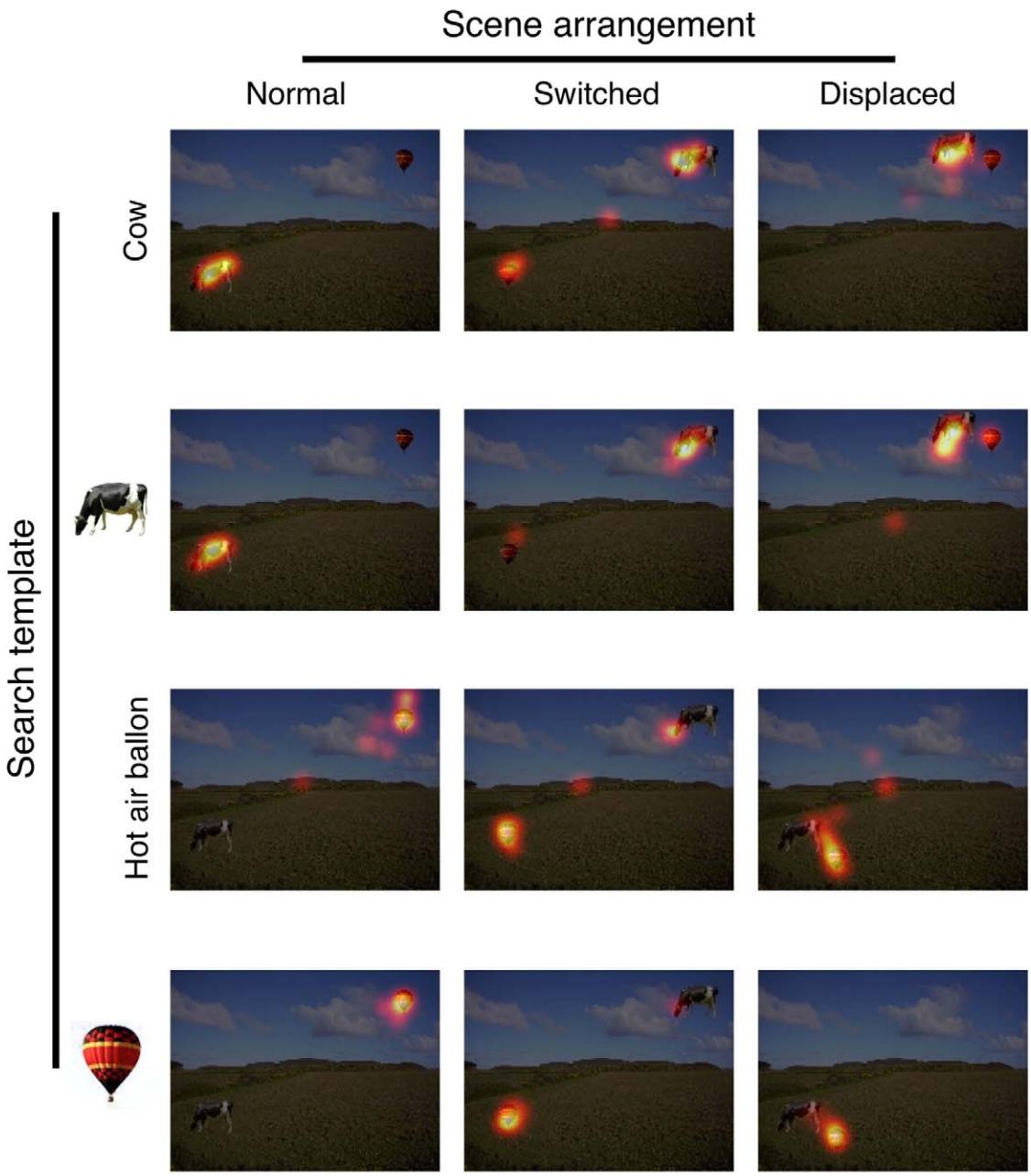

Figure 2. Fixation density distributions for each experimental condition for an example scene. Distributions comprise data across all search epochs from all participants and were created by iteratively adding Gaussians centered at each fixation location, each with full width at half maximum of 2° of visual angle. Hotter colors denote greater fixation density. The first fixation in each trial (which began on the central pretrial marker) is not included in these distributions.

experimental conditions. These distributions were created by iteratively adding Gaussians centered at each fixation location, each with full width at half maximum of 2° of visual angle. The first fixation in each trial was excluded because it was always central, as participants waited for the scene to appear. There are clear differences in viewing behavior between the experimental conditions. These differences are explored in the sections that follow.

Total trial duration

Because trials were terminated by the participant's button press to indicate that they had found the target, we can use total trial duration as a measure of overall search time, combining the three phases of search initiation, scene scanning, and target verification. All the effects reported below were large. There was a main effect of template type, $F(1, 21) = 72.76$, $p < 0.001$,

Scene arrangement	Normal				Switched				Target displaced			
Template type	Word		Picture		Word		Picture		Word		Picture	
	Mean	*SE*	Mean	*SE*	Mean	*SE*	Mean	*SE*	Mean	*SE*	Mean	*SE*
Total trial duration (ms)	763	(25)	652	(28)	939	(45)	680	(29)	821	(33)	707	(33)
Search initiation												
Probability of saccading (%)												
- Toward the target object	66.2	(3.3)	76.4	(3.3)	43.8	(3.8)	67.0	(4.1)	44.4	(4.3)	58.9	(5.0)
- Toward the target region					44.7	(3.3)	25.1	(3.8)	5.0	(2.2)	2.5	(1.2)
First saccade gain	0.83	(0.02)	0.85	(0.02)	0.84	(0.03)	0.87	(0.02)	0.85	(0.03)	0.84	(0.03)
First saccade latency (ms)												
- Toward the target object	198	(7)	196	(6)	200	(5)	205	(6)	209	(7)	202	(6)
- Toward the target region					208	(8)	188	(7)	-	-	-	-
Image scanning												
Scanning time (ms)	187	(9)	143	(8)	319	(28)	161	(13)	215	(15)	165	(16)
Number of fixations	1.76	(0.05)	1.68	(0.06)	2.47	(0.14)	1.72	(0.08)	1.92	(0.08)	1.69	(0.09)
Target verification time (ms)	377	(20)	311	(22)	413	(28)	321	(24)	396	(25)	334	(23)

Table 1. Results. Means and standard errors as a function of the two types of target templates and the three types of scene arrangements.

partial $\eta^2 = 0.78$, a main effect of scene arrangement, $F(1.53, 32.07) = 11.95$, $p < 0.001$, partial $\eta^2 = 0.36$. Mauchly's $W(2) = 0.690$, $p = 0.025$, and an interaction, $F(2, 42) = 11.61$, $p < 0.001$, partial $\eta^2 = 0.34$ (Table 1). For each of the three target position conditions, trial duration was shorter for picture than for word templates, all $ts(21) \geq -4.33$; all $ps \leq 0.001$. There were no differences between trial durations depending on the scene arrangement for picture templates (all $ts \leq 2.65$; all $ps \geq 0.135$). For word templates, trial durations were shorter when the target was in the high-probability location than when it was switched with the distractor object, $t(21) = -4.60$, $p < 0.001$. Trial durations also tended to be shorter when the target was in the high-probability location than when it was located near the distractor, $t(21) = -2.96$, $p = 0.072$ and in this latter condition than when the two objects were swapped, $t(21) = -3.02$, $p = 0.063$.

Search initiation

In order to investigate eye movement behavior during the first viewing epoch, we compared, for each scene arrangement and each type of template, the probability of first saccading toward the target high-probability region (that actually contained the target object only when the scene arrangement was normal) to the probability of first saccading toward the distractor high-probability region (that contained the target object in the switched and in the target displaced arrangement conditions). In all but one case the probability of saccading toward the target object was greater than the probability of saccading toward the other compared location, all $ts(21) \geq 5.52$, all $ps < 0.001$; Figure 3. The only exception was when the target was cued by its verbal label and the positions of the target object and of the distractor object were switched. In this case, participants were equally likely to direct the first saccade toward either the target object (43.83%), placed in the distractor plausible location, or the target plausible location, occupied by the distractor object (44.72%), $t(21) < 1$, $p = 0.896$.

In order to differentiate potential guidance effects arising from the target information and the scene context information, we conducted separate repeated-measure ANOVAs for selection with respect to the target object and expected target region (see Method).

With this logic, we first analyzed what influenced the probability of directing the first saccade toward the target object and the latency of the first saccades when launched in target direction (see the section Probability and latency of saccading toward the target object). As a supplement to analysis of direction, we also used an identical model of ANOVA to analyze gain for the first saccades directed toward the target object see the section First saccade gain toward the target object), in order to consider how close saccades directed to the target object landed to the center of the ROI enclosing the target object. Subsequently, we ran ANOVAs with the same design in order to examine what influenced the probability of directing the first saccade toward the expected target region and the latency of launching the first saccade toward this direction (see the section Probability and latency of saccading toward the target region).

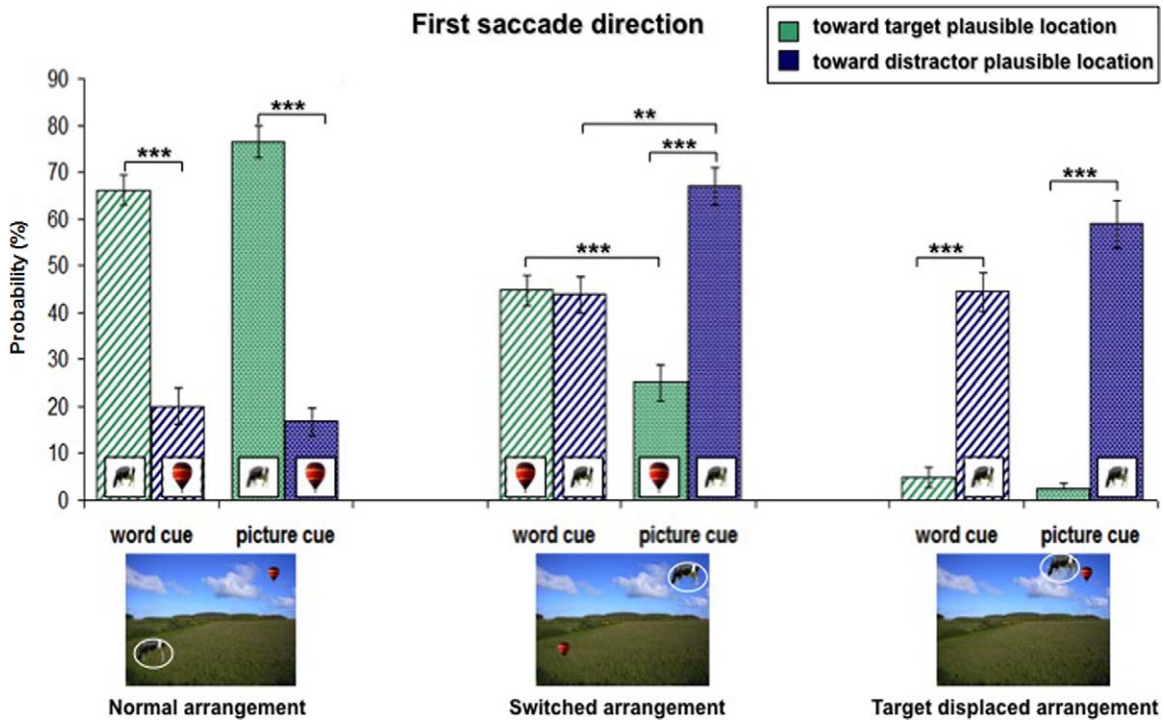

Figure 3. Search initiation. Probability that the first saccade is directed toward either the target plausible location (green bars) or the distractor plausible location (blue bars) as a function of location type, template type, and scene arrangement. Bars show condition means ± 1 *SEM*. ***: $p < 0.001$, **: $p < 0.01$ at Bonferroni corrected pairwise comparisons. Comparisons between scene arrangements are not shown. The objects depicted within the bars indicate the object toward which the first saccade was directed in each condition. The absence of an object (green bars in target displaced condition) indicates that the first saccade was directed toward an empty location. White circles in the inset depictions of the example scene indicate the target object and were not seen by participants.

Probability and latency of saccading toward the target object

One way of assessing the initial use of information in search is to consider how well participants were able to direct their first saccade toward the target object when provided with varying amounts of template information and differential plausibility of target object placement in the scene. By also analyzing the latency of the first saccade launched toward the target we were further able to consider whether there was evidence for different time courses of information assimilation and utilization to initiate search correctly.

For the first saccade direction (Figure 3 and Table 1), there was a large main effect of template type, $F(1, 21) = 34.49$, $p < 0.001$, partial $\eta^2 = 0.62$, with an higher probability of saccading toward the target object following a picture cue than a word cue ($M = 67.4\%$ vs. $M = 51.5\%$). There was also a large main effect of scene arrangement, $F(2, 42) = 12.04$, $p < 0.001$, partial $\eta^2 = 0.36$, with an higher probability of saccading toward the target object when it was in the expected location ($M = 71.3\%$) than when it was in an unexpected location, either alone, $M = 55.4\%$, $t(21) = 3.88$, $p = 0.003$, or near the distractor object, $M = 51.6\%$, $t(21) < 4.97$, $p < 0.001$. There was no difference in the probability of saccading toward the target object when it was in either of the two unexpected arrangements, $t(21) < 1$, $p > 0.999$. There was no significant interaction between template type and scene arrangement, $F(2, 42) = 1.45$, $p = 0.246$ (although the relative effect of the interaction could be considered of medium size: partial $\eta^2 = 0.065$). Despite the lack of significant interaction, we were a priori interested in breaking down the results for each of the three arrangement conditions in order to consider whether the impact of the template on saccade target selection depends upon the placement of the objects in the scene. Planned comparisons showed the probability of directing the first saccade toward the target object was greater with a picture cue than with a word cue only when the positions of the target object and the distractor object were switched, $t(21) = 4.01$, $p = 0.009$, while no differences were found depending on the type of template for the other scene arrangements (both $ts(21) \leq 2.70$, both $ps \geq 0.126$). We then considered how the arrangement of the objects in the scene influenced first saccade direction for the verbal

and the pictorial templates separately. For picture cues there were no differences in the probability of saccading toward the target object between the different scene arrangements, all $ts(21) \leq 2.89$, all $ps \geq 0.081$. For word cues the probability of saccading toward the target object was higher when it as in the expected location than when it was in an unexpected location, either alone, $t(21) = 4.18$, $p < 0.001$, or with the distractor object, $t(21) = 4.78$, $p < 0.001$, while it did not differ between these two latter arrangement conditions, $t(21) < 1$, $p = 0.914$.

When considering the latency of saccading toward the target object (Table 1), there was no main effect of either template type, $F(1, 21) < 1$, or scene arrangement, $F(2, 42) = 2.12$, $p = 0.132$. There was no interaction between template type and scene arrangement, $F(2, 42) < 1$, $p = 0.451$.

These findings indicated that when participants had a precise representation of the target object, they were likely to initiate search correctly toward the target object even when this was placed where they did not expect to find it. However, when the representation of the target object was abstract, switching the target with the distractor object interfered greatly with search initiation, with participants equally likely to direct the first saccade to the target or distractor object. The speed of initiation was not affected by our experimental manipulations.

First saccade gain toward the target object

The contribution of target template and spatial expectations to accurate saccade targeting might not be manifest solely in the direction of the first saccade but also in how close the saccade brings the fovea to the target. We therefore calculated the gain of the first saccade (if it was launched in the direction of the target object) relative to the center of the target: that is the ratio between the first saccade amplitude and the initial retinal eccentricity of the center of the target's ROI. Neither template type, $F(1, 21) < 1$, $p = 0.456$, nor scene arrangement, $F(1, 21) < 1$, $p = 0.712$, influenced the gain of the first saccade. The two factors did not interact, $F(1.58, 33.27) < 1$, $p = 0.655$, Mauchly's $W(2) = 0.738$, $p = 0.048$. These findings therefore indicate that neither the availability of precise information about the target nor the plausibility of object placement in the scene modulated the spatial accuracy of the landing points of saccades launched toward the target object. On average saccades undershot the target slightly, with a mean gain of 0.85 ($SD = 0.12$). That is first saccades toward the target object tended to cover approximately 85% of the distance from their launch site to the center of the ROI enclosing the target object.

Probability and latency of saccading toward the target region

The above measures do not fully address the question of how spatial expectations influence search initiation. In particular, they do not specify how the eyes are guided when participants rely initially on spatial expectations that overcome target appearance information in situation of conflict. Further insights in this respect are obtained by considering the probability and latency of directing the first saccade toward the location at which the target is expected to occur. Specifically, we compared the "baseline" nonconflicting condition of normal scene arrangement to the cases in which that location contains another object or no objects at all whilst the target is placed elsewhere.

The probability of saccading toward the location in which the target should occur was not influenced by the type of template, $F(1, 21) = 2.86$, $p = 0.106$. However, there was a main effect of scene arrangement, $F(2, 42) = 282.67$, $p < 0.001$, partial $\eta^2 = 0.93$, with a very large effect size, and a significant interaction between these two factors, $F(2, 42) = 14.56$, $p < 0.001$, partial $\eta^2 = 0.41$ (Figure 3 and Table 1), with a large effect size. Planned comparisons revealed that there was an effect of the scene arrangement when the target was indicated by either a picture or a word cue, all $ts(21) \geq 5.15$, all $ps < 0.001$. When the expected target location was occupied by a distractor (switched arrangement) it was more likely to be saccaded toward following a word cue than following a picture cue, $t(21) = 4.16$, $p < 0.001$. When the location in which the target was expected to occur was occupied by the target (normal arrangement) or was empty (target displaced arrangement) the type of search template did not influence the probability that this location would be saccaded toward, both $ts(21) \leq 2.29$, both $ps \geq 0.288$.

A complementary ANOVA was conducted on the direction of the first saccade with respect to the ROI that encompassed the entire region of the scene in which the target might be placed (see the section ROIs definition and data analyses). The pattern of results mirrored largely the one found when considering the object-based ROIs. The probability of saccading toward the region in which the target was expected was not influenced by the template type, $F(1, 21) < 1$, $p = 0.620$, but did differ across the three scene arrangements, $F(2, 42) = 145.58$, $p < 0.001$, partial $\eta^2 = 0.87$. Scene arrangement and template type interacted, $F(2, 42) = 3.35$, $p = 0.045$, partial $\eta^2 = 0.14$ (Table 1). Planned comparisons revealed an effect of the scene arrangement when the target was indicated by either a picture or a word cue, all $ts(21) \geq 3.19$, all $ps \leq 0.036$). However in this analysis there was no difference between the probability of saccading toward the target expected region following a word cue and that following a picture cue in any of the

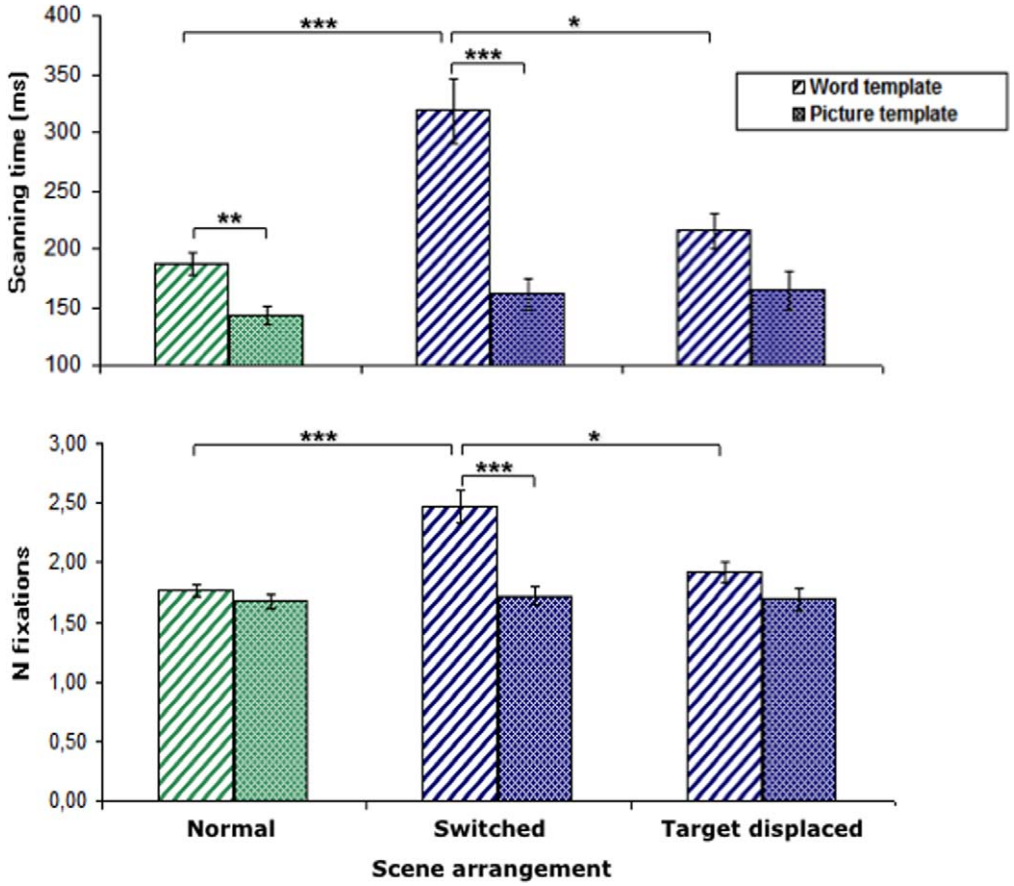

Figure 4. Scene scanning. Mean scanning time (top, in ms) and mean number of fixations until the first entry on the target object (bottom) as a function of template type and scene arrangement. Error bars indicate 1 *SEM.* ***: $p < 0.001$, **: $p < 0.01$, *: $p < 0.05$ at Bonferroni corrected pairwise comparisons. Green bars show cases when the target was in an expected location. Blue bars show cases in which the target was in an unexpected location.

three scene arrangements, all $ts(21) \leq 1.64$, all $ps \geq 0.999$.

When considering the latency of directing the first saccade toward the expected target location (Table 1), trials with the target displaced scene arrangement were excluded from analysis because there were too few cases in which the first saccade was launched toward the empty target plausible location. A weak tendency to significance, but with a relatively large effect size, was found for template type, $F(1, 21) = 3.47$, $p = 0.077$, partial $\eta^2 = 0.14$. Latencies tended to be shorter when the target was cued with a picture ($M = 192$ ms) than when it was cued with a verbal label ($M = 208$ ms). There was no main effect of scene arrangement, $F(1, 21) < 1$. There was a tendency to interaction, $F(1, 21) = 3.75$, $p = 0.066$, partial $\eta^2 = 0.15$. Despite the relatively large effect size, pairwise comparisons revealed no differences in latency depending on the type of the template or on the arrangement of the objects, all $ts(21) \leq 2.40$, all $ps \geq 0.234$.

Thus, the first saccade was rarely directed toward the expected target location, or the larger region in which the target might be expected, if this scene region was empty. However, when the expected target location was occupied by another object (but not the target) the probability of initially saccading toward this location depended upon the information supplied by the template. Fewer initial saccades were launched toward the expected target location when occupied by a distractor following a precise, pictorial cue than following an abstract word cue.

Scene scanning

Although our study was mainly focused on understanding how target information and spatial context information are used during the beginning of search to direct eye movements, we also considered how the visual system utilizes these two high-level sources of guidance during the subsequent search phases. We computed the scanning time and the mean number of fixations needed for locating the target during this second epoch of scene search (Figure 4 and Table 1).

These measures inform us of the time taken to locate the target and how this search process might be segmented into fixations.

Scanning time

There was a large main effect for both template type, $F(1, 21) = 43.24$, $p < 0.001$, partial $\eta^2 = 0.67$, and scene arrangement, $F(1.56, 32.74) = 9.30$, $p = 0.001$, partial $\eta^2 = 0.31$. Mauchly's $W(2) = 0.717$, $p = 0.036$. The two factors interacted, and the effect of the interaction had a large effect size: $F(2, 42) = 10.86$, $p < 0.001$, partial $\eta^2 = 0.34$ (Figure 4). Picture cues, compared to word cues, led to shorter scanning with a normal arrangement, $t(21) = -4.01$, $p = 0.009$, or with a switched scene arrangement, $t(21) = -5.69$, $p < 0.001$, but not when both the target object and the distractor object were placed in the highly plausible area for the distractor, $t(21) = 2.71$, $p = 0.117$. Moreover, in the case of a word cue, scanning was shorter either in the normal arrangement condition, $t(21) = -4.71$, $p < 0.001$, or in the target displaced arrangement condition, $t(21) = -3.22$, $p = 0.036$, than in the switched arrangement condition. No differences depending on the scene arrangement were found when the target was cued by a picture, all $ts(21) \leq 1.32$, all $ps \geq 0.999$.

Number of fixations

For the number of fixations needed to locate the target, the results followed to a large extent what was shown for scanning time. We found a large main effect for both template type, $F(1, 21) = 29.13$, $p < 0.001$, partial $\eta^2 = 0.58$, and scene arrangement, $F(2, 42) = 8.03$, $p = 0.001$, partial $\eta^2 = 0.28$. We also found a large two-way interaction, $F(2, 42) = 14.14$, $p < 0.001$, partial $\eta^2 = 0.40$ (Figure 4). The pattern of this interaction was the same as the one described for the scanning time, with the only exception that the difference due to the type of template was significant only with a switched scene arrangement, for which more fixations were needed to find the target when the object was cued by a word than when it was cued by a picture, $t(21) = 5.46$, $p < 0.001$. No differences due to the type of the template were found in the case of a normal or a target displaced arrangement (both $ts(21) \leq 2.50$, both $ps \geq 0.189$). In addition, the number of fixations during the scanning epoch was greater when the target and the distractor were switched than either when they were in their respective plausible locations, $t(21) = 5.36$, $p < 0.001$, or both were placed in the distractor high-probability region, $t(21) = 3.48$, $p = 0.018$. No difference was found between these two latter arrangements, $t(21) = 1.84$, $p = 0.720$.

Target verification

The last phase of search involves matching the currently inspected object with the target representation and, following sufficient positive evidence, accepting it as being the target object. We investigated whether having a specific representation of target features reduced the time needed to verify the target and also whether the plausibility of target location within scene context may affect target acceptance. It is worth to note that verification time always is a "mixed measure," as it also includes the time needed to plan and execute the manual response, once the decision upon the target has been made. However, it is reasonable to assume that this time component is constant across the experimental conditions; consequently, differences in verification time can be considered as reflecting genuinely the influence of the type of template or the scene arrangement.

An ANOVA showed that only template type had a large main effect, $F(1, 21) = 52.73$, $p < 0.001$, partial $\eta^2 = 0.71$, as verification time was shorter with picture ($M = 322$ ms) than with word ($M = 395$ ms) cues. A tendency to significance, with a medium effect size, was found for scene arrangement, $F(2, 42) = 2.84$, $p = 0.070$, partial $\eta^2 = 0.12$, for which planned comparisons showed that target verification tended to be quicker when the target object was in the plausible location ($M = 344$ ms) than when it was included in the same region than the distractor, $M = 367$ ms, $t(21) = -2.50$, $p = 0.063$, while no other tendency to significance was shown with other arrangements, both $ts(21) \leq 1.84$, both $ps \geq .237$. The interaction was not significant, $F(2, 42) = 1.09$, $p = 0.344$ (Figure 5 and Table 1).

Discussion

We investigated how knowledge about the target object and knowledge about where the target can be plausibly located are utilized to direct eye movements during search in real-world scenes. We focused in particular on the initial search epoch during the planning and execution of the first saccade, determining whether it is guided by both information sources simultaneously or preferentially by one source. We also analyzed the relationship between these two high-level sources of information across search, studying whether they interact or act independently, and whether this relationship varies across different phases of search (initiation, scene scanning, target verification).

Search initiation improved with a precise target template, following an exactly matching picture, compared to the case of a verbal (abstract) cue. It was also facilitated when the target was in an expected scene

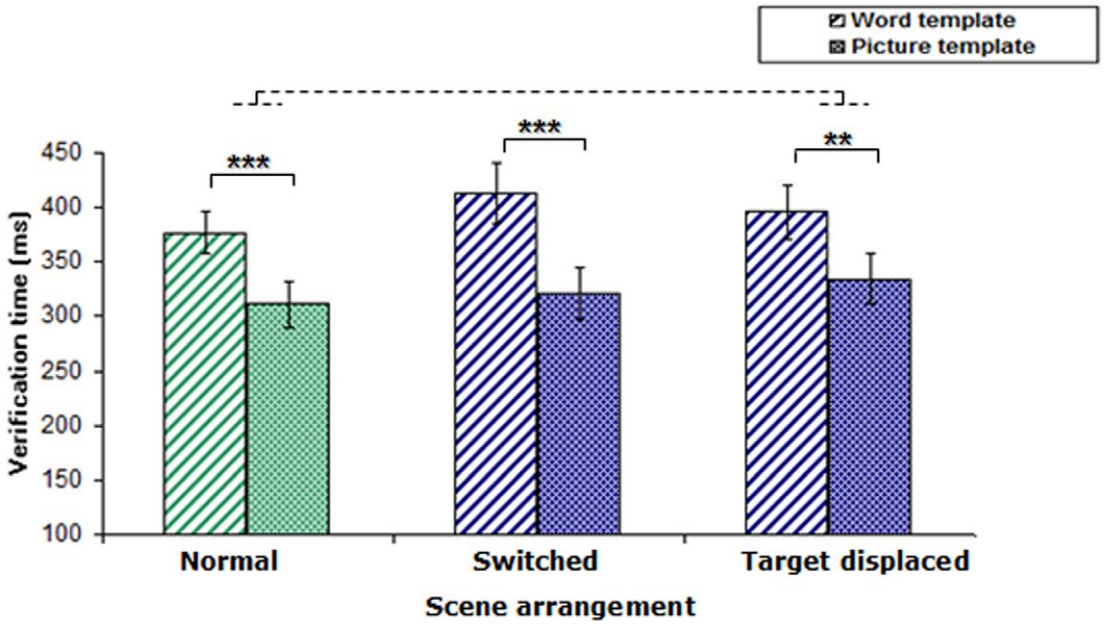

Figure 5. Target verification. Mean verification time (in ms) as a function of template type and scene arrangement. Error bars indicate 1 *SEM*. ***: $p < 0.001$, **: $p < 0.01$ at Bonferroni corrected pairwise comparisons. The dashed line indicates a tendency toward significance. Green bars show cases when the target was in an expected location. Blue bars show cases in which the target was in an unexpected location.

location compared to when it was in an unexpected location. These enhancements emerged in terms of a higher proportion of first saccades directed toward the target object, and not in terms of faster initiation. Thus the availability of information about the target appearance or its placement in the scene appears to influence the accuracy with which decisions to move the eyes are made, but not the time to make these decisions. Studies of search initiation are still rare and sometimes they have failed to report any effect of prior target information or scene context information (Hillstrom et al., 2012; Malcolm & Henderson, 2009, 2010; Võ & Henderson, 2009, 2011). However, previous findings seem to support the idea that first saccade direction may be a more sensitive measure than first saccade latency when studying target template guidance (Schmidt & Zelinsky, 2009, 2011) and spatial context guidance (Eckstein et al., 2006; Neider & Zelinsky, 2006).

Our findings allow us to specify further the conditions in which the type of target template and the reliability of spatial context guide the first saccade during real-world scene search. The results showed that information about the target object and information provided by the spatial context of the scene are integrated prior to initiating search (see also Eckstein et al., 2006; Ehinger et al., 2009; Kanan et al., 2009) and that the visual system can utilize both sources to constrain search from the beginning. We can also suggest that fixation selection is inherently object based. The fact that very few first saccades were directed toward an expected, but empty, location supports clearly an object-based account of saccade targeting and, by inference, of attentional selection during scene viewing (Egly et al., 1994). Thus, the visual system utilizes information provided by scene's context to direct the eyes toward plausibly-placed objects, not toward plausible regions per se. The early appearance of this effect provides evidence for rapid extrafoveal detection of object presence. Moreover, the fact that first saccades launched in the direction of the target object landed quite near the center of target's region of interest, regardless of the plausibility of the target's position within the scene or the specificity of the target representation, implies that once the object has been selected in peripheral vision, saccades are targeted with equal spatial precision. That is, the influences of spatial expectations and target template information are manifest in whether or not the target object is selected with the first saccade rather than how accurately the saccade reaches the target.

It should be noted, however, that the requirements of our task were explicitly to fixate the search target, and this instruction may have implications for the generalizability of our findings. By requiring participants to fixate the target we may have enforced suboptimal or unnatural viewing behavior. Najemnik and Geisler (e.g., 2005) demonstrated that viewers spontaneously adopt a nearly optimal strategy during search, selecting fixation placements that maximize information gathering about the target, and thus behaving very similarly to an ideal Bayesian observer. These locations may not

necessarily have the best matching with target features, but allow for optimization of information about target location. Our findings about target overt selection and foveation, therefore, should be considered carefully when it comes to generalize to searching in natural conditions. However, not fixating the target in visual search might be rewarding particularly when targets are placed in unpredictable and equally probable locations, as in Najemnik and Geisler's studies. When viewing natural scenes or exploring real-world setting, directly fixating an object that we are searching for is not an atypical behavior: In many behaviors we tend to bring the fovea to bear upon objects that we are searching for or using (see Ballard et al., 1992; Land & Tatler, 2009). Therefore, while our explicit instruction to foveate the target may have reduced the ecological validity of the findings, we do not feel that this imposes a behavior that is prohibitively unnatural. Whether or not fixating the target introduces some degree of unnaturalness to the task, our study crucially demonstrates that when required to do so, individuals may be highly effective and fast in directing the eyes to the target even in situations that are not characterized by the coupling of strong template guidance and strong contextual guidance.

We can use our findings to consider in what circumstances expectations about appearance and placement of objects facilitate search initiation. The availability of a specific search template facilitated initiation mainly when the target was in an unexpected region and a distractor was placed in an expected target location: a visual cue increased the probability of saccading toward the target object and reduced the probability of saccading toward the placeholder object. When only an abstract target representation was available, following a verbal cue, the same scene arrangement that put in conflict target template information and spatial expectations led to a similar proportion (around the 50%) of first saccades directed toward either the target or the distractor. This shows that both sources of guidance were utilized following a verbal cue and neither had a greater impact in winning the competition for attentional selection.

On the other hand, a plausible target position facilitated initiation mainly with an abstract target template, following a verbal cue. Observers tended to rely almost exclusively on local information when they had a precise target representation, with no significant difference in the probability of directing the first saccade toward the target object depending on where it was located. This means that knowing precisely what the target looks like may be sufficient to largely prevent interference due to unreliable spatial context. This result is somewhat surprising as it suggests that our previous experience with similar targets and similar scene contexts may be of marginal importance if precise information is available about the target's features.

Two main explanations can account for this pattern of results within an object-based framework of attention. Both involve a differential activation of two locations (one with the target, the other with the distractor) that becomes crucial in the case of conflicting high-level guidance. A first possibility is that the type of target template available influences the weighting of guidance sources before the scene appears. A real-world object representation is always likely to be, to some extent, an "object-in-context" representation, including object features together with memory of associations of that object with other co-occurring items and with typical contexts of occurrence in our experience (see Bar, 2004). We may speculate that when the template is visually detailed, the featural components of that representation may prime to a greater extent than the contextual components, leading to a relatively weaker influence of spatial expectations than in the case of an abstract target description. An abstract target description, conversely, could lead to the retrieval of a larger network of semantic knowledge linked to that target (see Kiefer & Pülvermüller, 2012), with a greater integration of short-term and long-term memory in the construction of the search template (Maxfield & Zelinsky, 2012; Schmidt & Zelinsky, 2009; Zelinsky, 2008). A stronger implication of the memory component of search following a verbal cue is also supported by the fact that in this case the observer has to look for any of the many possible items of interest that belong to the cued category. This may thus be considered as a form of hybrid task, involving both visual and memory searches (Wolfe, 2012).

An a priori source bias could also depend on a more active decision. When the information delivered by the target visual cue alone is enough to initiate search effectively, the visual system might actively reduce reliance upon expectations and contextual information. This would have the advantage of limiting the potential negative effects of any uncertainty due to a discrepancy between general semantic knowledge about objects in scenes and the specific episodic occurrence of the target in that given scene. The accessibility of template and context guidance from the start of search does not mean, necessarily, that both sources of information are always utilized to the same extent. If this criterion of usefulness is applied (see also Võ & Wolfe, 2012, 2013, for a discussion about distinguishing between availability and use of information in search), then in the case of a visually detailed template the activation of any location in the scene may depend essentially on its degree of matching with target features. Even though such activation could be potentially set to zero if none of the target features is matched, our results indicate that reliance on context or target features is likely to

follow a preferential bias along a continuum rather than an all-or-none mechanism.

A second alternative account of the outcome for first saccade direction does not include any former evaluation of usefulness, but posits that the visual system utilizes every source of available guidance in order to optimize oculomotor behavior. Consequently, all the decisions are taken online during search, depending on the combination between global information, delivered by scene context, and local information (Torralba et al., 2006) selected primarily according to matching with target appearance (see Ehinger et al., 2009; Kanan et al., 2009). When these sources are conflicting, each would lead to activation of a different location in the scene, so that saccade direction results finally from the online differential activation between the location that is implausible but contains the target and the location that would be plausible for the target but contains the distractor. In this situation, the precision of representation of target appearance following a picture template provides enough information to allow to saccade correctly toward the target in most of the cases. This results from greater activation at the target location due to more precise matching between information at this location and information represented from the target template. When the target has been described merely by its verbal label, information about its appearance is weaker and neither of the two competing locations primes clearly.

The present study does not allow us to distinguish between these two possible accounts. However, both accounts are consistent with a framework in which saccadic decisions derive from an object-based priority map (Nuthmann & Henderson, 2010) of the scene comprising (weighted) local information about objects and information about the likely placement of objects in scenes (Ehinger et al., 2009; Kanan et al., 2009). Our findings show clearly that target template guidance and scene guidance are coupled tightly in real-world image search. Moreover, context guidance never overrides target template guidance: In no cases were more initial saccades directed toward the target expected location when occupied by the distractor than toward the actual (implausible) location of the target. Future investigations will have to explore which specific properties of the target template are of key importance in guiding the eyes effectively, in particular when scene context is misleading.

While we interpret our findings in terms of what they may imply for how we search real world scenes it is important to consider the generalizability of our findings beyond the present study. Importantly, we created scenes with particular structure and content in order to test the relative reliance on target template and spatial expectations. These scenes are likely to be sparser than many real-world scenes that we encounter, and it is possible that the relative reliance on spatial expectations and target features may differ in more crowded scenes. We might predict that more crowded scenes make it harder to utilize target features effectively, due to disruptions to processes like figure/ground segregation and scene segmentation. This might therefore result in reduced overall search efficiency (Henderson, Chanceaux, & Smith, 2009; Neider & Zelinsky, 2011) and in greater reliance on spatial expectations in such scenes, especially during search initiation, when the short time available to viewers should render particularly challenging local information processing in crowded regions. On the other hand, we might expect that guidance provided by a picture cue would maintain much of its strength in more crowded scenes, without a shift of reliance to context information. With a visually precise template, matching with a single perceptual feature might be enough, in principle, to find the target even in absence of any explicit detection of objects. In support of this, recent evidence suggests that search for items perceptually defined by specific cues might be less affected by crowding. Asher, Tolhurst, Troscianko, and Gilchrist (2013) found overall weak correlations between a variety of measures of scene clutter and search performance, suggesting that this might arise from viewers searching for a specific scene portion, presented in a preview at the beginning of the trial. In this case searchers appeared to rely on target features equally, irrespective of scene clutter and, therefore, complexity. A greater search interference of clutter had been shown by Bravo and Farid (2008) utilizing one of the measures tested by Asher et al. and abstract target templates. It remains, therefore, uncertain how increasing scene complexity might influence the relative reliance upon target features and spatial expectations in the present study.

The pattern of results for search initiation is globally consistent with what we found during the next phase of search: scene scanning. Having the target object in an unexpected location and the distractor object in a location that would be plausible for the target led to longer scanning and more fixations before fixating the target. In contrast with search initiation, a visual template shortened scanning duration also when the target was plausibly placed. Previous research has shown that cueing the target with a precise picture (Castelhano & Heaven, 2010; Castelhano et al., 2008; Malcolm & Henderson, 2009, 2010; Schmidt & Zelinsky, 2009, 2011) and placing it in a consistent position (Castelhano & Heaven, 2011; Mack & Eckstein, 2011; Malcolm & Henderson, 2010; Neider & Zelinsky, 2006; Võ & Henderson, 2009, 2011) facilitated scene scanning. It is not clear why we found an interaction between expectations about target appearance and target placement while previous studies found

independent rather than interactive effects of these sources of guidance during scanning (Castelhano & Heaven, 2010; Malcolm & Henderson, 2010). It may be that the difference arises from the scenes we employed: Our scenes differed from those in previous studies by having two clearly differentiated regions and only two candidate target objects. In more cluttered scenes or scenes with less distinct regions to provide spatial guidance, other objects and scene-object relationships compete for attention. If the interaction between sources of guidance is subtle, the additional competitors for attention in more complex scenes might reduce the chance of detecting them.

Time to verify the target object once it has been fixated was affected significantly only by the type of prior information about the target. Quicker verification in the case of a visual cue than of a verbal cue shows that the target acceptance is easier when the representation of the target is visually precise (see also Castelhano & Heaven, 2010; Castelhano et al., 2008; Malcolm & Henderson, 2009, 2010). This is not surprising. However, the specificity of a verbal search target seems to influence the time it takes to verify the target once fixated: Basic category labels have been shown to be associated with faster verification than either subordinate or superordinate category labels. Both superordinate and subordinate labels, therefore, might require more processing to match the target with the template than basic category labels, but for opposite reasons: the need for constraining the type of characteristics to verify, when the information is generic; the need for checking for the numerous specific attributes that define the cued object, when the information is more specific (Maxfield & Zelinsky, 2012). Our verbal labels were predominantly basic category labels. Interestingly, a visual template shortened verification in all the object arrangement conditions, and its effect was only slightly larger when the positions of the target and the distractor were switched. Therefore, processes underlying verification appeared based essentially on feature matching, with at most only marginal consideration of the appropriateness of object position within the scene. Most previous studies have shown an effect of scene context knowledge on verification (Castelhano & Heaven, 2011; Henderson et al., 1999; Malcolm & Henderson, 2010; Neider & Zelinsky, 2006; Võ & Henderson, 2009, 2011), although some research did not find this influence (Castelhano & Heaven, 2010). In the present study, only a tendency to a quicker verification was obtained when comparing the normal arrangement condition with the case in which both objects were placed in a region plausible for the distractor.

There is another important aspect to take into account in interpreting our results and their generalizability to everyday situations. As in Malcolm and Henderson (2010), 75% of our stimulus set (including nonanalyzed extra scenes) had all objects placed at expected locations, in order to ensure that participants still considered scene context to be a reliable source of information. Nevertheless, the multiple occurrences of scenes with implausibly placed objects might have reduced the strength of context guidance, as participants might have relied less on their spatial expectations once they realized that targets sometimes could be in unexpected locations. Therefore, in everyday life misleading expectations might cause a greater reduction of search efficiency (see also Võ & Wolfe, 2013), even when viewers know the specific visual features of the target. However, it is worth noting that even when scene with contextual violations are more common—as common as 50% (e.g., Eckstein et al., 2006; Henderson et al., 1999; Underwood et al., 2008) or even 75% (e.g., Castelhano & Heaven, 2011; Võ & Henderson, 2009, 2011) of target-present trials—spatial expectations continue to play a role as search is still disrupted by such situations.

It is finally worth discussing whether this study may give some indications about the effect of object inconsistency on attentional allocation in scenes, which is a current matter of debate (see Spotorno, Tatler, & Faure, 2013). This study was not designed to consider effects of spatial inconsistency in scene viewing, but some suggestions may arise from what we found when only the target object was placed in an unexpected location while the distractor object was placed plausibly (i.e., with a target displaced arrangement). The target was initially saccaded to less in that case than when it was in an expected (consistent) location, and no significant differences were found between this arrangement and a "normal" object arrangement for search initiation time, scanning time, and number of fixations during scanning. Therefore, in this study no evidence of a extrafoveal detection of object inconsistency and an attentional engagement effect due to inconsistency was found. The tendency to a longer verification with a target displaced arrangement than with a normal arrangement might indicate that inconsistency processing leads to a longer involvement of attention once the object has been fixated. These findings are in agreement with several previous investigations (De Graef, Christiaens, d'Ydewalle, 1990; Gareze & Findlay, 2007; Henderson et al., 1999; Võ & Henderson, 2009, 2011).

Overall, we can conclude that our findings offer new insights into how we adapt oculomotor strategies in order to optimize the utilization of multiple sources of high-level guidance during search in naturalistic scenes. Even before we initiate the first saccade when searching a scene, information about the target's appearance and likely placement in the scene are being used to guide the eyes, maximizing the likelihood of initiating search

effectively. The fact that the specificity and reliability of these two sources of information does not influence first saccade latency suggests that these sources of information are extracted and used to set priorities for selection within the first 200 ms or so (the mean saccade latency in our experiment) of scene onset. The differences in accuracy of the first saccade direction suggest that the availability and reliability of information about the target's appearance and likely placement in the scene influence the weighting of local and spatial context information in setting priorities for fixation selection. This suggestion is consistent with recent framing of saccadic decisions as arising from priority maps that integrate information about object appearance and object placement (Eckstein et al., 2006; Ehinger et al., 2009; Kanan et al., 2009). Furthermore we can suggest that the priority map is likely to be an object-level description of the scene (Nuthmann & Henderson, 2010) because plausible regions that do not contain the target are only selected when occupied by a placeholder object. Prioritization depends on the reliability of scene context information and the specificity of prior target information. Priority weightings for the guidance sources appear to be dynamic. The balance between the use of context and the use of target template depends upon either an evaluation of usefulness before scene onset or an online competition between differentially co-activated object locations. However, having access to precise information about target appearance seems to supersede information about object placement in the scene. Thus if we have access to detailed information about the features of our search target, we can use this to find objects effectively even when they are not where we expect them to be.

Keywords: eye movements, visual search, target template, context information, spatial consistency

Acknowledgments

This research was funded by ESRC grant RES-000-22-4098 to BWT.

Commercial relationships: none.
Corresponding author: Sara Spotorno.
Email: s.spotorno@dundee.ac.uk.
Address: Active Vision Lab, School of Psychology, University of Dundee, Dundee, UK.

References

Asher, M. F., Tolhurst, D. J., Troscianko, T., & Gilchrist, I. D. (2013). Regional effects of clutter on human target detection performance. *Journal of Vision, 13*(5):25, 1–15, http://www.journalofvision.org/content/13/5/25, doi:10.1167/13.5.25. [PubMed] [Article]

Ballard, D. H., Hayhoe, M. M., Li, F., Whitehead, S. D., Frisby, J. P., Taylor, J. G., & Fisher, R. B. (1992). Hand eye coordination during sequential tasks. *Philosophical Transaction of the Royal Society B, 337,* 331–339.

Bar, M. (2004). Visual objects in context. *Nature Reviews: Neuroscience, 5,* 617–629.

Biederman, I. (1981). On the semantics of a glance at a scene. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual organization* (pp. 213–253). Hillsdale, NJ: Erlbaum.

Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: Detecting and judging objects undergoing relational violations. *Cognitive Psychology, 14,* 143–177.

Bravo, M. J., & Farid, H. (2008). A scale invariant measure of clutter. *Journal of Vision, 8*(1):23, 1–9, http://www.journalofvision.org/content/8/1/23, doi:10.1167/8.1.23. [PubMed] [Article]

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

Castelhano, M. S., & Heaven, C. (2010). The relative contribution of scene context and target features to visual search in real-world scenes. *Attention, Perception, & Psychophysics*, *72*(5), 1283–1297.

Castelhano, M. S., & Heaven, C. (2011). Scene context influences without scene gist: Eye movements guided by spatial associations in visual search. *Psychonomic Bulletin & Review*, *18*(5), 890–896.

Castelhano, M. S., & Henderson, J. M. (2007). Initial scene representations facilitate eye movement guidance in visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(4), 753–763.

Castelhano, M. S., Pollatsek, A., & Cave, K. R. (2008). Typicality aids search for an unspecified target, but only in identification and not in attentional guidance. *Psychonomic Bulletin & Review*, *15*(4), 795–801.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

De Graef, P., Christiaens, D., & d'Ydewalle, G. (1990). Perceptual effects of scene context on object identification. *Psychological Research, 52,* 317–329.

Eckstein, M. P., Drescher, B. A., & Shimozaki, S. S.

(2006). Attentional cues in real scenes, saccadic targeting, and Bayesian priors. *Psychological Science*, *17*(11), 973–980.

Egly, R., Driver, J., & Rafal, R. D. (1994). Shifting visual attention between objects and locations: Evidence from normal and parietal lesion subjects. *Journal of Experimental Psychology: General, 123,* 161–177.

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.

Eriksen, C. W., & Yeh, Y.-Y. (1985). Allocation of attention in the visual field. *Journal of Experimental Psychology: Human Perception and Performance, 11,* 583–597.

Findlay, J. M. (1997). Saccade target selection during visual search. *Vision Research*, *37*(5), 617–631.

Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology: General*, *141*(1), 2–18.

Gareze, L., & Findlay, J. M. (2007). Absence of scene context effects in object detection and eye gaze capture. In R. P. G. van Gompel, M. H. Fischer, W. S. Murray, & R. L. Hill (Eds.), *Eye movements: A window on mind and brain* (pp. 618–637). Amsterdam: Elsevier.

Greene, M. R., & Oliva, A. (2009). The briefest of glances: The time course of natural scene understanding. *Psychological Science, 20,* 464–472.

Henderson, J. M., Chanceaux, M., & Smith, T. J. (2009). The influence of clutter on real-world scene search: Evidence from search efficiency and eye movements. *Journal of Vision, 9*(1):32, 1–8, http://www.journalofvision.org/content/9/1/32, doi:10.1167/9.1.32. [PubMed] [Article]

Henderson, J. M., Weeks, P. A., & Hollingworth, A. (1999). The effect of semantic consistency on eye movements during complex scene viewing. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(1), 210–228.

Hillstrom, A., Scholey, H., Liversedge, S., & Benson, V. (2012). The effect of the first glimpse at a scene on eye movements during search. *Psychonomic Bulletin & Review, 19*(2), 204–210.

Hollingworth, A. (2009). Two forms of scene memory guide visual search: Memory for scene context and memory for the binding of target object to scene location. *Visual Cognition, 17,* 273–291.

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, *40*(10-12), 1489–1406.

Joubert, O., Rousselet, G., Fize, D., & Fabre-Thorpe, M. (2007). Processing scene context: Fast categorization and object interference. *Vision Research, 47,* 3286–3297.

Kanan, C., Tong, M. H., Zhang, L., & Cottrell, G. W. (2009). SUN: Top-down saliency using natural statistics. *Visual Cognition*, *17*(6/7), 979–1003.

Kiefer, M., & Pulvermüller, F. (2012). Conceptual representations in mind and brain: Theoretical developments, current evidence and future directions. *Cortex, 48,* 805–825.

Land, M. F., & Tatler, B. W. (2009). *Looking and acting: Vision and eye movements in natural behaviour*. Oxford, UK: Oxford University Press.

Mack, S. C., & Ekstein, M. P. (2011). Object co-occurrence serves as a contextual cue to guide and facilitate visual search in a natural viewing environment. *Journal of Vision, 11*(9):9, 1–16, http://www.journalofvision.org/content/11/9/9, doi:10.1167/11.9.9. [PubMed] [Article]

Mack, M. L., & Palmeri, T. J. (2010). Modeling categorization of scenes containing consistent versus inconsistent objects. *Journal of Vision, 10*(3):11, 1–11, http://www.journalofvision.org/content/10/3/11, doi:10.1167/10.3.11. [PubMed] [Article]

Malcolm, G. L., & Henderson, J. M. (2009). The effects of target template specificity on visual search in real-world scenes: Evidence from eye movements. *Journal of Vision, 9*(11):8, 1–13, http://www.journalofvision.org/content/9/11/8, doi:10.1167/9.11.8. [PubMed] [Article]

Malcolm, G. L., & Henderson, J. M. (2010). Combining top-down processes to guide eye movements during real-world scene search. *Journal of Vision, 10*(2):4, 1–11, http://www.journalofvision.org/content/10/2/4, doi:10.1167/10.2.4. [PubMed] [Article]

Maxfield, J. T., & Zelinsky, G. J. (2012). Searching through the hierarchy: How level of target categorization affects visual search. *Visual Cognition*, *20*(10), 1153–1163.

Najemnik, J., & Geisler, W. S. (2005). Optimal eye movement strategies in visual search. *Nature*, *434*(7031), 387–391.

Neider, M. B., & Zelinsky, G. J. (2006). Scene context guides eye movements during visual search. *Vision Research*, *46*(5), 614–621.

Neider, M. B., & Zelinsky, G. J. (2011). Cutting through the clutter: Searching for targets in evolving complex scenes. *Journal of Vision, 11*(14):

7, 1–16, http://www.journalofvision.org/content/11/14/7, doi:10.1167/11.14.7. [PubMed] [Article]

Nijboer, T. C. W., Kanai, R., de Haan, E. H. F., & van der Smagt, M. J. (2008). Recognising the forest, but not the trees: An effect of colour on scene perception and recognition. *Consciousness and Cognition*, *17*(3), 741–752.

Nuthmann, A., & Henderson, J. M. (2010). Object based attentional selection in scene viewing. *Journal of Vision, 10*(8):20, 1–19, http://www.journaofvision.org/content/10/8/20, doi:10.1167/10.8.20. [PubMed] [Article]

Oliva, A., & Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal in Computer Vision, 42,* 145–175.

Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory, 2,* 509–522.

Quattoni, A., & Torralba, A. (2009). Recognizing indoor scenes. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* 413–420.

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

Schmidt, J., & Zelinsky, G. J. (2009). Search guidance is proportional to the categorical specificity of a target cue. *Quarterly Journal of Experimental Psychology*, *62*(10), 1904–1914.

Schmidt, J., & Zelinsky, G. J. (2011). Visual search guidance is best after a short delay. *Vision Research, 51,* 535–545.

Scialfa, C. T., & Joffe, M. K. (1998). Response times and eye movements in feature and conjunction search as a function of target eccentricity. *Perception and Psychophysics, 60,* 1067–1082.

Spotorno, S., Tatler, B.W., & Faure, S. (2013). Semantic consistency versus perceptual salience in visual scenes: Findings from change detection. *Acta Psychologica*, *142*(2), 168–176.

Tatler, B. W., Hayhoe, M. M., Land, M. F., & Ballard, D. H. (2011). Eye guidance in natural vision: Reinterpreting salience. *Journal of Vision, 11*(5):5, 1–23, http://www.journalofvision.org/content/11/5/5, doi:10.1167/11.5.5. [PubMed] [Article]

Torralba, A., Henderson, J. M., Oliva, A., & Castelhano, M. (2006). Contextual guidance of eye movements and attention in real-world scenes: The role of global features on object search. *Psychological Review, 113,* 766–786.

Underwood, G., Templeman, E., Lamming, L., & Foulsham, T. (2008). Is attention necessary for object identification? Evidence from eye movements during the inspection of real-world scenes. *Consciousness and Cognition, 17,* 159–170.

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

Võ, M. L.-H., & Henderson, J. M. (2009). Does gravity matter? Effects of semantic and syntactic inconsistencies on the allocation of attention during scene perception. *Journal of Vision, 9*(3):24, 1–15, http://www.journalofvision.org/content/9/3/24, doi:10.1167/9.3.24. [PubMed] [Article]

Võ, M. L.-H., & Henderson, J. M. (2010). The time course of initial scene processing for eye movement guidance in natural scene search. *Journal of Vision, 10*(3):14, 1–13, http://www.journalofvision.org/content/10/3/14, doi:10.1167/10.3.14. [PubMed] [Article]

Võ, M. L.-H., & Henderson, J. M. (2011). Object-scene inconsistencies do not capture gaze: Evidence from the flash-preview moving-window paradigm. *Attention, Perception & Psychophysics, 73,* 1742–1753.

Võ, M. L.-H., & Schneider, W. X. (2010). A glimpse is not a glimpse: Differential processing of flashed scene previews leads to differential target search benefits. *Visual Cognition, 18*(2), 171–200.

Võ, M. L.-H., & Wolfe, J. M. (2012). When does repeated search in scenes involve memory? Looking at versus looking for objects in scenes. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(1), 23–41.

Võ, M. L.-H., & Wolfe, J. M. (2013). The interplay of episodic and semantic memory in guiding repeated search in scenes. *Cognition, 126,* 198–212.

Williams, D. E., & Reingold, E. M. (2001). Preattentive guidance of eye movements during triple conjunction search tasks: The effects of feature discriminability and stimulus eccentricity. *Psychonomic Bulletin and Review, 8,* 476–488.

Wolfe, J. M. (2012). Saved by a log: How do humans perform hybrid visual and memory search? *Psychological Science, 23,* 698–703.

Wolfe, J. M., Alvarez, G. A., Rosenholtz, R. E., & Kuzmova, Y. I. (2011). Visual search for arbitrary objects in real scenes. *Attention, Perception and Psychophysics, 73,* 1650–1671.

Wolfe, J. M., Horowitz, T. S., Kenner, N., Hyle, M., & Vasan, N. (2004). How fast can you change your mind? The speed of top-down guidance in visual search. *Vision Research, 44,* 1411–1426.

Wolfe, J. M., & Reynolds, J. H. (2008). Visual search. In A. I. Basbaum, A. Kaneko, G. M. Shepherd, & G. Westheimer (Eds.), *The senses: A comprehensive reference*. *Vision II* (Vol. 2, pp. 275–280). San Diego: Academic Press.

Wolfe, J. M., Võ, M. L.-H., Evans, K. K., & Greene, M. R. (2011). Visual search in scenes involves selective and non-selective pathways. *Trends in Cognitive Sciences*, *15*(2), 77–84.

Yang, H., & Zelinsky, G. J. (2009). Visual search is guided to categorically-defined targets. *Vision Research, 49,* 2095–2103.

Zelinsky, G. J. (2008). A theory of eye movements during target acquisition. *Psychological Review*, *115*(4), 787–835.

Zelinsky, G. J., & Schmidt, J. (2009). An effect of referential scene constraint on search implies scene segmentation. *Visual Cognition*, *17*(6), 1004–1028.