Available online at www.sciencedirect.com
ScienceDirect
Vision Research 46 (2006) 4118–4133
Vision Research
www.elsevier.com/locate/visres

Real-world visual search is dominated by top-down guidance

Xin Chen, Gregory J. Zelinsky *

Department of Psychology, Stony Brook University, Stony Brook, NY 11794-2500, USA

Received 3 May 2006; received in revised form 8 August 2006

Abstract

How do bottom-up and top-down guidance signals combine to guide search behavior? Observers searched for a target either with or without a preview (top-down manipulation) or a color singleton (bottom-up manipulation) among the display objects. With a preview, reaction times were faster and more initial eye movements were guided to the target; the singleton failed to attract initial saccades under these conditions. Only in the absence of a preview did subjects preferentially fixate the color singleton. We conclude that the search for realistic objects is guided primarily by top-down control. Implications for saliency map models of visual search are discussed.

Keywords: Top-down guidance; Bottom-up guidance; Real-world objects; Attention capture; Color singletons; Saliency maps; Eye movements; Overt visual search

1. Introduction

The visual search literature has long struggled with how best to characterize the contributions of top-down and bottom-up processes in guiding search behavior. Is search a primarily bottom-up process guided to regions of feature contrast in a scene, or is search guidance more usefully described as a top-down and task-specific process under the voluntary control of the searcher?

The suggestion that attention may be guided by bottom-up processes is deeply ingrained in our everyday introspections. Anyone who has looked at a busy city thoroughfare knows that attention seems drawn to the few colorfully dressed people mixed in with the rush-hour crowd wearing gray or black business suits (similar observations have been noted since Buswell, 1935). Instead of color, early experimental work focused on the role of pattern density and contrast in bottom-up guidance. Mackworth and Morandi (1967) showed that people prefer to inspect the regions of a scene having a high edge content, a finding later replicated by Mannan et al. (1996; see Henderson and Hollingworth, 1999, for a review). Evidence for contrast-based guidance in search dates back at least to Julesz (1981, 1986) and his formulation of *texton theory*, which states that local differences in the number of texture primitives or *textons* result in the near effortless guidance of search to these regions of feature discontinuity (but see Bergen & Adelson, 1983). A related phenomenon emerges when a visual pattern is globally distinctive in a scene. Whereas our attention may seem biased toward the handful of colorfully dressed people in a crowd, attention seems involuntarily captured by the solitary women wearing a red dress in the same crowd of drably attired business people. This attention capture effect was first noted in the search literature when a target, which was distinct from the distractors on at least one feature dimension, resulted in the parallel detection of the target and the phenomenal experience of pop-out (Treisman & Gelade, 1980). Similar capture effects have since been demonstrated for most globally distinct search items, not just the search target. Even when no target is designated and the task is simply to find the "odd item", featurally unique items or *singletons* pop-out of the display and into our awareness (Wolfe, Butcher, Lee, & Hyle, 2003).

Evidence for top-down guidance has likewise enjoyed a long history in the search literature. The earliest demonstrations of guidance appeared in the applied search community, where it was found that distractors sharing properties

* Corresponding author.
E-mail address: Gregory.Zelinsky@stonybrook.edu (G.J. Zelinsky).

0042-6989/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2006.08.008

with the target stood a higher likelihood of inspection during search (Luria & Strauss, 1975; Williams, 1967). Fueled by observations of highly efficient searches for conjunctively defined targets, Wolfe and colleagues (Wolfe, 1994; Wolfe, Cave, & Franzel, 1989) introduced *guided search theory* to formalize these guidance processes. According to this theory, preattentively available visual features, not yet bound by attention into objects, can be compared to high-level target descriptions to generate evidence for the target's location in a search display. Attention is then serially guided to those display locations indicating the greatest evidence for the target. As a result of this top-down guidance, targets and target-like objects would stand a higher probability of being visited by attention, resulting in increased search efficiency. The existence of top-down target guidance is now widely accepted in the search literature, with perhaps the most compelling evidence coming from studies of eye movements during search. Observers preferentially fixate targets, or distractors that share target features, regardless of whether these observers are humans (Findlay, 1997; Scialfa & Joffe, 1998; Williams & Reingold, 2001) or monkeys (Chelazzi, Duncan, Miller, & Desimone, 1998; Motter & Belky, 1998). However, this top-down guidance toward the target is not perfect (Zelinsky, 1996), leaving open the possibility that other top-down factors (e.g., scanning strategies) or purely bottom-up signals may also contribute to guiding search behavior.

With fairly incontrovertible evidence existing for both bottom-up and top-down processes affecting search, the focus of research has shifted to how these processes should be combined and their relative contributions to search guidance (Bacon & Egeth, 1994; for reviews, see Egeth & Yantis, 1997 & Yantis, 2000). A continuum of potential contributions have been considered, with some studies giving more weight to the bottom-up contribution and other studies preferentially weighting top-down control. A key debate in this literature is whether one process can override the other (e.g., Lamy & Egeth, 2003; Theeuwes, 2004). Studies advocating for a strong bottom-up weighting have argued that salient distractors can often override top-down control and capture spatial attention, even when these distractors are irrelevant to the ongoing task (e.g., Theeuwes, 1991, 1992, 1994, 2004; see also Posner, 1980). For example, Theeuwes (1992) asked observers to search for a green circle target in an array of green square non-targets. On some of the trials one of the squares would be replaced with a red distractor. Even though color was irrelevant to the search task, Theeuwes found that the appearance of this color singleton distractor reduced search efficiency, presumably due to the involuntary capture of attention. He concluded that attention may be guided by top-down factors, but only after bottom-up factors have run their course. When strong bottom-up signals are present, as would be the case with a color singleton distractor, top-down guidance signals are overwhelmed and attention is diverted first to the visually salient object. Researchers advocating for a stronger top-down weighting have a different interpretation of these capture effects (e.g., Folk, Leber, & Egeth, 2002; Folk & Remington, 1998; Folk, Remington, & Johnston, 1992; Folk, Remington, & Wright, 1994). Using a spatial cuing paradigm, Folk et al. (1992) found that singletons captured attention only when they shared a visual property with the target. Cues introducing an abrupt luminance change therefore interfered with task performance when the target was also defined by a luminance change, but interference arising from a task-irrelevant color singleton was successfully suppressed. These results are consistent with the use of "attention control settings" that can be tuned to allow only task-relevant features to capture attention. According to this idea, attention can be efficiently guided to a designated search target with the appropriate adoption of settings, or even captured by a singleton in an "odd man out" search task if a very broad control setting is designed to detect any salient visual discontinuity in a display (Wolfe et al., 2003). In the absence of any task-related top-down control, default settings might even be set reflecting attention biases and predispositions.

In the current study, we investigate the roles of top-down and bottom-up processes using more real-world search stimuli. There are good reasons to believe that visually complex objects may differ in their enlistment of top-down and bottom-up processes compared to the simpler types of stimuli typically used in the attention capture literature. One reason stems from the recent suggestion by Theeuwes (2004; also see Bacon and Egeth 1994) that display heterogeneity might differentially affect top-down and bottom-up guidance. Studies showing bottom-up guidance have typically used relatively homogenous arrays of non-target search items, whereas studies demonstrating top-down guidance have typically used more heterogeneous displays. By manipulating display heterogeneity, Theeuwes (2004) concluded that the presence of a strong saliency signal in a display determines the mode, feature or singleton, that observers will use in their search. When a display is heterogeneous, a strong bottom-up signal will not exist and observers will tend to adopt a feature search mode, thereby reducing the potential for singleton capture. However, when low-level processes reveal a strong saliency signal, observers will enter a singleton mode and their attention will be involuntarily drawn to the odd distractor, even when it is irrelevant to the search. Real-world objects offer the potential of pitting these two modes against each other in the same search task. Objects are visually complex and highly variable in their features, so a display consisting of real-world objects, even those presented without color (i.e., in grayscale), will have a high degree of heterogeneity. However, by adding color to one of these objects, a highly salient color singleton could be introduced to the search display. Under these conditions, will the presence of a color singleton prevail and cause observers to adopt a singleton search mode, or will the salient singleton be ignored, trumped by the heterogeneity of the display and the observer's top-down knowledge of the target?

Real-world objects are also visually and semantically rich, making them compelling search stimuli. Whereas the designation of a simple search target (e.g., a bar tilted 45°) makes minimal demands on working memory, guiding one's search towards a realistic object might require holding multiple visual features in memory. This greater demand on working memory might translate into a stronger, and more influential, top-down contribution to search guidance. Indeed, Chelazzi and colleagues (Chelazzi et al., 1998, Chelazzi, Miller, Duncan, & Desimone, 2001) trained monkeys to make an eye movement to a target object in a search array following exposure to a target preview. Recording from cells in areas IT and V4, they found that activity was initially undifferentiated in response to a target and non-target, but that attention soon modulated this activity such that only the object specified in the preview appeared to drive the cells. Chelazzi and colleagues interpreted this preview-defined "target effect" as evidence for a top-down process, perhaps originating in prefrontal cortex, that feeds back to areas along the visual pathway and biases the competition among search objects in favor of the target (also see Zelinsky, 1999; for another discussion of priming in relation to target previews). Target designation via a preview also removes a potential confound existing in some attention capture studies. For example, in van Zoest, Donk, and Theeuwes (2004), both the target (a right-leaning bar) and the distractor (a left-leaning bar) objects were featurally distinct from the non-targets (vertical bars), meaning that observers could have isolated both element types from the non-targets using a singleton search mode. Such an allocation of attention might have resulted in the underestimation of top-down guidance. In the case of real-world objects, all of the objects in the search display are comparable in their featural complexity and targets are designated using a preview. This problem of target distinctiveness is therefore minimized.

We examine the roles of top-down and bottom-up search guidance using pictures of common real-world objects as stimuli. As is typical in the attention capture literature, we had conditions in which top-down guidance was competing against bottom-up guidance, as well as conditions in which these two components were acting alone. We define bottom-up guidance in terms of color saturation; the distractor object appeared as a color singleton among the non-colored target and non-target search objects. We define top-down guidance in terms of the availability of a target preview, thereby tapping into a potentially rich source of memory-related guidance during search.[1] Evidence for guidance was assessed in terms of both manual reaction times (RTs) and saccadic eye movement behavior. To the extent that search is guided to objects by a bottom-up process, we expected that the majority of initial saccades would be directed to the color singletons and that RTs in a target identification task would be longer as a result of this attention capture. To the extent that search is guided by a top-down memory for the target, we expected faster RTs in the target preview condition and more initial saccades to the target object relative to the singleton distractor. By comparing the evidence for guidance when both top-down and bottom-up forms are available and one is pitted against the other, we can determine the relative contributions of each in guiding search behavior to real-world objects.

[1] Following Ullman (1984), we define a top-down process as any process that uses information that does not reside in the proximal stimulus, in our case the search display (see also Wolfe, Horowitz, Kenner, Hyle, & Vasan, 2004; for a related distinction).

2. Methods

2.1. Participants

Fourteen students from the Stony Brook University subject pool participated in the experiment for course credit. All reported normal or corrected to normal visual acuity and color vision.

2.2. Stimuli and apparatus

The stimuli were 2430 photorealistic images of objects selected from the Hemera® Photo-Objects collection. The objects varied in category, and an effort was made to minimize the categorical overlap between the objects appearing together on a given trial. No object was repeated throughout the experiment. Each search display consisted of nine objects appearing against a uniform white background. These objects were arranged on an imaginary circle such that the center of each object was equidistant (8.9° of visual angle) from the center of the display, which corresponded to the observer's initial gaze position (Fig. 1). Individual objects naturally varied in size, but all were scaled to fit snugly into a 2° × 2° (90 × 90 pixel) bounding box. Stimuli were displayed on a 19 in. ViewSonic CRT color monitor at a refresh rate of 100 Hz using a GeForce FX 5200 graphics card. The screen resolution was set at 1280 × 960 pixels, which described a 28° × 21° field of view. The gaze position of each observer was recorded throughout the experiment using an SR Research EyeLink® II eye tracker system. Eye position was sampled at 500 Hz in pupil-only mode. The observer's head was stabilized using a chinrest, which was located 72 cm from the computer monitor. Manual responses were collected using a Microsoft SideWinder gamepad which was connected to a USB port on the computer. Custom C++ software incorporating DirectX functionality and running under Microsoft Windows 2000 was used to control the experiment.

All objects, including the target preview, were displayed in grayscale, with the following exception. One object in the search display, which we will refer to as the *distractor object*, was presented in one of three color conditions. In the 100% color condition, the distractor object was displayed in full 24-bit RGB color, making it a salient color singleton in the context of the other eight grayscale objects in the display (Fig. 1C). In the 50% color condition, the color saturation of the distractor object was reduced by 50% (−50) using Adobe Photoshop v7.1, making the color singleton visibly less salient compared to the full color version (Fig. 1B). In the 0% color condition, the distractor object was presented in grayscale (Fig. 1A). All objects in the search display were therefore of comparable salience in the 0% color condition relative to the singleton conditions.

Superimposed over one of the search items in each display was a small "+" or "×" character which served as the target for the search task. The character was 0.25° × 0.25° and was constrained to appear within a 0.4° radius of the object's center. For the sake of clarity of exposition, the ×/+ character will be referred to as the *search target* and the search display object holding the ×/+ character will be referred to as the *target object*. Given this perfect spatial correlation between the search target and the target object, knowledge of the target object's identity and appearance might therefore be useful in locating the search target. If an observer knows from a target preview that the target object was a yellow rubber ducky, they could guide their search to this object and look for the embedded "+" or "×" character in order to make their identification decision. The target object was chosen at random from the eight non-distractor

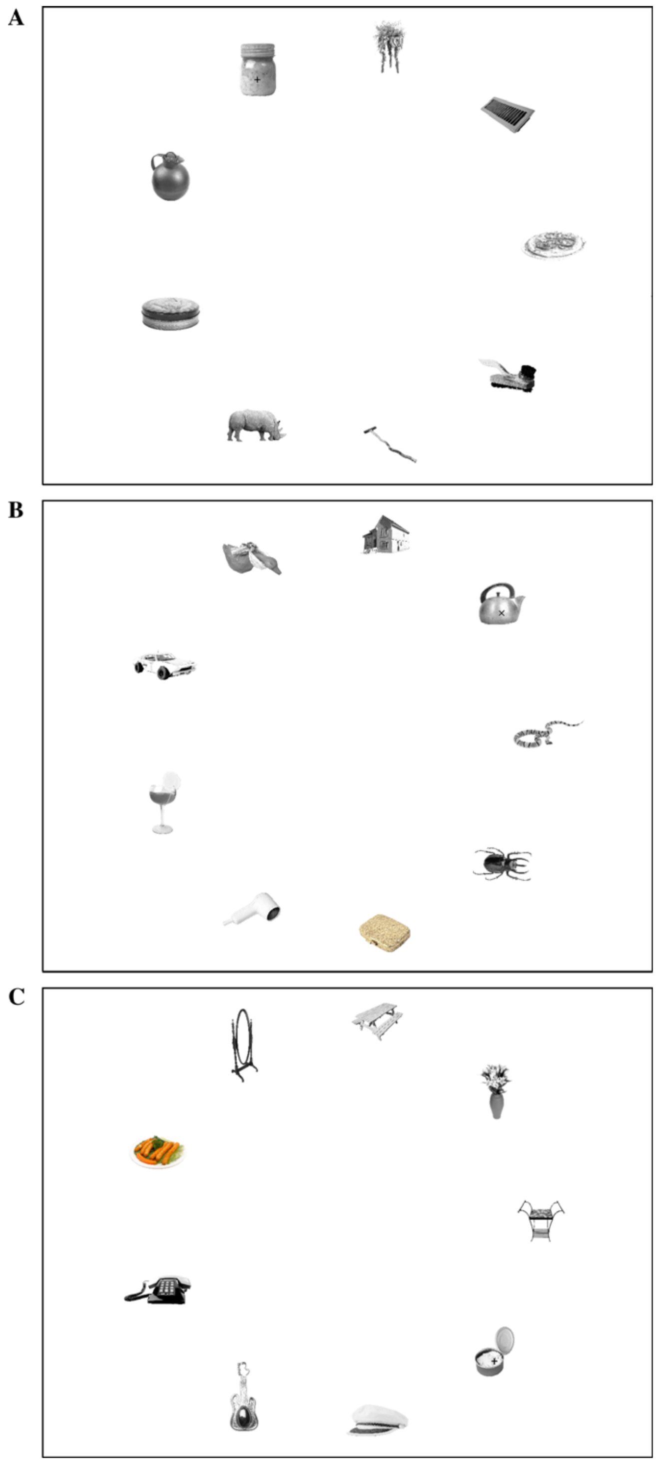

Fig. 1. Representative search displays used in the experiment, drawn to scale. Note the small + or × search target on one object in each display (e.g., the kettle in B). (A) 0% color saturation condition. All objects appeared in grayscale; no color singleton. (B) 50% color saturation condition. There was a color singleton distractor, but it was not vividly colored. (C) 100% color saturation condition. The display contained a vividly colored singleton distractor. (For interpretation of color mentioned in this figure the reader is referred to the web version of the article.)

objects in the search display. Following Yantis and Egeth (1999; also Theeuwes and Godijn 2001), the target object was prevented from coinciding with the distractor object so as to maintain a clear separation between bottom-up and top-down contributions to guidance. If search is captured by the color singleton distractor object, we can therefore be certain that this evidence for bottom-up guidance was not contaminated by a top-down signal. Likewise, if we find that search is guided to the target object, we can be similarly certain that this behavior was due to a top-down process and was not caused by a bottom-up direction of attention to the target.

2.3. Procedure and design

Prior to the experiment, the observer performed a 9-point calibration procedure needed to map eye position to screen coordinates. A calibration was considered "valid" if the maximum spatial error was less than 1° and the average error was less than 0.5°. This calibration procedure was supplemented by a drift correction procedure prior to the start of each trial. Each trial began with a fixation point displayed at the center of the screen. Observers were instructed to fixate this target carefully, then to press a button on the gamepad using their right thumb. This button press was used by the EyeLink® II to correct for any head movement that might degrade the system's estimate of eye position (Stampe, 1993), then by the display program to initiate the search trial. Depending on the experimental condition, the observer might then see a preview of the target object. There were three target preview conditions. In the long preview condition the target object was displayed at the center of the screen for 1 s; in the short preview condition the target object was similarly displayed for only 100 ms. There was also a no-preview condition in which the search display appeared immediately following a successful drift correction. Note that it was the existence of this no-preview condition that necessitated the use of the ×/+ target in this study; even in the absence of a target preview, the identification task remained well defined.

Upon presentation of the search display, observers were instructed to indicate, as quickly and as accurately as possible, the identity of the search target. They were to press the right trigger of the gamepad if the target was the "×" character, and they were to press the left trigger if the target was the "+" character. A search target was present on every trial. The search display remained visible for 2 s after the target judgment so that we might determine where in the display the observer looked immediately following target identification. Accuracy feedback was provided after every trial via a tone cue. A 2000 Hz tone signaled a correct response; a 500 Hz tone signaled an incorrect response. Search trials were self paced; observers could rest at any time during the experiment simply by delaying the start of the next trial. The entire experiment lasted approximately 1 h and was completed in one session.

The experiment was a 3 × 3 within-subjects design, with three levels of distractor object color saturation (100%, 50%, and 0%) and three levels of target preview duration (1000, 100, and 0 ms) randomly interleaved throughout the experiment. The 270 trials per observer were evenly divided between these conditions, leaving 30 trials per cell of the design. The "+" or "×" characters were used equally often as search targets, creating a 50% baseline level of accuracy for random response.

3. Results and discussion

To the extent that search is guided by bottom-up processes in this task, we would expect RTs and the percentage of initial saccades to the distractor object to increase with this object's color saturation. The underlying assumption is that a vividly colored object (100% saturation) will be more salient and therefore a better attractor of attention compared to a less vivid (50% saturation) or grayscale object (0% saturation). RTs should therefore increase with salience due to the more effective guidance of search towards the singleton distractor and away from the target. No effect of target preview on initial saccade direction is predicted. However, if search is predominately a top-down process, the availability of a target preview should affect both RTs and initial saccade direction. Specifically, RTs should decrease as preview duration increases from 0 to 1000 ms. Similarly, this creation of a strong top-down guidance signal should result in the percentage of initial saccades to the target increasing with preview duration.

3.1. Manual errors and reaction times

Observers incorrectly identified the search target on an average of 5.3% of the trials. Errors did not vary systematically with color saturation or target preview condition, nor did they depend on target type ("+" or "×"). These data were excluded from all further analysis.

Fig. 2A shows the mean RT data as a function of target preview and distractor saturation. It is clear from this figure that the time needed to identify the search target in this task depended only on the availability of a target preview. Our conclusion is supported by a two-way ANOVA combining the three levels of preview duration and distractor saturation. Only the main effect of preview duration proved significant, $F(2,26) = 412.56$, $p < 0.001$. This finding was qualified using pairwise *t*-tests with Bonferroni correction,

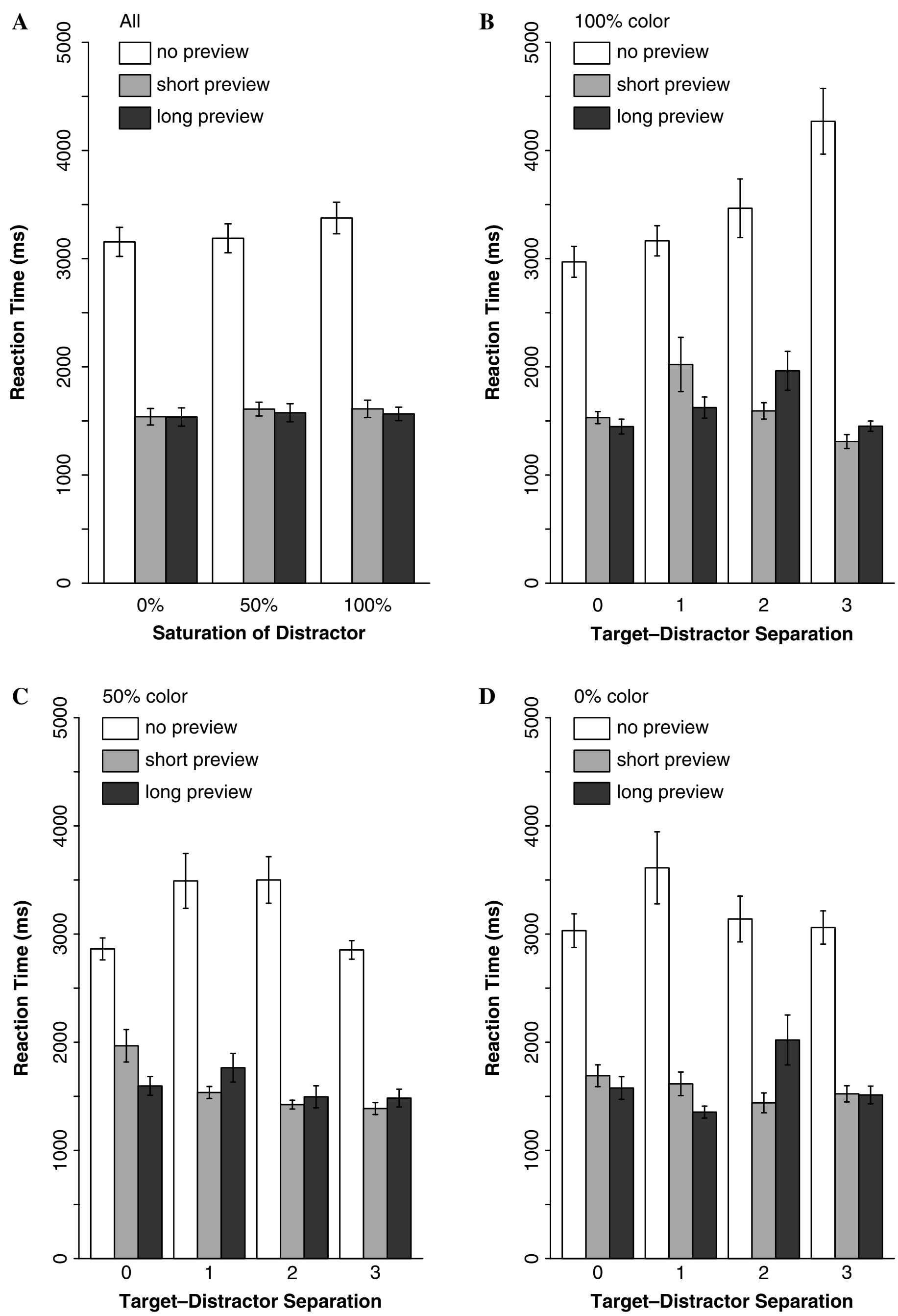

Fig. 2. Mean manual reaction time data. (A) RT data grouped by target preview condition and color saturation of the distractor. (B–D) Data from the 100%, 50%, and 0% color saturation conditions, grouped by target preview and plotted as a function of target-distractor separation in the display. Error bars indicate one standard error of the mean.

which revealed significant differences only between the no-preview condition and the two preview conditions, $t(1,13) \geqslant 20.18$, $p < .001$. There was no effect of distractor saturation, $t(1,13) = 1.15$, $p > 0.25$, nor was there a reliable preview × saturation interaction, $F(2,26) = 1.12$, $p = 0.36$.

In terms of our predictions, these data provide unambiguous support for guidance by a top-down process. When observers were allowed to see a preview of the target object, regardless of the specific duration of this preview, they used this information to speed their search. Relative to the no-preview condition, this top-down contribution was substantial, resulting in a roughly two-fold decrease in RT. No comparable effect of bottom-up guidance was observed in these data. Indeed, even in the no-preview condition when top-down target guidance was unavailable, the saliency of the distractor object failed to meaningfully affect search behavior.

The degree of bottom-up guidance in a difficult search task has been found to vary with the distance between the distractor object and the target object (Kim & Cave, 1999; Lamy, Tsal, & Egeth, 2003; Turatto, Galfano, Gardini, & Mascetti, 2004), and this may explain why a stronger bottom-up contribution was not observed in our data. A performance benefit resulting from a bottom-up guidance signal originating near to the target might be offset by a performance cost as a result of a bottom-up guidance signal originating far from the target. To address this possibility we analyzed our RT data as a function of target-distractor separation (Fig. 2B–D). The three panels segregate the data by distractor object color saturation; the x-axes in each panel indicate the number of non-target objects separating the target object from the distractor object in the circular display configuration. Data were analyzed using a 4 (separation) × 3 (target preview duration) × 3 (distractor saturation) repeated-measures ANOVA, which produced a significant three-way interaction between these factors, $F(12,156) = 7.03$, $p < 0.001$. Separate pairwise t-tests, with Bonferroni correction, revealed evidence for a relationship between preview duration, distractor saturation, and target-distractor separation only under full-color conditions (Fig. 2B). Specifically, RTs in the no-preview data were faster at target-distractor separations of 0 (2970 ms) and 1 (3165 ms) compared to a 3-object separation (4270 ms; $p < 0.01$ in both cases). The results from this target-distractor separation analysis qualify the data patterns shown in Fig. 2A. When a distractor object is vividly colored, and when top-down information is unavailable to guide search to the target, search is affected by bottom-up salience. Consistent with previous studies (Kim & Cave, 1999; Lamy et al., 2003; Turatto et al., 2004), RTs were likely slower for large target-distractor separations due to the color singleton pulling search away from the target object. However, the more prominent pattern emerging from these data is the clear dominance of top-down information on search performance. In those conditions where top-down and bottom-up guidance processes were pitted one against the other, the presence of a color singleton failed to meaningfully affect RTs in this search task.

3.2. Direction of initial saccades

Although the manual RT data clearly indicated an effect of top-down processing on search, these data did not perfectly conform to our predictions regarding the target preview manipulation. Our expectation was that the contribution of the top-down component should increase with the duration of the target preview, but this was not the case. There was a large difference between the no-preview and the preview conditions, but search times in the short and long preview conditions did not differ. One explanation for this failure to find an effect of preview duration may be that the top-down guidance process is extremely efficient, even with a very short preview. If this were the case, perhaps the manual RT dependent measure was simply not sensitive enough to discern these small differences in target guidance.

To address this possibility we analyzed the direction of the initial saccades as a function of object type (i.e., target object, distractor object, or non-target). Each circular 9-object search display was divided into nine equally sized (40°) pie-shaped sectors. Individual sectors were created by extending a line from the center of the display to the center of the bounding box enclosing an object, then demarcating a ±20° region relative to this line. Using this method, every initial saccade could be identified with a particular display object. Trials in which the initial saccade was less than 2° were excluded from this analysis (5.7% of the correct trials). We hypothesized that longer durations of the target preview would result in more initial saccades to the target, and that the number of initial saccades directed to previewed targets would be above chance and far outnumber those directed to the distractor objects.

Fig. 3 shows the results from this analysis, with the data grouped by target preview condition (no-preview, short preview, long preview) and distractor object saturation (0%, 50%, and 100%). The nine leftmost bars show the percentages of initial saccades directed to the distractor object in these conditions; the nine rightmost bars show the percentages of initial saccades directed to the target objects. The dashed line running horizontally through these data indicate the 11.1% level of guidance expected from a random direction of initial saccades to the nine search objects. Turning first to cases in which a target preview was available, we found a large and above-chance level of initial saccades guided to the target in both the short, mean $(M) = 20.7\%$, $t = 5.47$, $p < 0.001$, and long, $M = 16.1\%$, $t = 3.62$, $p = 0.003$, preview conditions, collapsing across distractor saturation condition. Further analysis revealed that these preview effects were significant at all three levels of distractor saturation, $t \geqslant 2.31$, $p \leqslant 0.04$. The percentages of initial saccades to the distractor object failed to differ from chance for any of the six combinations of short/long preview duration and distractor object saturation, $t \leqslant 1.52$, $p \geqslant 0.15$. Consistent with the manual RT data, we interpret these data as further evidence for a dominate role of top-down processes in guiding search behavior. However, and

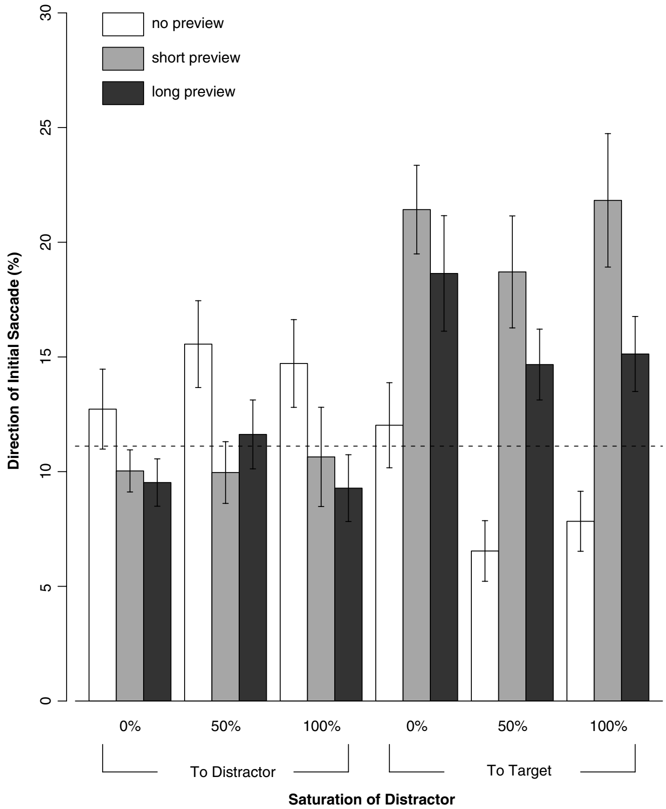

Fig. 3. Percentages of initial saccades directed to the target object (nine rightmost bars) and the distractor object (nine leftmost bars), grouped by preview condition and color saturation of the distractor. The dashed line indicates the level of saccade selectivity defined by chance. Error bars indicate one standard error of the mean.

contrary to our predictions, we did not find an increase in the percentage of initial saccades to the target with longer target preview durations. Indeed, we observed the opposite pattern; more initial saccades were directed to the target in the short preview condition compared to the long preview condition, $t \geqslant 2.44$, $p \leqslant 0.03$. We will defer until the next section our explanation for why this expected data pattern failed to emerge.

Analyses of the no-preview data revealed qualitatively different patterns. The proportion of initial saccades to the target was at chance when the distractor object was grayscale, $t = 0.50$, $p = 0.63$, but well below chance in the 50% saturation, $t = 3.45$, $p = 0.004$, and in the 100% saturation, $t = 2.50$, $p = 0.03$, conditions. Conversely, the percentages of initial saccades to the distractor object were significantly above chance in the 50% saturation condition, $M = 15.6\%$, $t = 2.36$, $p = 0.03$, and approached significance in the 100% saturation condition, $M = 14.7\%$, $t = 1.90$, $p = 0.08$. Initial saccades to the distractor object failed to differ from chance when the object was grayscale, $M = 12.7\%$, $t = 0.93$, $p = 0.37$. Based on this more direct analysis of search guidance, we can conclude that color singleton objects did capture gaze in this search task but at a reduced level relative to target guidance, and more importantly only in the absence of a target preview. When top-down and bottom-up guidance signals were put in competition, the top-down signals clearly dominated search guidance as evidenced by the greater percentages of initial saccades directed to the search target.

3.3. Initial saccade latency

Several questions remain unanswered from the previous analysis of initial saccade orienting. First, why were initial saccades directed to the target objects more frequently in the short preview condition compared to the long preview condition? Our working hypothesis is that the latency of these saccades may have been longer in the short preview condition, thereby explaining their greater accuracy. Second, and related to the first question, can a similar relationship between latency and accuracy describe initial saccades to the distractor objects in our task. Previous studies have shown that short latency initial saccades were more likely to be directed to singleton targets (van Zoest & Donk, 2006; van Zoest et al., 2004; see also Ludwig & Gilchrist, 2002). Although initial saccades to color singletons were relatively uncommon in our task, they did occur with an above-chance level of frequency in the absence of a target preview. If these initial saccades to distractor objects had a comparatively short latency, then this might indicate a more prominent role for bottom-up guidance at a very early stage in the search process.

To answer these questions, we analyzed the latency of the initial saccades in our task. Initial saccade latency was defined as the time between search display onset and the onset of the initial saccade, based on a 30°/s velocity criterion and a 9500°/s/s acceleration threshold. So as to keep this analysis consistent with the saccade direction analysis, we excluded initial saccades having amplitudes less than 2°. Given that a longer target preview might allow for a better match to the target object in the search display, we predicted that initial saccade latencies should decrease with longer target preview durations. We also predicted that longer latency initial saccades should be more accurately directed to the target object, again reflecting an increased opportunity for top-down processes to guide search. Finally, based on previous work we expected to find an inverse relationship between latency and initial saccades to color singleton distractors under no-preview conditions; the shorter the latency of the initial saccade, the more likely it should be to land on the distractor object.

Fig. 4 shows initial saccade latency plotted as a function of target preview condition and the color saturation of the distractor object. Initial saccade latencies in this task varied only with the duration of the target object preview. A 3×3 repeated-measures ANOVA confirmed this relationship, showing only a significant main effect of preview condition, $F(2,26) = 40.90$, $p < 0.001$. As predicted, latencies in the short preview condition ($M = 228$ ms) were longer than those in the long preview condition ($M = 173$ ms), $t = 7.85$, $p < 0.001$. The mystery of why more initial saccades were directed to the target object under short preview conditions

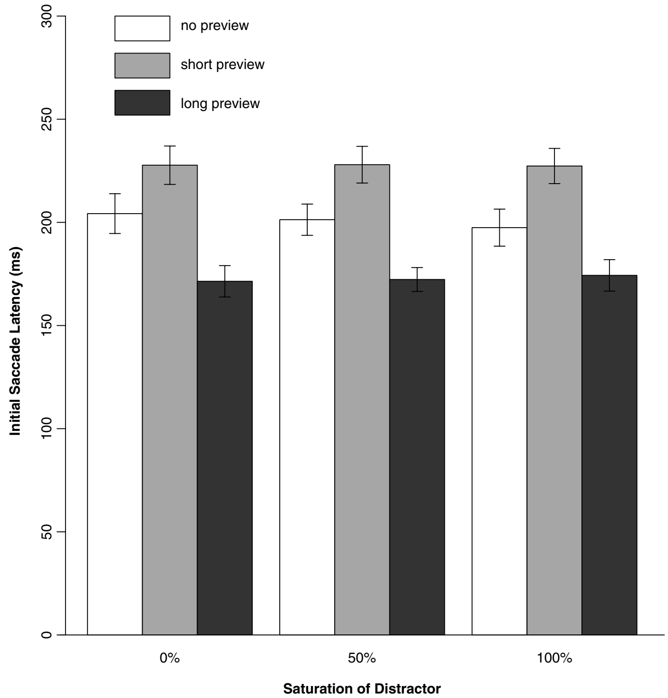

Fig. 4. Mean initial saccade latency, grouped by preview condition and color saturation of the distractor. Error bars indicate one standard error of the mean.

can therefore be explained in terms of a simple speed-accuracy tradeoff. With a long preview, observers would likely form a relatively complete description of the target object in their visual working memory. If they then trusted in top-down guidance to efficiently direct their gaze to the target, they might have settled into the habit of initiating short-latency saccades under long preview conditions. Of course, if this “go signal” came before accurate guidance information was available; a misdirected initial saccade would result. With a short target preview, observers would likely be far less confident in the quality of this top-down guidance signal. Indeed, the target’s description in visual working memory may not even be completely formed following a 100 ms target preview. Under these conditions, observers might choose to delay launching their initial saccade until the top-down guidance signal becomes available, a strategy that would result in more initial saccades accurately directed to the target. Consistent with this interpretation is the fact that intermediate initial saccade latencies were found in the no-preview data ($M = 201$ ms). These latencies were longer than those in the long preview condition ($p = 0.05$) because there was no top-down guidance signal to speed the selection of the saccade target; these latencies were marginally shorter than those in the short preview condition ($p = 0.07$) because the initial saccades were not being delayed so as to maximize the potential for top-down guidance.

Fig. 5 explores this relationship between initial saccade latency and direction in more detail. Initial saccade latencies were segregated into five bins and plotted as a function of the proportion of initial saccades directed to either the target object or the distractor object. The two panels show different breakdowns of the data. Fig. 5A shows data from guidance trials, meaning trials in which either a top-down or bottom-up guidance signal was available. Excluded from this analysis are data from the no-preview and 0% distractor saturation conditions. A clear pattern can be seen between initial saccade latency and the direction of the initial saccade. When latencies were very short (<200 ms), initial saccades to the target and distractor objects were equally likely, and neither targeting behavior differed from chance. However, clear trends emerged with longer initial saccade latencies (>200 ms); the proportion of initial saccades to the target object increased with saccade latency ($r = 0.32$), and the proportion of initial saccades to the distractor object decreased with saccade latency ($r = -0.37$). Both trends were statistically different from zero, $p < 0.01$. Our earlier conclusions must therefore be qualified; search can indeed use target preview

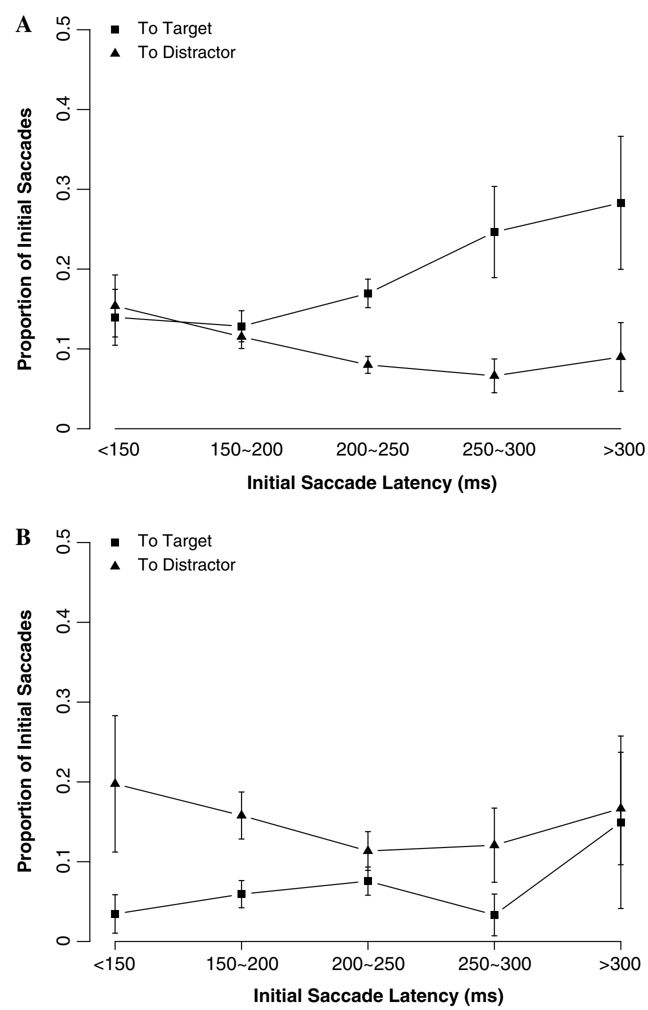

Fig. 5. Proportions of initial saccades to the target and distractor objects plotted as a function of initial saccade latency. (A) Data from trials in which there was a target preview (short or long) and a color singleton (50% or 100% saturation). (B) Data from color singleton trials (50% or 100% saturation) without a target preview (no-preview condition). Error bars indicate one standard error of the mean.

information when top-down and bottom-up guidance signals are placed in competition, but only if the initial saccade latencies are sufficiently long to allow this information to bias the targeting behavior.

What is the relationship between initial saccade latency and saccade direction in the absence of top-down guidance? Fig. 5B shows this analysis conducted on the no-preview data from the two color singleton conditions. Although the paucity of cases lends considerable noise to the analysis, there is the suggestion of inverse trends in the short latency data (<250 ms). Initial saccades to the distractor object are relatively common at very short latencies, but tend to decrease with longer latency saccades ($r = -0.25$; $p = 0.05$); initial saccades to the target object are rare (below chance) at very short latencies, but tend to increase in frequency with saccade latency ($r = 0.17$; n.s.). These analyses concur with similar relationships reported previously in the literature (e.g., van Zoest et al., 2004); purely bottom-up factors can capture our attention during search, with the influence of these processes being greatest very shortly after display onset. Our data contribute to this unfolding story by showing that this bottom-up influence is largely negated when top-down processes are factored into the mix. More specifically, when a target preview is available the relationship observed in the no-preview data between initial saccade latency and direction is reversed, with the proportion of initial saccades to the target now increasing with saccade latency and dramatically outnumbering those to the distractors.

3.4. Distribution of initial saccade endpoints

Although the analyses of initial saccade direction and latency did reveal some evidence for bottom-up guidance to color singletons, we expected to find larger singleton capture effects when the bottom-up component was not competing with target guidance. Why were initial saccades not directed more often to the color singleton in the absence of a target preview? Fig. 6 provides a partial answer to this question. Plotted are the initial saccade landing positions from no-preview trials in which a color singleton appeared in the display (i.e., same as in Fig. 5B). Fig. 6A–C shows these data for three individual observers; Fig. 6D shows a combined scatterplot for all observers. What is clear from these plots is that observers were biased to look initially to the upper-left quadrant of the display. This behavior is consistent with observer reports obtained during post-experiment debriefing. Because observers knew that the target and distractor objects would never coincide, on no-preview trials they frequently adopted a systematic strategy of serially searching neighboring objects on the circular display, with the starting position of this search often being an upper-left object. A similar upper-left search bias was reported in Zelinsky (1996) using a very different display configuration and stimulus set, and our suggestion of a relationship between eye movement biases and search guidance is broadly consistent with a study by Peterson, Kramer, Irwin, and Hahn (2002), which showed that observers demonstrating a systematic bias in their initial

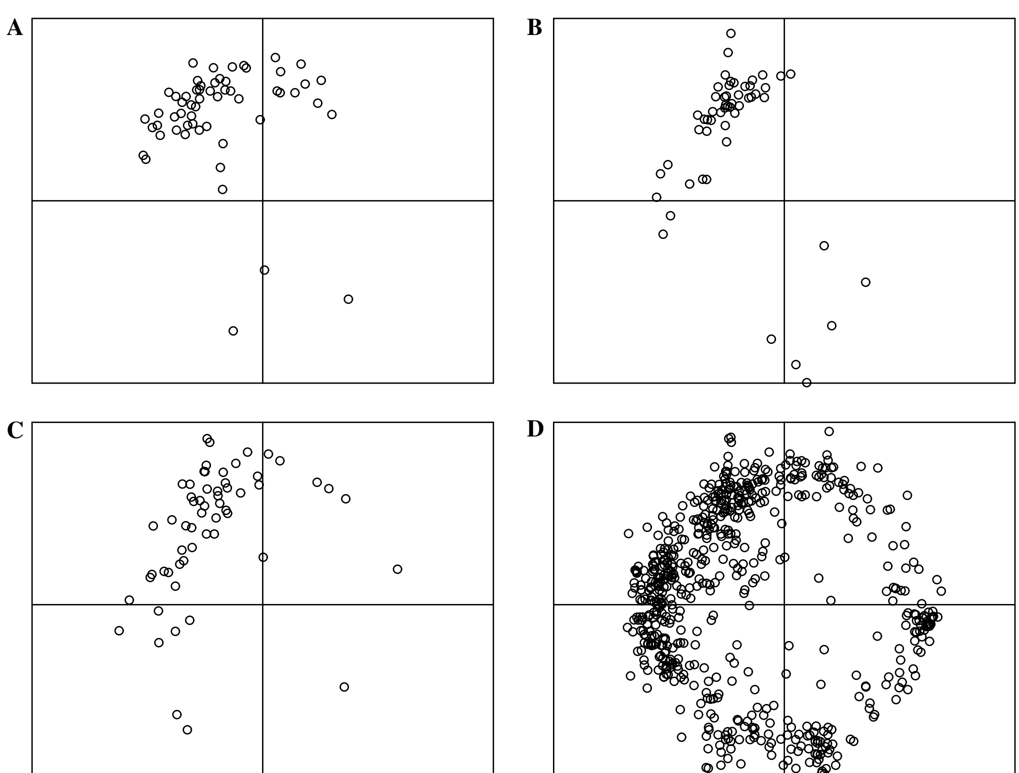

Fig. 6. Scatterplots of initial saccade landing positions for trials without a target preview (no-preview condition) but with a color singleton (50% or 100% saturation). (A–C) Representative data from three observers. (D) Data from all observers.

saccade direction were less likely to have their eyes captured by abrupt onsets.

The existence of this position bias has important implications for top-down and bottom-up search guidance. Even in the absence of a top-down guidance signal originating from the target preview, other top-down factors in the form of position biases and scanning strategies may affect search behavior. The fact that these target non-specific top-down factors were instrumental in directing gaze provides additional and converging evidence for the very limited role of bottom-up guidance in this task. Apparently, the bottom-up guidance signals were so weak that they could often be overridden by the observer's intention to start their systematic search in a particular display quadrant. However, observers were not always successful in completely ignoring the color singleton distractor. As indicated in Fig. 5, initial saccades were often captured by the distractor object when their latencies were very short, presumably because of an insufficient opportunity for top-down factors to redirect search.

3.5. Initial post-target fixations

So far we have discussed how successful top-down processes are in overriding competing bottom-up guidance signals, but does this weighting change once the target is found? One possibility is that top-down control signals are highly specific to the search task and very strong, ordinarily overwhelming the weaker but omnipresent bottom-up guidance signals that are less tied to the ongoing task. Under these assumptions, one might expect bottom-up signals to drive gaze to the salient color singleton once a search judgment is rendered and the top-down control is lifted. Another possibility is that there are multiple levels of top-down control in this search task, and that once the task is completed control simply passes to the next level. Under this scenario, observers might choose to look at semantically interesting non-targets following their search judgment, or perhaps they will simply move their gaze to the center of the display in anticipation of the next trial.

To explore these alternatives we analyzed the percentages of post-target saccades directed immediately to the distractor object. Recall that our paradigm had the search display remain visible to the observers for 2 s following their manual response, thereby allowing us to determine where gaze moved after leaving the target. Given the difficulty of the identification task, observers invariably fixated the target object prior to making their judgment. We define the initial post-target fixation as the first fixation made outside of the bounding region enclosing the target following the initial target object fixation. These fixations were assigned to distractors if they fell within a 3° radius of the distractor object's center, a distance corresponding roughly to half the center-to-center distance between two neighboring objects.

Fig. 7 plots the percentages of initial post-target fixations directed to the distractor object, as a function of distractor salience (0%, 50%, or 100% saturation) and

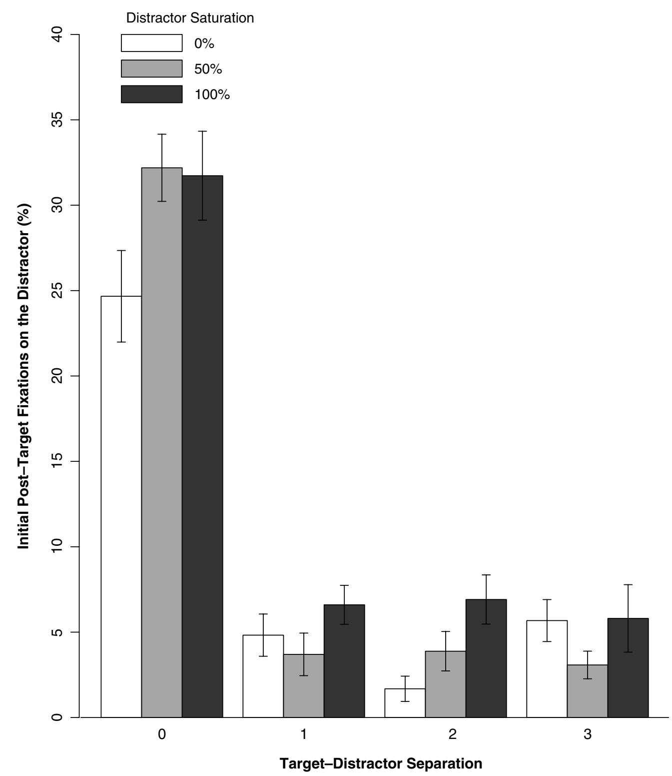

Fig. 7. Percentages of eye movements from the target object to the distractor object, plotted as a function of target-distractor separation. Error bars indicate one standard error of the mean.

separation from the target (0, 1, 2, or 3 intervening non-targets). Note that these data are shown collapsed over target preview condition after a three-way (preview × saturation × separation) ANOVA failed to reveal any reliable effect of this factor ($p > 0.05$). The main finding from this analysis is that observers often shifted their gaze to an object neighboring the target object following their search judgment. This tendency appeared as a highly significant main effect of target-distractor separation, $F(3, 39) = 165.21$, $p < 0.001$, even when the neighboring distractor was grayscale (~25%) and therefore unremarkable relative to the other non-targets in the display. However, this separation effect was limited to only the target's nearest neighbors in the display. Distractors that were not adjacent to the target object attracted very few initial post-target fixations. Importantly, there was also a main effect of distractor saturation, $F(2, 26) = 7.14$, $p = 0.003$, as well as an interaction between saturation and target-distractor separation that approached significance, $F(6, 78) = 2.07$, $p = 0.06$. When a color singleton distractor appeared next to the target object in the display, observers shifted their gaze away from the target and tended to look to this singleton more frequently than they would a grayscale object. This trend provides some support for the suggestion that top-down processes serve to stifle bottom-up guidance signals, and that these signals can exert more of an influence on behavior once this top-down control is lifted.

4. General discussion

Our goal in this study was to explore the relative contributions of top-down and bottom-up processes as they relate to oculomotor guidance in a search task using visually complex objects. Our findings can be summarized as follows:

(1) When top-down and bottom-up guidance signals are placed in competition, top-down guidance clearly prevails. This dominant role of top-down guidance was evidenced in both manual RTs and in the direction of initial saccades. Previous studies have argued for a similarly dominant role of top-down processes during search (e.g., Folk et al., 2002, 1992, 1994; see also Bacon & Egeth, 1994). Our study extends this earlier work by quantifying top-down guidance using visually complex and semantically rich objects as stimuli, and by pitting top-down guidance from a target-preview signal against a clear and opposing bottom-up signal. Except under no-preview and 0% saturation (grayscale) conditions, on each trial our observers were faced with a choice, follow the top-down guidance signal to the target object or follow the bottom-up guidance signal to the distractor object. Our data indicate that observers, when faced with this choice in our task, were successful in overriding the guidance signal originating from the color singleton distractor, thereby allowing their search to be efficiently directed to the target. Because the target was defined by a preview and did not systematically differ in salience from the non-target objects in the display, this evidence for top-down guidance could not have resulted from bottom-up processes. These results are broadly consistent with the *contingent orienting hypothesis* (Folk et al., 1992) and theories of visual search that assume the potential for task-relevant feature dimensions to be weighted by top-down processes prior to the onset of a search stimulus (Bacon & Egeth, 1994; Müller, Heller, & Ziegler, 1995; Pashler, 1988; Wolfe et al., 2003).

(2) The dominant role of top-down guidance does not depend critically on the strengths of the bottom-up or top-down guidance signals. Color saturation of the distractor object, our measure of bottom-up guidance, was manipulated over a substantial range, from 50% to 100%. Similarly, the strength of the top-down component was varied over a full order of magnitude in our task, from 100 to 1000 ms. Yet, despite these sizeable ranges of variation, neither manipulation meaningfully impacted the observer's reliance on top-down processes to guide their search. Manual RTs did not vary with the saliency of the distractor object, nor did they reliably differ between the short and long preview conditions. Likewise, when top-down guidance information was available, the percentages of initial saccades to the distractor object were at chance and did not depend on the specific level of saturation of the color singleton. Initial saccades to the target object did vary with preview duration, but this was likely due to a tradeoff between initial saccade latency and accuracy; if the initial saccades were of comparable latency, it is unclear whether differences in initial saccade accuracy would have been observed between the short and long preview conditions. These observations may place constraints on suggestions in the attention control literature that people are highly sensitive to top-down and bottom-up guidance signals and that relatively small differences in these signals might affect the mode of guidance used during search (Lamy & Egeth, 2003; Theeuwes, 2004; Yantis, 2000).

There are at least two explanations for why search in our task is relatively insensitive to bottom-up and top-down manipulations. One possibility is that our manipulations simply failed to produce meaningful changes in the top-down or bottom-up guidance signals. We consider this possibility unlikely. Although we would not expect the strength of a guidance signal to vary linearly with changes in distractor saturation and target preview duration, we certainly would expect that these effects would be discernable to the search guidance system given the wide ranges over which saturation and preview duration were varied in this experiment. Moreover, we have positive evidence for observers being sensitive to the preview duration manipulation in that they adjusted their initial saccade latencies depending on the duration of the preview. We therefore opt for a second explanation of our data; that observers were indeed sensitive to the top-down and bottom-up manipulations, but that the top-down guidance signal was so heavily weighted that any contribution from a bottom-up signal was negated. This explanation is again consistent with the *contingent orienting hypothesis* (Folk et al., 1992), which suggests that attention guidance is determined primarily by top-down control settings. As for why such an extreme top-down weighting was adopted in this experiment, we believe that there were two contributing factors. First, detection of the actual search target (the ×/+ character) would require an extremely effortful search in the absence of preview information regarding the target object. Observers were therefore highly motivated by the difficulty of the task to use top-down target preview information when it was available. Indeed, even when this information was incomplete (i.e., in the short preview condition), observers chose to nurture this top-down signal, as evidenced by their longer initial saccade latencies, rather than to allow their search to be guided by bottom-up salience. Second, observers were well aware that the search target would not be found on the distractor object. They would therefore be motivated to de-weight bottom-up information in the creation of a search guidance signal. Of course neither of these two practices is particularly surprising, and indeed they amount to perfectly rational behavior on the part of the observer. However, the adoption of this behavior implies that the observer has the capacity to control these attention settings, and that when a high top-down weighting is set, the appearance of a color singleton in a display has little or no impact on search. In this sense, our findings might inform the lively debate in the attention control literature regarding the flexibility of the attention control process and the inevitability of singleton capture effects (e.g., Folk et al., 2002, 1992, 1994; Theeuwes, 1991, 1992, 1994, 2004; see Egeth & Yantis, 1997 & Yantis, 2000; for reviews).

(3) When top-down information is unavailable, bottom-up signals exert more of an influence on search guidance. Although our data are unequivocal regarding the dominant role of top-down guidance in the reported search task, this does not mean that bottom-up signals were non-existent or completely ineffectual in guiding search behavior. Quite the contrary, we found clear evidence for singleton-induced attention capture in the absence of a target-related top-down guidance signal. Observers in the no-preview search trials responded faster to targets under full-color conditions when these targets were in close spatial proximity to the singleton distractor (see also Kim & Cave, 1999), and an above chance percentage of initial saccades were directed to the color singletons in the no-preview trials. Moreover, observers tended to shift their gaze from the target to a neighboring color singleton distractor after making their identification judgment. All of these findings are consistent with previous work arguing for a relationship between bottom-up saliency and attention capture effects during search (Baldassi & Burr, 2004; Nothdurft, 2002; Theeuwes, 1991, 1992, 1994, 2004). The comparatively small size of these bottom-up guidance effects, relative to top-down guidance, is likely due to the previously discussed bias against inspecting singletons in this task. Although target-specific control settings cannot be used in no-preview trials, search might still be biased away from the distractor object as a result of a de-weighting of this item in the guidance computation. We believe that the attenuated singleton capture effects observed in this study may be due in part to this de-weighting of the color feature dimension (Müller et al., 1995), as well as the contribution of a secondary top-down position bias, as evidenced in Fig. 6.

(4) The influence of top-down and bottom-up guidance signals depends in part on when these signals are used in the search task. Top-down guidance signals appeared to increase in strength and influence with the passage of time during search. This relationship exists in our data as a significant positive correlation between initial saccade latency and the proportion of initial saccades directed to the target. The longer observers delayed making their initial saccades, the more likely these eye movements were to land on the target objects (Fig. 5A). As expected, this positive relationship produced a concomitant negative correlation for color singleton distractors under target preview conditions; the proportion of initial saccades to singletons decreased with increasing saccade latency. Importantly, a similar relationship between saccade latency and initial saccades to the singleton distractor was found when target preview information was unavailable. Although the trend was not as clear, more initial saccades were directed to the color singleton under these no-preview conditions when saccade latencies were short. Taken together, these findings are broadly consistent with models of oculomotor control suggesting that a stimulus-driven saccade generation process is biased by top-down information that accumulates gradually over time (e.g., Godijn & Theeuwes, 2002; Trappenberg, Dorris, Munoz, & Klein, 2001; see also Nakayama & Mackeben, 1989). Our data are also largely consistent with recent work by van Zoest and colleagues (van Zoest et al., 2004; van Zoest & Donk, 2006; see also Ludwig & Gilchrist, 2002; Tatler, Baddeley, & Gilchrist, 2005) who showed that shorter latency saccades tend to be stimulus-driven and longer latency saccades tend to be goal-driven. However, our findings differ from this earlier work in that we observed less evidence for attention capture in the presence of top-down information. We attribute this minor discrepancy to our use of a target preview to manipulate the top-down guidance signal, as well as our use of a more visually and semantically complex stimulus set.

4.1. Implications for image-based search theories

Most theories of visual search explicitly acknowledge (e.g., Treisman & Sato, 1990; Wolfe, 1994) or tacitly assume (e.g., Duncan & Humphreys, 1989) the contribution of both bottom-up and top-down processes in guiding search behavior. In the case of Wolfe's (1994) *guided search theory*, these processes are even defined in terms of a computational model, thereby enabling one to quantify the relative contributions of top-down and bottom-up factors for a given search task. However, even the most descriptive of these theoretical frameworks are poorly equipped to address questions of guidance as they apply to realistic and visually complex objects, the search stimuli of interest in our study. For example, in his search simulations conducted using oriented color bar stimuli, Wolfe (1994) hand-picked orientation and color features to match the stimulus dimensions and used these to compute estimates of bottom-up feature contrast and top-down target guidance. The problem arises, however, in generalizing this approach to visual search tasks in which the relevant feature dimensions are not known. How does the feature contrast of a plate of eggs compare to that of a child's rag doll? What features should be used to guide your search to a favorite duck in a busy pond; color, orientation, shape, or some high-dimensional combination of all of these? Several recent models of search have developed representational frameworks that are sufficiently complex to allow these questions to be addressed. We will refer to these as image-based search theories to highlight the fact that they can accept as input arbitrarily complex images.[2] Search stimuli can therefore range in complexity from colored oriented bars to fully realistic scenes.

In addition to their ability to accommodate realistic objects, image-based theories are useful in that they typically provide a control signal for each point in the stimulus image that the search process might use for guidance. For example, in the Itti and Koch (2000, 2001) model,

[2] Note that this characterization of the representation as "image-based" does not mean that these models are unable to capture the importance of objects in a search task. We assume that an object-based representation can be constructed from an image, although this has not been the central focus of the theories under consideration here.

center-surround receptive field mechanisms are used to derive luminance, color, and orientation contrast difference signals at multiple spatial scales within a pyramid. Combining these feature contrast signals and plotting them for each point in the search image produces a topographic map of feature discontinuity, what these authors refer to as a *saliency map* (see also, Koch & Ullman, 1985; Parkhurst, Law, & Niebur, 2002; see Itti, 2005; for a review of more recent implementations of saliency maps, and Nothdurft, 2002; for a review of the behavioral evidence for local saliency affecting search). Winner-take-all competition is used to isolate the region of maximum saliency on this map, which then becomes the target for a saccade or shift of attention. Using this model, Itti and Koch were able to account for a key finding in the search literature; that feature singleton targets can be detected very quickly and independently of set size whereas less efficient search is typically observed for targets defined by a conjunction of features (Treisman & Gelade, 1980). This behavior of the model stems from the fact that singletons create a localized region of high feature contrast on the saliency map, thereby making them a strong attractor of attention. However, because the features of conjunction targets are distributed more uniformly throughout the search scene, signals arising from these features will tend to cancel each other, resulting in reduced target salience and a more difficult search.

With regard to our discussion of top-down and bottom-up processes in search, the Itti and Koch (2000) model is strictly bottom-up. Feature contrast signals are defined entirely within the search image; there is no provision for top-down guidance in this model. The absence of a top-down component in this model means that it will necessarily be unable to account for evidence of top-down guidance in the search literature (e.g., Findlay, 1997; Williams & Reingold, 2001; Motter & Belky, 1998). This limitation of the Itti and Koch (2000) model was first demonstrated by Turano and colleagues (Turano, Geruschat, & Baker, 2003) in a study designed to evaluate how well gaze behavior in a goal-directed navigation task can be predicted by several bottom-up and top-down models. They found that purely bottom-up approaches, as exemplified by the Itti and Koch (2000) model, provided a poor fit to the behavioral data. The best fit was obtained for a model that combined both coarse feature and geographic information (their *feature-geographic model*), with the latter source of information introducing a form of top-down control.

Itti and Koch (2000) acknowledged this limitation of their model and suggested that their approach may be most useful in describing shifts of attention and gaze that occur very early in search, before top-down factors can exert an influence. However, even this more modest claim is not easily reconciled with the data from our study. The left panels in Fig. 8 show the initial saccade landing positions from our

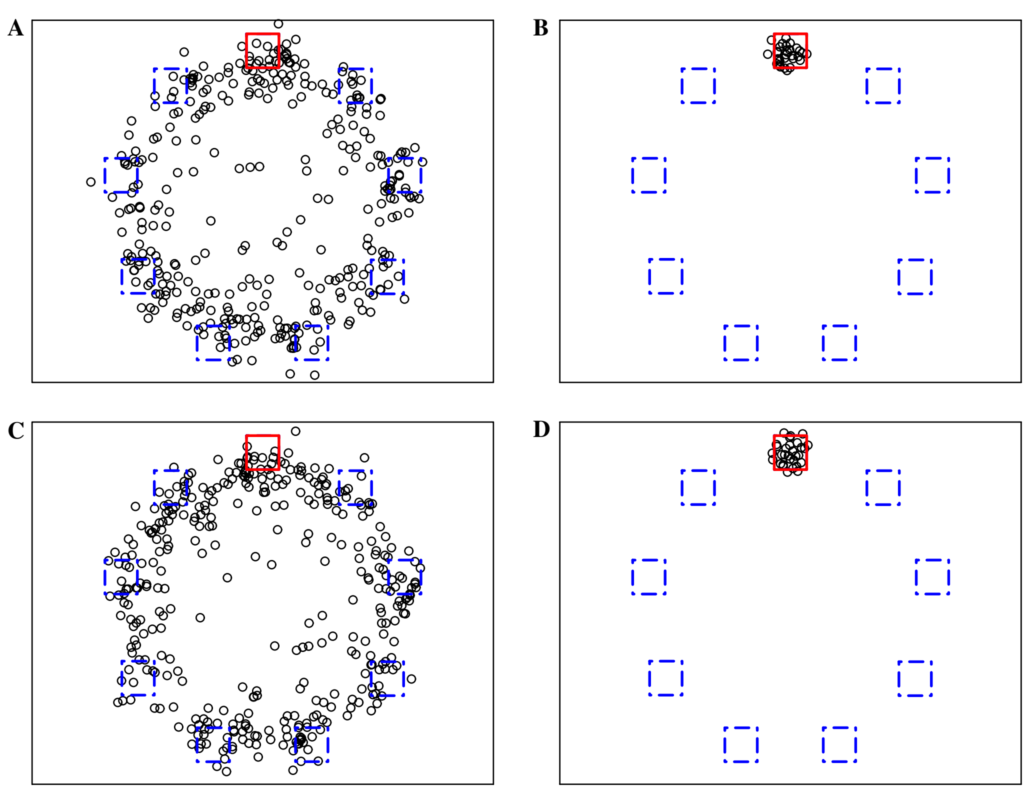

Fig. 8. Scatterplots of initial saccade landing positions from human observers and a model of bottom-up guidance in the absence of a target preview (no-preview condition). Note that the data were spatially transformed so that the distractor object always appeared in the 12o'clock display position (solid box). (A) Human data; 100% saturation condition. (B) Model data; 100% saturation condition. (C) Human data; 50% saturation condition. (D) Model data; 50% saturation condition.

14 observers in the absence of target preview information. Fig. 8A shows data from the 100% distractor saturation condition; Fig. 8C shows data from the 50% saturation condition. Note that the data from each trial was rotated so that the distractor object always appeared in the 12 o'clock display position. What is clear from this analysis is that, although observers did tend to direct their initial saccades to the color singletons under no-preview conditions (see Fig. 3), this tendency was certainly not pronounced. Fig. 8B and D show the corresponding analysis for simulated initial saccades generated by the Itti and Koch (2000) model. To obtain these data we screen captured the 30 no-preview search displays from the 50% distractor saturation condition and the 30 no-preview displays from the 100% saturation condition, input these 60 images to the model, then recorded for each the maximally salient image coordinate. As expected from a bottom-up model, the color singletons produced strong color contrast signals, causing the distractor object to attract, without exception, the initial saccade on every trial. Contrasting this behavior of the model with human behavior reveals an obvious and profound discrepancy. Although our observers did occasionally direct their initial saccades to the distractor object, suggesting that color singletons could successfully override a top-down bias against their fixation, this did not happen on every trial. Rather, initial saccades were dividing fairly evenly between the display objects even on the no-preview trials. Many initial saccades were even directed between two objects (Zelinsky, Rao, Hayhoe, & Ballard, 1997), meaning that an empty region of the display was often a more attractive saccade target than the color singleton distractor. Purely bottom-up models of guidance during search cannot explain the human search behavior reported in this study, not even in an analysis limited to only the initial eye movements.

The bottom-up image-based model of Itti and Koch (2000) can be usefully contrasted with the primarily top-down model by Rao, Zelinsky, Hayhoe, and Ballard (1996, 2002). As in the case of the Itti and Koch model, the Rao et al. model also decomposes a search image into a high-dimensional array of visual features. However, rather than computing difference signals between spatially neighboring features to generate a bottom-up guidance signal, these image features are compared to a search target to create a *top-down saliency map*.[3] Because activity on this map indicates evidence for the target, using this map to guide search results in the preferential direction of gaze to the target or target-similar objects in the search scene. In the context of the current study, this theory of overt visual search suggests that observers might extract a target feature vector from the target preview, then compare these features to those in the search image to generate a top-down guidance signal. This theory is supported by the current data in two key respects: first, in that we obtained strong evidence for target-related top-down search guidance, and second in that this guidance signal was mediated by exposure to a target-preview. However, because the Rao et al. model is primarily top-down, it would not be able to explain our evidence for bottom-up guidance on no-preview trials, nor would it explain why our observers tended to look at the singleton distractor after making their judgments.

The Itti and Koch (2000) and the Rao et al. (2002) models are both extreme in the sense that they focus exclusively on either bottom-up or top-down contributions to search guidance. More recent image-based models attempt to integrate these two sources of guidance into a single framework. This effort is nicely exemplified in a recent model by Navalpakkam and Itti (2005) in which bottom-up saliency is combined with information about the ongoing task or goal. Task constraints are introduced by user-supplied keywords interacting with knowledge in a long-term memory network. Importantly, this task-specific guidance works by biasing the visual features used in the bottom-up representation, essentially changing the saliency of a scene to reflect the goals of the task (see also Wolfe et al., 2003). This model is interesting in that it adopts a very broad and flexible definition of top-down information, extending well beyond visual search tasks.[4] It might also better describe the current data than either a purely bottom-up or top-down approach, particularly the shift to a singleton detection task that apparently occurs following a target judgment (Fig. 7). As for the dominant role played by top-down guidance in our task, Navalpakkam and Itti's model can account for this behavior by giving the bottom-up contribution a very small weight, thereby minimizing the effect of low-level salience on the search process.

A very different integrative approach was adopted in another recent image-based model by Zelinsky and colleagues (Zelinsky, Zhang, Yu, Chen, & Samaras, 2006). In this model, the relative contributions of top-down and bottom-up processes were systematically explored in the context of a realistic search task. Rather than biasing individual bottom-up features to reflect task demands, these authors computed separate bottom-up and top-down saliency maps, then combined the two in various mixtures to derive a guidance signal. The bottom-up model was similar in type to the Itti and Koch (2000) saliency model. The top-down model extended the Rao et al. (2002) model in a number of respects (see also Zelinsky, 2005), but still quantified top-down information in terms of a correlation between the features of a target vector and the features of a search scene. A range of mixture maps were created by combining the top-down and bottom-up components in different proportions. In comparing simulated guidance to human behavior at each of these mixtures, these authors found that only a pure top-down model could adequately describe the behavioral data. Specifically, the addition of

[3] Note that it would be incorrect to describe this model as entirely top-down as it still critically depends on the bottom-up extraction of visual features in the search image.

[4] Note however, that this breadth of focus comes with a price; task information must be manually inputted to the model.

even a 0.25 bottom-up component to the proportional mixture resulted in an excessive 36% miss rate and an unrealistically large number of eye movements to locate the target on trials in which the target was ultimately acquired. These authors concluded that, when specific target appearance information is available for top-down guidance, people largely ignore bottom-up feature contrast signals when searching realistic scenes. Our current data support this conclusion by Zelinsky et al. (2006) by showing no evidence for singleton capture on trials in which a target preview was available. Less clear is the more general implication of our findings for models combining top-down and bottom-up processes. Regardless of whether specific bottom-up features were de-weighted in favor of top-down features (e.g., Navalpakkam & Itti, 2005), or the entire bottom-up component was suppressed en masse (e.g., Zelinsky et al., 2006; see also Bacon & Egeth, 1994), our data suggest a very lopsided combination, one heavily weighted towards top-down guidance. Although there are undeniably situations in which more equitable combinations exist, the circumstances described by our reported task and stimulus set did not constitute one of these situations.

5. Conclusion

We conducted an experiment to determine the relative contributions of top-down and bottom-up processes in a search task using realistic objects as stimuli. We found that top-down guidance dominates bottom-up processes when the two sources of information are put in competition. Evidence for bottom-up guidance was also observed, but only in the absence of a target preview. We conclude that previewing a real-world target can result in an extremely potent form of top-down guidance, one that can largely overwhelm competing guidance signals. This relatively narrow focus on preview-related top-down guidance, however, also highlights an obvious weakness of our investigation. There are many dimensions along which we could have manipulated top-down and bottom-up guidance other than the preview and color saturation dimensions that we chose to use in this study. Had we picked a potentially stronger bottom-up cue (e.g., sudden onsets; see Theeuwes, Kramer, Hahn, Irwin, & Zelinsky, 1999 & Yantis & Jonides, 1990), or a weaker top-down cue (e.g., scene context; see Neider & Zelinsky, 2006), our results might have been very different. We certainly believe that under different testing conditions and with simpler stimuli, evidence for different combinations of top-down and bottom-up contributions to search guidance would emerge.

Acknowledgments

This work was supported by grants from the Army Research Office (DAAD19-03-1-0039) and the National Institute of Mental Health (R01 MH63748) to G. J. Z. We thank Mark Neider, Hyejin Yang, and Joseph Schmidt for their thoughtful comments throughout this project.

References

Bacon, W. F., & Egeth, H. E. (1994). Overriding stimulus-driven attentional capture. *Perception & Psychophysics, 55*(5), 485–496.

Baldassi, S., & Burr, D. C. (2004). "Pop-out" of targets modulated in luminance or colour: the effect of intrinsic and extrinsic uncertainty. *Vision Research, 44*, 1227–1233.

Bergen, J., & Adelson, E. (1983). Early vision and texture perception. *Nature, 333*, 363–364.

Buswell, G. T. (1935). *How people look at pictures*. Chicago: University of Chicago Press.

Chelazzi, L., Duncan, J., Miller, E. K., & Desimone, R. (1998). Responses of neurons in inferior temporal cortex during memory-guided visual search. *Journal of Neurophysiology, 80*, 2918–2940.

Chelazzi, L., Miller, E., Duncan, J., & Desimone, R. (2001). Responses of neurons in Macaque area V4 during memory-guided visual search. *Cerebral Cortex, 11*, 761–772.

Duncan, J., & Humphreys, G. (1989). Visual search and stimulus similarity. *Psychological Review, 96*, 433–458.

Egeth, H. E., & Yantis, S. (1997). Visual attention: control, representation, and time course. *Annual Review of Psychology, 48*, 269–297.

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

Folk, C. L., Leber, A. B., & Egeth, H. E. (2002). Made you blink! contingent attentional capture produces a spatial blink. *Perception & Psychophysics, 64*(5), 741–753.

Folk, C. L., & Remington, R. W. (1998). Selectivity in distraction by irrelevant feature singletons: evidence for two forms of attentional capture. *Journal of Experimental Psychology: Human Perception and Performance, 24*, 847–858.

Folk, C. L., Remington, R. W., & Johnston, J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology: Human Perception and Performance, 18*, 1030–1044.

Folk, C. L., Remington, R. W., & Wright, J. H. (1994). The structure of attentional control: contingent attentional capture by apparent motion, abrupt onset, and color. *Journal of Experimental Psychology: Human Perception and Performance, 20*, 317–329.

Godijn, R., & Theeuwes, J. (2002). Programming of endogenous and exogenous saccades: evidence for a competitive integration model. *Journal of Experimental Psychology: Human Perception and Performance, 28*, 1039–1054.

Henderson, J. M., & Hollingworth, A. (1999). High-level scene perception. *Annual Review of Psychology, 50*, 243–271.

Itti, L. (2005). Quantifying the contribution of low-level saliency to human eye movements in dynamic scenes. *Vision Cognition, 12*, 1093–1123.

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

Itti, L., & Koch, C. (2001). Computational modeling of visual attention. *Nature Reviews Neuroscience, 2*(3), 194–203.

Julesz, B. (1981). Textons, the elements of texture perception and their interactions. *Nature, 290*, 91–97.

Julesz, B. (1986). Texton gradients: the texton theory revisited. *Biological Cybernetics, 54*, 245–251.

Kim, M., & Cave, K. R. (1999). Top-down and bottom-up attentional control: on the nature of interference from a salient distractor. *Perception & Psychophysics, 61*, 1009–1023.

Koch, C., & Ullman, S. (1985). Shift in selective visual attention: towards the underlying neural circuitry. *Human Neurobiology, 4*, 219–227.

Lamy, D., & Egeth, H. E. (2003). Attentional capture in singleton-detection and feature-search modes. *Journal of Experimental Psychology: Human Perception and Performance, 29*(5), 1003–1020.

Lamy, D., Tsal, Y., & Egeth, H. E. (2003). Does a salient distractor capture attention early in processing? *Psychonomic Bulletin & Review, 10*, 621–629.

Ludwig, C. J. H., & Gilchrist, I. D. (2002). Stimulus-driven and goal-driven control over visual selection. *Journal of Experimental Psychology: Human Perception and Performance, 28*, 902–912.

Luria, S. M., & Strauss, M. S. (1975). Eye movements during search for coded and uncoded targets. *Perception & Psychophysics, 17*, 303–308.

Mackworth, N. H., & Morandi, A. J. (1967). The gaze selects informative details within pictures. *Perception & Psychophysics, 2*, 547–552.

Mannan, S., Ruddock, K., & Wooding, D. (1996). The relationship between the locations of spatial features and those of fixation made during visual examination of briefly presented images. *Spatial Vision, 10*, 165–188.

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

Müller, H. J., Heller, D., & Ziegler, J. (1995). Visual search for singleton feature targets within and across feature dimensions. *Perception & Psychophysics, 57*, 1–17.

Nakayama, K., & Mackeben, M. (1989). Sustained and transient components of focal visual attention. *Vision Research, 29*, 1631–1647.

Navalpakkam, V., & Itti, L. (2005). Modeling the influence of task on attention. *Vision Research, 45*, 205–231.

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

Nothdurft, H. C. (2002). Attention shifts to salient targets. *Vision Research, 42*, 1287–1306.

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

Pashler, H. (1988). Cross-dimensional interaction and texture segregation. *Perception & Psychophysics, 43*, 307–318.

Peterson, M. S., Kramer, A. F., Irwin, D. E., & Hahn, S. (2002). Modulation of oculomotor capture by abrupt onsets during attentionally demanding visual search. *Visual Cognition, 9*, 755–791.

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology, 32*, 3–25.

Rao, R., Zelinsky, G., Hayhoe, M., & Ballard, D. (1996). Modeling saccadic targeting in visual search. In D. Touretzky, M. Mozer, & M. Hasselmo (Eds.), *Advances in neural information processing systems 8* (pp. 830–836). Cambridge, MA: MIT Press.

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

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

Stampe, D. M. (1993). Heuristic filtering and reliable calibration methods for video-based pupil-tracking systems. *Behavioral Research Methods, Instruments and Computers, 25*(2), 137–142.

Tatler, B. W., Baddeley, R. J., & Gilchrist, I. D. (2005). Visual correlates of fixation selection: effects of scale and time. *Vision Research, 45*, 643–659.

Theeuwes, J. (1991). Exogenous and endogenous control of attention: the effect of visual onsets and offsets. *Perception & Psychophysics, 49*, 83–90.

Theeuwes, J. (1992). Perceptual selectivity for color and form. *Perception & Psychophysics, 51*, 599–606.

Theeuwes, J. (1994). Stimulus-driven capture and attentional set: selective search for color and visual abrupt onsets. *Journal of Experimental Psychology: Human Perception and Performance, 20*, 799–806.

Theeuwes, J. (2004). Top down search strategies cannot override attentional capture. *Psychonomic Bulletin & Review, 11*, 65–70.

Theeuwes, J., & Godijn, R. (2001). Attentional and oculomotor capture. In B. S. Gibson (Ed.), *Attraction, Distraction, and Action: Multiple Perspectives on Attentional Capture* (pp. 121–149). Elsevier Science.

Theeuwes, J., Kramer, A., Hahn, S., Irwin, D., & Zelinsky, G. (1999). Influence of attentional capture on oculomotor control. *Journal of Experimental Psychology: Human Perception and Performance, 25*, 1595–1608.

Trappenberg, T. P., Dorris, M. C., Munoz, D. P., & Klein, R. M. (2001). A model of saccade initiation based on the competitive integration of exogenous and endogenous signals in the superior colliculus. *Journal of Cognitive Neuroscience, 13*, 256–271.

Treisman, A. M., & Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology, 12*, 97–136.

Treisman, A. M., & Sato, S. (1990). Conjunction search revisited. *Journal of Experimental Psychology: Human Perception and Performance, 16*, 451–478.

Turano, K. A., Geruschat, D. R., & Baker, F. H. (2003). Oculomotor strategies for direction of gaze tested with a real-world activity. *Vision Research, 43*(3), 333–346.

Turatto, M., Galfano, G., Gardini, S., & Mascetti, G. G. (2004). Stimulus-driven attentional capture: an empirical comparison of display-size and distance methods. *Quarterly Journal of Experimental Psychology, 57A*, 297–324.

Ullman, S. (1984). Visual routines. *Cognition, 18*, 97–159.

van Zoest, W., & Donk, M. (2006). Saccadic target selection as a function of time. *Spatial Vision, 19*, 61–76.

van Zoest, W., Donk, M., & Theeuwes, J. (2004). The role of stimulus-driven and goal-driven control in saccadic visual selection. *Journal of Experimental Psychology: Human Perception and Performance, 30*(4), 746–759.

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

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

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

Wolfe, J. M., Butcher, S. J., Lee, C., & Hyle, M. (2003). Changing your mind: on the contributions of top-down and bottom-up guidance in visual search for feature singletons. *Journal of Experimental Psychology: Human Perception and Performance, 29*, 483–502.

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

Wolfe, J. M., Horowitz, T., 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.

Yantis, S. (2000). Goal directed and stimulus driven determinants of attentional control. In S. Monsell & J. Driver (Eds.), *Attention and Performance* (Vol 18, pp. 73–103). Cambridge: MIT Press.

Yantis, S., & Egeth, H. E. (1999). On the distinction between visual salience and stimulus-driven attentional capture. *Journal of Experimental Psychology: Human Perception and Performance, 25*, 661–676.

Yantis, S., & Jonides, J. (1990). Abrupt visual onsets and selective attention: voluntary versus automatic allocation. *Journal of Experimental Psychology: Human Perception and Performance, 16*, 121–134.

Zelinsky, G. J. (1996). Using eye saccades to assess the selectivity of search movements. *Vision Research, 36*, 2177–2187.

Zelinsky, G. J. (1999). Precueing target location in a variable set size "nonsearch" task: dissociating search-based and interference-based explanations for set size effects. *Journal of Experimental Psychology: Human Perception and Performance, 25*, 875–903.

Zelinsky, G. J. (2005). Specifying the components of attention in a visual search task. In L. Itti, G. Rees, & J. Tsotsos (Eds.), *Neurobiology of attention* (pp. 395–400). Elsevier.

Zelinsky, G. J., Rao, R. P. N., Hayhoe, M. M., & Ballard, D. H. (1997). Eye movements reveal the spatiotemporal dynamics of visual search. *Psychological Science, 8*, 448–453.

Zelinsky, G. J., Zhang, W., Yu, B., Chen, X., & Samaras, D. (2006). The role of top-down and bottom-up processes in guiding eye movements during visual search. In Y. Weiss, B. Scholkopf, & J. Platt (Eds.), *Advances in Neural Information Processing Systems 18* (pp. 1609–1616). Cambridge, MA: MIT Press.