Searching in dynamic displays: Effects of configural predictability and spatiotemporal continuity

George A. Alvarez
Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA

Talia Konkle
Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA

Aude Oliva
Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA

A visual search task was used to probe how well attention can operate over a dynamically changing visual display. Participants searched for a target item among an array of distractor items while the items either shifted location several times per second or remained stationary. Not surprisingly, Experiment 1 showed that shifting display items slowed search. However, search was faster if the shift preserved the global, *configural structure* of the display. The results of Experiment 2 suggest that the benefit of maintaining configural structure comes from improved spatial predictability: Knowing where the searchable items will be at any given moment enables faster search. Finally, Experiment 3 shows that, given *spatiotemporal continuity,* attention can operate just as efficiently over a dynamically changing display as it can over a stationary display. In the real world, objects often move, but they do so in a predictable way. The current findings suggest that the mechanisms underlying search can capitalize on configural predictability and spatiotemporal continuity to enable efficient search in such dynamic situations.

Keywords: visual search, dynamic displays, motion, global configuration, global context, configural predictability, spatiotemporal continuity, object persistence

Citation: Alvarez, G. A., Konkle, T., & Oliva, A. (2007). Searching in dynamic displays: Effects of configural predictability and spatiotemporal continuity. *Journal of Vision,* 7(14):12, 1–12, http://journalofvision.org/7/14/12/, doi:10.1167/7.14.12.

Introduction

During everyday visual perception, people spend a great deal of time searching for objects. The search process tends to be very efficient when salient or unique features distinguish the search target from surrounding distractor items (Egeth, Jonides, & Wall, 1972; Treisman & Gelade, 1980). Even when the target does not possess a unique and salient feature, combinations of features can be used to guide the deployment of attention toward the target item (for a review, see Wolfe & Horowitz, 2004). In the absence of reliable guiding features, the search process tends to be slow, as if requiring the allocation of attention to one item at a time, sequentially until the target item is located (Treisman & Gelade, 1980). Thus, whether looking for a friend in a crowd or trying to find an elusive link on a Web site, it may take several seconds to find the target item.

A variety of models have been proposed to explain the factors limiting visual search speed (e.g., Duncan & Humphreys, 1989; Kinchla, 1974; Palmer & McLean, 1995; Treisman & Gelade, 1980; Wolfe, Cave, & Franzel, 1989). These models are based mostly on laboratory studies employing visual search on stationary displays. In the classic search study, observers are required to find a target item among distractors, say the letter T among Ls. These items are typically displayed on a 2-dimensional computer display, and the items do not move during a trial, except when motion is investigated as a target defining visual feature (e.g., McLeod, Driver, Dienes, & Crisp, 1991; Royden, Wolfe, & Klempen, 2001). In the real world, however, our view of a scene changes continuously due to shifts in our point of view as well as changes in the layout of objects as they move around. It is likely that under these conditions, search can capitalize on prior knowledge and statistical regularities, as it does in static conditions (Oliva, Wolfe, & Arsenio, 2004; Torralba, Oliva, Catelhano, & Henderson, 2006).

In the current study, we examine the extent to which attention benefits from spatial and temporal regularities, under dynamic search conditions. Specifically, we modified the typical laboratory search task to investigate two potential factors which previous research suggest have an

doi: 10.1167/7.14.12 Received January 17, 2007; published December 13, 2007 ISSN 1534-7362

impact on the deployment of visual attention: *spatial configuration* and *spatiotemporal continuity*.

During visual search, attentional deployment appears to be sensitive to global, configural regularities learned from previous exposures. For example, visual search is faster when the target location is predicted by the global configuration of items (Chun, 2000; Chun & Jiang, 1998; Jiang, Song, & Rigas, 2005; Kunar, Flusberg, & Wolfe, 2006; Torralba et al., 2006). This work suggests that long-term memory and learning of contextual information can help guide attention early on during the search process. However, to our knowledge, there is no comparable research examining how we use global configural information while searching within a single trial, when there is no opportunity for long-term, implicit learning.

Research using a very different paradigm, the multiple object tracking task, has been used to explore limits on attentional selection in dynamic displays (Pylyshyn & Storm, 1988). For example, objects that implode or explode in an unnatural way are difficult to attentively track, whereas objects following the exact same trajectories but disappearing and reappearing with occlusion cues are trackable (Scholl, Pylyshyn, & Feldman, 2001). This work suggests that spatiotemporal continuity is necessary to track something as "the same" persisting object. However, whether spatiotemporal continuity and persisting objecthood are necessary to search efficiently in a dynamic display is an open question.

Combined, these previous studies suggest that consistent global configuration across trials can improve the speed of visual search in the long term, and that attention operates more efficiently when it can link a currently attended object with its history. In the current study, we investigated the extent to which these two factors, configural predictability and spatiotemporal continuity, influence the speed of visual search when displays are changing dynamically.

In three experiments, we used a visual search task to probe how well attention can operate over dynamically changing visual displays similar to those used by Horowitz and Wolfe (1998). While Horowitz and Wolfe focused on analysis of search slopes (reaction time × set size), we were interested both in search slope and in overall reaction time. In all experiments, participants searched for a target letter among an array of distractor letters while the items either shifted location several times per second or remained stationary. The results of Experiment 1 show that shifting display items slowed search, however, search was faster if the shift preserved the global, *configural structure* of the display. The results of Experiment 2 suggest that the benefit of maintaining configural structure comes from improved spatial predictability: Knowing where the searchable items will be at any given moment enables faster search. Finally, Experiment 3 shows that, given *spatiotemporal continuity,* attention can operate just as efficiently over a dynamically changing display as it can over a stationary display. Combined, these findings suggest that attention capitalizes on spatial and temporal regularities (configural predictability and spatiotemporal continuity) when searching in displays that change dynamically. In real-world conditions, the global layout of a scene tends to be stable, and when objects move they do so in a continuous and predictable way. Thus, these cues seem to be well suited for the visual system to rely upon when deploying attention and searching for objects in real world, dynamic situations.

General method

Participants

All participants were between the ages of 18 and 45, reported having normal or corrected-to-normal vision, gave informed consent, and were paid $10 per hour for their participation. There were 12 participants in each experiment (separate, but partially overlapping groups across experiments).

Apparatus

The experiments were written in MATLAB using the Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997) and were conducted on a PC computer with a 35° × 28° display, which was viewed without restraint from approximately 57 cm.

Stimuli

In each experiment, the target was a T (rotated from vertical either 0° or 180°) and the distractor set consisted of Ls (rotated randomly to 0°, 90°, 180°, or 270° clockwise from vertical). The total number of letters in the display ("set size") was 4, 6, or 8. Each letter subtended 1° × 1° and was drawn in black (1.8 cd/m2) on a gray background (34.5 cd/m^2).

Dynamic search displays were created by cycling between two sets of item positions (frame 1, mask, frame 2, mask, repeat) until the participant responded (see Figure 1). The frame time was 100 ms and the mask time was 300 ms. When the target was present, it was present on both frame 1 and frame 2 and in the same orientation within a trial. There were three conditions: *stationary, random shift,* and *translational shift*. In each condition, items were initially plotted in a random position within a 16° × 16° region, with a minimum spacing of 1.5° between items. In the *random shift* condition, new positions were selected for the second frame by moving each item 3.75°, with the shift direction

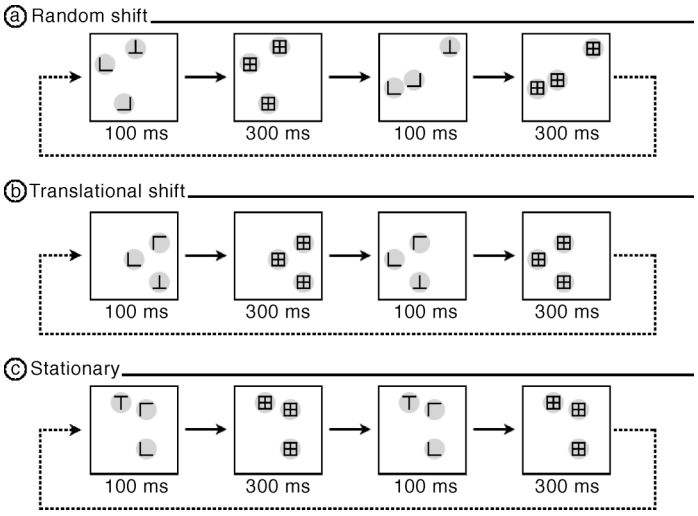

Figure 1. Displays for Experiment 1. The search display consisted of four frames that were repeatedly displayed in a loop until subjects made a response. (a) In the random shift condition, the target locations changed by a fixed distance, where the direction of each target shift was random. (b) In the translational shift condition, the locations of the target changed by the same fixed distance, but the direction of each target shift was the same, resulting in a global translation. (c) In the stationary condition, the target locations remained the same throughout the trial.

chosen independently and randomly for each item (Figure 1a). In the *translational shift* condition, new positions were selected by moving each item 3.75°, with a randomly chosen shift direction that was similar for each item, jittered by ±5° from the base direction (Figure 1b). In the *stationary* condition, the item positions were the same for frame 1 and frame 2 (Figure 1c). A total of 720 displays of initial item positions were generated (frame 1 s), and for each of these displays, a random shift, translation, or stationary display of positions was also generated (frame 2 s). This set of displays was used for each experiment.

Procedure

At the beginning of each trial, participants pressed a key and a display of 4, 6, or 8 items was immediately presented. The task was to find the letter T and indicate its orientation as quickly and as accurately as possible. Each participant completed 720 trials (240 each in the random shift, translational shift, and stationary condition) with the order of conditions randomly mixed.

Data analysis

In each experiment, error rates were low (typically less than 10% collapsed across participants and conditions), and there was no evidence for speed accuracy tradeoffs across conditions (errors were positively correlated with reaction time). Therefore, analyses focused on reaction time for correct trials, but Appendix A provides a more detailed analysis of error rates for interested readers.

For all effects, we report a standard measure effect size, partial eta squared (η_p^2), which can be interpreted as the proportion of variance accounted for by the factor or interaction between factors (Cohen, 1973). This measure estimates the contribution of each factor or interaction as if it were the only factor. Thus, it is possible for the sum of the η_p^2 values for an analysis to be greater than 1.

Error bars in each figure represent the 95% inferential confidence intervals (Tryon, 2001), which graphically depict whether means are reliably different form each other. If these confidence intervals do not overlap, the difference between means is significant at the .05 level.

Experiment 1: Role of global configuration in dynamic search

The current study investigates whether preserving the global structure of a display will enable observers to search more efficiently in a dynamically changing display. There were three conditions that varied in the extent to which they preserved individual item position and the global configuration of the display. In the *stationary condition,* the items remained in the same positions throughout the trial, preserving both individual item position and global configuration. In the *random shift condition,* both individual position and global configuration were disrupted. Finally, in the *translational shift condition,* individual item position was disrupted by the same distance as in the random condition, but the global configuration of items was preserved.

If visual search is sensitive to changes in individual item position, then overall reaction time to find the target should be slower in both conditions where the items are relocating (random and translational conditions) than when the items remain in place (stationary condition). Such an effect would be expected based on previous experiments employing a similar dynamic search paradigm (Horowitz & Wolfe, 1998). Second, to the extent that there is a benefit to maintaining the *global configuration* of items, reaction time should be faster for the translation condition than the random condition.

Results

As shown in Figure 2, reaction time is slower for the both shift conditions (random and translational) than in the stationary condition. Moreover, reaction time is slower

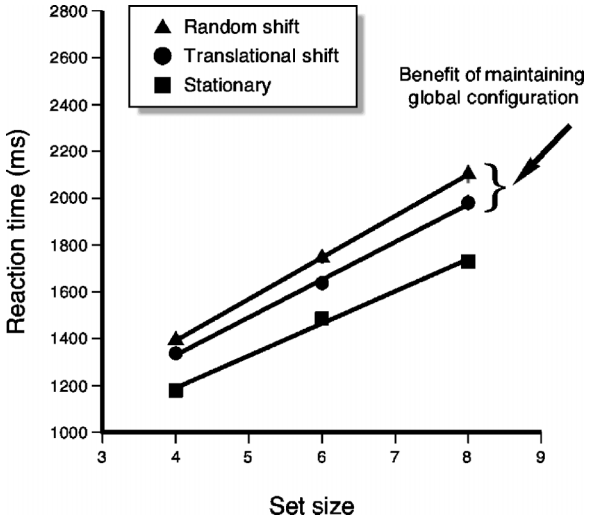

Figure 2. Results of Experiment 1. Average reaction time as a function of set size is shown for each condition. Searching in a dynamic display took more time than searching in a static display. When the global configuration of the targets was maintained, search was easier than when the targets shifted in a random direction with respect to each other.

for the random shift condition than for the translational shift conditions. These effects were verified with a series of ANOVAs on reaction time for trials in which a correct response was given.

An omnibus ANOVA on reaction time was run with set size (4, 6, or 8) and condition (random shift, translational shift, and stationary) as factors. Reaction time slowed monotonically as set size increased ($F(2, 22) = 40.4$, $MSE = 89588$, $p < .001$, $\eta_p^2 = 0.79$). The main effect of condition was significant ($F(2, 22) = 7.2$, $MSE = 103122$, $p < .01$, $\eta_p^2 = 0.39$), but the interaction between set size and condition was not significant ($F < 1$). Given the significant effect of condition, we conducted several more focused ANOVAs to determine which conditions significantly differed from each other.

Focused ANOVAs were conducted on reaction time with set size and condition as factors. Reaction time was faster in the translational shift condition than in the random shift condition (main effect of condition, $F(1, 11) = 18.4$, $MSE = 8766$, $p < .01$, $\eta_p^2 = 0.63$; set size × condition interaction not significant, $F < 1$), indicating a benefit to maintaining the global configuration of items (see Figure 2). Reaction time was also significantly slower in the translational shift condition than the stationary condition ($F(1, 11) = 5.1$, $MSE = 124306$, $p < .05$, $\eta_p^2 = 0.32$; set size × condition interaction not significant, $F < 1$). Finally, reaction time was slower in the random shift condition than in the stationary condition (main effect of condition, $F(1, 11) = 8.10$, $MSE = 176294$, $p < .05$, $\eta_p^2 = 0.42$; interaction not significant, $F(2, 22) = 1.3$, $MSE = 31517$, $p > .05$, $\eta_p^2 = 0.10$).

Overall error rates were low for each condition (random shift $M = 11.5\%$, $SEM = 2.8\%$; translational shift $M = 10.6\%$, $SEM = 2.5\%$; stationary $M = 7.0\%$, $SEM = 2.4\%$; see Appendix A for a more detailed analysis of error rates).

Discussion

In a dynamically changing search display, unsurprisingly, search is faster over a stationary display than one in which items change positions randomly. Thus, as expected from Horowitz and Wolfe (1998), overall reaction time in search is sensitive to manipulations that change the positions of items.

However, there was a savings for search times when the global configuration of items was preserved. Reaction time was faster when items changed position translationally than when items changed in a random direction. Critically, in both conditions the shift in the individual object locations was the same distance *locally*. The only difference was that in the translational shift condition the global configuration of objects was preserved more than in the random shift condition. So, while previous research has shown that long-term, implicit memory processes can build up a representation of global configuration that speeds search (Chun, 2000; Chun & Jiang, 1998; Jiang et al., 2005; Torralba et al., 2006), the current results suggest that short-term memory for the configuration of a display can also increase the speed of search.

While the results of this experiment suggest that the global configuration of a display plays an important role in visual search, there is an alternative explanation that does not require any explicit encoding of the global configuration. Alternating between the two stimulus frames creates a motion signal and it is possible that increasing the number of motion directions in the display increases the amount of motion noise in the display. On this view, the random shift condition would have the most motion noise (many motion directions), the translation condition less (one main motion direction), and the stationary condition would have none (no motion). To the extent that motion noise slows search performance, we would expect search to be slow in the random condition, faster in the translation condition, and fastest in the stationary condition. This is exactly the pattern observed in Experiment 1.

An alternative explanation that does rely on encoding the global configuration concerns the spatial predictability of item locations in the next frame. One strategy when performing this task might be to check an item and plan where attention should go next, taking into account that the items will appear in new locations in the next frame. On this view, the easier it is to predict where items will be located on the next frame, the easier it will be to plan the next shift of attention. In the random shift condition, there is the most uncertainty as to where an object will be in the

next frame; whereas in the translational shift condition, there is more predictability of where items will be in the next frame. Of course, in the stationary condition, the location of items is completely predictable. Thus, to the extent that spatial predictability of item location across frames aids search performance, we would expect search to be fastest in the stationary condition, slowest in the random condition, and somewhere in between for the translational condition (again mirroring the pattern observed in Experiment 1).

Experiment 2 addresses these alternative explanations.

Experiment 2: Motion noise vs. configural predictability

To distinguish between the motion hypothesis and the configural predictability hypothesis, we equated the random shift condition and translational shift condition in terms of predictability by adding placeholders to the display (see Figures 3a and 3b). For both conditions, every frame consisted of several gray discs, half of which were filled with letters and the other half of which were empty. The empty disks marked the locations in which the letters would appear on the subsequent frame. With these continuously visible placeholders, it is possible to know exactly where the items will appear on the next frame in both conditions. If the advantage for translational shifts over random shifts in Experiment 1 is explained by a benefit for greater predictability of subsequent locations, then the advantage should be eliminated by the addition of these placeholders. However, if the advantage for translational shifts over random shifts has to do with motion noise, then the advantage should be observed even with presentation of placeholders because the placeholders do not change this factor.

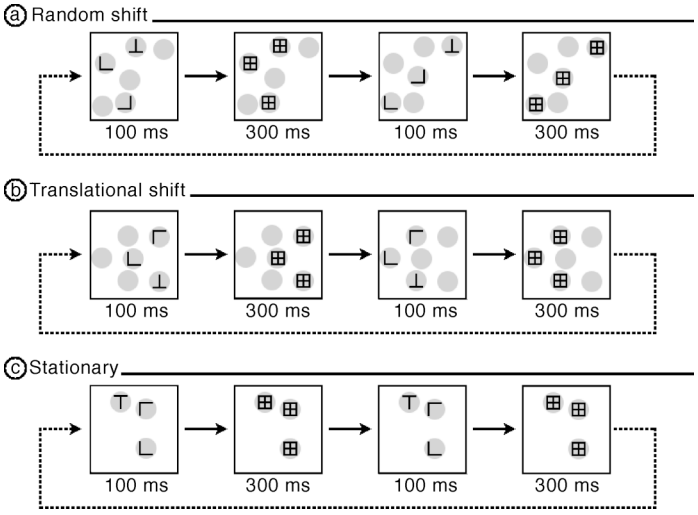

Figure 3. Displays for Experiment 2. The search displays were the same as in Experiment 1, with one addition. In the (a) random shift condition and (b) translational shift condition, each frame contained gray disks as placeholders for the items in the unshown frames. Gray disks were always present, some of which contained search items, and some of which were empty placeholders indicating the locations of the search items in the next frame. (c) In the stationary condition, the displays were exactly the same as in Experiment 1.

Method

Stimulus

The stimuli are shown in Figure 3. The timing was identical to Experiment 1, but the appearance of the random shift and translational shift displays changed. As in Experiment 1, all items appeared within a set of dark gray disks (20.2 cd/m^2). However, the random shift and translational shift conditions also contained empty gray disks on each frame (see Figures 3a and 3b). The empty disks marked the location of the letters for the *following* frame. Thus, these displays consisted of a fixed set of gray disks with letters alternating between one half of the disks and the other half, eliminating any uncertainty about where the items would appear on each frame. The sets of locations used in this experiment were identical to those used in Experiment 1. The only difference was the addition of the extra placeholders in the random shift and translational shift conditions. The stationary condition was the same as in Experiment 1 (see Figure 3c).

Procedure

The procedure was identical to that of Experiment 1.

Results

There was no difference between the random shift and translational shift condition in the current experiment (see Figure 4). The following analyses focused on reaction time for trials in which a correct response was given.

An omnibus ANOVA on reaction time was run with set size (4, 6, or 8) and condition (random shift, translational shift, and stationary) as factors. Reaction time slowed monotonically as set size increased ($F(2, 22) = 169.3$, $MSE = 30743$, $p < .001$, $\eta_p^2 = 0.94$). The main effect of condition was significant ($F(2, 22) = 24.0$, $MSE = 44523$, $p < .01$, $\eta_p^2 = 0.69$), but the interaction between set size and condition was not significant ($F < 1$). Given the significant effect of condition, we conducted several more focused ANOVAs to determine which conditions significantly differed from each other.

Focused ANOVAs were conducted on reaction time with set size and condition as factors. In the comparison of

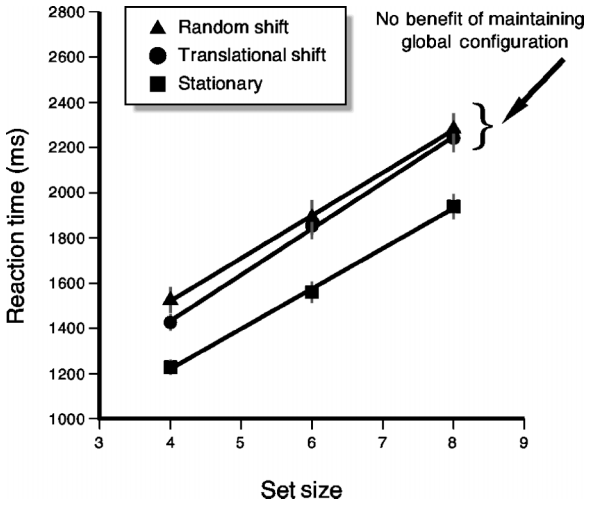

Figure 4. Results for Experiment 2. Average reaction time as a function of set size is shown for each condition. Searching in the dynamic displays (random, translation) was slower than searching in the static display. Here, the translational shift condition showed no difference from the random shift condition.

interest, reaction times were not significantly different between the random shift and translational shift conditions (main effect of condition not significant, $F(1, 11) = 3.0$, $MSE = 21006$, $p > .05$, $\eta_p^2 = 0.21$; set size × condition interaction not significant, $F < 1$). However, reaction time was significantly slower for both the translational shift and the random shift conditions compared to the stationary condition (translational vs. stationary: $F(1, 11) = 35.4$, $MSE = 35597$, $p < .001$, $\eta_p^2 = 0.76$; set size × condition interaction not significant, $F(2, 22) = 1.18$, $MSE = 16931$, $p > .05$, $\eta_p^2 = 0.10$; random vs. stationary: main effect of condition, $F(1, 11) = 24.47$, $MSE = 76967$, $p < .001$, $\eta_p^2 = 0.69$; interaction not significant, $F < 1$).

Overall error rates were low for each condition (random shift $M = 6.6\%$, $SEM = 1.6\%$; translational shift $M = 6.3\%$, $SEM = 1.7\%$; stationary $M = 5.0\%$, $SEM = 1.5\%$; see Appendix A for a more detailed analysis of error rates).

Discussion

In Experiment 2, unlike Experiment 1, there was no significant effect of the type of shift: Search was the same speed whether the items shifted randomly in the random shift condition, or as a group in the translational shift condition. The only change to the displays compared to Experiment 1 was the addition of empty placeholders to the random shift and translational shift conditions. These placeholders eliminated any uncertainty in where the items would appear on each frame. Eliminating this uncertainty appears to have eliminated any difference between the random shift and translational shift conditions, suggesting that the advantage for translation in Experiment 1 was based on greater predictability of item positions when the items translate and there is continuity in the global configuration from frame to frame.

There is more motion noise in the random shift displays than in the translational shift displays. However, the same difference in motion noise is present in the Experiment 1 and Experiment 2, yet there is no difference between the random and translation conditions once visible placeholders were added. These results rule out the hypothesis that greater motion noise results in slower search in the random shift displays.

An important alternative explanation for the current result concerns the possible interfering effects of the empty placeholders. If there were an interfering effect from the placeholders, and if this interference was greater for translational displays, then interference alone could eliminate the difference between the random and translation conditions. To address this concern, we ran a control experiment in which the items were always static, but there were empty, irrelevant placeholders present. The key manipulation was that the empty placeholders were positioned either as they were in random shift condition of Experiment 2 or in the translational shift condition of Experiment 2. It is important to emphasize that in this control experiment, the items never shifted between positions.

The results of this control experiment showed that overall, reaction time was slowed by 56 ms when empty placeholders were present in a static search task (no placeholders = 1282 ms vs. placeholders present = 1338 ms, $t(7) = 2.4$, $p < .05$). This supports the hypothesis that the placeholders interfere with processing the letters. The critical question for our purposes was whether reaction time would be slower when the placeholders were from the translation displays than when they were from random displays. In fact, there was a small effect in the opposite direction. For static search tasks with empty, irrelevant placeholders, reaction time was about 40 ms slower for the random placeholder condition than the translation placeholder condition (1360 ms vs. 1316 ms, $t(7) = 3.2$, $p < .05$). This interfering effect works against our prediction that adding the placeholders would speed reaction time in the random condition relative to the translation condition. Thus, interference from placeholders cannot explain why searching in random displays was not significantly different from translational displays in Experiment 2.

Although the difference between the translational shift condition and the random shift condition appears to be explained by differences in spatial predictability, there still remains a large difference between the two dynamic conditions and the stationary condition in the current results. Thus, spatial predictability alone is not sufficient to eliminate the difference between dynamic conditions and the stationary condition. In Experiment 3, we investigated the possibility that the shift conditions are more difficult than the stationary condition because there is no spatiotemporal continuity to link items across frames

in the shift conditions. Because the items essentially "jump" from one location to another all at the same time, it is often difficult to link an object in one frame to the corresponding object in the next frame. In Experiment 3, we introduced smooth motion in between these locations during the masking interval to provide stronger links between objects and their locations over time. We hypothesized that this spatiotemporal continuity would speed search in the random and translational shift conditions.

Experiment 3: Effect of spatiotemporal continuity

Here we investigated the role of spatiotemporal continuity in dynamic visual search. Again we tested search performance in the same 3 conditions (stationary, translational shift, and random shift), using the same set of locations as in Experiments 1 and 2. However, in the translational shift and random shift conditions, the locations for the two frames were spatially linked with "tracks" and the stimulus masks moved along these tracks during the masking interval. Thus, there were two explicit cues to object continuity in these conditions, and the motion made it particularly clear which items were linked across frames. If search was slower in the shift conditions of Experiments 1 and 2 because of the difficulty of linking items across frames, then addition of the tracks and motion cues in this experiment should reduce or eliminate the difference between the shift conditions and the stationary condition. An alternative possibility is that shifting items introduces motion noise into the display, and that this increased noise slows search in the shift conditions. On this view, adding stronger motion cues in the current study should make it even more difficult to search in the shift conditions relative to the stationary condition.

Method

Stimulus

The stimuli are shown in Figure 5. The timing was identical to Experiments 1 and 2, but the appearance of the displays was different in the random shift and translational shift conditions. As in Experiment 2, all items appeared within a set of dark gray disks and placeholders marked the item locations for the following frame. However, we also provided two cues that would directly link items from their position in frame 1 to their position in frame 2. First, we added a gray pathway (20.2 cd/m^2) linking the item positions from frame 1 and frame 2. Second, the masks moved along the gray pathway during the interval between frame 1 and frame 2

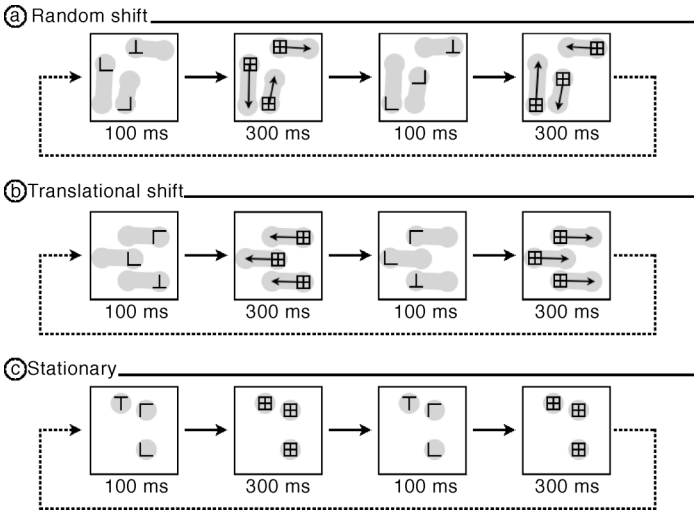

Figure 5. Displays for Experiment 3. The search displays were the same as shown in Experiment 2, with the addition of gray tracks between the pairs of placeholders. In the (a) random shift and (b) translational shift conditions, the items moved smoothly between the two locations during the masked frames. (c) In the stationary condition, the displays were exactly the same as in Experiments 1 and 2.

(see Figures 5a and 5b). Thus, these displays eliminated any uncertainty about which items moved to which position on each frame. The sets of locations used in this experiment were identical to those used in Experiments 1 and 2. The only difference was the addition of the gray pathways and the motion of the masks between frames in the random shift and translational shift conditions. The stationary condition was the same as in the previous experiments (see Figure 5c).

Procedure

The procedure was identical to that of Experiment 1.

Results

Adding the gray tracks and mask motion between item positions eliminated any differences between conditions. There appears to be no difference between the random shift and translational condition, suggesting no effect of changing global configuration. Surprisingly, there was also no difference between either shift condition and the stationary conditions, suggesting there was no cost for changing item position at all in this experiment (see Figure 6).

An omnibus ANOVA was run on reaction time for trials in which a correct response was given, with set size (4, 6, or 8) and condition (random shift, translational shift, and stationary) as factors. Reaction time slowed monotonically

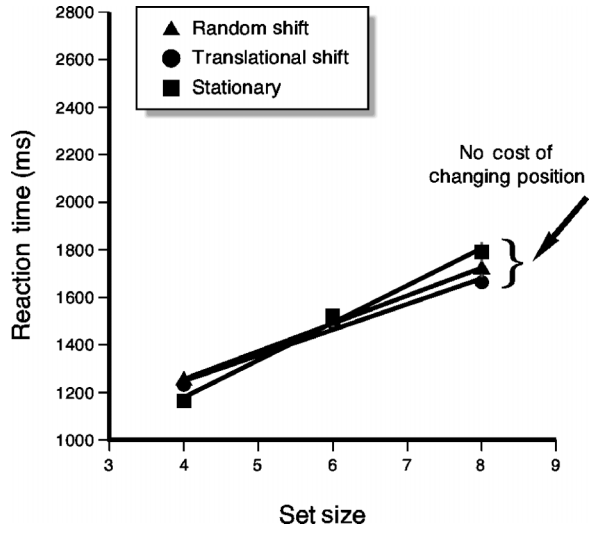

Figure 6. Results for Experiment 3. Average reaction time as a function of set size is shown for each condition. Searching in dynamic displays (translational shift, random shift), in which the items smoothly move from location to location, was indistinguishable from searching in a static display (stationary).

as set size increased ($F(2, 22) = 124.6$, $MSE = 18819$, $p < .001$, $\eta_p^2 = 0.92$). However, unlike previous experiments, the main effect of condition was not significant ($F < 1$, $p > .05$).

Overall error rates were low for each condition (random shift $M = 3.4\%$, $SEM = 0.7\%$; translational shift $M = 2.8\%$, $SEM = 0.7\%$; stationary $M = 2.6\%$, $SEM = 0.7\%$; see Appendix A for a more detailed analysis of error rates).

Discussion

Experiment 3 shows that with clear cues linking objects across frames, there is no difference in search speed for the static, random shift, and translational shift conditions. In the static condition, items merely blinked off and on, maintaining a constant global layout and posing no correspondence problem across frames. In the other conditions, the objects changed location from frame to frame, with each object moving to an independently chosen position in the random shift condition, and in roughly the same direction for all items in the translational shift condition. In Experiments 1 and 2, there was no information to explicitly link items between their positions on frame N, and frame $N + 1$, and reaction time was significantly slower in these shift conditions than in the static condition. However, in the current experiment, visible tracks linked the positions and item masks moved from their positions on frame N to their positions on frame $N + 1$, providing strong spatiotemporal continuity and unambiguously linking objects across frames. With this continuity to link objects across frames, it was possible to search just as efficiently in a dynamically changing visual display as in a stationary display.

It is possible that the tracks alone are sufficient to speed search in the random shift and translational shift conditions. Phenomenologically, the presence of the tracks alone serves to disambiguate which pairs of discs are linked, and they unambiguously indicate where a particular letter will appear next. To determine the role of the tracks alone, we ran a control experiment ($N = 8$ observers) that was identical to Experiment 3, except that the items were not shown moving from frame to frame. Instead, the items simply jumped from location to location as in Experiments 1 and 2, except with tracks linking pairs of locations. We found that overall, reaction time in the static condition was over 300 ms faster than either the random condition ($t(7) = 4.2$, $p < .01$) or the translation condition ($t(7) = 3.4$, $p < .05$), and there was no difference between the random and translation conditions ($t(7) = 1.5$, $p = .18$). Thus, even though the tracks unambiguously link letters across frames, this information was not sufficient to speed search in the dynamic conditions. The improved visual search in the random and translation conditions in Experiment 3 (see Figure 6) is due to the benefits of seeing items move from one location to the next. Thus, motion is not simply serving as a cue to inform observers of where a letter will appear next but seems to actually play a role in carrying attention from location to location.

General discussion

During natural perception, our view of a scene changes constantly as we move and as objects around us move. How does attention cope with these dynamic changes during visual search? Experiment 1 showed that search is slowed when items dynamically change position from moment to moment, as expected from previous research (measured by overall reaction time, not slope, see Horowitz & Wolfe, 1998). Surprisingly, this cost was reduced if the *global configuration* of the display remains the same between successive frames (Experiment 1). These results suggest that attention can capitalize on regularities in the global structure of a scene to adjust to changes that occur from moment to moment. Experiment 2 suggested that the advantage for maintaining global configuration comes from the spatial predictability provided by continuity in the global structure. Finally, Experiment 3 showed that spatiotemporal predictability provided by motion cues completely eliminated any cost associated with dynamic scene changes. This finding suggests that the ability to perceive items as persisting objects is necessary for efficient visual search. Both configural predictability and spatiotemporal continuity appear to place important constraints on our ability to search rapidly in a dynamically changing scene.

It is important to note that the effects observed in the current study occurred in overall reaction time, and not in the slope of the reaction time by set size function. Classically search slope is used as a measure of "search efficiency" because slope reflects the cost in reaction time for each additional distractor in the display (Wolfe, 1998). However, other researchers, particularly those who have investigated the role of contextual learning in visual search, have focused on overall reaction time as the critical measure (e.g., Chun, 2000; Chun & Jiang, 1998). We think that both overall reaction time and search slope are important measures of search efficiency, as each captures a different aspect of performance.

In a theory-neutral description, search slope is a measure of the reaction time cost for adding an additional distractor to the display. It appears that configural stability and spatiotemporal continuity do not affect this measure of search performance. A possible explanation for this null result is that search slopes are determined by similarity between targets and distractors, and by the similarity between distractors and other distractors (see Duncan & Humphreys, 1989). The more similar the target is to distractors, and the more dissimilar distractors are from each other, the greater the search slope. Critically, our manipulations of configural predictability and spatiotemporal continuity did not affect target–distractor or distractor–distractor similarity. Thus, a similarity account of visual search predicts that our manipulations would not influence search slopes. However, it is possible that slope effects might arise at set sizes larger than those tested in the current experiments (see Kristjánsson, 2000). Importantly, such effects would not change the conclusion that configural stability and spatiotemporal continuity speed search, which is based on effects in overall reaction time.

Overall reaction time is determined by the time to isolate the searchable items, find the target among the distractors, make a decision about the target, and execute a motor response. We assume that decision processes and motor response time were the same for random, translation, and static conditions in each experiment because observers were always performing the same task in each condition (e.g., find the target, report its orientation) and giving the same type of motor response (pressing one key or another on the keyboard). Thus, our finding that configural predictability and spatiotemporal continuity speed overall reaction time appears to indicate that these factors influence the time isolate the searchable items and locate the target among them. One can find a search target and report its appearance faster when searchable items appear in a stable configuration over time, and when changes in location are accompanied by clear motion cues.

Remarkably, adding motion between the frames of a dynamic display speeds search to the point that there is no difference in reaction time relative to searching in a stationary display. We suggest that the motion cues create a continuous link between objects that are changing location, and that this link facilitates the accumulation of object-identity information over time. Without this link, the accumulation of identity information can be interrupted forcing identification to begin anew on the next frame, or a false link between items can be made which could cause interference in the identification process. The control experiment reported in Experiment 3 indicates that motion per se is important for this process, given that simply linking objects with tracks did not speed search in random shift and translational shift conditions.

The current results can be important for models of visual search. Previous research suggests that visual attention operates over a set of pre-attentive object files (Wolfe, 1996; Wolfe & Bennett, 1997) or proto-objects (Rensink, 2002). For example, it appears that prior to attentional selection, it is possible to know that an object has the attributes "red" and "vertical" without knowing exactly how those features fit into the overall shape of the object. In this sense, "red" and "vertical" are part of a "proto-object" because there exits a representation in which they are bundled together without an explicit representation of *how* they are bundled together. Graduation from a proto-object to full "objecthood" requires binding features into a coherent, integrated unit, which requires selective attention (Treisman, 1996; Treisman & Gelade, 1980). While this previous work suggests that attention operates over such a map of proto-objects, the current work suggests that this map might include information about the relative spatial relationships between proto-objects (i.e., the global configuration) and about object history (e.g., *this* object is the same as the one that was *over there* previously).

The current study follows recent work in connecting research on visual search and multiple object tracking (Alvarez, Horowitz, Arsenio, DiMase, & Wolfe, 2005). The current work shows that spatiotemporal continuity improves search over dynamic displays, and previous work has shown that it improves the ability to attentively track objects (Scholl et al., 2001). This suggests that other factors which play a role in the ability to track objects, such as *grouping* of the moving items (Yantis, 1992), the *spacing* between items (Alvarez & Franconeri, 2007; Intriligator & Cavanagh, 2001), the relative position of items in the visual field (Alvarez & Cavanagh, 2005; Carlson, Alvarez, & Cavanagh, 2007), or the *cohesiveness* of the moving items (van Marle & Scholl, 2003), will also play an important role in the ability to search in dynamically changing displays. If any or all of these factors play a role in dynamic search, it would strongly suggest that the same mechanisms underlying attentive tracking play an important role in visual search under dynamic conditions.

Finally, the ultimate goal is to understand how attention operates over dynamic changes in a real-world setting. Two important differences between our displays and real-world displays are the amount and variability of visual clutter as well as contextual familiarity. Clutter can make

it difficult to isolate the target object, slowing search (Bravo & Farid, 2004; Rosenholtz, Li, Mansfield, & Jin, 2005; Wolfe, Oliva, Horowitz, Butcher, & Bompas, 2002), while contextual familiarity can limit search to regions likely to contain a target (Eckstein, Drescher, & Shimozaki, 2006; Torralba et al., 2006). Future work will be needed to explore how each of these real-world factors will interact with configural predictability and spatiotemporal continuity.

Conclusions

This work presents an important first step in a larger project aimed at investigating the limits on attention and visual search in dynamic situations. While still artificial, the dynamic visual search paradigm employed here enables us to isolate and investigate two important real-world factors with careful control of the timing and appearance of displays. Here we determined that *configural predictability* and *spatiotemporal continuity* increase the speed of visual search in dynamic displays. In fact, search operates equally well for static displays and for predictably and continuously changing dynamic displays. By using quickly learned configural regularities and spatiotemporal continuity that are present in the real world, attentional processes can overcome the challenges of searching in a dynamical world.

Appendix A

Error rate analysis

This appendix contains a detailed analysis of the error rates in each experiment. Please note that in some cases, error rates were significantly influenced by search task condition (random shift, translation, or stationary), but the effects were always in the same direction as the reaction time data (errors were positively correlated with reaction time) indicating that reaction time differences between conditions cannot be explained by a speed accuracy tradeoff. Thus, the pattern of errors reported here has no impact on the conclusions drawn from the reaction time data presented in the main body of the paper. However, they are presented for readers interested specifically in error rates.

Experiment 1

Error rates were fairly low overall (about 10% collapsed across participants and conditions). An ANOVA on error rates with set size (4, 6, or 8) and condition (random shift, translation, and stationary) as factors showed that error rates increased as set size increased ($F(2, 22) = 12.4$, $MSE = 16.1$, $p < .001$, $\eta_p^2 = 0.53$) and there was a main effect of condition ($F(2, 22) = 11.5$, $MSE = 17.6$, $p < .001$, $\eta_p^2 = 0.51$). Error rates in the random shift, translation, and static condition were 11.5%, 10.7%, and 7.1%, respectively. The difference between the random shift and the translation condition was not significant ($t(11) = 1.2$, $p > .05$, $r^2 = .12$), but error rates were lower in the static condition than in the random shift condition ($t(11) = 3.6$, $p < .01$, $r^2 = .54$) or the translation condition ($t(11) = 3.6$, $p < .01$, $r^2 = .54$). Finally, the interaction between set size and condition was significant ($F(4, 44) = 3.64$, $MSE = 10.3$, $p < .05$, $\eta_p^2 = 0.25$), indicating that the difference in accuracy across conditions was greater for larger set sizes.

Experiment 2

Error rates were low overall (about 6% collapsed across participants and conditions). An ANOVA on error rates with set size (4, 6, or 8) and condition (random shift, translation, and stationary) as factors showed that error rates increased as set size increased ($F(2, 22) = 5.0$, $MSE = 7.2$, $p < .05$, $\eta_p^2 = 0.31$), and there was a main effect of condition ($F(2, 22) = 6.0$, $MSE = 4.5$, $p < .01$, $\eta_p^2 = 0.35$). Mean error rates in the random shift, translation, and static condition were 6.6% ($SEM = 1.56$), 6.3% ($SEM = 1.66$), and 5.0% ($SEM = 1.46$), respectively. The difference between the random shift and the translation condition was not significant ($t < 1$), but error rates were lower in the static condition than in the random shift condition ($t(11) = 3.0$, $p < .05$, $r^2 = .45$) or the translation condition ($t(11) = 2.4$, $p < .05$, $r^2 = .35$). The interaction between set size and condition was not significant ($F < 1$).

Experiment 3

Error rates were low overall (about 3% collapsed across participants and conditions). An ANOVA on error rates with set size (4, 6, or 8) and condition (random shift, translation, and stationary) as factors showed no significant effects (main effect of set size, $F(2, 22) = 1.3$, $MSE = 10.7$, $p > .05$, $\eta_p^2 = 0.11$; main effect of condition, $F(2, 22) = 2.6$, $MSE = 2.8$, $p > .05$, $\eta_p^2 = 0.19$; interaction, $F < 1$).

Acknowledgments

For helpful suggestions and comments on these experiments, we thank Jeremy M. Wolfe, Todd S. Horowitz, as

well as two reviewers. This research was supported by NIH/NEI fellowship #F32 EY016982 to G.A.A., by an NDSEG fellowship to T.K, and an NSF CAREER award (# 0546262) to A.O.

Commercial relationships: none.
Corresponding author: George A. Alvarez.
Email: alvarez@mit.edu.
Address: 77 Massachusetts Avenue, Building 46-4078c, Cambridge, MA, 02128.

References

Alvarez, G. A., & Cavanagh, P. (2005). Independent resources for attentional tracking in the left and right visual hemifields. *Psychological Science, 16,* 637–643. [PubMed]

Alvarez, G. A., & Franconeri, S. L. (2007). How many objects can you track?: Evidence for a resource-limited attentive tracking mechanism. *Journal of Vision, 7*(13):14, 1–10, http://journalofvision.org/7/13/14/, doi:10.1167/7.13.14. [PubMed] [Article]

Alvarez, G. A., Horowitz, T. S., Arsenio, H. C., DiMase, J. S., & Wolfe, J. M. (2005). Do multielement visual tracking and visual search draw continuously on the same visual attention resources? *Journal of Experimental Psychology: Human Perception and Performance, 31,* 643–667. [PubMed]

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10,* 433–436. [PubMed]

Bravo, M. J., & Farid, H. (2004). Search for a category target in clutter. *Perception, 33,* 643–652. [PubMed]

Carlson, T. A., Alvarez, G. A., & Cavanagh, P. (2007). Quadrantic deficit reveals anatomical constraints on selection. *Proceedings of the National Academy of Sciences of the United States of America, 104,* 13496–13500. [PubMed] [Article]

Chun, M. M. (2000). Contextual cueing of visual attention. *Trends in Cognitive Sciences, 4,* 170–178. [PubMed]

Chun, M. M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology, 36,* 28–71. [PubMed]

Cohen, J. (1973). Eta-squared and partial eta-squared in fixed factor ANOVA designs. *Educational and Psychological Measurement, 33,* 107–112.

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

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

Egeth, H., Jonides, J., & Wall, S. (1972). Parallel processing of multielement displays. *Cognitive Psychology, 3,* 674–698.

Horowitz, T. S., & Wolfe, J. M. (1998). Visual search has no memory. *Nature, 394,* 575–577. [PubMed]

Intriligator, J., & Cavanagh, P. (2001). The spatial resolution of visual attention. *Cognitive Psychology, 43,* 171–216. [PubMed]

Jiang, Y., Song, J. H., & Rigas, A. (2005). High-capacity spatial contextual memory. *Psychonomic Bulletin & Review, 12,* 524–529. [PubMed]

Kinchla, R. A. (1974). Detecting targets in multi-element arrays: A confusability model. *Perception & Psychophysics, 22,* 19–30.

Kristjánsson, A. (2000). In search of remembrance: Evidence for memory in visual search. *Psychological Science, 11,* 328–332. [PubMed]

Kunar, M. A., Flusberg, S. J., & Wolfe, J. M. (2006). Contextual cuing by global features. *Perception & Psychophysics, 68,* 1204–1216. [PubMed]

McLeod, P., Driver, J., Dienes, Z., & Crisp, J. (1991). Filtering by movement in visual search. *Journal of Experimental Psychology: Human Perception and Performance, 17,* 55–64. [PubMed]

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

Palmer, J., & McLean, J. (1995). *Imperfect, unlimited-capacity, parallel search yields large set-size effects.* Paper presented at the Society for Mathematical Psychology, Irvine, CA.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10,* 437–442. [PubMed]

Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision, 3,* 179–197. [PubMed]

Rensink, R. A. (2002). Change detection. *Annual Review of Psychology, 53,* 245–277. [PubMed]

Rosenholtz, R., Li, Y., Mansfield, J., & Jin, Z. (2005). Feature congestion: A measure of display clutter. *SIGCHI 2005,* 761–770.

Royden, C. S., Wolfe, J. M., & Klempen, N. (2001). Visual search asymmetries in motion and optic flow fields. *Perception & Psychophysics, 63,* 436–444. [PubMed] [Article]

Scholl, B. J., Pylyshyn, Z. W., & Feldman, J. (2001). What is a visual object? Evidence from target merging in multiple object tracking. *Cognition, 80,* 159–177. [PubMed]

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

Treisman, A. (1996). The binding problem. *Current Opinion in Neurobiology, 6,* 171–178. [PubMed]

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

Tryon, W. W. (2001). Evaluating statistical difference, equivalence, and indeterminacy using inferential confidence intervals: An integrated alternative method of conducting null hypothesis statistical tests. *Psychological Methods, 6,* 371–386. [PubMed]

van Marle, K., & Scholl, B. J. (2003). Attentive tracking of objects versus substances. *Psychological Science, 14,* 498–504. [PubMed]

Wolfe, J. M. (1996). Extending guided search: Why guided search needs a preattentive "item map." In A. F. Kramer, M. G. H. Coles, & G. D. Logan (Eds.), *Converging operations in the study of visual attention* (pp. 247–270). Washington, DC: American Psychological Association.

Wolfe, J. M. (1998). Visual search. In H. Pashler (Ed.), *Attention* (pp. 13–74). Hove, East Sussex, UK: Psychology Press Ltd.

Wolfe, J. M., & Bennett, S. C. (1997). Preattentive object files: Shapeless bundles of basic features. *Vision Research, 37,* 25–43. [PubMed]

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

Wolfe, J. M., & Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews, Neuroscience, 5,* 495–501. [PubMed]

Wolfe, J. M., Oliva, A., Horowitz, T. S., Butcher, S. J., & Bompas, A. (2002). Segmentation of objects from backgrounds in visual search tasks. *Vision Research, 42,* 2985–3004. [PubMed]

Yantis, S. (1992). Multielement visual tracking: Attention and perceptual organization. *Cognitive Psychology, 24,* 295–340. [PubMed]