Research Note

Visual Search in Continuous, Naturalistic Stimuli

Jeremy M. Wolfe*†

Received 9 June 1993; in revised form 24 September 1993

In standard visual search experiments, subjects look for a target item among distractor items. Targets and distractors are generally presented as isolated entities on a blank background. Do theories based on results of such experiments have any application to the real world of spatially continuous stimuli where “items” are not so conveniently delineated? Experiment 1 examines search for conjunctions of color with form and with orientation in novel stimuli designed to resemble aerial views. As with artificially isolated items, conjunction search with these complex stimuli is highly efficient, holding out the promise that models of visual search based on laboratory data can be extended to real-world situations. Experiment 2 shows that some “serial” searches are made significantly less efficient when items are embedded in these specific naturalistic backgrounds. This suggests that even classic serial searches rely on some parallel processing—the processing required to locate the “items” to be serially searched.

Attention Visual search Parallel processing Feature conjunction

The world presents us with an endless variety of tasks that have visual search as a central component. Where in this pile of junk are the car keys? Is there a tumor in this radiological image of a liver? Has anything important changed in this satellite image of Bosnia? Because it is such a basic part of visual behavior, visual search has been extensively studied in the laboratory (Duncan, 1985; Green, 1991; Treisman, 1986; Wolfe, 1992b). In standard laboratory visual search experiments, subjects look for a target item among distractor items as illustrated in the upper half of Fig. 1. The target might be uniquely defined by a single basic feature [as in Fig. 1(a)—size—a big item in an array of smaller items] or by a conjunction of basic features [Fig. 1(b)—color x size—a big black item among small black and big white items] or by some more complex property [Fig. 1(c)—a “2” among “5”s]. In one standard paradigm, the number of distracting items (set size) is varied and the response time (RT) is measured. Changes in RT as a function of set size are varied in making inferences about the underlying mechanisms of visual search. Average data from 30 subjects are shown in the lower half of Fig. 1. In Fig. 1(a), a feature search for a single size, RT increases little or not at all with set size, suggesting that all items are processed in parallel.

*Harvard Medical School, Cambridge, MA 02138, U.S.A.
†Address for correspondence: Center for Ophthalmic Research, Brigham and Women’s Hospital, 221 Longwood Avenue, Boston, MA 02115, U.S.A.
‡In the actual experiment, colors were red and green, not black and white.

(Treisman & Gelade, 1980; Treisman & Souther, 1985). There is a limited set of visual operations that can be performed in parallel over all or most of the visual field at once. Many other visual operations can be performed only over much more restricted portions of the field at any one time. Examples include face recognition and reading. Attention needs to be deployed to the face or letter before it is processed. Thus, looking at Fig. 1(c), in the search for a “2” among “5”s, RTs increase in a linear fashion with set size. The slope of the blank trial RTs is about twice that of the target trials. This is consistent with a serial, self-terminating search where subjects examine each “5” in random sequence until the “2” is found or until all the “5”s have been rejected (Kwak, Dagenbach & Egeth, 1991). Figure 1(b) shows a sample search for a conjunction, a target defined by the joint presence of two or more basic features; here, color and size. These searches are generally more efficient than strictly serial searches and somewhat less efficient than the easiest feature searches (Nakayama & Silverman, 1986; Treisman & Sato, 1990; Wolfe & Cave, 1989; Wolfe, Yee & Friedman-Hill, 1992). We describe these results in the context of the Guided Search model (Cave & Wolfe, 1990; Wolfe, 1992b, 1994). According to this model, efficient search for conjunctions reflects the ability of parallel feature processors to guide attention. Thus, in Fig. 1(b), no one feature process can find big black items but a color processor can activate the locations of black items and a size processor can activate the locations of all big items. The combination of these two sources of activation can guide attention toward big...
black items. There are other ways to describe these results (e.g. Duncan & Humphreys, 1989; Humphreys & Müller, 1993). This paper will use the language of Guided Search.

A large body of data has been obtained from experiments of this sort and from related paradigms that measure accuracy rather than response speed (see Kinchla, 1992; Shiffrin, 1988; Treisman, 1986 for fairly recent reviews). However, like many areas of inquiry in vision research and, indeed, in psychology, the study of visual search has been nagged by the suspicion that it has strayed far from the real world questions that began the work. Visual search in the real world is a search for a target item in a spatially continuous visual field containing other items arrayed irregularly on what is usually a non-homogeneous background. Visual search in the lab is a search for a target item among a set of discrete distractor items, all floating serenely on the homogeneous black or gray background of the computer screen. Often these items are presented in highly regular arrays (e.g. circles of items equidistant from fixation or items placed in orderly rows and columns). At the very least, the usual laboratory visual search task has greatly simplified the task of determining the set of items to be searched. At worst, it is possible that the task has become so artificial that the body of sophisticated theory mustered to explain laboratory search tasks will have nothing to say about real world tasks.

The present experiments have two purposes. Experiment 1 provides some encouraging evidence that the classic results obtained in standard laboratory studies are still obtained with a new class of more naturalistic stimuli. Experiment 2 shows that these stimuli can be used to extend our understanding of visual search behavior. Some previous work has been done with natural images; e.g. eye movement studies (Enoch, 1959; Yarbus, 1967) and studies of the supporting effects of natural context on object identification (Biederman, Glas & Stacy, 1973). However, it is difficult to do traditional visual search tasks of the sort illustrated in Fig. 1 with natural images because the traditional search paradigm involves presenting the same stimuli hundreds of times with random placement of target and distractor stimuli. With these experimental demands, stimulus generation with truly natural scenes would be daunting. In the present experiments, we have created artificial stimuli that share a number of important properties with real world stimuli. Displays are continuous rather than containing isolated items on a blank background. They contain regions of various colors, orientations, luminances, and sizes. The images can be assigned some meaning, in this case as aerial views of terrain. At the same time, the stimuli are composed of interchangeable "tiles", allowing for random placement of target and distractors, which prevents learning about the overall structure of one specific scene. With these stimuli, in
Expt 1, we show that visual searches for conjunctions of color × form and color × orientation remain highly efficient even in cases where the target is partially occluded (here by "clouds"). In Expt 2, we show that some "serial" searches become significantly less efficient when the targets and distractors to be searched are embedded in a complex background.

**EXPERIMENT 1 CONJUNCTION SEARCHES**

**Methods**

Construction of the stimuli is shown in Fig. 2. Square "tiles" of terrain were created with rivers leaving each tile at the midpoint of each side. Figure 2(a) shows the blue lake condition of Expt 1. Each target tile contained a blue lake. Half the time, the lake was partially occluded by clouds. Distractor tiles included a "polluted", yellow lake, and tiles with other configurations of blue rivers. Additionally, there were cities of the approximate size of the lake in one-third of the tiles. Clouds similar to those occluding the lake were present over another third of the distractor tiles. Brown "hills" punctuated the green background in all tiles. The stimuli for each trial were created by presenting 9, 16, or 25 tiles in square arrays (3 × 3, 4 × 4, or 5 × 5) with no space between tiles. Tiles were chosen randomly from the distractor set. On 30% of the trials, one of the target tiles was substituted for a distractor tile. On a given trial, subjects searched for a blue lake but did not know if the lake would be partially occluded.

With tiles placed one next to the other, the segments of blue rivers in each tile formed a continuous network of waterways across the entire stimulus. Rather than being isolated "items" on a homogeneous background, the lakes were continuous with the rivers. Thus on each trial subjects saw a square "map" of some admittedly rather odd terrain. Odd or not, Fig. 2 shows that the stimuli do not look like isolated items on a blank background.

The blue lake task shown in Fig. 2(a) is clearly a conjunction search. No single feature provides information adequate to identify the target. The display contains yellow lakes of the same size and shape as the blue lake and blue rivers of the same color. Beyond these obvious distractors, there are other regions of green, brown, or gray of the same approximate size. It is difficult with stimuli of this sort (and more difficult with real-world stimuli) to specify the "set size". A conservative estimate would be to assume one distracting element per component tile; that is, to assume that the yellow lake tiles contain one lake distractor and to ignore the presence of blue river distractors in those tiles. Using the number of tiles as the set size, these experiments had set sizes of 9, 16, and 25 items. If this is an underestimate of the true number of distractors, then RT × set size slopes will be overestimated.

Two other tasks were run in Expt 1. Figure 2(b) shows the polluted river task—a color × orientation conjunction search. Subjects searched for a "polluted" yellow river region that ran north–south (NS—roughly vertical) in a scene that contained polluted, yellow east–west (EW) river regions and clean, blue NS regions. Scenes also contained "clean" EW regions, green vertical and horizontal contours and a variety of other possibly distracting elements. The polluted regions had clear terminations that the continuous blue rivers did not, so isolated NS blue river segments (oxbow lakes) with clear terminators were added to the scene in one third of the tiles. Thus, the polluted river task is best described as a conjunction search for yellow vertical among a variety of distractors that are either yellow or vertical but not both.

The third condition, the terminator task, was a "simple" feature search. Here the target was an oxbow lake [an isolated NS river segment as shown in Fig. 2(b)] embedded in a scene in which all other river regions were continuous. Evidence for the status of terminators as features comes from studies of isolated items (Julesz, 1984; Julesz & Bergen, 1983). It seemed to be a particularly interesting feature to test with stimuli where the most obvious distractors (the other blue rivers) formed a continuous structure. There are other possible terminators in these stimuli; specifically parts of the low contrast brown blobs that denoted "hills" and the terminators of the rivers at the edge of the stimulus.

Again, for all tasks, set size can be conservatively defined as equal to the number of tiles making up each stimulus. In all three tasks, stimuli were square, continuous arrays of tiles. The examples in Fig. 2 are not square but otherwise representative. Each tile subtended 3.2 deg. Subjects fixated at the center of a 24 × 17 deg field and the stimulus was presented at a random location within this field. The target element was equally likely to appear at all locations in the field. Stimuli were presented on a MacIntos running VSearch software (Enns, Ochs & Rensink, 1990). Eleven subjects were tested on 330 trials. The first 30 were discarded as practice. Nine subjects were naive as to the purposes of this experiment though four of these had been in other visual search experiments. All subjects had given informed consent.

**Results and discussion**

The results are summarized in Fig. 3. The blue lakes condition produces very efficient search with essentially no change in target trial RT as a function of set size. Color × size conjunctions are among the easiest conjunction tasks in standard search paradigms (Treisman & Sato, 1990; Wolfe, 1992a) and such tasks remain easy with these more naturalistic stimuli. The 1.6 msec/item target trial slope in this task is comparable to traditional "parallel" searches. The polluted river task produces somewhat steeper slopes that are comparable with those seen for color × orientation tasks in standard search paradigms (Treisman & Sato, 1990; Wolfe et al., 1992). The feature search for terminators produces slopes comparable to those in the polluted river task. The search is reasonably efficient but the slopes are steeper than...
FIGURE 2. Construction of stimuli for the present experiment. The stimulus that a subject saw on each trial was composed from pieces of terrain here called tiles. Each tile has rivers entering and exiting at the midpoints of the sides of the tile. As a result, when the tiles are placed next to one another, the rivers form a continuous network. Different scenes are created for each trial, by randomly choosing tiles for each position in the stimulus on each trial. On 50% of the trials, a target tile is placed at a random location. Thus, we can create continuous, naturalistic stimuli and still have the random placement of stimulus elements that is an important aspect of the traditional visual search paradigm. (a) Sample stimulus. The blue lake target is embedded in blocks drawn at random from the distractor set. Stimulus blocks fit together to form a continuous image. (b) The polluted river condition. Target: yellow NS river segment; Distractors: yellow EW segment blue NS segments and lakes. Actual stimuli were, of course, in color.
those that one would see with a highly salient feature (e.g. a unique and salient color).

It is worth reiterating that these RT x set size slope estimates may be overly steep. If there is more than one distractor “item” in each component tile of the stimulus, then the real set size is larger than the nominal set size shown in Fig. 3 and the real slopes should be proportionally shallower. Nevertheless, even with the possible overestimation of slopes, it is clear that efficient search is possible with the stimuli used here.

In standard search stimuli, the ratio of blank trial slopes to target trial slopes tends to average 2:1 over a wide range of search tasks (Chun & Wolfe, 1994). With the present stimuli, the slope ratios are larger. One-tailed t-tests (d.f. = 10) support the hypothesis that the slope ratios are greater than 2:1 for all three conditions (P < 0.02 in each case). This probably reflects subjects' caution in terminating a blank trial with these crowded and complex stimuli.

In the *blue lake* condition, the target was occluded by clouds on 50% of the trials. If the response to the two target types are compared, we find that the unoccluded lake is found an average of 81 msec faster (ANOVA: F = 78.3, P < 0.0001) but subjects can still search very efficiently for the occluded lakes. The average slope for the search for the occluded lake is still only 3.1 msec/item. This supports the finding that occluded objects are treated as though continuous by the mechanisms of visual search (Enns & Rensink, 1992) even if they take somewhat longer to process (in parallel) than unoccluded items.

*Control experiments*

Several control experiments were run. All were versions of the *blue lake* condition. When lakes and pieces of rivers were presented in isolation on a blank background, target and blank trial slopes were 1.5 and 1.2 msec/item respectively. For target trials, RTs were 80–100 msec faster with isolated items than with the stimuli shown in Fig. 2. The blank trials were 150–200 msec faster. In the original *blue lake* condition, it could be argued that the network of rivers formed a single item. The control experiment result indicates that it was not critical that the rivers formed a single network in the original experiment. In a second control, the square terrain patches shown in Fig. 2 were presented in random positions with intervening space on the screen rather than in the continuous square arrays of the original experiment. The background was either gray, in which case the borders of the squares could be seen, or green, in which case the borders were not seen. These manipulations had little impact on either the slope or the RTs for target trials. Blank trials were somewhat slower, particularly when the squares were placed on a green background. These controls experiments demonstrate that the search for the blue lake can be efficient whether or not the lake is embedded in this particular naturalistic context. In this case, the “natural” context neither hinders nor helps search, suggesting that the underlying processes are relatively insensitive to this information.

**EXPERIMENT 2—SERIAL SEARCHES**

How can conjunction searches of the sort described in Expt 1 be performed so efficiently? Elsewhere, we have argued that parallel processes guide the serial deployment of attention (Wolfe, 1992b, 1994). Taking the example of the *polluted river* condition of Fig. 2(b), a parallel color process would locate “yellow” while an orientation process would locate the NS (vertical) river segments. Suppose that no such guidance were available. In a standard search, the result is a serial, self-terminating search through all items (Bergen & Julesz, 1983; Kwak et al., 1991; Wolfe, Cave & Franzel, 1989). It is obvious, but usually not noted, that even such a serial search requires some parallel processing.
FIGURE 4. Sample stimuli for the continuous condition of Expt 2. The target is a river segment in the form of a letter $S$. The nominal distractors are segments in the form of a mirror-reversed $S$. It is clear that this is a difficult search. The difficulty arises in part from our inability to preattentively locate items for subsequent serial attention.
Subjects do not search where there are no items. Even if there is no preattentive information about the identity of specific items, the items themselves are found in parallel. The importance of this overlooked parallel step may be seen when naturalistic stimuli are used. These stimuli do not offer up their "items" so readily.

Methods

Subjects searched for an “S” among mirror-reversed “S”s. This task usually produces standard serial, self-terminating results as shown in Fig. 1(c). In Expt 2, these items were embedded to a greater or lesser degree in the naturalistic background used in Expt 1. A black and white example is shown in Fig. 4. Actual stimuli were in color. In this continuous condition, the S and mirror S items are part of the network of rivers. Even without the benefit of data, it is clear that this is a difficult search. In the isolated natural condition, the S and mirror S items were placed on a similar naturalistic background of rivers, hills, etc. However, the S and mirror S items were isolated “lakes” not connected to the river network. They were not connected with the network of rivers. In the standard condition, S and mirror S items were presented on a blank, green background as in a standard visual search experiment. Black and white illustrations of the components of these two control conditions are shown in Fig. 5. All other methods were similar to those in Expt 1.

Results and discussion

Group average RT x set size functions are shown in Fig. 6. Beginning with the standard condition, when items are isolated on a blank background, the resulting search slopes are comparable to other tasks that are thought to require serial, self-terminating search (target trials—22.3 msec/item; blank—71.0 msec/item). However, uncharacteristically, the blank trial to target trial slope ratio is significantly steeper than 2:1 \[ r(9) = 4.7, P < 0.001 \]. Placing these stimuli on a naturalistic background increases both target and blank trial slopes (to 45.8 and 95.3 msec/item respectively—isolated natural condition). Adding the extraneous rivers, hills, etc. makes the search harder.

It is in the continuous condition that performance becomes truly awful. Slopes are very steep (101.8 and 206.3 msec/item). Error rates go up (reaching 50% for one subject at set size 25). Intercepts are significantly higher than in the other two conditions (ANOVA, F-tests, \( P < 0.001 \) for both target and blank trial intercepts). Why is the continuous condition so much worse than the isolated natural condition? In the continuous condition, subjects may spend significant time attending to the areas between and around items. Moreover, it is likely that the problems in the continuous condition are a by-product of crowding or some related phenomenon (Andriesen & Bouma, 1976; Arditi & Cagenello, 1993; Toet & Levi, 1992). One would expect the same crowding effects in the isolated natural condition where the distribution of “S”s and other contours is similar. In sum, it seems that the difference between the continuous and isolated natural conditions is that preattentive processing can better identify items in the isolated natural condition than in the continuous condition. The number of contours is similar in the two conditions. However, the number of contours entertained as candidate targets may be very different.

These results shed new light on the role of location in visual search (see, e.g. Cohen & Ivry, 1991; Johnston & Pashler, 1990; Tsal & Lavie, 1988). Preattentive processing of location is required if spatially-limited attention is to be intelligently deployed. That location information can be developed from any of many

FIGURE 5. Components of the stimuli for the isolated natural and standard conditions of Expt 2.
parallel feature processes. In the blue lake condition of Fig. 2, color and form processing give the likely locations of targets. In a standard serial search task, location information can come from any of a number of feature processors. The items are the only things in the field of a particular color, size, etc. This parallel information is of only limited use because it identifies all items as candidate targets in this case. However, at least it identifies the relevant set of items. In the continuous condition of Expt 2, the S and mirror S items cannot be preattentively located. If these items are not uniquely defined on any attribute, they may not be attended to in any orderly manner. The result is an experimental version of the needle-in-a-haystack class of visual search.

**GENERAL DISCUSSION**

Do the rules of visual search, established with very artificial stimuli in the laboratory have any application to the real world of continuous stimuli? The results of the present experiments support a qualified positive answer. With one set of spatially continuous, naturalistic stimuli, the standard results are replicated. Feature searches are easy. Performance for the conjunction tasks employed here is at least as good as performance on comparable tasks as reported in the standard visual search literature. Given the efficiency of these searches, it may be that the structure of these particular continuous stimuli or of natural stimuli in general can actually aid the search process in ways that we have not yet investigated. Certainly this would be consistent with previous work on context effects and object recognition (Biederman et al., 1973; Weisstein & Harris, 1974). Performance for a "serial" task became much worse than a standard "serial" search when the items were embedded in a continuous background. There are a number of explanations for this result but all of them point to the role of preattentive processing even in the most serial of standard search tasks. The "serial" searches of the
standard visual search literature may well be serial but, in that serial search, attention is deployed over a set of preattentively located items.

Obviously, the generality of the present results is limited by the specificity of the stimuli (a chronic problem with studies of anything resembling real world stimuli). The stimuli described here are still artificial. This allows them to be manipulated in ways demanded by the paradigms of laboratory visual search experiments. At the same time, they have enough in common with continuous, naturalistic stimuli to hold out the hope that they may provide the bridge between the controlled world of the laboratory and the real world of searches for lost keys, airfields in the Balkans, or tumors in the liver.

REFERENCES


**Acknowledgements** I thank Julie Sendell, Anne Treisman, and Ron Rensink for comments on drafts of this paper and I thank Alexander Bilsky and Patricia O’Neill for help with all aspects of the project. This work was supported by NIH-NEI EY05087 and AFOSR grant F49620-93-1-0407.