Cognitive Science Technical Report #26, October 2001
Center for Cognitive Science
The Ohio State University
220C Page Hall
1810 College Road
Columbus, OH 43210, USA
PH: 614-292-8200; FX: 614-292-0321
EML: sharon.3@osu.edu
URL: www.cog.ohio-state.edu

Efficient Visual Search without Top-down or Bottom-up Guidance:

A Putative Role for Perceptual Organization

DeLiang Wang[†]
Arni Kristjansson[‡]
Ken Nakayama[‡]

[†]*Center for Cognitive Science and Department of Computer and Information Science*
The Ohio State University
2015 Neil Avenue
Columbus, OH 43210, U.S.A.

[‡]*Vision Sciences Laboratory*
Department of Psychology
Harvard University
33 Kirkland Street
Cambridge, MA 02138, U.S.A.

(Email addresses: dwang@cis.ohio-state.edu; {kristjan,ken}@wjh.harvard.edu)

Correspondence should be directed to DLW: Department of Computer and Information Science, The Ohio State University, 2015 Neil Avenue, Columbus, OH 43210. Phone: 614-2926827, Fax: 614-292-2911, URL: www.cis.ohio-state.edu/~dwang/.

Abstract

Two types of mechanisms have dominated theoretical accounts of efficient visual search. First are bottom-up processes related to the hypothesized characteristics of retinotopic feature maps. Second are top-down mechanisms related to feature selection. Little effort has been made to understand visual search in terms of perceptual grouping despite its acknowledged importance in general visual perception. To examine the possible role of perceptual grouping we employ a new search paradigm whereby a target is defined only in a context-dependent manner by multiple conjunctions of feature dimensions. Because targets in a multi-conjunction task cannot be distinguished from distractors either by bottom-up guidance or top-down guidance, current theories of visual search predict inefficient search. While inefficient search does occur for the multiple conjunctions of orientation with color or luminance, we found efficient search for multiple conjunctions of luminance with size, shape, and topological properties. We also show that repeated presentations of either targets or a set of distractors result in much faster performance. Our results suggest that perceptual organization can play a decisive role in visual search, and theories of visual attention need to take this into account. Furthermore, multi-conjunction search may provide a new vehicle for investigating perceptual grouping and scene analysis.

Introduction

The feature integration theory of Treisman and Gelade (Treisman & Gelade, 1980), proposed twenty years ago, continues to dominate the study of visual attention (Pashler, 1998; Palmer, 1999). According to the theory, the visual system first analyzes a scene in parallel by separate retinotopic feature maps. Neurophysiological evidence supports the notion of early visual analysis by feature maps (Zeki, 1993; Allman et al., 1985). Focal attention then integrates analyses in different feature maps to produce a coherent perceptual object, and perceiving multiple objects in a scene requires sequential shift of attention from one location to another (Treisman & Gelade, 1980; Koch & Ullman, 1985). The theory has been confirmed by extensive visual search experiments; for example, it successfully explains efficient search for targets that distinguish from distractors in a single feature dimension, such as a red item in a field of green distractors. By efficient search we mean that the slope of reaction time (RT) with respect to the set size (number of items) in a display is small (e.g. 5 ms per item) when the target is present, and inefficient search corresponds to a substantial search slope, say 20 ms per item. A major prediction of the theory concerns conjunction search: the search for a target that is defined by a conjunction of features, say a red vertical target in a field of red horizontal and green vertical distractors. According to the theory, conjunction search requires serially attending to individual items and thus should be inefficient because the target can be distinguished from distractors only by examining multiple features simultaneously. In many cases experimental results are just what the theory predicts.

However, subsequent experiments reveal that conjunction search can be relatively efficient (under 10 ms per item). For example, it has been shown that conjunctions of depth plane with color or motion produce very efficient search with a flat search slope (Nakayama & Silverman, 1986), and even a conjunction of color and orientation can produce efficient search (Wolfe et al., 1989). These conflicting results to the feature integration theory have led to modern revisions of the theory by Treisman and others (Treisman & Sato, 1990). The so-called attentional engagement theory is composed of three components: a parallel description of individual items in a scene, matching between individual descriptions with a target description, and finally, weight linkage that serves to spread suppression, caused by a mismatch, between like items (Duncan & Humphreys, 1992). The attentional engagement theory is motivated by the observation that search performance depends on how similar distractors are to each other and how dissimilar a target is from distractors (Duncan & Humphreys, 1989). The most notable revision is the guided search model by Wolfe (Wolfe et al., 1989; Wolfe, 1994). The model hypothesizes two stages of processing and the first stage, consisting of parallel feature maps, is very similar to the feature integration theory. The difference lies in the second stage, where the serial deployment of attention is guided by combined bottom-up activation and top-down activation. Bottom-up activation is computed as the difference of an item from others in a local neighborhood, or in general by local spatial filters. Top-down activation results from selectively triggering the feature dimensions that characterize a target object. Thus, searching for a conjunctively defined red vertical item, for instance, benefits from simultaneous top-down activation of a color map and an orientation map, which can explain efficient search performance. The guided search model not only avoids the pitfall of conjunction search in the feature integration theory, but also makes underlying mechanisms explicit, thus allowing for successful computer simulations of a range of empirical data (Wolfe, 1994; Wolfe, 1998).

Largely missing from these theories are visual proceses that occur at an intermediate level of visual analysis or representation. Processes that are related to perceptual organization (Koffka, 1935; Kanizsa, 1979) have received much less attention in providing an account for visual search. Perceptual organization refers to processes by which visual elements are organized into global perceptual units. This state of affairs is somewhat surprising considering the acknowledged importance of such intermediate processes for visual perception (Marr, 1982; Nakayama et al., 1995). In addition, numerous studies provide strong hints that perceptual organization is at play. The so-called subset search shows that with explicit instruction to focus on a subset of items defined by a feature dimension (for instance, red), the search for a conjunction target can be significantly facilitated (Egeth et al., 1984; Bacon & Egeth, 1997). Duncan and Humphreys (Duncan & Humphreys, 1989) showed that the use of a set of homogeneous distractors facilitates search performance, suggesting the role of similarity-based grouping of the distractor set. A recent study of conjunction search using form and motion shows that adding a set of dots moving together with either a target or distractors improves search performance significantly (Kingstone & Bischof, 1999). Furthermore, it has been shown that three-dimensional layout plays a crucial role in facilitating visual search (Ramachandran, 1988; Enns & Rensink, 1990; He & Nakayama, 1992), and in each of these studies search becomes much more efficient if the distractor set can be seen as a group of elements lying on a smooth surface.

Despite these indications in the experimental literature for an important role of perceptual organization, one of the problems in analyzing possible roles of intermediate level processing is the lack of appropriate experimental paradigms. We suggest that in any given visual search task, multiple visual processes are at play, which could influence search efficiency. It is likely that low-level feature extraction, mid-level grouping, and top-down selection are all present in any given search experiment. Indeed, in the above-cited studies the potential involvement of top-down selection considerably complicates the picture; in particular, the prior specification of a fixed target permits not only top-down guidance but also allows for target/distractor priming (see Experiment 3 below), both of which can improve search efficiency. What are critically needed are more clearly designed paradigms to isolate specific processes, in particular those that would depend on possible visual grouping.

As such, the working assumption of our study was that perceptual grouping processes are likely to be important for visual search, but new paradigms are needed to examine its characteristics in isolation. Our effort was to construct a visual search paradigm that would eliminate influences from both local bottom-up processing and top-down selection, thus isolating possible grouping mechanisms.

In this context, we developed a search paradigm called multiple conjunction search. This was extended from Treisman's conjunction task, in a way that the target is not uniquely defined on its own, but defined only with respect to other items on a display. In a typical experiment, items are varied along two dimensions, e.g. luminance (black or white) and topology (disk or annulus); see Figure 1. A target is defined as an odd-luminant disk *or* an odd-luminant annulus. Thus, one of four possible targets may appear on any trial: a black annulus among white annuli and black disks (Figure 1a), a white annulus among black annuli and white disks (Figure 1b), a black disk among white disks and black annuli (Figure 1c), or a white disk among black disks and white annuli (Figure 1d). In each of the four scenarios, the target is defined as a conjunction of luminance and topology (a disk and an annulus differ only in the presence/absence of a hole), hence a multi-conjunction. Like in typical conjunction search, the target appears randomly on 50% of trials. Items are equally distributed in each dimension, and are randomly placed on a computer screen. This layout eliminates bottom-up guidance, since, as in conjunction search tasks, local comparisons among items do not yield salient activations that can be used to guide visual attention (see Figure 1). More importantly, because every item in the display can on its own be either a target or a distractor, there can be no top-down guidance or expectation prior to each trial. In short, multi-conjunction search eliminates both bottom-up and top-down guidance.

For multi-conjunction search tasks, standard theories of visual attention predict inefficient search with substantial search slopes, because, lacking strong target saliency, attention would have to treat target and distractors alike in serial examination.

Experiment 1 shows that search for multi-conjunctions of color and orientation is indeed inefficient. In Experiment 2, we report that three stimulus combinations: luminance and topology, luminance and shape, and luminance and size, yield surprisingly efficient search. Experiment 3 further shows that a multi-conjunction task that combines two cases in Experiment 2 still yields efficient performance. . In Experiment 4, we modified a multi-conjunction task by simplifying the response procedure and found flat search slopes for both target-present and target-absent trials. In Experiment 5, we report that the multi-conjunctions of luminance and orientation also lead to inefficient search, demonstrating that luminance alone cannot explain efficient performance in Experiment 2. Finally, Experiment 6 demonstrates that priming for both targets and distractor sets leads to dramatic reduction of reaction times. The role of perceptual organization in visual search is further discussed in light of the six experiments in the general discussion section.

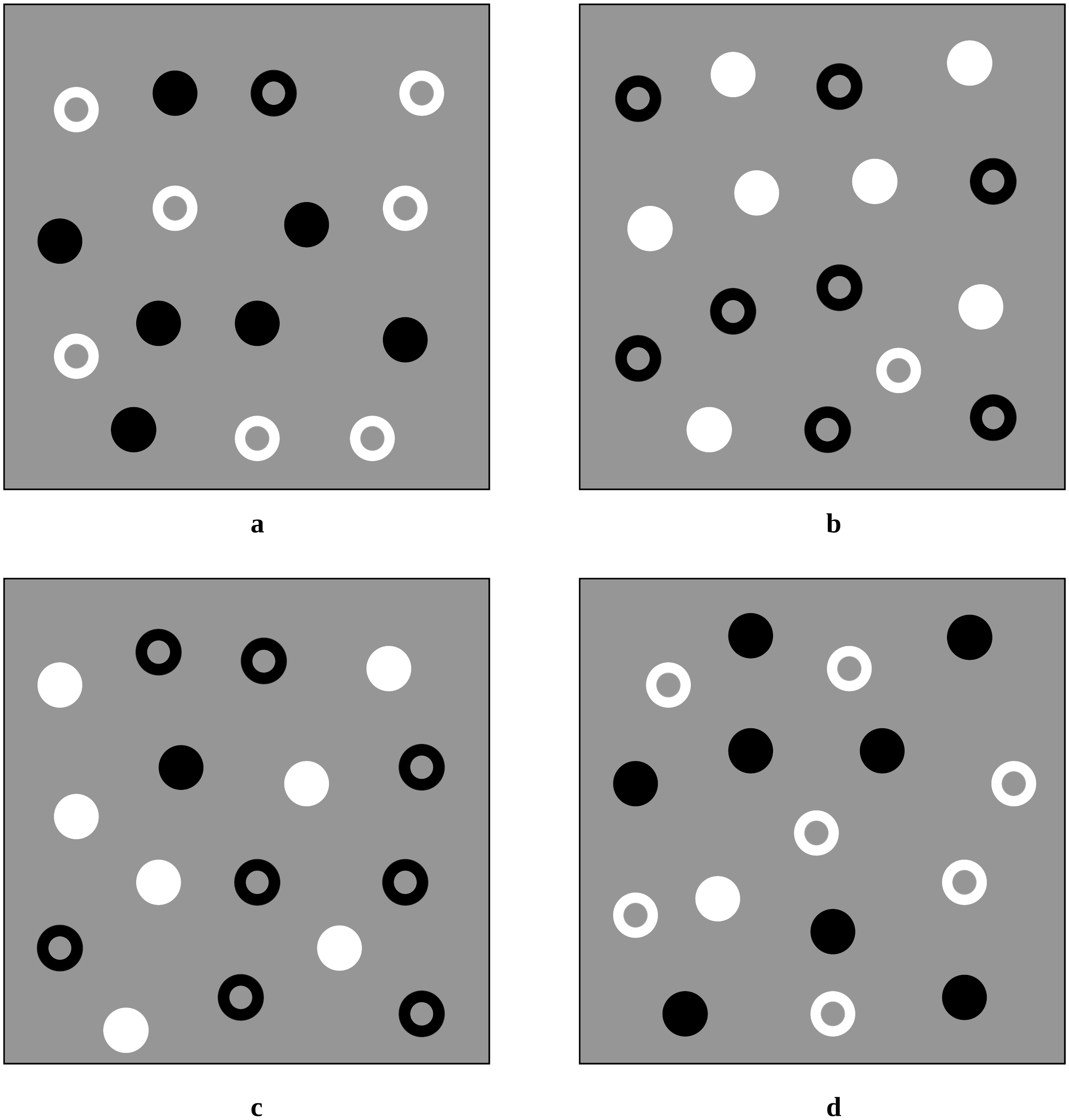


Figure 1. Four stimulus arrays illustrating different ways for a target to appear in multi-conjunction search: (a) the target is the black annulus, (b) the target is the white annulus, (c) the target is the black disk, and (d) the target is the white disk.

Experiment 1: Inefficient Multi-conjunction Search

The main purpose of this experiment was to draw a performance comparison between standard conjunction search and multi-conjunction search. Because color and orientation are the two feature dimensions commonly used in conjunction search, this experiment was to test search performance of the multi-conjunctions of color and orientation.

Methods

Apparatus and procedure. Stimuli were presented on a 75Hz monitor controlled by a Macintosh G3 computer. The display area lied at the center of the monitor, and was divided into 6x6 squares. The size of each square was 2.2°x2.2°. Each stimulus item fit within a square and was centered on one of nine random positions to produce some layout irregularity. Viewing distance was 57cm. On any given trial, the display array contained either 8, 16, 24, or 32 items, and a target occurred with a 50% chance. When a target occurred, it always replaced one of the distractors. Set size, locations of targets and distractors, and presence or absence of target were randomly varied across trials.

Subjects were told to maintain fixation at a small cross at the center of the monitor, but eye movements were not checked. Subjects were instructed to respond by pressing two designated keys to whether the target was present or absent as quickly as possible, while maintaining a high degree of accuracy. The key corresponding to the target-present case needs to be pressed by the right hand, and to the target-absent case by the left hand. The display remained on the screen until the subject responded. A feedback tone was given to subjects right after each trial to inform whether their response was correct or not. Data associated with mistakes were discarded.

Each subject participated in a total of 330 trials. The first 30 trials were for practice purposes, and the associated data were discarded. The remaining trials were divided into 3 blocks of 100 trials each. Subjects were allowed to take a short break between blocks.

Stimuli. Each stimulus item was a rectangular bar of the size 1.1°x0.28°, oriented either horizontally or vertically. The color of each item was either red (12.7 cd/m^2) or green (11.8 cd/m^2), and the background was black (0.5 cd/m^2). On a given display, except for a target item, half of the items were red and the other half were green; red items had one orientation and green items had the other orientation. A target was defined as an odd-colored horizontal item or an odd-colored vertical item. See Figure 1 for a general relationship between a target and distractors in a typical multi-conjunction task.

Subjects. Six subjects recruited from Harvard University participated in the experiment. They all had normal or correct-to-normal visual acuity and normal color vision. None of the subjects were aware of the purpose of the experiment.

Results and Discussion

Figure 2 shows the RT results for both target-present and target-absent cases. For the target-present case, the search slope is 24 ms/item, while for the target-absent case, the slope is 49.3 ms/item. Such search performance is consistent with a typical serial search, whereby attention presumably processes items sequentially until the target is identified. Moreover, the performance is significantly worse than in conjunction search. For conjunction search with a very similar stimulus condition of color and orientation, Friedman-Hill and Wolfe (Friedman-Hill & Wolfe, 1995) obtained a slope of 7 ms/item for the target-present case and 8.9 ms/item for the target-absent case. Comparable performances were obtained by Wolfe et al. (Wolfe et al., 1989), Treisman and Sato (Treisman & Sato, 1990), and also by us in a recent study (Kristjansson et al., 2001). Not only are slopes steeper, but intercepts are also significantly higher in multi-conjunction search. Table 1 gives percentage error rates for both target-present and target-absent cases and for all set sizes; the average error rate in this experiment is 9.7%.

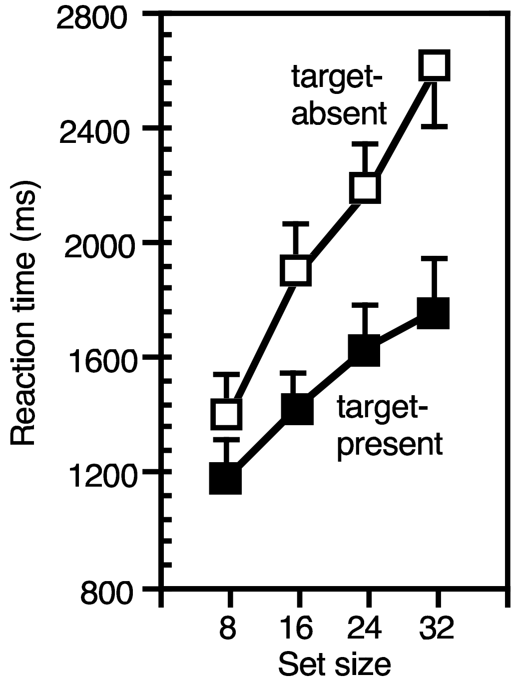


Figure 2. Mean reaction times plus SEM with respect to set size in the multi-conjunction search of color and orientation. The slope for the target present case is 24 ms/item, and for the target absent case is 49.3 ms/item. The average error rate is 9.7%.

Table 1. Error rates in percentages for Experiment 1

	Set size			
	8	16	24	32
Target present	5.5	5.1	7.8	8.1
Target absent	3.1	4.3	4.9	6.3

As analyzed earlier, standard search theories predict inefficient search for multi-conjunction tasks, because, lacking top-down or bottom-up guidance, attention would have to treat all items alike in serial examination. The prediction is confirmed by the data from this experiment.

Many other pairs of feature dimensions could be explored in the multi-conjunction paradigm, however, and counter-examples to the predictions of extant theory would be particularly interesting, since they would reveal the likelihood of perceptual grouping mechanisms at work in visual search. Towards this end, the next experiment tested on the following three conditions: luminance and topology, luminance and shape, and luminance and size.

Experiment 2: Efficient Multi-conjunction Search

This set of experiments was designed to address the question of whether efficient search was at all possible in multi-conjunction tasks. In the experiments, subjects searched for targets defined by the multi-conjunctions of luminance with topology, shape, and size, respectively.

Methods

The apparatus and procedure part of this experiment was identical to Experiment 1.

Stimuli. As illustrated in Figure 1, each stimulus item was either black (0.5 cd/m^2) or white (50.9 cd/m^2), and the background was gray (33 cd/m^2). Three stimulus conditions were tested. For the condition of luminance and topology (see Figure 1), disks and annuli were used. The diameter of a disk and the outer circle of an annulus is 0.82°, and that of the inner circle of an annulus is 0.43°. The target in this condition was defined as an odd-luminant disk or an odd-luminant annulus. For the condition of luminance and shape, circles and horizontal bars were used. The circles are of the same size as those used in the first condition, and the size of a rectangular bar is 1.4°x0.35° so that the area is the same for the two shapes. The target in this condition was defined as an odd-luminant circle or an odd-luminant bar. For the condition of luminance and size, squares of two sizes were used. The size of a large square is 0.82°x0.82°, and that of a small square is 0.41°x0.41°. The target in this condition was defined as an odd-luminant small square or an odd-luminant large square. The stimulus patterns used in the three conditions are illustrated in the inset of Figure 3.

Subjects. Six new subjects recruited from the Ohio State University graduate student population participated in the experiment. They all had normal or correct-to-normal vision. They were not only naive to the purpose of the experiment, but also had never participated in any psychological experiment before. Each subject participated in all three stimulus conditions, for a total of 930 trials.

Results

Figure 3 shows the RT results for the three stimulus conditions. As in Figure 2, both target-present and target-absent data were given. For the multi-conjunctions of luminance and topology, the search slope is near zero for the target-present case. For luminance/shape and luminance/size, the search slopes are 3.3 ms and 7.6 ms per item, respectively. It is clear from Figure 3 that search is efficient for all three conditions, much more so than the condition with orientation and color using the identical paradigm. Search is less efficient when target is absent, but this pattern of target-absent performance in Figure 3 seems too efficient to support a serial rejection process. The target-absent case will be examined in Experiment 4, and we focus on the target-present case for now.

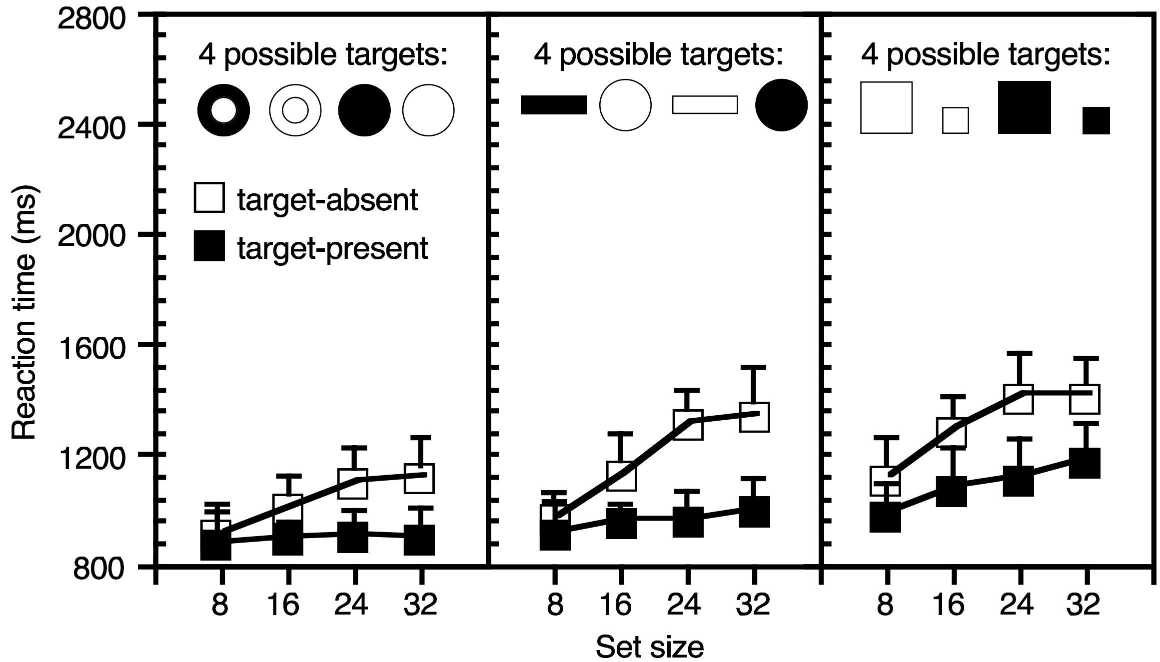


Figure 3. Mean reaction times plus SEM with respect to set size in multi-conjunction search for three different conditions: topology (discs and annuli), shape (bars and discs), and size (small and large squares). The three stimulus conditions are illustrated in the figure inset. Each item is either black or white on a gray background. For the target-present cases the slopes for the three conditions were: 0.7 ms/item, 3.3 ms/item, and 7.6 ms/item, respectively, and for the target-absent cases the slopes were 9.1 ms/item, 12.6 ms/item, and 16.1 ms/item, respectively. The average error rates are 8.1%, 6.8%, and 9.1%, respectively.

Table 2 gives percentage error rates for the three multi-conjunction tasks for both target-present and target-absent cases and for all set sizes. The average error rate is 8.1% for the topology task, 6.8% for the shape task, and 9.1% for the size task. Unlike the multi-conjunctions of color and orientation, error rates are much higher for target-present cases than for target-absent cases. This will be discussed after Experiment 3 is reported below. We found little correlation between error rates and set sizes.

Table 2. Error rates in percentages for the three conditions of Experiment 2

		Set size			
		8	16	24	32
Topology	Target present	15.6	15.9	14.7	16.7
	Target absent	3.1	4.7	0.8	0.9
Shape	Target present	14.6	12.3	14.6	15.4
	Target absent	5.2	3.1	4.6	3.6
Size	Target present	13.1	16.2	15.2	15.7
	Target absent	2.7	2.3	2.1	2.7

Search efficiencies in Figure 3 differ dramatically from those of Figure 2. Contradicting what is predicted by standard theories of visual search, for these stimulus conditions efficient search can take place without either top-down or bottom-up guidance. Before we discuss the implications of these surprising results, let us first report on a subsequent experiment.

Experiment 3: Efficient Combined Multi-conjunction Search

Typical search tasks involve sessions of many consecutive trials (e.g. 100) where target and distractor identities do not vary from trial to trial. This experimental protocol leaves open the question whether earlier trials have a facilitory influence, or priming, on the present trial, as it has been demonstrated that priming can improve search performance considerably (Maljkovic & Nakayama, 1994; see also Kristjansson et al., 2001). In the present situation, such priming is unlikely to be a significant factor because one of four random conjunctions defines the target and the distractor set in each trial. To argue even more strongly against this possibility, we conducted an experiment that combined two conditions examined in Experiment 2: the luminance and topology condition and the luminance and size condition.

Methods

The apparatus and procedure part of this experiment was identical to Experiment 1.

Stimuli. The stimulus patterns used in this experiment were the same as those used in the luminance/topology and the luminance/size conditions of Experiment 2. The target was defined as an odd-luminant disk, an odd-luminant annulus, an odd-luminant small square, or an odd-luminant large square. As illustrated in Figure 4a, distractors on a display are either small and large squares of different luminance, or disks and annuli of different luminance, that are consistent with the target item.

Subjects. They were the same six subjects who participated in Experiment 2. The subjects were naive to the purpose of the experiment.

Results

In this experiment, not even the dimensions along which features can vary are constant. On each trial, a target, when present, can be one of the *eight* conjunctively defined items: four with disks and annuli, and another four with small and large squares. Like a target, a distractor on a given trial can be one of the eight possible items. Figure 4b shows the RT results of this experiment. When a target is present, search is very efficient, with a search slope less than 4 ms/item. Table 3 gives percentage error rates for both target-present and target-absent cases and for all set sizes; the average error rate in this experiment is 7.7%. The overall pattern of results for this multi-conjunction task is similar to that in Experiment 2. This experiment shows that efficient search can occur with negligible repetition of either targets or distractor sets, thus reinforcing the results of Experiment 2.

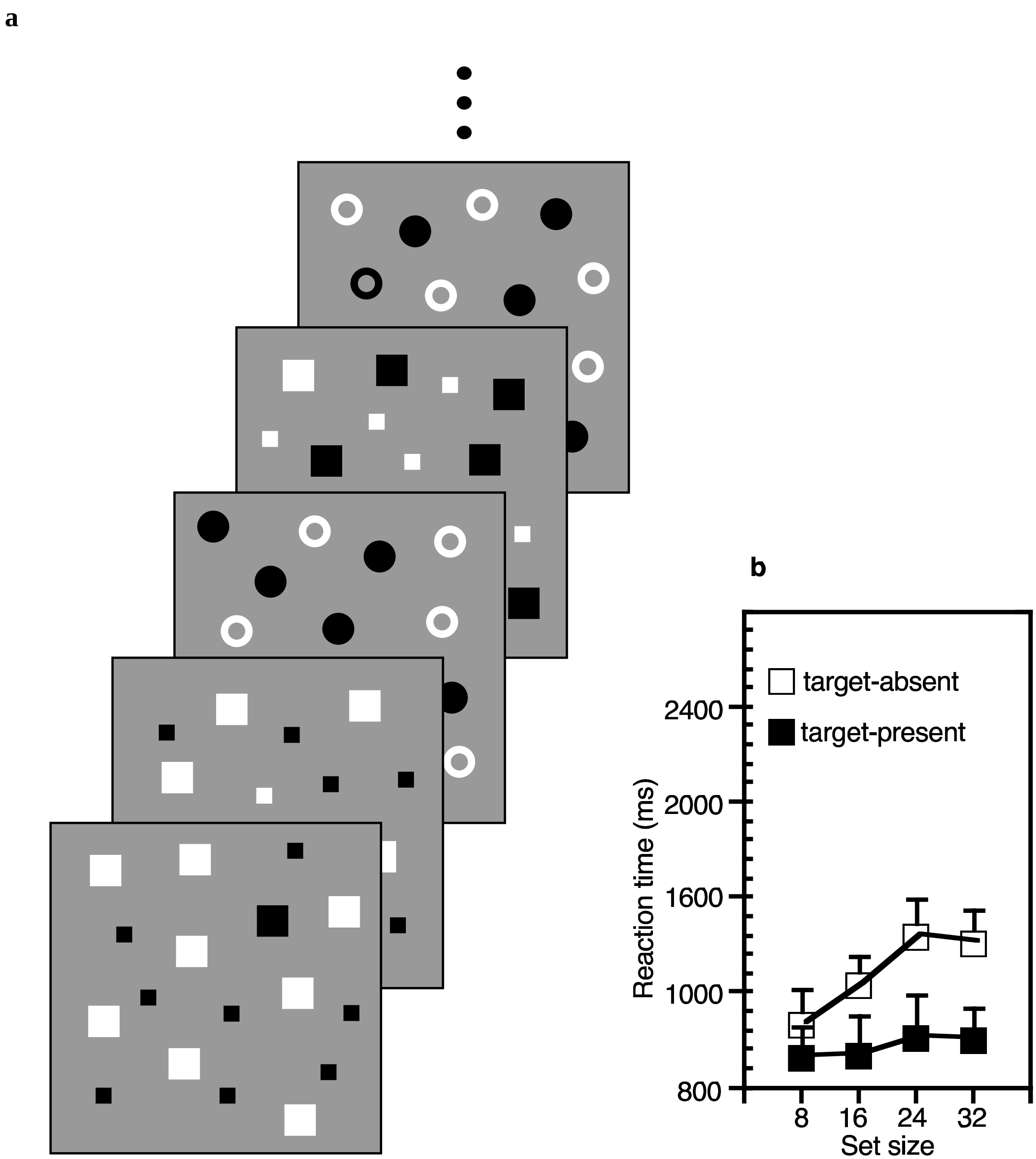


Figure 4. Multi-conjunction search for a combination of topology and size together with luminance. (a) A sequence of possible stimulus arrays illustrating multi-conjunction search for a combination of disk/annulus and small/large squares. The target is absent in array 3 and present in the other four arrays. (b) Mean reaction time plus SEM with respect to set size. The slope is 3.8 ms/item for the target-present case, and 15.5 ms/item for the target-absent case. The average error rate is 7.7%.

Table 3. Error rates in percentages for Experiment 3

	Set size			
	8	16	24	32
Target present	13.8	15.0	14.8	13.5
Target absent	3.6	5.3	3.3	5.3

Discussion of Experiments 2 and 3

These two experiments clearly demonstrate that efficient search can occur without either top-down or bottom-up guidance, and therefore contradict a basic prediction of the feature integration theory and its modern revisions that efficient search requires either top-down or bottom-up guidance. The multi-conjunction search paradigm diminishes the potential role of priming for either targets or distractors. In particular, Experiment 3 provides strong evidence for ruling out priming benefits as a cause for the observed efficient search performance. In other words, the subjects appear to perform efficient search for targets on the basis of each individual display alone.

With top-down or bottom-up guidance eliminated, almost flat target-present search slopes in these multi-conjunction tasks point to perceptual grouping as the cause for efficient performance. In other words, subjects appear to deal with items on a display as groups. More specifically, items in a visual scene appear to be organized into perceptual groups or segments, and each segment behaves as a basic unit for attentional processing. A segment could be a single item, or a group of items. Our results suggest that, for the examined stimulus conditions, items of like luminance and like topology, shape, or size may be grouped into single segments and they may be separated from those with different stimulus values. Consider Figure 1 for example. Items may be grouped together based on luminance, topology, or both. When a target is present, perceptual organization would produce a small number of segments, one of which would be the target itself. Attention would then process each segment serially, and should be able to identify the target efficiently because of the small number of groups that need to be examined. This account implies that perceptual grouping takes place on the basis of each display alone and not from priming built up from previous trials, consistent with our data.

As mentioned earlier, in the target-absent case there are positive search slopes in both Experiments 2 and 3. How an observer reaches the no-target conclusion in a search task is not well understood, and possibly involves a complex decision process. For instance, in the context of the guided search model, Chun and Wolfe (Chun & Wolfe, 1996) introduced an activation threshold, below which no distractor will be examined in the serial examination process of a display array. Therefore, some distractor items with very weak activation are eliminated from the search process altogether. They further assumed that this threshold is an adaptive one, increasing following correct responses (thus examining less of the display screen subsequently) and decreasing following incorrect responses (thus examining more of the screen subsequently).

Notice that there is a task asymmetry between the target-present case and the target-absent case. The fact that an observer in a visual search experiment is instructed to look for a target gives the target-present case precedence over the target-absent case. The observer sets out to find a target on a display and, only when failing to do so, switches to decide on target absence. As specified in the Methods section of Experiment 1, the keys corresponding to these two cases require two different hands to press. This task asymmetry appears to be supported by a prominent trend of the error rates in Tables 2 and 3 that a large majority of search errors occurred for the target-present case. That is, the subjects missed a target much more frequently than mistaked a distractor as target. Given this fact, the Chun-Wolfe analysis on target-absent trials suggests that the subjects tend to be less certain when they do not see a target and thus examine more of the display before they press the target-absent key. It is possible that the amount of time spent on this extra screen examination is related to set sizes, yielding the upward slopes for target-absent cases in Experiments 2 and 3. Our next experiment attempts to minimize this task asymmetry by asking subjects to search for target presence and target absence separately.

Experiment 4: Go/No-Go Search

This experiment was conducted to further examine the multi-conjunctions of luminance and topology. In particular, trials were divided into sessions, within each of which subjects were instructed to search for either target presence or target absence, but not both; this is called a "Go/No-Go" task. For example, within a target-presence session, subjects were instructed to respond if a target was present and do nothing otherwise; likewise for target-absence sessions. This way the search response is made simpler and the task asymmetry is removed.

Methods

Apparatus and procedure. Like in previous experiments, stimuli were presented on a 75Hz monitor of a Macintosh G3 computer. In a Go/No-Go target-presence session subjects were instructed to respond with a key press only if the target was present (on 50% of trials), but to "sit and wait" if the target was absent, in which case the trial ended after 1,500 ms. In a Go/No-Go target-absence session, subjects were instructed to respond with a key press only if the target was absent. As a control condition, subjects also performed a standard multi-conjunction search (indicating whether the target was present or absent by pressing corresponding keys with two hands), repeating the luminance and topology condition of Experiment 2.

Each subject participated in three sessions: one for multi-conjunction, one for Go/No-Go target-presence, and one for Go/No-Go target-absence. The order of the sessions is randomly chosen for each subject. Each session consisted of 330 trials, and the first 30 trials were for practice purposes and their associated data were not collected.

Stimuli. The stimulus patterns used in this experiment were the same as those used in the luminance and topology condition of Experiment 2. The target was defined as an odd-luminant disk or an odd-luminant annulus, as illustrated in Figure 1.

Subjects. Nine observers from the Harvard University undergraduate student population participated in this experiment for course credit. They all had normal or correct-to-normal vision, and were naive to the purpose of the experiment.

Results and Discussion

Figure 5 shows the RT results of this experiment. For the multi-conjunction search, the slope for the target-present case is -1.1 ms/item and for the target-absent case is 8.3 ms/item. This pattern of results is very similar to that reported in Experiment 2. The higher intercepts in Figure 5 likely reflect the fact the subjects participated in this experiment were different from those participated in Experiment 2.

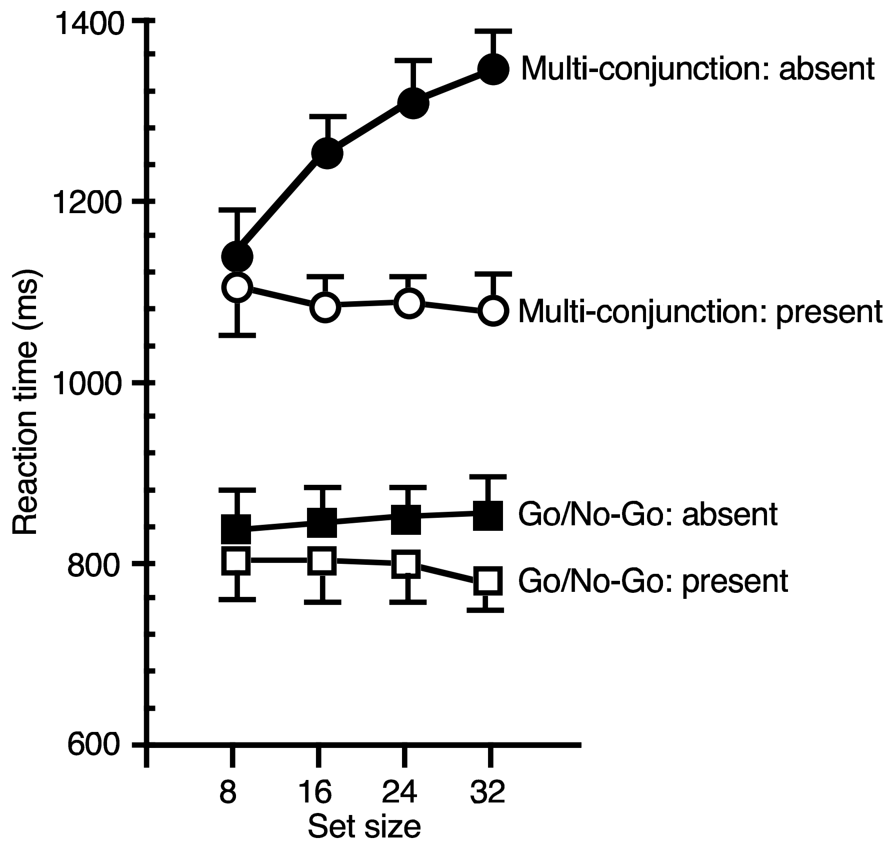


Figure 5. Mean reaction times plus SEM with respect to set size in multi-conjunction and Go/No-Go search tasks for the luminance and topology condition. For the multi-conjunction task the slope is -1.1 ms/item for the target-present case and 8.3 ms/item for the target-absent case. The average error rate for this task is 6.8%. For the Go/No-Go task the slope is -1 ms/item for the target-present case and 0.9 ms/item for the target-absent case. The average error rate for this task is 2.4%.

For the Go/No-Go procedure, the slope for the target-present case is -1 ms per item and for the target-absent case is 0.9 ms per item. Comparing the Go/No-Go data with those of multi-conjunction, the most significant difference lies in target-absent performance. In the new procedure, search slopes are flat for target-present as well as for target-absent cases. Additionally, there is a large drop of RT for the Go/No-Go procedure, which has a lower task demand: the subjects did not need to switch between the two cases with two different motor responses.

Table 4 gives percentage error rates for the above two experimental conditions. Again, error rates are provided for both target-present and target-absent cases and for all set sizes. The average error rate for the multi-conjunction task is 6.8%, and for the Go/No-Go task is 2.4%. Note that for the Go/No-Go task there are no appreciable differences in error rates for the target-present and the target-absent case.

Table 4. Error rates in percentages for two conditions of Experiment 4

		Set size			
		8	16	24	32
Multi-conjunction	Target present	9.4	10.1	11.6	9.5
	Target absent	5.1	4.1	4.6	3.7
Go/No-Go	Target present	1.7	2.3	2.4	3.0
	Target absent	2.1	2.4	2.8	2.7

The results of this experiment demonstrate that after target-present and target-absent tasks are made simpler and more symmetrical with the Go/No-Go procedure, the pattern of RT results for both tasks follows the same trend: the time that subjects take to perform either task does not change as the set size varies. Even when the stimulus array does not contain a target, the subjects can process the array in parallel to perceive target absence. Because features that define distractor sets vary randomly across trials, the flat search slope for the target-absent case provides strong evidence against activation-based models of visual attention. Such models, in the absence of top-down guidance and with items from two distractor sets spatially interleaved, must resort to a serial deployment of attention from one item to another. On the other hand, the results lend further support to our explanation on the basis of perceptual organization (see the Discussion of Experiments 2 and 3).

Experiment 5: Inefficient Multi-conjunctions of Luminance and Orientation

So far, all efficient multi-conjunction tasks involve luminance as a stimulus feature. It has been shown that luminance can be more conducive to efficient search than color. For example, Theeuwes and Kooi (Theeuwes & Kooi, 1994) compared RT performance between luminance and color in a conjunction task involving the letters of Xs and Os, and they found efficient conjunction search with luminance but not with color. They attributed efficient search to dedicated neural pathways for luminance processing, suggesting a special status for luminance that is not shared by color. Thus, an important question in the present context is whether luminance alone is responsible for efficient multi-conjunction search. To investigate this question we performed the present experiment on the multi-conjunctions of luminance and orientation.

Methods

The apparatus and procedure part of this experiment was identical to Experiment 1.

Stimuli. Each stimulus item was a rectangular bar of the size 1.1°x0.28°, oriented either horizontally or vertically. The size of the rectangular bar was identical to that used in Experiment 1. Each item was either black (0.5 cd/m^2) or white (50.9 cd/m^2), and the background was gray (33 cd/m^2). This luminance layout was identical to that in Experiment 2. A target was defined as an odd-luminant horizontal item or an odd-luminant vertical item. See Figure 1 for a general relationship between a target and distractors in a typical multi-conjunction search.

Subjects. Six new subjects, plus one who participated in Experiments 2 and 3, were tested. The subjects were recruited from the Ohio State University graduate student population. They all had normal or correct-to-normal vision, and they were naive as to the purpose of the experiment. Five out of the seven never participated in any psychological experiment before.

Results and Discussion

Figure 6 shows the RT results for both target-present and target-absent cases. For the target-present case, the search slope is 23.8 ms/item, while for the target-absent case, the slope is 70 ms/item. Table 5 gives percentage error rates for both target-present and target-absent cases and for all set sizes. The pattern of the search performance is similar to that of Experiment 1, indicating serial deployment of attention. The higher intercepts in Figure 6 compared to Figure 2 likely reflect the fact that a new subject group participated in this experiment. Furthermore, unlike those in Experiment 1, many of the subjects in this experiment never had any experience with psychological experiments.

Table 5. Error rates in percentages for Experiment 5

	Set size			
	8	16	24	32
Target present	3.6	3.8	4.6	6.1
Target absent	3.3	4.0	4.1	5.1

The results of this experiment demonstrate that luminance alone cannot explain efficient performance in multi-conjunction search. The use of luminance could facilitate some multi-conjunction tasks. Such a facilitating role is supported by our informal observation that the search efficiency for the multi-conjunctions of color and shape, involving circles and bars (see Figure 3), is somewhat reduced compared to those of luminance and shape. Taken together, these observations indicate that luminance and other stimulus features, such as topology and shape, combine to yield efficient multi-conjunction search. From the perspective of perceptual organization, this implies that grouping of items may integrate multiple feature dimensions, rather than relying on a single individual dimension.

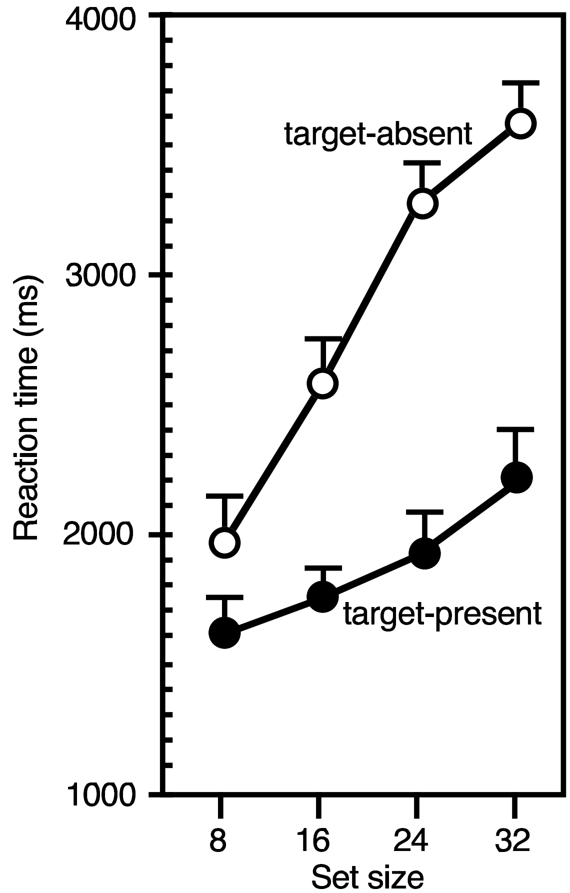


Figure 6. Mean reaction times plus SEM with respect to set size in multi-conjunction search of luminance and orientation. The slope for the target present case is 23.8 ms/item, and that for the target absent case is 70 ms/item. The average error rate is 4.3%.

The results of this experiment demonstrate that luminance alone cannot explain efficient performance in multi-conjunction search. The use of luminance could facilitate some multi-conjunction tasks. Such a facilitating role is supported by our informal observation that the search efficiency for the multi-conjunctions of color and shape, involving circles and bars (see Figure 3), is somewhat reduced compared to those of luminance and shape. Taken together, these observations indicate that luminance and other stimulus features, such as topology and shape, combine to yield efficient multi-conjunction search. From the perspective of perceptual organization, this implies that grouping of items may integrate multiple feature dimensions, rather than relying on a single individual dimension.

Experiment 6: Priming in Multi-conjunction Search

For efficient search based on individual features or a conjunction of features as in efficient conjunction search, RT intercepts typically range from 400 ms to 700 ms for target-present cases (Treisman & Gelade, 1980; Nakayama & Silverman, 1986; Wolfe et al., 1989; Treisman & Sato, 1990; Theeuwes & Kooi, 1994; Friedman-Hill & Wolfe, 1995). In comparison, intercepts in

Figures 2 and 3 are relatively high. Multi-conjunction search is more demanding than feature or conjunction search, and subjects would be expected to take longer to respond. Experiment 4 demonstrates that a significant drop in RT intercepts results when the response procedure is simplified.

As alluded to earlier, one potential benefit not available in multi-conjunction search is priming, which can occur when targets or distractors repeat their identities in consecutive trials. Typical exprimental protocals employed in feature or conjunction search are potentially subject to priming, which can significantly improve overall search performance (Maljkovic & Nakayama, 1994; Kristjansson et al., 2001). In this experiment, we modified the experimental procedure of a multi-conjunction task in order to make it subject to potential priming, and tested whether priming could bring down RT intercepts of multi-conjunction search.

Methods

Apparatus and procedure. Unlike typical multi-conjunction search where the target identity changes randomly between four possible ones from one trial to the next, in this experiment the target retained its identity for longer "streaks" than would be expected from a random choice of the target identity. This allows for documenting the effects of the repetition of the target (and distractor) identity across a number of trials - up to 7 in a row in this experiment. Specifically, streaks were generated in the following way. The probability that the target on the previous trial repeats in the current trial was set to $1–N(0.1–0.01N)$, where N indicates how many times the same target has been used in a streak. It is important to note that the target identity determines the corresponding distractor sets whether a target actually occurs on the display or not. Hence, the target identity must be decided for each trial. With the repeating target identity, distractor identities repeat even when a target does not occur on a trial within a streak; in other words, N counts whether or not target occurs on a given trial. For example, assume that the target in the current trial is a small square and the target in the previous trial is a different one. According to the probability formula, for the next trial the small square is used as target, whether it is present or not, with a probability of 0.91 and for the subsequent trial the same target is used with a probability of 0.84, and so on. The probability was set to the asymptote of 0.75 (when $N = 5$), and it was further set to 0 when $N = 7$ since the maximum streak length was set to 7. As a result, the target identity was never completely predictable from one trial to the next, but overall, the probability that the same target would repeat from one trial to the next was considerably greater than 1/4.

Each subject participated in total of 830 trials. The first 30 trials were for practice only and their associated data were not collected. The remaining 800 trials were divided into 4 blocks of 200 trials each, and subjects were allowed to take a short break between blocks. Compared to typical multi-conjunction search, more trials were needed for analyzing various streak lengths.

The rest of the apparatus and procedure part is the same as in Experiment 1.

Stimuli. The stimulus patterns used in this experiment were the same as those for the size/luminance condition of Experiment 2. On each trial, the target could be one of the following: a small black square, a small white square, a large black square or a large white square. Half of the distractors on each trial would differ from the target in size and the other half would differ from the target in luminance (either white or black).

Subjects. Six observers from the Ohio State University graduate student population participated in this experiment, and four of them had also participated in Experiments 2 and 3. They all had normal or correct-to-normal vision, and were naive to the purpose of the experiment.

Results and Discussion

We tested for potential priming for both targets and distractor sets by employing a streak condition, under which the same target and distractor sets in the multi-conjunctions of luminance and size repeated for long stretches of consecutive trials (see the above Methods). If there exist priming benefits, one expects to see response times drop as the number of repetitions increases. Figure 7a presents reaction times as a function of the position of a trial within a streak for both target and distractor repetitions. The data for the target-present case in Figure 7a show striking priming benefits. The negative slopes between RT and number in streak indicate that target repetitions speed up search by more than 60 ms for each repetition. The benefits are more dramatic for the first few repetitions than for later ones. The pattern of improvement is similar for different set sizes, consistent with our finding that the RT slope for this multi-conjunction task is small (see Figure 3). The right panel of Figure 7a presents data for distractor repetitions in a similar fashion. Again, there are strong priming effects for repetitions of distractor sets. Greater improvements occur for larger set sizes, and this correlates with the data in Experiment 2 that reaction times are higher for larger set sizes in target-absent trials. Strong priming effects for both targets and distractor sets suggest that priming in multi-conjunction search works on the whole search array rather than being bound only to individual items. This observation is consistent with our recent study on priming in conjunction and subset search paradigms (Kristjansson et al., 2001), and provides further evidence for the involvement of perceptual organization. Figure 7b provides the RT results averaged across different set sizes for both target-present and target-absent cases. The overall result of this experiment strongly demonstrates that priming can play a large role in visual search that is often ignored in the literature (see Kristjansson et al., 2001 for further discussions)

Unlike multi-conjunction search, stimulus dimensions that define targets and distractors in feature search or conjunction search do not vary from trial to trial. That is, there are constant repetitions of target and distractor identities, which correspond to an infinitely long streak situation. Thus, it is reasonable to conclude that feature search and conjunction search enjoy maximal priming benefits. Adding such benefits at the maximum streak length in Figure 7b, about 400 ms for the target-present case and 250 ms for the target-absent case, to our data on the multi-conjunctions of luminance and size brings down reaction times of this task to the typical range associated with efficient feature and conjunction search tasks. In other words, reaction times in multi-conjunction search are more or less comparable to regular search, after priming is taken into consideration.

General Discussion

The results of the experiments presented in this paper clearly demonstrate that very efficient visual search, typically associated with popout, can occur in the absence of either top-down or

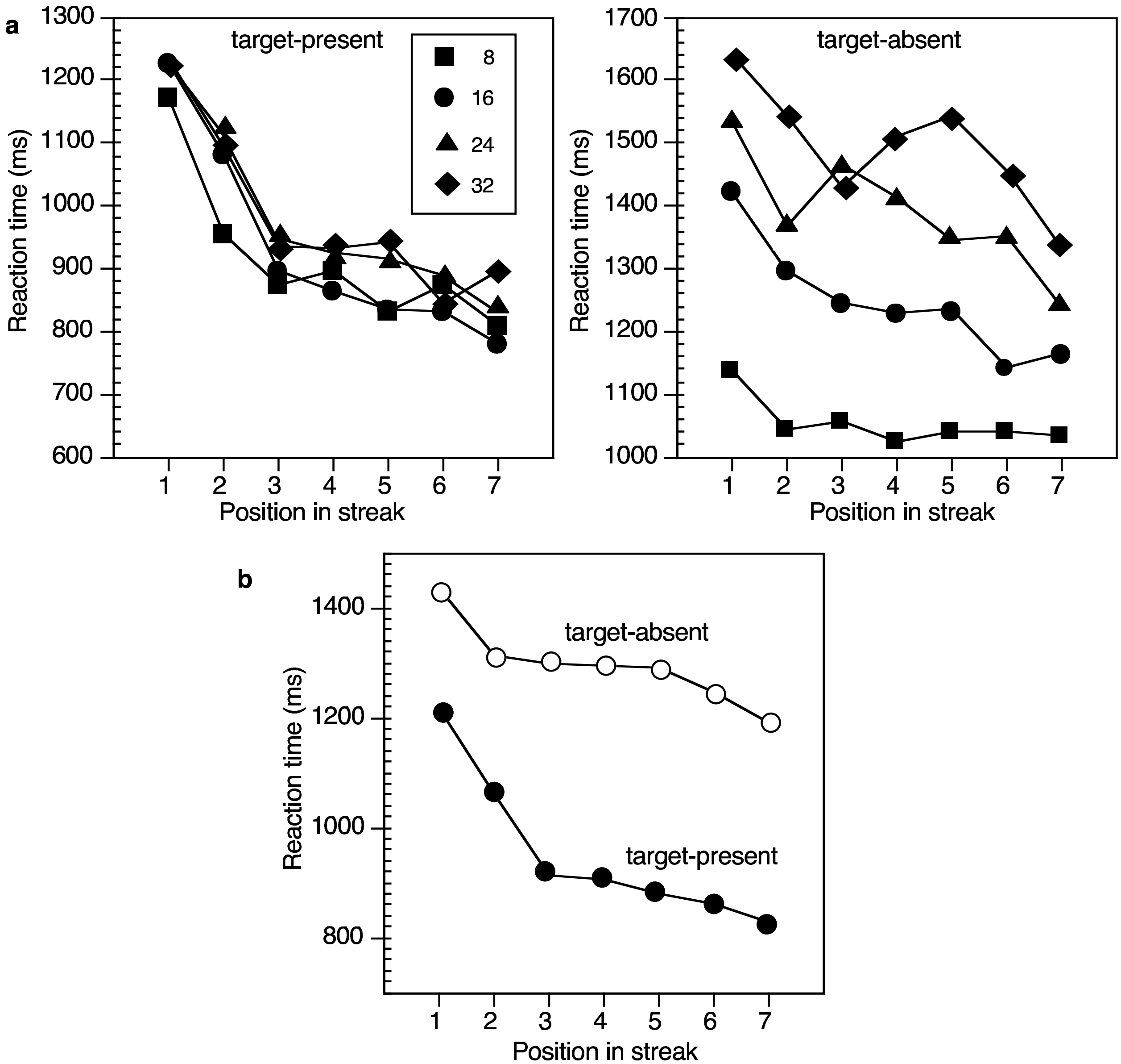


Figure 7. Priming effects for multi-conjunctions of luminance and size. (a) Mean reaction times as a function of position in a streak (see text). The target-present case is shown on the left and the target-absent case on the right. Data for different set sizes are displayed separately. (b) Mean reaction times with respect to position in a streak, collapsed across all the set sizes.

bottom-up guidance. Standard theories of visual attention rely on such guidance to explain efficient search, and are thus untenable. Our results instead suggest an important role for perceptual organization in visual search, which may be a decisive factor for efficient search performance. Two main reasons lead us to this suggestion. First, after top-down and bottom-up guidance are eliminated, there are no other credible explanations known to us except for an explanation at the mid-level of visual processing. Second, perceptual organization, which occurs at an intermediate level of visual analysis, has been hinted in a number of previous studies that reported efficient search beyond simple two-dimensional features (Ramachandran, 1988; Enns &

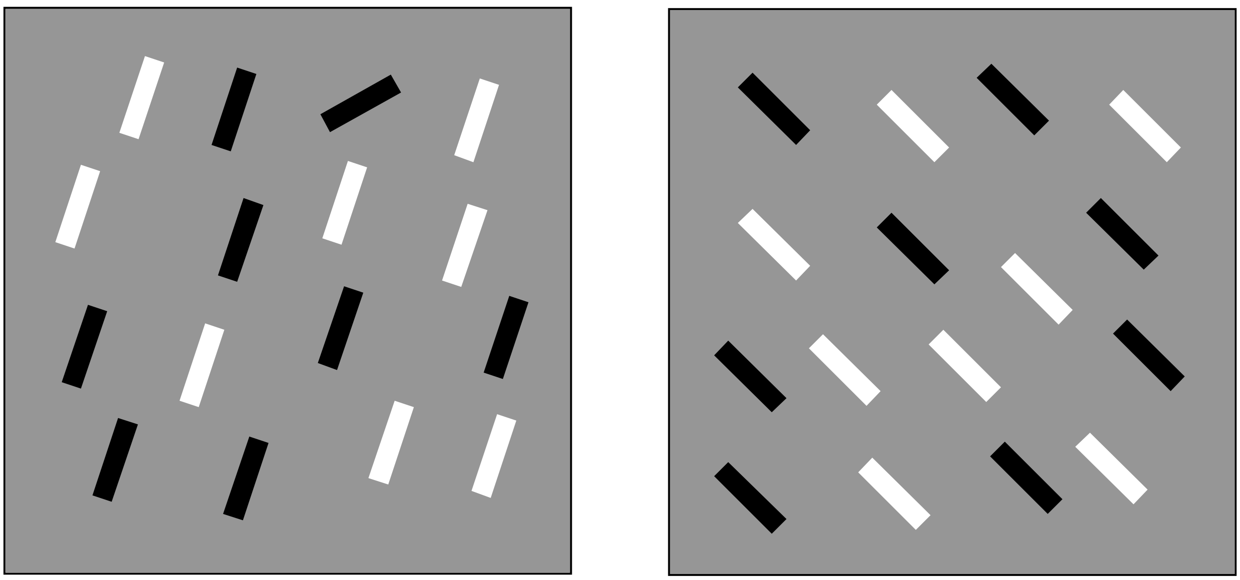
Rensink, 1990; He & Nakayama, 1992). Because conventional search paradigms do not separate bottom-up, intermediate level, and top-down processes, previous results are subject to different interpretations. The multi-conjunction search paradigm introduced in this paper makes it possible to eliminate potential influences from top-down and bottom-up processes. Thus, our efficient search results provide much stronger evidence for the involvement of perceptual organization.

Although perceptual organization is based on characteristics of local scene elements, we stress that it differs from bottom-up guidance (Wolfe, 1994; Niebur & Koch, 1998). First, bottom-up guidance is generated by local computation via spatial filters. In contrast, perceptual organization yields global structures via a dynamical process of grouping and segmentation, or emergent grouping that occurs across an entire display. Second, bottom-up guidance serves to guide attention to a particular item or location; perceptual organization does not directly give rise to such guidance. Our explanation is consistent with neural network models that emphasize emergent grouping based on local connectivity between model neurons (Grossberg et al., 1994; Wang, 1999).

One might interpret efficient multi-conjunction search as two subsequent subset searches in the following way, say on the basis of luminance. When searching for black items, white ones "fade" into the background and subset search would result in a feature search, and vice versa. This interpretation would imply that multi-conjunctions of luminance and any other feature lead to efficient search, contradicting inefficient search observed in Experiment 5 for the multi-conjunctions of luminance and orientation. Since color has been shown to be an effective feature for subset search (Egeth et al., 1984), one would also predict, by the same token, efficient search for the multi-conjunctions of color and orientation, incompatible with the data of Experiment 1. Our results suggest instead that a combination of features underlies efficient multi-conjunction, indicative of emergent grouping.

From the perspective of perceptual organization, the results of Experiment 2 suggest that items of the same luminance and the same topology, shape, or size are grouped to form an perceptual unit. In Figure 1a, for example, this entails that black circles are grouped together over spatially interleaved white annuli. On the other hand, for Experiments 1 and 5 involving orientation, grouping between like orientation items appears to be blocked by items of different orientation. There are indeed indications that orientation-based grouping is sensitive to spatial alignment between items. In particular, psychophysical evidence supports the notion of an association field that groups neighboring items of similar orientation that are aligned to form a curvilinear contour (Field et al., 1993; Kovacs & Julesz, 1993). Neurobiological findings also show that neurons in the primary visual cortex are interconnected by long-range horizontal connections, which link neurons with similar orientation responses (Rockland & Lund, 1983; Gilbert & Wiesel, 1989). Such alignment is not present in most visual search experiments (including ours). As such, it may provide an explanation as to why items of the same orientation do not group across interleaved items of different orientation.

In an early study linking perceptual grouping and visual search, Treisman (Treisman, 1982) found evidence suggesting that attention shifts serially between groups rather than individual items. In her experiments, groups were spatially well separated and items of different groups were not interleaved as those in Figure 1. Thus, spatial proximity can be an effective grouping cue. In the experiments of Humphereys and Duncan (Duncan & Humphreys, 1989), distractors were arranged either homogeneously or heterogeneously. In their homogeneous conditions, which produced efficient search, distractors can again be grouped based on proximity. Unlike these experiments, putative perceptual grouping in our experiments extends across items of different groups. That is, perceptual grouping among similar items is possible in the presence of interleaving, dissimilar ones. Previous studies have shown that such grouping is possible in conjunction with depth, motion, or color, all having dedicated neural filters in the visual system. Our findings in Experiments 2 and 3 are striking in that a combination of luminance with other features, which do not seem to have dedicated neural filters, could also yield such grouping. This makes the findings very hard to dismiss as preattentive exceptions because of the existence of neural pathways, a view that has persisted since early counterexamples surfaced for the feature integration theory.



a **b**

Figure 8. Orientation-based feature search. Items are either red (indicated by black) or green (indicated by white). (a) An example of the target-present case. The orientation of the target is different from that of other items. (b) An example of the target-absent case.

The results of Experiment 4, which demonstrates that a simpler Go/NoGo procedure yields efficient search not only for target-present trials but also for target-absent ones, are particularly revealing. The feature integration theory and the guided search model have never offered a clear explanation on efficient search for target-absent or blank trials. In the multi-conjunction paradigm, the activation map formed from combined bottom-up and top-down guidance cannot yield much guidance for visual attention, and futile but serial examination of the entire search array would have to follow. Because no top-down guidance is available, there would be no basis for setting a threshold on the activation map to exclude items from serial examination, as would be suggested from the analysis of Chun and Wolfe (Chun & Wolfe, 1996). In fact, this predicament exists even for feature search. For example, Figure 8 shows a feature search experiment conducted by Friedman-Hill and Wolfe (Friedman-Hill & Wolfe, 1995), where subjects searched for an odd-oriented bar in an array of homogeneously-oriented bars with two different colors (red and green). For both target-present and target-absent trials, search performance is efficient with flat search slopes. With target presence, the result fits well the feature integration theory. Without a target, however, how do subjects perceive target absence efficiently? Note that prior to each trial subjects have no idea about target orientation, and yet they can reject all the items in parallel! From a different perspective, perceptual organization based on orientation would readily explain the results for both target-present and target-absent cases.

A consistent picture that emerges from the six experiments described here is that visual attention serially examines perceptual groups rather than individual items. It is possible that, under laboratory conditions, groups are individual items themselves. But in many cases, a group emerges from a set of items. Our results suggest that an adequate theory of visual attention need to include perceptual organization as an important component. The inclusion of perceptual organization could provide a natural link for integrating visual search into the larger picture of visual perception.

Conclusion

Real-world scenes are composed of highly structured objects, and perceptual organization plays a fundamental role in the remarkable efficiency of human scene analysis. Imagine yourself driving on a busy street. Our visual perception can decide rapidly and accurately whether there are obstacles lying ahead. Such an ability manifests perceptual grouping that rapidly organizes a massive array of pixels sensed by our retina into a small number of objects for attentional examination. The results presented here are more in line with such mechanisms and we do not think that our results constitute rare exceptions to current theories of visual search, which put a great deal of emphasis on analyzing individual items.

Our results demonstrate that some stimulus conditions are more conducive to efficient multi-conjunction search than others, and there are processes independent of filtering and selection that can determine whether search is efficient or laborious. We expect that more conditions, and perhaps even principles, will be revealed in the future. Given its selectivity to stimulus conditions, multi-conjunction search could provide an effective tool for studying perceptual organization in visual search. Conversely, properly designed visual search experiments could also shed new light on scene analysis mechanisms.

Acknowledgements. Part of this work was undertaken while DLW was a visiting scholar in the Harvard Vision Sciences Laboratory. DLW was supported in part by an ONR YIP award, an NSF grant, and an AFOSR grant. AK was supported by a Fulbright graduate award, and AK and KN were supported in part by an AFOSR grant.

References

Allman, J., Miezen, F., & McGuiness, E. (1985). Stimulus specific responses from beyond the classical receptive field: Neurophysiological mechanisms for local-global comparisons in visual neurons. *Annual Review of Neuroscience*, **8**, 407-430.

Bacon, W. F., & Egeth, H. E. (1997). Goal-directed guidance of attention: evidence from conjunctive visual search. *Journal of Experimental Psychology: Human Perception and Performance*, **23**, 948-961.

Chun, M. M., & Wolfe, J. M. (1996). Just say no: How are visual searches terminated when there is no target present? *Cognitive Psychology*, **30**, 39-78.

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

Duncan, J., & Humphreys, G. W. (1992). Beyond the search surface: visual search and attentional engagement. *Journal of Experimental Psychology: Human Perception and Performance*, **18**, 578-588.

Egeth, H. E., Virzi, R. A., & Garbart, H. (1984). Searching for conjunctively defined targets. *Journal of Experimental Psychology: Human Perception and Performance*, **10**, 32-39.

Enns, J. T., & Rensink, R. A. (1990). Influence of scene-based properties on visual search. *Science*, **247**, 721-723.

Field, D. J., Hayes, A., & Hess, R. F. (1993). Contour integration by the human visual system: Evidence for a local "association field". *Vision Research*, **33**, 173-193.

Friedman-Hill, S., & Wolfe, J. M. (1995). Second-order parallel processing: visual search for the odd item in a subset. *Journal of Experimental Psychology: Human Perception and Performance*, **21**, 531-551.

Gilbert, C. D., & Wiesel, T. N. (1989). Columnar specificity of intrinsic horizontal and corticocortical connections in cat visual cortex. *Journal of Neuroscience*, **9**, 2432-2442.

Grossberg, S., Mingolla, E., & Ross, W. D. (1994). A neural theory of attentive visual search: interactions of boundary, surface, and object representations. *Psychological Review*, **101**, 470-489.

He, Z. J., & Nakayama, K. (1992). Surfaces versus features in visual search. *Nature*, **359**, 231-233.

Kanizsa, G. (1979). *Organization in vision: Essays on Gestalt psychology*. New York: Praeger.

Kingstone, A., & Bischof, W. F. (1999). Perceptual grouping and motion coherence in visual search. *Psychological Science*, **10**, 151-156.

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

Koffka, K. (1935). *Principles of Gestalt psychology*. New York: Harcourt.

Kovacs, I., & Julesz, B. (1993). A closed curve is much more than an incomplete one: Effect of closure in figure-ground segmentation. *Proceedings of the National Academy of Sciences of USA*, **90**, 7495-7497.

Kristjansson, A., Wang, D. L., & Nakayama, K. (2001). The role of priming in conjunctive visual search. *submitted*.

Maljkovic, V., & Nakayama, K. (1994). Priming of pop-out: I. Role of features. *Memory & cognition*, **22**, 657-672.

Marr, D. (1982). *Vision*. New York: Freeman.

Nakayama, K., He, Z. J., & Shimojo, S. (1995). Visual surface representation: A critical link between lower-level and higher-level vision. In S. M. Kosslyn & D. N. Osherson (Ed.), *An invitation to cognitive science* (pp. 1-70). Cambridge MA: MIT Press.

Nakayama, K., & Silverman, G. H. (1986). Serial and parallel processing of visual surface conjunctions. *Nature*, **320**, 264-265.

Niebur, E., & Koch, C. (1998). Computational architectures for attention. In R. Parasuraman (Ed.), *The attentive brain* (pp. 163-186). Cambridge, MA: MIT Press.

Palmer, S. E. (1999). *Vision science*. Cambridge MA: MIT Press.

Pashler, H. (Ed.). (1998). *Attention*. Hove UK: Psychology Press.

Ramachandran, V. S. (1988). Perception of shape from shading. *Nature*, **331**, 163-165.

Rockland, K. S., & Lund, J. S. (1983). Intrinsic laminar lattice connections in primate visual cortex. *Journal of Comparative Neurology*, **216**, 303-318.

Theeuwes, J., & Kooi, F. L. (1994). Parallel search for a conjunction of contrast polarity and shape. *Vision Research*, **34**, 3013-3016.

Treisman, A. (1982). Perceptual grouping and attention in visual search for features and for objects. *Journal of Experimental Psychology: Human Perception and Performance*, **8**, 194-214.

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

Treisman, A., & Sato, S. (1990). Conjunction search revisited. *Journal of Experimental Psychology: Human Perception and Performance*, **16**, 459-478.

Wang, D. L. (1999). Object selection based on oscillatory correlation. *Neural Networks*, **12**, 579-592.

Wolfe, J. M. (1994). Guided search 2.0: A revised model of visual search. *Psychological Bulletin & Review*, **1**, 202-238.

Wolfe, J. M. (1998). Visual search. In H. Pashler (Ed.), *Attention* Hove UK: Psychology Press.

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.

Zeki, S. (1993). *A vision of the brain*. Oxford, England: Blackwell Scientific.