Treisman, A. (1991) Search, similarity and integration of features between and within dimensions. *J. Exp. Psychol. Hum. Percept. Perform.* 17: 652–676.

Treisman, A. (1993) The perception of features and objects. In *Attention: Selection, Awareness, and Control*, edited by A. Baddeley, and L. Wieskrantz. Oxford: Oxford University Press.

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

Treisman, A. and Schmidt, H. (1982) Illusory conjunctions in the perception of objects. *Cognitive Psychol.* 14: 107–141.

Ungerleider, L. G. and Haxby, J. (1994) What and where in the human brain. *Curr. Opin. Neurobiol.* 4: 157–165.

Ungerleider, L. G. and Mishkin, M. (1982) Two cortical visual systems. In *Analysis of Visual Behavior*, edited by J. Ingle, M. A. Goodale, and R. J. W. Mansfield, pp. 549–586. Cambridge, MA: MIT Press.

Wolfe, J. M., Cave, K. R., and Franzel, S. L. (1989) Guided search: An alternative to the feature integration model of visual search. *J. Exp. Psychol. Hum. Percept. Perform.* 15: 419–433.

13 Attention, Pattern Recognition, and Pop-Out in Visual Search

Ken Nakayama and Julian S. Joseph

ABSTRACT Theories of visual search inspired by neurophysiological investigations of early vision have postulated built-in visual primitives which determine whether visual search will occur in parallel (without attention) or whether it will require serial attentional sampling. Against this dualistic view, we argue that attention is required for all search tasks but that the spatial scale over which attention is allocated differs. Easy (often mistakenly called parallel) search can be regarded as pattern recognition, requiring attention to be distributed globally, spreading preferentially across emergent perceptual segmentations such as surfaces. Pop-out, which can accompany this easy global recognition, is a separate and distinct process, involving the automatic narrowing of attention to an odd item. Pop-out can be primed such that the deployment of attention is enhanced for repeated positions and features.

The topic of visual search has received an unusual amount of interest over the past 15 years. Several reasons explain this popularity. First, because the stimuli are plainly suprathreshold, it provides a connection to everyday life in ways that studies of early vision do not. In visual search tasks, the observer is required to find or identify a target in a multi-element array, a task not unlike the spotting of friends in a group or finding one's car in a crowded parking lot.

Second, while being complex, these displays can be varied in numerous quantifiable ways. Each display has a specific number of separate elements, with specified color, shape, and contrast at defined spacings. Thus, performance, in terms of reaction time (RT) or accuracy, can be measured in relation to these variables. What captured the most initial interest was the fact that search behavior appeared to fall into two separate classes and each was thought to exemplify a different underlying process.

The term *serial search* was attributed to cases where RTs increased with distractor number. This suggested that the observer was required to process each target one at a time by moving attention or by making saccadic eye movements. A much sought after signature to confirm this underlying hypothetical process was the 2:1 difference in RT increase for target-absent versus target-present conditions as distractor number increased. This difference was presumed to reflect the need to exhaustively sample the full display when the target was absent and, on average, to sample just half the display when the target was present.

Parallel search was deemed to occur when RTs did not increase with the number of distractors. This suggested that the underlying process was mediated by many independent detecting mechanisms, all requiring a certain amount of time but acting in parallel. The term "preattentive" (following Neisser, 1967) was used to indicate that all of these processes occurred prior to visual attention.

THEORY INSPIRED FROM THE CHARACTERISTICS OF EARLY VISION

Owing mainly to the theories of Treisman and Gelade (1980) and of Julesz (1984), visual search achieved even greater prominence. These theories promised the beginnings of a low-level, image-based explanation for visual search. Their underlying assumptions were closely related to neurophysiological ideas regarding the organization of early vision. Just as physiological descriptions of receptive fields suggested that neurons in the cortex were analyzers, specific to color, bar orientation, spatial frequency, binocular disparity, motion, and so forth, these theories of visual search suggested that there existed over the visual field feature analyzers that were arranged in a parallel array, each feature array comprising a retinotopic map. Of obvious attraction was the implicit yet ambitious linkage to the whole edifice of findings associated with the receptive fields of the visual cortex.

Although Treisman and Gelade's and Julesz's theories were inspired by neurophysiological findings, they maintained a certain distance from these results, preferring to define the characteristics of these hypothetical units a priori or to let them be characterized by the search experiments themselves. Treisman introduced the concept of feature maps, using their implied properties to explain the data from simple visual search experiments. An observer could easily find a target defined by a single unique feature—say a red target in a field of green distractors—because in the map of red features, only one locus of "red" activity would be evident and would thus "pop out." This explained the flat search functions and appeared to support the view of distributed parallel processing in early vision. Equipped with these views one could also use the presence of flat search functions as a diagnostic method to determine which features were elementary, which elements constituted the basic building blocks of perception. With more complex displays, such as Treisman's conjunctive paradigm, parallel search was not possible because the target was not unique in any simple feature map. Thus, targets had to be processed item by item.

Julesz's theory was similar, although it provided a more principled account as to why particular features pop out. Julesz postulated that there were canonical elementary particles of perception called textons. Julesz suggested that early preattentive discrimination was based on texton densities and that when densities of such textons became sufficiently inhomogeneous, then

there would be a corresponding inhomogeneity in the texton map, leading to effortless selection of an item without attention.

Characteristic of both theories was the emphasis on the independence of activities in parallel channels, akin to neurons with receptive fields. The properties of the neurons themselves were deemed adequate to explain the class of visual search results in which performance did not vary with distractor number. Visual processing, therefore, proceeded more or less automatically and did not require the higher intervention of focal attention. This led to flat search functions, which in turn provided a confirming signature of an independent parallel process.

The popularity of these theories was immediate and widespread. First, they seemed to provide a satisfactory account of visual search by explaining a seemingly complex visual phenomenon in terms of something very primitive: features or textons. Second and following from this first point, the theories suggested that one might even discover new visual primitives via clever psychological experimentation. Visual search experiments by themselves might provide a powerful technique to identify new and perhaps unsuspected visual elements. Julesz's program, for example, raised the possibility that line terminators might act as basic elements.

In sum, these theories were bold and promising, accounting for complex phenomena and providing a new way to understand vision in terms of its constituents. Yet, as is occasionally the case with the most popular scientific theories, they initiated a line of research that began to undermine their own foundations. In the effort to find elementary units of vision, new stimuli were created that questioned the most basic idea, that the properties of primitive parallel array of retinotopically organized analyzers could explain these complex visual phenomena.

First, the work of Ramachandran (1988) indicated that concave depressions derived from shading easily segregated from a field of top-lit shaded spheres and could be easily detected with RT little affected by distractor number (Kleffner and Ramachandran, 1992). In a closely related study Enns and Rensink (1990) found that subjects could easily find countershaded cubes among top-lit shaded cubes. Again, set size had no appreciable effect on detection speed. Even more telling were the experiments of Wang, Cavanagh, and Green (1994), who showed that an observer could not easily detect a ᑌᑎ among ᑎᑌ, but that as soon as the stimuli were rotated 90 degrees, detection became much easier because the stimuli, ㄹ and Ƨ, then looked very much like the numerals 2 and 5. A related phenomenon was shown during searches for an И among normal N distractors. Performance was excellent. Interestingly, the converse was not the case. Searching for an N among И distractors was far more difficult.

These results and others can be considered in two ways. First, Ramachandran's original results can be interpreted, as he did, by considering shaded sphere-like bulges to be yet another texton or feature. One encounters many shaded figures in the world, and perhaps it is not far-fetched to think of passive analyzers for convex shapes reproduced, distributed, and tiled over the whole visual field. The explanation, however, becomes more problematic as the list of putative textons increases. Familiar patterns like ᘔ and ᔕ, as opposed to unfamiliar patterns having almost identical texton differences (simply rotated by 90 degrees), evoke very different reactions in visual search tasks. Although dense feature maps may exist for simple features such as color and orientation, it becomes much more difficult to conceive of an exhaustive set of maps for various letters, surface shapes, and so forth. Moreover, with the report of each new example of an element supporting rapid visual search, yet another map of primitives is needed, each also represented densely at different retinotopic locations and scales.

A complementary class of experiments is also relevant. He and Nakayama (1992) found that the search for a reversed *L* among normal *L*s was easy in a multi-element search array. RT did not increase with distractor number. Thus, one could conceive of the task as being mediated at a featural level. However, performance could be severely degraded by manipulating binocular disparity such that the elements would appear to perceptually complete behind occluders, rendering them less clearly distinguishable as targets and as distractors. This indicates that even with features intact, a higher-level representation of surfaces is decisive in determining whether a visual search task can be performed easily. Analogous results were found by Suzuki and Cavanagh (1995), who showed that feature differences do not support rapid visual search when embedded in a face representation. These two experiments suggest that we only have access to higher-order representations. We do not have access to image features.

Thus, there are two sets of evidence against early vision accounts of visual search. First, the number of primitive features on the list is looming too large. Second, there is evidence that there is no response to features at all per se. For these and for other reasons, described in the following section, we argue for the need to abandon or, even more strongly, to exorcise the early vision metaphor. In its place, we suggest that easy visual search tasks be regarded as requiring a higher-order process, that of a global pattern recognition at the scale of the full search array.

ATTENTION REQUIRED FOR ALL VISUAL SEARCH TASKS

To begin anew, we first need to step back and build upon some of the major findings in perception, including those from the older Gestalt tradition and, more recently, from cognitive psychology. Phenomenological studies indicate the existence of nonlocal organizing principles that operate in vision to determine whether widely separated portions of an image are grouped together or are segregated as different units (Koffka, 1935; Kanizsa, 1979; Nakayama, He, and Shimojo, 1995). In addition, another tradition of attentional research has been spawned more recently, in the earliest days of cognitive psychology.

First, consider the properties of visual attention. One of the earliest and most obvious facts to reemerge with the birth of cognitive psychology was the inherent selectivity of attention (Broadbent, 1958). Such selection implies capacity limits and these limits were understood to vary in a graded quantitative manner (Kahneman, 1973; Sperling and Melchner, 1978). For example, it seemed reasonable to assume that with more attentional effort, performance would increase. Yet, this is not always the case.

In an important theoretical contribution, Norman and Bobrow (1975) outlined with unusual clarity a plausible relationship between many seemingly disparate domains—incoming sensory information (data), attentional effort (resources), task difficulty, and performance. Figure 13.1A depicts their postulated relationship between performance and attentional effort. Also labeled in this diagram is the customary range over which attention can be

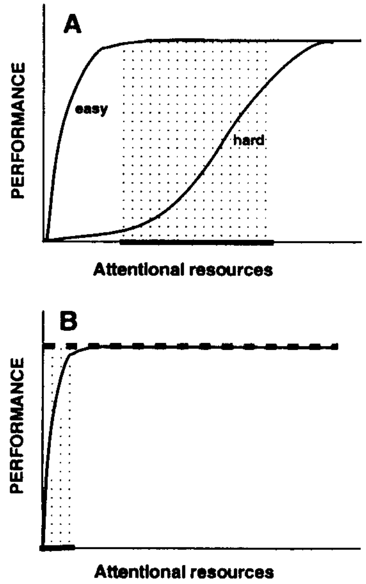

Figure 13.1 (A) Performance versus attentional resource allocation curves (redrawn from Norman and Bobrow, 1975). Note that for a hard task, added expenditures of attention improve performance. Contrast this to easy tasks in which very little attention is needed and variations of attention over the experimental range will have no effect. The thickened line on abscissa represents the range of attentional variation customarily achieved in laboratory studies of attention. (B) Comparison of two hypothetical situations in Norman and Bobrow's coordinate framework. The solid line represents a hypothetical very easy visual search task that requires only very small amounts of attention to reach asymptotic performance. The dashed line represents the hypothesized behavior of parallel search implied by early vision theories attention. The thickened line on the abscissa represents the range of attention needed to show necessity of attention on very easy search tasks.

varied in the usual laboratory experiments (denoted by the thickened line on the abscissa). From this formulation it should be clear that attention can influence performance dramatically, yet its influence occurs only within a restricted range. Above a certain level of attention (where the performance versus attention curve reaches asymptote), there is little or no effect of attention on performance. Task difficulty is also highly relevant. Added attention can improve performance of all tasks, both easy and hard, but the specific range over which this occurs differs. For example, greater attentional effort can increase performance for difficult tasks but will not have any effect on easy tasks. The latter require much less attention and cannot be further improved with greater resources or effort. In this range, Norman and Bobrow indicate that performance is data limited but not resource limited. For example, degrading the stimulus here might reduce performance, whereas reducing attention would not. One can conceive of even easier tasks in which even less attentional resources are required. Such rising curves would be shifted even further to the left, showing attentional influences on performance only for very small allocations of attention (as in the solid curve in figure 13.1B).

Norman and Bobrow's conception of attention and performance contrasts sharply with the views on visual search described in the previous section (Treisman and Gelade, 1980; Julesz, 1984). These latter theories assume two categorically distinct processes and divide visual search tasks into those that require visual attention (serial search) and a special class (parallel search) that does not. Because such theories claim that the latter processes require no attention, performance is constant in Norman and Bobrow's coordinates (refer to the dashed line in figure 13.1B). Contrast this to the presumed dependence of simple search on attention if only a very small amount of attention were necessary (represented by the rising solid curve in figure 13.1b).

From this graphic formulation, it should be clear that only the most drastic reductions of attention are capable of demonstrating the role of attention in "easy" visual search tasks. Reducing attention by arbitrarily large amounts is not enough. One needs to reduce it to the level at which its absence will have obvious and deleterious consequences. Thus, in figure 13.1B, attention needs to be reduced to the range denoted by the solid line on the abscissa. Consider an analogy with low-temperature physics. Just because temperature can be lowered dramatically, by hundreds of degrees, does not mean that all heat (kinetic energy) has been removed. In fact, almost heroic measures were required to reduce temperature to near absolute zero. Eventually, the effort succeeded and new and unexpected properties of matter were discovered—for example, superconductivity. So too with attention. We argue that the usual competing tasks in dual task studies do not consume sufficient resources and allow for small but significant amounts to be allocated elsewhere, particularly in experiments where task demands are clear. Thus, until recently we have lacked powerful methods to ensure that attention is reduced to almost zero. This technical inability seems in part to have led experimenters to conclude that attentional effort is not needed for the simplest of perceptual tasks, including pop-out tasks.

For example, Braun and Sagi (1990) argued that attention is not necessary for visual pop-out based on orientation differences. Using a dual task procedure, they varied attentional allocation between two tasks, orientation pop-out and letter discrimination, such that the letter task showed improved performance with increasing atttentional allocation to it with no corresponding decrement in performance for the pop-out task. Braun (1993) found similar results for Ramachandran's (1988) shape-from-shading array. From this Braun and colleagues drew the conclusion that attention is not required for the pop-out task. Referring to Norman and Bobrow's diagram (figure 13.1), however, note that withdrawing attentional effort from a region in which attention is not limiting will not have any effect on performance. It simply moves leftward along the curve in the region of constant performance. Thus, visual search tasks can be strongly dependent on attention but will not reveal such a dependence unless attention is sufficiently reduced. Kowler, Anderson, Dosher, and Blaser (1995) made an analogous argument against the claim that attention is not required for saccadic eye movements. In an important study employing a dual task paradigm, they showed that normal saccades required a measurable amount of attention. The programming and execution of saccades may be categorized as an easy task in Norman and Bobrow's family of curves, however. As such, it also explains why it has been so hard to actually prove the necessity of attention for saccades even though there has been much circumstantial evidence to establish this linkage (Fischer, 1987; Fischer and Weber, 1993; Mackeben and Nakayama, 1993).

Attention is also needed for even the simplest visual search tasks. Severe drops in performance can be seen in an orientation pop-out task when the demands of an additionally imposed task are very high (Joseph, Chun, and Nakayama, 1997). In this situation, a highly demanding task that proved adequate to consume required resources was used. Observers monitored a rapid serial visual presentation (RSVP) stream of letters and were required to identify a differently colored letter in the stream while performing a simultaneous task of oddball detection in an orientation pop-out task (figure 13.2). In comparison to control conditions in which performance on the pop-out task was very high using an accuracy measure, performance dropped almost to levels of chance with the addition of the RSVP letter task. The performance drop was also dependent on the asynchrony between the target letter presentation and the onset of the visual search array. Shortest lags led to the greatest interference. Subsidiary experiments indicated that the same stimuli yielded flat search functions of set size with RT as the measure, indicating that the task met the criteria of so called preattentive or parallel search. Similarly severe impairments of performance occured in a shape-from-shading oddball detection task (Joseph et al., 1996).

In a related series of studies, Rock, Linnett, Grant, and Mack (1992) and Mack, Tang, Tuma, and Kahn (1992) showed that even the simplest visual

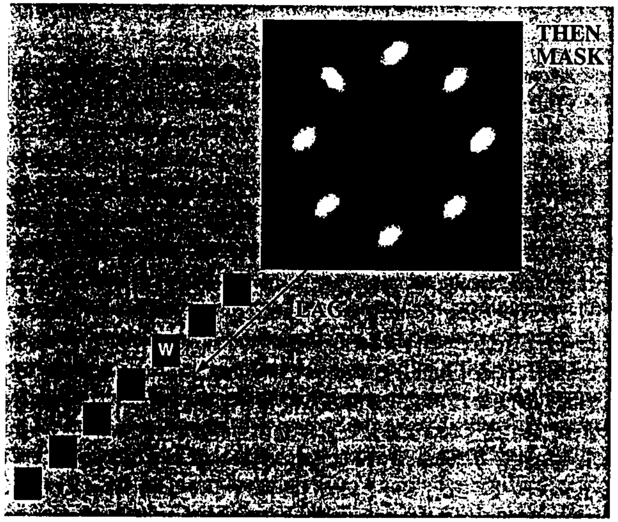

Figure 13.2 A pop-out task based on orientation differences requires visual attention. An array of oriented Gabor patches can appear with or without an orientation oddball; the task is to report whether an oddball is present. A competing task at fixation is to attend to the central region and identify the white letter in a stream of black letters. Lag refers to the time asynchrony between the presentation of the white target letter and onset of the visual search array. Oddball detection accuracy is severely impaired by the letter task. (From Joseph et al., 1997.)

tasks are compromised when attention is taken up elsewhere. They employed a clever, unconventional design in which there was only one test trial per subject. Thus, there was no reason for the unsuspecting observer to allocate attention away from the primary task on the single trial because the observer was unaware of any other task. As a consequence, performance failure in the otherwise very easy secondary task was dramatic. Rock et al. dubbed the phenomenon inattentional blindness to underscore its importance.

These studies share an important conclusion. If attention is largely removed through effective methods, either by allocating it more fully to a primary task or by not allocating it efficiently to an unexpected secondary task, the conclusion is the same: Almost all of what is considered to be conscious vision cannot occur without attention.

ATTENTION DEPLOYED TO SURFACES NOT FEATURES

Norman and Bobrow's (1975) formulation provides the basis for additional understanding about visual search, particularly when it is coupled with a more explicit description of what underlies visual task difficulty. The issue of capacity limits in attention invites an exploration of quantitative factors that might play role in visual search difficulty. What first comes to mind are the elementary notions outlined by information theory: redundancy, coding, data compression, and so forth. Stimulus coding leads to issues of perceptual organization, and it also opens the door to consideration of a greater role for perceptual and other forms of visual learning.

Norman and Bobrow also made an explicit connection between their analysis and its relationship to learning and practice. They suggested that as learning proceeds and tasks become easier, the rising portion of the performance versus attention curve should shift progressively to the left. Employing the language of information theory, we hypothesize that with extended practice information regarding the display becomes, to use Miller's (1956) phrase, chunked. Thus, in analogy to the chess master who codes the seemingly complex displays on a chessboard according to his deep knowledge of the strategy of the game, our perceptual systems expertly chunk information in visual search displays. Chunking reduces the information load on the system, and progressively less and less attention is required as chunking increases. This is the reason that the curve shifts to the left. In terms of visual search tasks, if the stimulus could become more easily codable (with fewer bits) through practice, then less attentional resources would be required to achieve the same level of performance. How is such chunking achieved in vision? Efficient coding, of course, requires redundancy and, broadly speaking, one can conceive of the code as removing much of the natural redundancy in everyday scenes (Attneave, 1954). Frequently occurring scene patterns, therefore, could be coded with fewer bits. Thus, one codes familiar faces efficiently and can spot subtle blemishes on them much more quickly than on a stranger's face.

Of course, redundancy reduction is not restricted to high-level visual patterns; it is perhaps even more relevant when considered at the mid-level organization of vision, particularly when applied to visual search tasks. Here, the whole field of perceptual psychology, including the earlier Gestalt tradition, is clearly relevant, and organizational factors spanning large retinal distances become important. Thus, similarity in color, shape, motion, and so forth all contribute to perceptual grouping, as do certain configurational patterns such as collinearity and cocurvature. Also important are processes of surface completion that either segregate or join distinct patches of images as surfaces in depth (Kanizsa, 1979; Nakayama et al., 1995).

It is likely, therefore, that efficiencies that are developed for normal scene encoding also influence visual search tasks. Visual search tasks that allow the most efficient coding of the distractors and target as separate entities are at a distinct advantage. Displays with identically colored distractors will have advantages over displays with more variegated colors because of the ability to group identical colors. This idea has been offered as an alternative to Treisman's theory because it also predicts the difficulty of finding odd targets in Treisman's conjunctive search task (Duncan and Humphreys, 1989).

It should be noted, however, that grouping is not determined by the linkage of low-level features but is a relational process determined by whether the elements conform to regularities in the world. Experiments on the spread of attention in three-dimensional space are instructive. He and Nakayama (1995) set up a binocular depth display consisting of targets at three stereoscopic depths of near, middle, and far. Observers were required to find a single odd-colored target in the array at the middle distance. The number of same-colored targets in the near and the far planes and the local slant of the individually colored target rectangles were varied. They could be slanted forward, backward, or in the same plane implied by the middle depth array. Observers performed this task easily when each of the targets in the middle depth plane were coplanar, that is, were not slanted backward or forward with respect to the middle plane (figure 13.3a,b). Thus, grouping is not de-

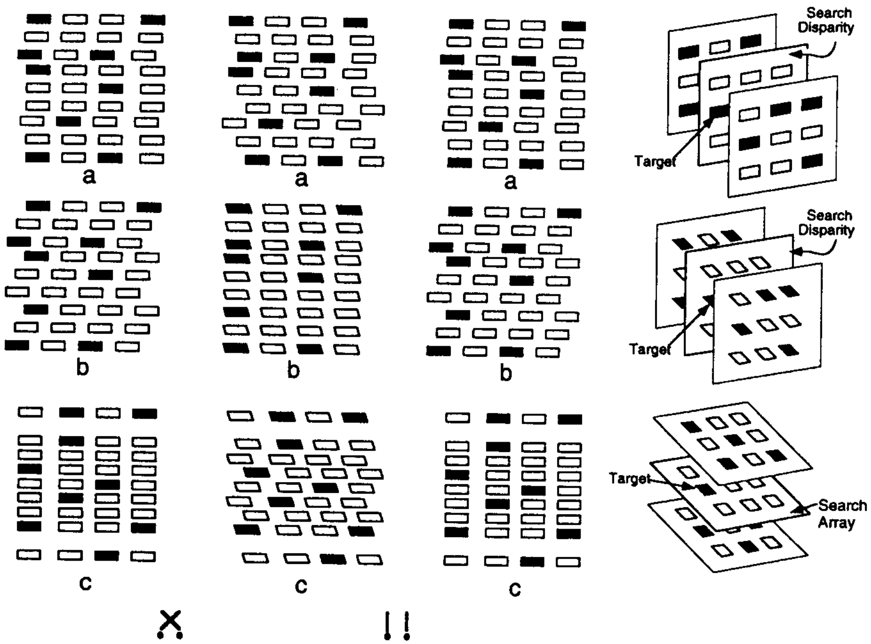

Figure 13.3 Three-dimensional search arrays showing the importance of surfaces in the deployment of attention. On the left three columns are stereograms containing elements at three distances. Left two half-images should be fused for crossed eyes, and the right two should be fused for divergent fusion. Visual search for an odd color confined to the middle distance is rapid in (a) because the elements lie in the same front parallel plane, whereas searching at the same middle distance in (b) is much more slow because the elements do not lie in that plane. Search is also easy when a subject looks for an odd target in a depth array in the middle of three horizontal stacks because elements in the middle horizontal plane are all coplanar. Because distractors of the same color as the target are also in the flanking planes, the task requires the efficient spreading of focal attention in the plane to be searched. Diagrams at right provide a pictorial description of the depth relations in the perceived display. (Redrawn with permission from He and Nakayama, 1995.)

termined by low-level cortical factors such as a common binocular disparity, as was originally hypothesized by Nakayama and Silverman (1986). Rather, the results suggest that surfaces with coplanar elements can effectively support the spread of attention to spatially distributed groups of elements. Supporting and extending this conclusion is the fact that observers were able to selectively search for targets in a set of rectangles having coplanar elements that spanned the most extreme range of binocular disparities and stereoscopic depths (see figure 13.3c). Thus, analogous to the Gestalt principle of good continuation in two dimensions, He and Nakayama (1995) showed that such a principle also operates in three dimensions, allowing coplanar elements to emerge as a surface over which selective attention can easily spread (see also Nakayama and He, 1995).

VISUAL SEARCH AS ATTENTIVE PATTERN RECOGNITION AT VARIOUS SCALES

Having argued against an early vision approach to visual search, it is worth reflecting on what we are proposing in contrast to what we are rejecting, and also commenting on what has been gained and lost. Early vision theories based their explanatory power on the elementary properties of built-in units. Lost, therefore, is the immediate hope of reducing higher-level vision to the presumed properties of single units in early cortical structures. What is gained? At worst, perhaps only the sober realization that terms like parallel processing, preattentive, features, and so forth, can no longer be used so confidently. More positively, we think our proposal opens the door for a wider range of theoretical accounts, a restructured descriptive vocabulary, and the opportunity to observe new phenomena. The remaining portion of this chapter is devoted towards those ends.

Acknowledging the severe capacity limits of visual attention, Nakayama (1990) proposed a close relationship between attention at different scales and pattern recognition. Attentional fixations, at varying loci and spatial scales, allow selected portions of the image to be matched with templates in visual memory. Due to the attentional bottleneck, the full richness of the visual scene cannot be sampled. Attentive sampling, therefore, represents a compromise between scale (the area to be sampled) and resolution (detail). Thus, a global sampling of a large portion of an image can be accomplished but only at low spatial resolution. Higher-resolution sampling can also occur but at the expense of limiting the area. Thus, to recognize the details of a scene requires narrow focusing of attention. To apply these constraints to visual search tasks, Nakayama (1990), in agreement with previous views, accepted the notion that for difficult visual search tasks, focal attention is necessary to inspect items serially. That is, each attentive fixation allows pattern recognition to occur in a restricted local area.

The divergence in thinking came mainly with the interpretation of easy search tasks, or so-called parallel search. This we also regard as pattern recognition but on a larger scale, with a concomitant loss of spatial resolution. Thus, in easy visual search, usually mistakenly called parallel search, global pattern recognition boils down to a binary decision based on a coarse sampling of the image—does the overall array correspond to one with the target present? Or does it correspond to the pattern with the target absent? As mentioned previously, this too requires attention (Joseph et al., 1996a).

POP-OUT ACCOMPANIES BUT DOES NOT ACCOUNT FOR EASY VISUAL SEARCH

As theories change, even the most basic vocabulary, apparently so descriptive at one time, can lose or change its meaning. This is true for the terms associated with visual search. If the dualistic notion of visual search with and without attention is not accepted, then the terms "preattentive vision" and "parallel processing" lose their specific referents. Thus, the term "parallel search," in particular, should be discarded. Other terms, however, cannot be so easily abandoned but need re-analysis and redefinition. One of the most commonly used terms is "pop-out." This seemingly descriptive term has very different meanings under the different theoretical perspectives.

If one adopts the early vision view of visual search, pop-out is both a phenomenological term and a mechanistic and theoretical construct. It is phenomenological because it describes the psychic fact that an odd item becomes more salient in a display. It is theoretical because it seemed to provide a mechanistic description of how, in a retinotopic array of feature or texton analyzers, only one site is active. That mechanism, in turn, played an essentially causal role in allowing simple or easy visual search to occur. According to this way of thinking, rapid search independent of distractor number occurs because of a sole active element in a parallel array of analyzers (for example, see Treisman and Gelade, 1980).

Even after rejecting the theoretical notions associated with Treisman's theory, the phenomenological term "pop-out" at first glance seems appropriate because it accompanies simple search displays. As such, we have used it earlier in this paper to denote easy visual search tasks because of the common usage of the term. Yet, this usage can also be very misleading because it is too closely tied to the notion of parallel search and glosses over the empirical characteristics of pop-out. Not being tied to an early vision conception of pop-out, we need a notion of pop-out that is more descriptive of its phenomenological characteristics. Most distinctive about the experience is the strong involuntary awareness of the odd target. In pop-out, attention is jerked suddenly to that locus. The target becomes more distinct, and fine details about its shape become more discernible. Thus, we define pop-out more descriptively as the involuntary narrowing of attention to an odd item in a field of elements.

However, it may now strike the reader that our new definition of pop-out appears to contradict our hypothesis regarding simple (easy) visual search tasks. We have argued that rapid visual search is based not on the narrowing of attention but on its opposite, a distributed spread of attention over the whole array, allowing pattern recognition at a larger scale. So how can we reconcile the spread of attention, which is required for rapid search, with its opposite, the narrowing of attention, which also accompanies it? Our view, based on the work of Bravo and Nakayama (1992), is that the two processes are distinct and occur sequentially. Pop-out occurs with the presence of an odd target but only after the global matching process required for detection. Thus, pop-out, or the narrowing of attention to the odd target, has no direct causal role in detection of the presence of a target. Thus, easy visual search tasks generally lead to two separate allocations of attention in a customary sequence: a global attentional allocation to the whole array (useful to do the rapid search task) followed by a narrowing of attention to the target (unnecessary for the same detection task).

To show that pop-out can be experimentally dissociated from visual search performance, Bravo and Nakayama (1992) employed two different tasks using the same visual search display, one requiring the global pattern match needed for visual search, the other requiring the narrowing of attention associated with pop-out. The display consisted of a set of diamonds, either red or green, and each diamond was randomly truncated, either on the right or left side (figure 13.4A). In addition, there were two types of trial

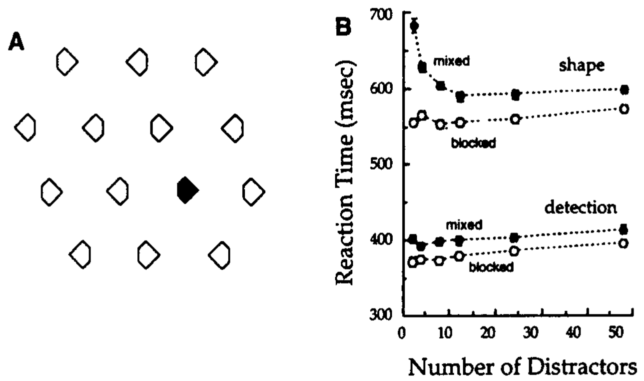

Figure 13.4 Pop-out and flat search functions do not reflect the same process. (A) Visual display for all tasks is identical, consisting of an odd-colored target (green) among distractors (red) or, vice versa, an odd red target among green distractors. In a given set of trials, target and distracter color can either stay the same for each trial (blocked) or can switch randomly from trial to trial (mixed). The two tasks are as follows. Simple search (labeled detection) consists of detecting an odd target. Pop-out task (labeled shape) requires the discrimination of the shape of the odd-colored target (whether it is truncated on the left or the right). Note the unusual relationship between distractor number and reaction time, with lower reaction times for increasing numbers of distractors. This occurs only in the mixed popout case. (Redrawn with permission from Bravo and Nakayama, 1992.)

blocks. In the blocked case, targets and distractors each had the same color from trial to trial. For example, targets would remain red and distractors would remain green within a particular block. In the mixed case, target and distractor color would reverse on a random basis from trial to trial within a block.

The first task used the usual visual search task, and subjects were asked to respond quickly as to the presence or absence of an odd-colored target. Here it should be clear that the detailed shapes of the elements were irrelevant. Not surprising, and in close agreement with the literature, search times were relatively fast and did not vary with distractor number. This result is shown in figure 13.4B as the shorter latency pair of curves denoted by the label "detection." Note that within the pair mixed trials had a small but consistently longer RT for all set sizes.

Second, to characterize pop-out and using the exact same display, the authors selected a task that required the narrow focusing of attention to the odd target. They asked observers whether the odd-colored diamond was truncated on the right or on the left. Curves labeled "shape" in figure 13.4B show that performance on this shape discrimination task was very different than simple detection performance. First, RTs were much longer. Furthermore, there was a pronounced difference between the mixed and blocked condition. In the blocked condition RTs were faster and constant across distractor number. In the mixed condition and going strongly against the usual trend for visual search tasks, slopes were negative. Increasing distractor number reduced RTs dramatically. This negative slope function has been replicated under a variety of other stimulus conditions (Bravo and Nakayama, 1992; Maljkovic and Nakayama, 1994).

Taken together, it should be clear that the behavior in the two tasks was very different. The usual visual search task had flat search functions as expected. The pop-out task, requiring a shape discrimination, can have a search function with a steep negative slope. The longer RTs for pop-out and the negative slope argue against a causal relationship between pop-out and so-called parallel search.

At this point, one might raise an objection regarding our measure of pop-out. We have assumed that the measurement of RT during shape discrimination reflects the speed of the deployment of attention to the peripheral target site. Is that warranted? We think it is because we assume that the discrimination requires focal attention. To strengthen our case, however, it would be worthwhile to have a very different indicator of attention, one that does not require visual discrimination or manual reaction. We have found such an indicator in the measurement of saccadic eye movement latencies. Earlier, we noted the need for small amounts of attention to be directed to a target prior to a saccade (Kowler et al., 1995). This leads to a simple prediction. If we measure saccadic eye movement latencies to the same visual displays, we should see the same signature that Bravo and Nakayama (1992) identified for pop-out. First, there should be flat saccadic latencies functions for cases in which the distractors and target colors remained unchanged from trial to trial. Second, saccadic latencies should be slower overall when the target and distractor colors are mixed. Most critically, saccadic latencies should also decrease with increasing distractor number. That exact pattern of results was obtained in a measurement of human saccadic eye movement latencies (McPeek and Nakayama, 1995b). This pattern of results confirms the relationship between attention and saccades and adds independent support for our distinction between pop-out and the hypothesized global pattern recognition required for easy visual search.

At this point we need to comment on the reason for the large qualitative difference between pop-out as revealed in the discrimination task and simple pattern matching as revealed in the detection task. First, the decrease in RTs with increasing distractor number in the pop-out task are predicted, at least implicitly, by several mechanistic theories of attentional deployment (Julesz, 1986; Koch and Ullman, 1985). Each theory has a slightly different emphasis, one stressing gradients of feature differences, the other, inhibitory interactions between distractors. Our own bias is that the phenomenon may be better understood at a higher level, possibly related to surface formation. At present, however, there is insufficient evidence to distinguish between these alternatives. A second, more tractable question deals with the large and consistent difference between mixed and blocked conditions. Why are RTs so much faster in the blocked condition?

PRIMING OF POP-OUT

There are several possible accounts. The most obvious explanation is the possible role of expectancy, or knowledge of the upcoming trial. In the blocked condition, observers might be expected to utilize the temporal regularities in the sequence of displays to predict what would come up on the next trial, and thereby performance would benefit. This view was quickly dispelled by Maljkovic and Nakayama (1994), who manipulated the probability of a color switch of targets and distractors, p(switch), within a block of trials. Under those circumstances, predictability was minimal when the trials were presented randomly and at a maximum when target color either remained the same on each trial or alternated on each trial. Contrary to a predictability hypothesis, RT were not lowest for maximum predictability. RT were highest when the target was completely predictable, that is, when it alternated on each trial (p(switch)=1.0). These results argue strongly against expectancy and leave only one likely alternative: priming. On each trial, it seems that some small beneficial residue of the previous trial accumulated from previous trials of the same color, such that RTs on subsequent trials of the same color will be faster.

To examine this priming in greater detail, Maljkovic and Nakayama (1994) developed a new method, memory kernel analysis, to measure the effects of a single trial over time. They looked at a sequence of many independent trials

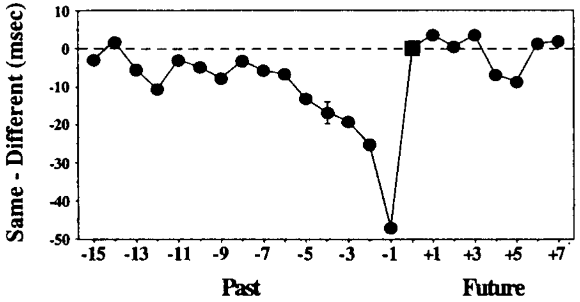

Figure 13.5 Priming of pop-out. Each point represents the influence on the current trial reaction time occasioned by the same versus different target color trials in the past (left) and future (right). Approximately 5–8 trials in the past have an influence on current trials, shortening reaction times appreciably as one compares same versus different color trials. (Reproduced with permission from Maljkovic and Nakayama, 1994.)

in which the probability of, for example, a red target among green distractors (and vice versa) was random. Then they took a given trial in the sequence, trial *n*, and analyzed it repeatedly, separately tallying RT as a function of whether on a neighboring trials, the color of the target was the same or different. They did this for essentially all preceding trials, up to fifteen. Because each trial was presented at intervals of approximately 2–3 s, they could examine the influence of events over the past 45 s. They also examined the dependence of current trials on each of next seven upcoming trials. Those trials had not yet occurred and would not be expected to influence the results. As such, they provided an estimate of the variability of the data. Because the color of the targets and distractors were randomly presented, it should be clear that the method isolated the effect of any arbitrary trial in the past because on average an equal number of intervening trials of both colors occurred; thus, the effects of those intervening trials would cancel.

Figure 13.5 plots the differences between the same-color and different-color trials for a single subject. Each point in the curve represents (whether it be the past or the future) the influence of same versus different colors on the current trial. The square symbol represents the grand mean of all the RT. Negative values indicate the speeding up of RT for same-color trials. Note that the effect is substantial: a single trial can alter the RT for the next trial by as much as 40 ms. Remarkably, single trials presented many trials earlier (over a span of 30 s) influenced behavior. The monotonically decreasing influence seen here indicates that this memory lasted for 5–8 trials, or for up to approximately 30 s. Additional experiments indicated that priming of consecutive same-color trials is cumulative and can account for the main difference in RT seen for the mixed versus blocked pop-out (shown earlier in figure 13.4B).

Maljkovic and Nakayama (1993) found that priming is involuntary and is a form of short-term implicit memory, distinguishable from explicit memory. They also showed that an analogous priming occurs for position, such that a previous deployment of attention to a given position, many trials earlier, speeds attention to that same position (Maljkovic and Nakayama, 1996). In addition, the primed position need not be in retinal coordinates but rather in the coordinates of the stimulus configuration. Finally, McPeek and Nakayama (1995a) have confirmed these results by showing a similar effect with saccadic eye movement latencies. Taken together, the results indicate that attentional deployment to an odd target is greatly influenced by past events, revealing the existence of a short-term implicit memory system.

CONCLUSIONS

Early vision theories of visual search have suggested two types of vision, one not requiring attention and responsible for "parallel" visual search, the other requiring attention and mediating more deliberate serial search. So far, we have presented three reasons to reject this early vision metaphor. First, the number of primitive features emerging is too large. With such a list, it becomes very difficult to imagine how all such patterns, including letters of the alphabet, are reproduced in all positions at all spatial scales in early cortical maps. Second, there is psychological evidence that we do not respond to elementary features at all in rapid vision, but that visual search works on a representation that is of a higher order. Visual search has no access to these putative earlier representations. Third, we have provided evidence that even the easiest, so-called parallel visual search tasks require attention.

In addition, we suggest two more reasons. Mounting evidence indicates that higher-level perceptual representations mediate most visual functions, even those traditionally thought to rely on low-level features. For example, motion perception, texture segregation, and object recognition may all be mediated by a surface level, not an image feature level (Nakayama et al., 1995). Most of this work is based on the importance of perceptual surfaces in determining whether we see image fragments as separate pieces or as connecting portions of surfaces that perceptually complete either in front of or behind occluders. Visual search is no exception, falling into line with other visual functions that depend on an intermediate surface level of representation.

Finally, we suggest a strategic reason to abandon or, even more strongly, to exorcise the early vision metaphor. It comes from a full acknowledgment of the metaphor's resilient strength, its continuing ability to define the vocabulary and, thus, even the phenomenology of visual search. Again, terms like parallel processing, preattentive vision, and so forth are powerful evocative terms that are laden with theoretical and physiological meaning at a time when such meaning should be skeptically regarded. What is needed is an alternative, more neutral vocabulary that sticks closer to the psychological facts and that opens the door to a range of new phenomena. Hopefully, we have at least partially convinced the reader that older ideas from perceptual and cognitive psychology remain alive and can form the foundation for further advances. In this regard, we have described new, unexpected facts about visual search in the second part of this chapter, relating visual search to surface representation, eye movements, and short-term memory.

How will these new psychological results be understood in terms of the rapid growth of knowledge about the brain? Most obvious is the fact that we can no longer rely on the properties of visual receptive fields to understand attention. We need an understanding of how higher-level vision (surfaces, objects, and so forth) are represented in neural circuits: a daunting challenge for the future.

ACKNOWLEDGMENTS

This work is supported in part by grants from the McKnight Foundation, Air Force Office of Scientific Research grant to K. N., and NEI grant F32-EY06531 to J. S. J. Special thanks to Marvin Chun, Charles Stromeyer, and Robert McPeek for comments on an earlier version of the manuscript.

REFERENCES

Attneave, F. (1954) Some informational aspects of visual perception. *Psychol. Rev.* 61: 183–193.

Braun, J. (1993) Shape-from-shading is independent of visual attention and may be a "texton." *Spat. Vis.* 7: 311–322.

Braun, J. and Sagi, D. (1990) Vision outside the focus of attention. *Percept. Psychophysiol.* 48: 45–58.

Braun, J. and Sagi, D. (1991) Texture-based tasks are little affected by second tasks requiring peripheral or central attentive fixation. *Perception* 20: 483–500.

Bravo, M. and Nakayama, K. (1992) The role of attention in different visual search tasks. *Percept. Psychophysiol.* 51: 465–472.

Broadbent, D. (1958) *Perception and Communication.* Oxford: Pergamon Press.

Duncan, J. and Humphreys, G. W. (1989) Visual search and stimulus similarity. *Psychol. Rev.* 96: 433–458.

Enns, J. T. and Rensink, R. A. (1990) Sensitivity to three-dimensional orientation in visual search. *Psychol. Sci.* 1: 323–326.

Fischer, B. (1987) The preparation of visually guided saccades. *Rev. Physiol. Biochem. Pharmacol.* 106: 2–35.

Fischer, B. and Weber, H. (1993) Express saccades and visual attention. *Behav. Brain Sci.* 16: 553–610.

He, Z. J. and Nakayama, K. (1992) Surface vs. features in visual search. *Nature* 359: 231–233.

He, Z. J., and Nakayama, K. (1995) Visual attention to surfaces in 3-D space. *Proc. Natl. Acad. Sci. U S A* 92: 11155–11159.

Joseph, J. S., Chun, M. M., and Nakayama, K. (1996) Attention plays a role in the perception of three-dimensional structure in shaded cube stimuli [Abstract]. *Invest. Ophthalmol. Vis. Sci.* 37: S213.

Joseph, J. S., Chun, M. M., and Nakayama, K. (1997) Attentional requirements in a "preattentive" feature search task. *Nature* 387: 805–807.

Julesz, B. (1984) Toward an axiomatic theory of preattentive vision. In *Dynamic Aspects of Neocortical Function*, edited by G. M. Edelman, W. E. Gall, and W. M. Cowan. New York: Neurosciences Research Foundation.

Julesz, B. (1986) Texton gradients: The texton theory revisited. *Biol. Cybern.* 54: 245–251.

Kahneman, D. (1973) *Attention and Effort.* Englewood Cliffs, NJ: Prentice-Hall.

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

Kleffuer, D. A., and Ramachandran, V. S. (1992) On the perception of shape from shading. *Percep. Psychophys.* 52: 18–36.

Koch, C. and Ullman, S. (1985) Shifts in selective visual attention: Towards the underlying neural circuitry. *Hum. Neurobiol.* 4: 219–227.

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

Kowler, E., Anderson, E., Dosher, B., and Blaser, E. (1995) The role of attention in the programming of saccades. *Vis. Res.* 35: 1897–1916.

Mack, A., Tang, B., Tuma, R., and Kahn, S. (1992) Perceptual organization and attention. *Cognitive Psychol.* 24: 475–501.

Mackeben, M. and Nakayama, K. (1993) Express attentional shifts. *Vis. Res.* 33: 85–90.

Maljkovic, V. and Nakayama, K. (1993) Priming of popout: An example of implicit short-term memory. *Soc. Neurosci. Abstr.* 19: 439.

Maljkovic, V. and Nakayama, K. (1994) Priming of popout: I. Role of features. *Mem. Cognition* 22: 657–672.

Maljkovic, V. and Nakayama, K. (1996) Priming of popout: II. Role of position. *Percept. Psychophys.* 58: 977–991.

McPeek, R. M. and Nakayama, K. (1995a) Linkage of attention and saccades in a visual search task. *Invest. Ophthalmol. Vis. Sci.* 36: S354.

McPeek, R. M. and Nakayama, K. (1995b) Repetition of target color affects saccadic latency and accuracy. Paper presented at the eighth European Conference on Eye Movements, Derby, United Kingdom.

Miller, G. A. (1956) The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychol. Rev.* 63: 81–97

Nakayama, K. (1990) The iconic bottleneck and the tenuous link between early visual processing and perception. In *Vision: Coding and Efficiency*, edited by C. Blakemore, pp. 411–422. Cambridge, MA: Cambridge University Press.

Nakayama, K. and He, Z. J., (1995) Attention to surfaces: Beyond a Cartesian understanding of visual attention. In *Early Vision and Beyond*, edited by T. V. Papathomas. Cambridge, MA: MIT Press.

Nakayama, K., He, Z. J., and Shimojo, S. (1995) Visual surface representation: A critical link between lower-level and higher level vision. In *Visual Cognition*, edited by S. M. Kosslyn and D. N. Osherson, pp. 1–70. Cambridge, MA: MIT Press.

Nakayama, K. and Silverman, G. H. (1986) Serial and parallel processing of visual feature conjunctions. *Nature* 320: 264–265.

Neisser, U. (1967) *Cognitive Psychology*. New York: Appleton-Century-Crofts.

Norman, D. A. and Bobrow, D. G. (1975) On data-limited and resource-limited processes. *Cognitive Psychol.* 7: 44–64.

Ramachandran, V. S. (1988) Perception of shape from shading. *Nature* 331: 163–166.

Rock, I., Linnett, C. M., Grant, P., and Mack, A. (1992) Perception without attention: Results of a new method. *Cognitive Psychol.* 24: 502–534.

Sperling, G. and Melchner, J. J. (1978) The attention operating characteristic: Some examples from visual search. *Science* 202: 315–318.

Suzuki, S. and Cavanagh, P. (1995) Facial organization blocks access to low-level features: An object inferiority effect. *J. Exp. Psychol. Hum. Percept. Perform.* 21: 901–913.

Treisman, A. (1985) Preattentive processing in vision. *Comp. Vis. Graph. Image Proc.* 31: 156–177.

Treisman, A. and Gelade, G. (1980) A feature-integration theory of attention. *Cognitive Psychol.* 12: 97–136.

Wang, Q., Cavanagh, P., and Green, M. (1994) Familiarity and pop-out in visual search. *Percept. Psychophys.* 56: 495–500.

14 Attention and Visual Object Segmentation

Jon Driver and Gordon C. Baylis

ABSTRACT The debate between space-based versus object-based accounts of visual attention is discussed. At issue is the extent to which scene segmentation can take place prior to visual selection, and whether that selection takes place within a spatial medium. Recent studies with both healthy and brain-injured persons suggest that a range of segmention processes can influence selection, leading to a variety of senses in which visual attention may be object-based. It is concluded that all these phenomena remain consistent with selection operating on a spatial array, and that different types of object-based attention must be carefully distinguished in future work on the neural substrates involved.

The last decade has seen many papers (see Kanwisher and Driver, 1992) on the issue of whether covert visual attention is directed to segmented objects, to regions of space, or perhaps to both, as we would argue. At the heart of this contemporary issue lies the old question of how much processing can take place prior to attentional selection. In the past (e.g., Broadbent, 1958), the question was posed in terms of whether or not stimulus categorization could precede attention. More recent disputes over the extent of preattentive processing concern better-specified image segmentation processes. At issue is whether or not those processes operate preattentively to allow selection of segmented objects for further attentional processing. The emerging consensus is that visual attention can indeed be object-based.

However, attention has been characterized as object-based in several subtly different ways, and the issue remains controversial. On the one hand, Baylis and Driver (1992) recently concluded that "visual attention is directed to groups derived from a preattentive segmentation of the scene according to Gestalt principles." On the other hand, in the same year, Mack, Tang, Tuma, and Kahn argued that "*no* perception of either texture segregation or Gestalt grouping" (1992, p. 488) takes place prior to attention, apparently in direct contradiction to Baylis and Driver's claim. This chapter aims to resolve such conflicts, while raising further issues.

IS THERE ANYTHING OUT THERE? NAIVE SPOTLIGHT METAPHORS FOR ATTENTION

Many theorists have likened covert visual attention to a spotlight (e.g., Posner, 1980). That metaphor can have various implications, depending on