Psychological Review
1994, Vol. 101, No. 3, 470–489

0033-295X/94/$3.00

A Neural Theory of Attentive Visual Search: Interactions of Boundary, Surface, Spatial, and Object Representations

Stephen Grossberg, Ennio Mingolla, and William D. Ross

Visual search data are given a unified quantitative explanation by a model of how spatial maps in the parietal cortex and object recognition categories in the inferotemporal cortex deploy attentional resources as they reciprocally interact with visual representations in the prestriate cortex. The model visual representations are organized into multiple boundary and surface representations. Visual search in the model is initiated by organizing multiple items that lie within a given boundary or surface representation into a candidate search grouping. These items are compared with object recognition categories to test for matches or mismatches. Mismatches can trigger deeper searches and recursive selection of new groupings until a target object is identified. The model provides an alternative to Feature Integration and Guided Search models.

Section 1. Introduction: Attentive Spatial and Object Search of Three-Dimensional (3-D) Boundaries and Surfaces

In recent psychophysical experiments on visual search (Nakayama & Silverman, 1986; Pashler, 1987; Treisman & Gelade, 1980; Wolfe, Cave, & Franzel, 1989), a target item is predefined, either verbally or by visual exposure, and an observer is required to determine whether it is present in a scene that includes distractor items. Response times for scenes with various numbers of items can then be compared to determine whether search is accomplished by parallel processing of the entire visual field or by serial investigation of each item (Figure 1). The alternatives of parallel processing and serial search among items or locations are not, however, exhaustive. Although discrete target and distractor items can be arbitrarily defined in the construction of test scenes, it does not follow that each item will preattentively give rise to a distinct perceptual object or region for further analysis. Typical laboratory-generated items (e.g., colored bars, Xs, and Os) bear little resemblance to the scenic objects that draw attention during naturally occurring tasks. Indeed, such items may occur as part of textures, in which they are grouped with other items into emergent perceptual objects (Figure 2), in accord with the suggestion that effortless texture segregation and pop-out during visual search reflect the same underlying processes (Julesz, 1984). Wolfe (1992) has challenged the claim of unified processing by showing that certain juxtapositions of features in conjunctive items can support effortless texture segregation but not rapid visual search or, conversely, rapid search but not texture segregation. This article develops a neurally based computational model of how context-sensitive emergent units, including textural groupings, are integrated into the search process.

Growing experimental evidence supports the view that the perceptual units that engage visual attention during search are emergent units that are the outcome of considerable preattentive visual processing rather than merely the outputs of early filters (Bravo & Blake, 1990; Enns & Rensink, 1990; He & Nakayama, 1992; Humphreys, Quinlan, & Riddoch, 1989). These processes include the formation of 3-D emergent boundary segmentations that combine information about scenic edges, texture, shading, and depth, and the filling-in of surface representations that combine information about surface brightness, color, depth, and form. A neurally based theory of how such emergent boundary and surface representations are formed in striate and extrastriate cortex is called FACADE theory because it suggests how representations that combine Form-And-Color-And-DEpth information are generated within the visual cortex (Grossberg, 1987a, 1987b, 1994; Grossberg & Mingolla, 1985, 1987; Grossberg, Mingolla, & Todorović, 1989; Grossberg & Todorović, 1988). Our model follows up the observation in Grossberg (1987a) that 3-D boundary segmentation and surface representations are the inputs to an attentive visual object recognition system with which they reciprocally interact during visual search. The present algorithm uses FACADE-like boundary and surface representations as its "front end" and suggests how spatial attention mechanisms in the parietal cortex and object recognition categories in the inferotemporal cortex may deploy attentional resources to reciprocally interact with 3-D boundary and surface representations in the prestriate visual cortex. These interactions can modulate and reorganize perceptual units after they are preattentively formed. We suggest, therefore, that mechanisms of preattentive boundary segmentation and surface filling-in, including preattentive mechanisms

---

Stephen Grossberg, Ennio Mingolla, and William D. Ross, Center for Adaptive Systems and Department of Cognitive and Neural Systems, Boston University.

This study was supported in part by grants from the Air Force Office of Scientific Research and the Office of Naval Research to Stephen Grossberg, Ennio Mingolla, and William D. Ross; a grant from the Northeast Consortium for Engineering Education to Ennio Mingolla; and grants from British Petroleum and the National Science Foundation to William D. Ross.

We wish to thank Robin Locke for her valuable assistance in the preparation of this article.

Correspondence concerning this article should be addressed to Stephen Grossberg, Department of Cognitive and Neural Systems, Boston University, 111 Cummington Street, Room 244, Boston, Massachusetts 02215.

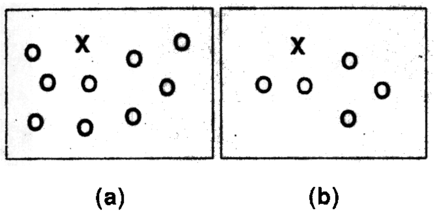


Figure 1. Search scenes: (a) The target (black X) is distinguishable by a single feature (the form X). Response time is fast and does not vary with the number of distractor items. (b) The target is distinguishable only by the conjunction of two features (the form X and the color black).

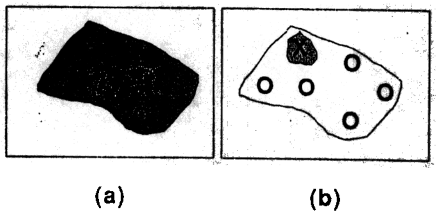


Figure 3. Illustration of search by recursive grouping: (a) Initially, a region of target color that includes many items is separated as a candidate region for target presence. (b) Next, this region is searched for target form. Thus, fast search times can be achieved without parallel processing of feature conjunctions.

of figure–ground pop-out, help to define the perceptual units that attentive spatial and object recognition mechanisms can further modulate and reorganize during visual search.

This interaction may be modeled as an attentively organized search that may be recursively applied to emergent groupings that may contain multiple items. For targets defined by a conjunction of color and form (Figure 1b), search could proceed as a two-step recursive process (Figure 3). Initially, a multi-item grouping defined by a single target feature, such as color, would be separated in parallel from the rest of the scene. Next, spatial registration of the other target feature within that multi-item candidate grouping would guide target search. For appropriate parameter ranges, this two-step recursive process would yield fast search times that within an *item = object* paradigm would be interpreted as evidence for simultaneous or parallel processing of feature conjunctions. Our theory thereby seeks to explain how seemingly minor experimental manipulations, such as variations in item spacing, can differentially modulate identical underlying processes in a manner that makes those processes *appear* at times serial and at times parallel.

Section 2. A Review of Psychophysical Studies and Models of Visual Search

If search were to be conducted by retinotopic sets of specialized object detectors, a different detector for every ecological object in each possible visual position would be required. Given the many thousands of objects that one encounters daily, this arrangement would result in a paralyzing combinatorial explosion (Tsotsos, 1990). It is fortunate that ecological objects can be distinguished on the basis of variation in a much smaller set of boundary and surface properties. There is now accumulating experimental evidence concerning how such properties are represented by the visual cortex in interactive, retinotopic neural maps (DeYoe & van Essen, 1988). Over the past several decades, much research concerning visual search has also been devoted to discovering how these visual representations interact with the attentive processes whereby they are detected and bound into visual objects. This section reviews and evaluates key data and hypotheses about how this process may occur.

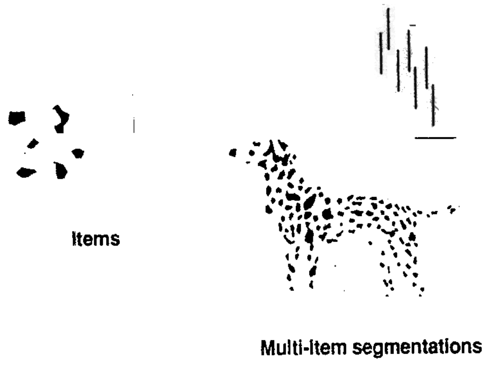


Figure 2. Whereas search items can be arbitrarily defined, camouflaged, and partially occluded, environmental objects automatically yield multi-item perceptual segmentations.

Parallel Search for Stimulus Features

Early psychophysical experiments using tachistoscope displays established that in a field of simple items (e.g., oriented lines, curves, Xs, or Os, of various colors) a *disjunctive* target, distinguished from distractors by a single stimulus feature (e.g., orientation or color), could be detected by parallel processing of the scene (Treisman & Gelade, 1980). That is, regardless of the number of distractors, a correct decision on whether the target is present in the scene can be made in approximately the same amount of time (Figure 1a). By contrast, it was found that a target distinguished only by a conjunction of certain features demands an effortful search for which detection time increases linearly with the number of distractors (Figure 1b). With increases in the number of distractors, the average amount of time needed to recognize the absence of a target increases at twice the rate as the average amount of time needed to recognize the presence of a target. This would occur if a serial search took place because, on average, the presence of a target is detected after investigation of about half of the items in the display, whereas certainty of absence requires investigation of all the items.

These early results suggested that search could at times be accomplished by parallel registration of elementary stimulus features. If parallel registration failed to discern the target, search could proceed by a serial process that links or binds feature conjunctions for recognition. This rule was embodied in the original version of a theory of visual search called Feature Integration Theory (Treisman & Gelade, 1980). This theory predicts serial search for conjunctive targets and parallel search for disjunctive targets.

Parallel Conjunctive Search

The paradigms used in early studies involved simple form–color items. Later work showed that targets distinguishable by the conjunction of depth (binocular disparity) and color or depth and motion (Nakayama & Silverman, 1986) and even motion and color could be detected at a rate independent of the number of distractors. One hypothesis offered by Nakayama and Silverman was that when a particular depth plane can be segregated from the rest of a scene it can be quickly searched for the presence of another feature. These findings contradict the conclusion that fast search can only be conducted for features.

The results of Wolfe et al. (1989) directly challenged the interpretation of early findings concerning parallel versus serial search. They reported that searches for form–color conjunctive targets can occur in parallel provided that feature salience is great enough. By using cathode ray tube (CRT) displays, they showed that whereas at low color contrasts a red vertical bar among red horizontals and green verticals cannot be found by parallel processing, an increase in item–item and item–background color contrasts in otherwise identical scenes can result in a seemingly parallel search. They in fact reported a continuum of flat to steep slopes resulting from varying saliency factors (Wolfe et al., 1989). This work led to the hypothesis that parallel conjunctive search is a normal capability that is hindered by a low signal-to-noise ratio.

Wolfe et al.'s (1989) hypothesis is designed to address "a curious characteristic of the standard feature integration model [which] is that the model holds that the parallel and serial processes are autonomous. The serial process cannot use information collected by the parallel process" (Wolfe et al., 1989, p. 427). They suggested that if feature saliency is strong enough, parallel processes can distinguish regions of a scene worth searching and pass that information on to the serial search mechanism. In a conjunctive search, the intersection of multiple regions, each distinguished by a different target feature, could unambiguously specify the target as the item most worth searching. This model has been named *Guided Search* because it postulates that information from two or more retinotopically registered feature dimensions is combined to guide item selection. This hypothesis is supported by data on search for triple conjunctive targets that differ from each distractor in two out of three features. Search times in these scenes increase more gradually with the number of distractors than search times in simple conjunctive scenes.

Revising Feature Integration Theory

In response to recent data and the Guided Search model, Treisman and Sato (1990) have proposed an updated feature integration hypothesis. Specifically, they have introduced a feature inhibition mechanism by which the simultaneous inhibition of all distractor features can yield fast conjunctive searches. This version of the theory is similar to Guided Search, except that inhibition of nontarget features rather than activation of target features is the mechanism by which item locations are rated.

Both Guided Search and the revised Feature Integration theories must account for the serial searches seen in the multitude of early experiments as the failure of the parallel conjunctive search mechanisms because of a low signal-to-noise ratio. However, the early physical stimuli that consisted of tachistoscopically displayed painted scene cards represent more ecologically realistic stimuli than do CRT displays. Phosphor displays emit rather than reflect light and therefore they do not support an unambiguous surface interpretation. Moreover, they can be used to generate scenes of bright saturated colors that exceed the surface color contrast ranges typically seen on reflecting environmental surfaces (Arend, 1991). It is possible that the increases in stimulus saliency by the use of CRT displays may result in qualitative changes in the organization of the perceptual scene that are best explained by a theory concerned with how grouping processes organize a visual search.

The possibility that feature integration could be augmented by grouping processes was noted by Treisman (1982, 1985). Treisman and Sato (1990) further discussed the possibility that conjunctive searches could be facilitated by the segregation of scenes into item groupings based on a single target feature. They suggested that these groups could then be treated as units, reducing the number of investigations required to determine target presence. Subjects are often aware of global shapes composed by items sharing a single feature. In Treisman and Sato's Experiment 1, subjects were, on some blocks of trials, asked to search for a target and, on some other blocks of trials using the same stimuli, asked to match the shape of multi-item regions formed by similarity on some featural dimension to luminance-defined standards. A strong positive correlation (.92) was found between reaction time for same/different judgments on the shape-matching task and time for conjunctive search (Figure 4). The shape task is evidently mediated by grouping mechanisms. Treisman and Sato (1990) concluded that there must be "some shared factor determining the efficiency of segregation and global matching and the speed of search for a conjunctive target" (p. 464). They did not, however, specify how the segregation of multi-item groupings might be quantified and integrated into the original theory. Instead, they focused on the feature inhibition hypothesis, citing data that suggested additivity of feature effects. An explanation of these data as a consequence

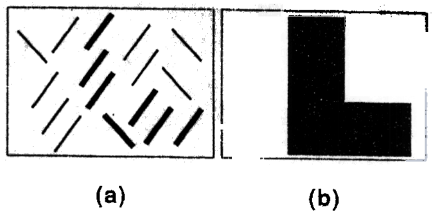

(a) (b)

Figure 4. Thin lines stand for red and thick for green; the target is a green line sloping down from left to right. In separate trials, Treisman and Sato (1990) asked subjects to determine match or mismatch of the shape defined by, for example, green items in (a) with a luminance defined shape such as the L of (b). Subjects matched global shapes with errors of 6% or less. They found a strong correlation (.92) between ease of segregation and global matching and the speed of conjunctive search and concluded that some shared factor determines the ease of segregation and conjunctive search. From "Conjunction Search Revisited" by A. Treisman and S. Sato, 1990, *Journal of Experimental Psychology: Human Perception and Performance, 16,* p. 463. Copyright 1990 by the American Psychological Association. Adapted by permission of the authors.

rather than a contradiction of the interaction of grouping and search are addressed in a subsequent section of this article.

Another theory that includes a provision for the effects of grouping on search is that of Bundesen (1990). The provisions are ancillary, however, rather than fundamental, to a theory that is primarily a computational analysis of constraints on filtering and categorization of item featural dimensions. Because the Bundesen (1990) model includes a provision that "groups [of items] compete for attention on a par with those individuals that make up the groups" (p. 536), and search efficiency is affected by group or item saliency, the theory can account for data such as those of Mordkoff, Yantis, and Egeth (1990), in which the minimal response time in conjunctive scenes with two targets is faster than of scenes with one target. Our theory's analysis of the Mordkoff et al. (1990) data appears later in this article. Bundesen (1990) also assumed, however, that "by being embedded in a group, an individual suffers a general loss in eta values" (p. 536), where "$\eta(x,i)$ is the strength of the sensory evidence that element x belong to category i" (p. 524). The assumption is opposite in effect to what we argue is the consequence of grouping of target and nontarget items.

Duncan and Humphreys (1989) have presented a theory of visual search based on grouping that is much closer in spirit than the Bundesen (1990) model is to our own. Their theory includes an elegant analysis of how similarity and disimilarity of targets and nontargets affects response times in visual search. Perceptual grouping can speed up rejection of clusters of nontargets by "spreading suppression," a process which is related to, though mechanistically and functionally distinct from, our own analysis of grouping effects. In addition, as with Bundesen's theory, a key divergence of the Duncan and Humphreys model from our own concerns the treatment of groups containing both target and distractor items. Although they argue that "selection will be enhanced by . . . decreasing grouping between target and nontarget" (p. 450), our own approach admits for either facilitation or interference with efficient search through such groupings. Although similarity of targets and distractors across all feature dimensions can act to camouflage a target, a grouping of, for example, a single target with several distractors by their value on one featural dimension (e.g., color) may be immediately reorganized into two subgroups on another, sufficiently salient dimension (e.g., orientation). One of the subgroups will then contain only distractors, which can thereby be rapidly rejected, and the other subgroup will contain only the singleton target.

Multi-Item Grouping in Search

Results of many recent experiments support the hypothesis that multi-item boundary and surface representations influence visual search. Humphreys et al. (1989) used Ts among upside-down Ts, and reported that search was greatly facilitated when the items were arranged to form a coherent global shape (Figure 5). They concluded that when multiple items can be organized into a familiar shape they can be treated as a whole. Otherwise, each item must be investigated separately. In the authors' words, "visual search cannot be understood independently of the processes mediating grouping and segmentation. The processes that we bring to bear in simple search tasks seem predicated on more general visual processes which are crucial for object recognition" (Humphreys et al., 1989, p. 278).

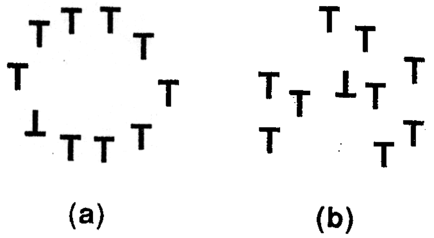


Figure 5. Organizing items into coherent shapes (e.g., a ring) speeds search. Finding an inverted T in scenes such as (a) is faster than in scenes such as (b), with equal numbers of items. From "Grouping Processes in Visual Search: Effects With Single- and Combined-Feature Targets" by G. W. Humphreys, P. T. Quinlan, and M. J. Riddoch, 1989, *Journal of Experimental Psychology: General, 118,* p. 261. Copyright 1989 by the American Psychological Association. Adapted by permission of the authors.

Bravo and Blake (1990) dissociated items from features by demonstrating that search for orientation defined by a group of items yielded fast times that were largely independent of the number of oriented groups (Figure 6). This experiment makes clear that perceptual features may be more complex than experimenter-defined items; indeed, *features* may be better viewed as outputs rather than inputs of segmentation processes. Bravo and Blake (1990) quoted Neisser (1967) as follows: "[T]he units of preattentive vision are not local features but objects produced by perceptual grouping" (Bravo & Blake, 1990, p. 522).

Parallel Search for Surface Properties

More recent research has suggested that parallel processing can accomplish search for more than just elementary stimulus features. Enns and Rensink (1990) reported that parallel search

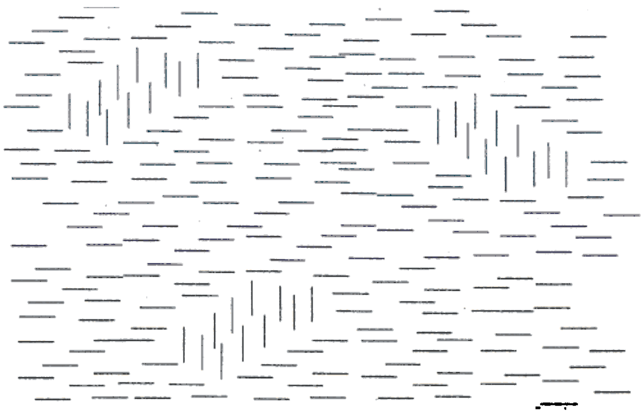
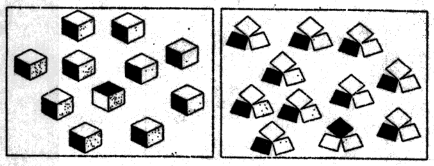


Figure 6. Parallel searches can occur for features defined by multi-item segmentations (i.e., diagonally oriented regions of vertical line segments). Bravo and Blake (1990) argued that perceptual groupings, not local features, are the units of preattentive vision. From "Preattentive Vision and Perceptual Groups" by M. Bravo and R. Blake, 1990, *Perception, 19,* p. 517, Figure 1a. Copyright 1990 by Pion Limited, London. Adapted by permission.



(a) (b)

Figure 7. The perception of scene-based properties like direction-of-lighting or surface color can guide search. In (a) the cube whose darkest face is on top "pops out," though the control in (b) does not. Search is efficient when items (polygons) can be organized into multi-item objects of a lighted scene. From "Influence of Scene-Based Properties on Visual Search" by J. T. Enns and R. A. Rensink, 1990, *Science, 247,* p. 722. Copyright 1990 by the American Association for the Advancement of Science. Adapted by permission.

can be conducted on the basis of properties such as surface color. Using 3-D shapes for items, they demonstrated that fast search for surface properties can result from the perception of direction of scenic lighting (Figure 7). Ramachandran (1988) has shown that the emergent perceptual organization of depth (disparity) that may take seconds to develop can form the basis for fast search for targets of a particular depth. Even more indicative of the importance of surface properties are recent studies indicating that shape from shading or shadows can form the basis for parallel search (Aks & Enns, 1992; Kleffner & Ramachandran, 1992). These new data suggest that parallel detection may operate at multiple levels of the perceptual hierarchy. Although elementary stimulus features may support fast search, so can surface properties, such as surface color, shading, shadows, or depth.

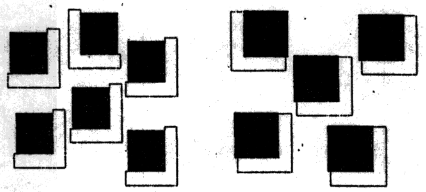
He and Nakayama (1992) have shown how 3-D surface properties involved in figure–ground separation can influence search. In their experiments, search was for Ls among backward Ls. The items were displayed stereoscopically and given relative depth by comparison to adjacent squares of contrasting color. If the items were displayed to appear in front of the squares, then search was relatively fast. If, instead, the items were displayed to appear behind the squares so that they could indicate occluded squares rather than Ls, search was difficult (Figure 8). The perception of an array of occluded squares was immediate and hard to ignore. This experiment demonstrates that 3-D surface representations and figure–ground separation properties input to the search process and are at least partially buffered against top-down influence. It is clear that search does not occur across scenic locations independent of grouping. Nor can the specification of arbitrary target items be expected to define the emergent surface organization. He and Nakayama (1992) concluded that "visual search has little or no access to the processing level of feature extraction but must have as an input a higher level process of surface representation" (p. 231).

In contrast, both the Feature Integration model with feature inhibition and the Guided Search model assume that somehow the mechanisms of search compute high target likelihood in the intersection of the sets of items that are similar to the target along single featural dimensions. For these models, a source of error and therefore of steep slopes in plotting response time as a function of number of distractors is the high evaluation of items in the union rather than the intersection of these sets. It may be that candidate groupings can be ordered by such calculations. However, the data reviewed here show that items are not always separated before search. In fact, it is often when items easily form multi-item groupings that search is easiest.

Parallel Recognition

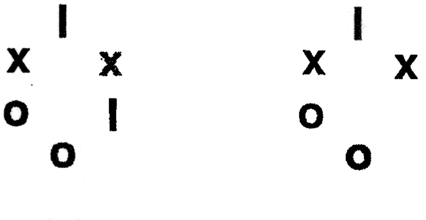
Revised Feature Integration Theory (Treisman & Sato, 1990) and Guided Search and Guided Search 2.0 (Wolfe, in press; Wolfe, et al., 1989) explain search as the result of the interaction of parallel and serial mechanisms. In both models, conjunctive recognition can only operate on retinotopic input from the feature and property arrays at single-item locations. Thus, although the models differ in their parallel mechanisms, they share the property that the problem of binding together information from multiple featural dimensions is solved by a serial mechanism.

Another line of research has pursued the hypothesis that under restricted conditions binding can be accomplished for multiple items at the same time. Several researchers have proposed that all visual searches are parallel but that there is a limited parallel processing capacity that is easily exceeded in scenes with many items. Pashler (1987) has proposed that parallel



(a) (b)

Figure 8. Surface organization affects search speed. The forward white L is found more readily in (a) than in (b), where forward and backward Ls look like occluded squares. From "Surface Features in Visual Search" by Z. J. He and K. Nakayama, 1992, *Nature, 359,* p. 231. Copyright 1992 Macmillan Magazines Limited. Adapted by permission.



(a) (b)

Figure 9. The fastest search times for sparse scenes containing two targets (e.g., bold X) in (a) are faster than for scenes with a single target, as in (b), violating any model postulating a serial search among items. This suggests that multi-item grouping controls recursive search. From "Dividing Attention Between Color and Shape: Evidence of Coactivation" by J. T. Mordkoff and S. Yantis, 1993, *Perception and Psychophysics, 53,* p. 361. Copyright 1993 by the Psychonomic Society, Inc. Adapted by permission.

binding of form and color has a maximum capacity of eight items.

This observation finds support in recent research showing that in six-element displays of colored Xs, Os, and Is, the fastest response to scenes including multiple targets (e.g., two red Xs instead of one) is faster than the fastest response to scenes including single targets (Figure 9; Mordkoff et al., 1990). These data are a striking contradiction of the serial binding hypothesis, which predicts that the fastest responses to scenes should not vary with the number of targets. That is, whereas a serial search among items would yield a greater number of fast search times for double-target scenes, as opposed to single-item scenes, the fastest search times for each scene type would be expected to be equivalent because the fastest possible response would correspond to felicitously choosing a target item first, regardless of whether one or two targets appeared. The results suggest, instead, that multiple items are evaluated simultaneously and that greater evidence for target presence is combined across items, supporting faster recognition.

Further support for the simultaneous evaluation of multiple items comes from recent data on monkey visual search (Chelazzi, Miller, Duncan, & Desimone, 1993). Monkeys were presented with scenes including a randomly located target shape and up to four distractor shapes. They were rewarded for foveating the target shape. During the task, inferotemporal (IT) cells that had previously been found to respond selectively to the target or to the distractor shapes were continuously monitored. Initially, IT cells responsive to both target and distractor shapes fired as if a multi-item group was being attended. Just before a saccade to the target item, the firing rate for the IT cells selectively responsive to the distractor shape decreased to the baseline level. These data are inconsistent with theories that require serial deployment of attention across items. Instead, they suggest that search occurs by a recursive narrowing of attention within multi-item groups.

In other research, conjunctive searches have been found to be easier in scenes of well-separated items than in scenes in which items are clumped together (Cohen & Ivry, 1991). These authors hypothesized that coarse location binding can occur in parallel across a scene. However, as we show quantitatively, the distance between scenic items can be an important factor influencing the spatial segregation of item groups for reasons other than resolution. This occurs, for example, when segmentation processes generate qualitatively different groupings, or different numbers of groups, as item spacing is varied.

Section 3. What Are the Perceptual Units of Search?

We claim that multi-item boundary and surface groupings not only influence visual search but are, in fact, the perceptual representations on which the search process is based. In certain experimental paradigms, single experimenter-defined items may be the only functional "groups." More generally, however, the identification of a grouping that includes multiple items speeds search by reducing the total number of candidate visual regions (N) that have to be serially investigated. Factors that influence boundary and surface grouping—such as featural contrasts, item spacing, and spatial arrangement—alter N, yielding variations in search time.

Our argument has until now emphasized the usefulness of the multi-item grouping hypothesis in giving unified explanation to a wide range of psychophysical search data. However, this hypothesis is also compatible with key functional demands that constrained the evolution and development of the visual system. Even without the benefit of a priori knowledge, the visual system must be able to segment ecological objects whose detection and recognition is necessary for survival. Segmentation mechanisms must therefore be equipped to deal with the environmental objects and scenic conditions that are typically encountered (Figure 2). Mottled surface coloration, natural animal camouflage, and partial object occlusion are all obstacles to the perception of ecological objects because they result in featural discontinuities that do not signal object boundaries. Successful segmentation demands mechanisms that can group similar featural regions that are spatially separated and segregate whole areas including these regions. In experimental scenes, these same mechanisms may encourage the segmentation of regions that group together what the experimenter considers to be distinct items. Individual items may, in fact, often be treated as textures or surface groupings by a system that has evolved to understand naturally occurring scenes.

If it is indeed true that bottom-up mechanisms drive the formation of these emergent perceptual units, then limits must exist on the capacity of semantic or even visual definitions of target items to exert top-down influence over preattentive grouping mechanisms. The ability of bottom-up processing to accurately distinguish ecological objects depends on a certain amount of autonomy or resistance to top-down interference. Otherwise, expectations would routinely result in perceptual illusions (Kanizsa & Luccio, 1987). We do not mean by these remarks to deny that perceptual grouping can be guided by top-down processes (Grossberg & Mingolla, 1985; Yantis, 1992) but only to point out that some groupings *emerge* from the structure of scenic input without the help of top-down influences.

Section 4. Attentive Interactions of Boundary and Surface Representations With Object and Spatial Representations

Our model of visual search, specified computationally in Section 5 and simulated to quantitatively explain search data in Section 6, is based on three types of neural network models that have been undergoing development for some time: models of 3-D visual boundary and surface representation; models of attentive category learning, recognition, priming, and memory search; and models of attentive spatial representation. Our search model shows how properties of these boundary, surface, object, and spatial processes can be incorporated into a computational algorithm that is capable of providing a unified quantitative explanation of many experiments about visual search. The search model is specified algorithmically, rather than neurally, because it has not yet been possible to define a neural architecture that combines all these elements and their interactions, which in vivo are spread across visual cortex, temporal cortex, and parietal cortex, in interaction with frontal cortex, among other structures. Our computational model describes the types of interactions among boundary, surface, object, and spatial representations that are capable of explaining many search data. The model hereby imposes additional design constraints on the ongoing development of the model neural architecture. The predictive success of the model also provides additional evidence that 3-D boundary and surface representations that interact reciprocally with mutually interacting object and spatial representations are sufficient to explain search data. In this sense, visual search phenomena are reduced to an exploration of how these four types of representations interact.

Our proposal that 3-D boundary and surface representations, possibly multi-element ones, are the inputs to the visual search process has its theoretical basis in the FACADE theory of preattentive 3-D vision, whose mechanisms model how visual inputs to the eyes are transformed, by the time they reach extrastriate visual cortex, into emergent boundary segmentations and surface representations that are capable of achieving figure–ground separation of scenic data (Grossberg, 1987a, 1987b, 1992, 1994; Grossberg & Mingolla, 1985, 1987; Grossberg & Todorović, 1988). This conclusion is supported not only by a large number of perceptual experiments, many of them performed after the corresponding FACADE predictions were published, but also from a theoretical analysis of how such representations, taken together, overcome various computational uncertainties that could otherwise undermine the perceptual process.

Within FACADE theory, a Static Boundary Contour System (BCS) models how the parvocellular cortical processing stream (LGN–Interblob–Interstripe–V4) generates 3-D emergent boundary segmentations from combinations of edge, texture, shading, and stereo image data (Figure 10). A Feature Contour System (FCS) models how the parvocellular cortical processing stream (LGN–Blob–Thin Stripe–V4) discounts the illuminant and fills-in surface properties of brightness, color, depth, and form within the boundary segmentations that are defined by the BCS. In our conception of visual search, the entire FACADE model forms the front end to the attentive search processes that are engaged during the visual object recognition process. In particular, the grouping properties that we have described so far in this article are among those that have been analyzed in the various articles on FACADE theory. Likewise, the object recognition and spatial localization properties that are used in our algorithmic search model have previously been derived from neural models of these processes. Throughout this article, we often use the term *grouping* to denote either a multi-element BCS boundary segmentation or an FCS surface representation and reserve the latter, more specialized words for expressly BCS or FCS processes, respectively.

The structure of FACADE preattentive boundary and surface representations clarifies how perceptual units are defined and manipulated during attentive visual research of static scenes. For example, properties of BCS-generated, emergent, multi-item boundary segmentations help us to understand how the global shapes studied by Humphreys et al. (1989) facilitate search. Properties of the FCS surface representations clarify how 3-D surface shape can influence search in the Enns and Rensink (1990) study. These FCS surface representations separate different combinations of color and depth into distinct slabs, or filling-in domains (Grossberg, 1987b, 1994). Such a slab organization helps to explain how unique conjunctions of color and depth are rapidly searched in the Nakayama and Silverman (1986) study. The boundary and surface representations are, moreover, organized to facilitate figure–ground separation and recognition of occluding and occluded objects (Grossberg, 1994), thereby helping to explain results of He and Nakayama (1992).

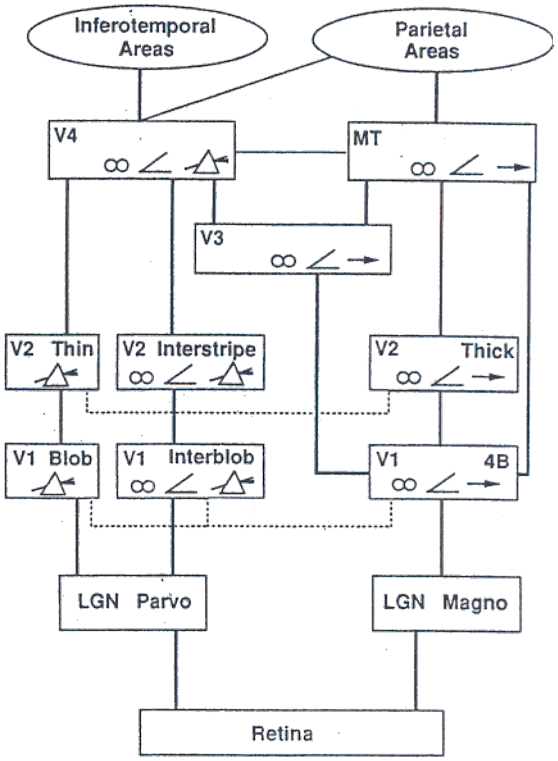


Figure 10. Schematic diagram of anatomical connections and neuronal selectivities of early visual areas in the macaque monkey. LGN = lateral geniculate nucleus (parvocellular and magnocellular divisions). Divisions of visual areas V1 and V2: blob = cytochrome oxidase blob regions; interblob = cytochrome oxidase-poor regions surrounding the blobs; 4B = lamina 4B; thin = thin (narrow) cytochrome oxidase strips; interstripe = cytochrome oxidase strips; V3 = visual area 3; V4 = visual area(s) 4; MT = middle temporal area. Areas V2, V3, V4, and MT have connections to other areas not explicitly represented here. Area V3 may also receive projections from V2 interstripes or thin stripes. Heavy lines indicate robust primary connections, and thin lines indicate weaker, more variable connections. Dotted lines represent observed connections that require additional verification. Icons: rainbow = tuned and/or opponent wavelength selectivity (incidence at least 40%); angle symbol = orientation selectivity (incidence at least 20%); spectacles = binocular disparity selectivity and/or strong binocular interactions (V2; incidence at least 20%); arrow = direction of motion selectivity (incidence at least 20%). From "Concurrent Processing Streams in Monkey Visual Cortex" by E. A. DeYoe and D. van Essen, 1988, *Trends in Neuroscience, 11,* p. 223. Copyright 1988 by Elsevier Trends Journals. Adapted by permission.

As noted earlier, two types of attentive processes are hypothesized to interact with preattentive boundary and surface representations during visual search of a static scene (Figure 11). One process explicates concepts of object attention (Duncan, 1984), and the other process explicates concepts of spatial attention (Posner, 1980). In neurobiological terms, these processes model part of the "What" cortical processing stream for object learning, categorization, and recognition, and the "Where" cortical processing stream for spatial localization and

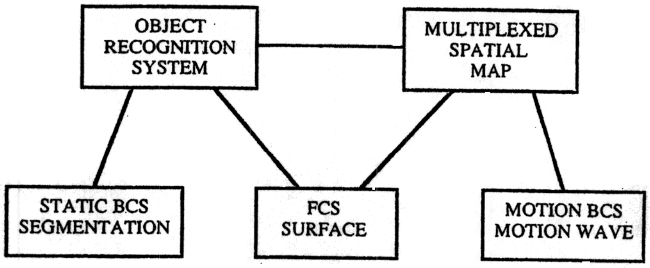


Figure 11. The reciprocal interactions of the Object Recognition System (ORS) with the Boundary Contour System (BCS) and the Feature Contour System (FCS) are supplemented by reciprocal interactions with an attentive Spatial Map. These object-based and spatial-based interactions are used to coordinate attentive object recognition, spatial orientation, and visual search. Expressed somewhat more technically, the Static BCS and FCS (which models aspects of the parvocellular cortical processing streams) interact reciprocally with the ORS (which models aspects of inferotemporal cortex) for purposes of attentive visual object recognition. The FCS and a Motion BCS (which models aspects of the magnocellular cortical processing stream) interact reciprocally with a Multiplexed Spatial Map (which models aspects of parietal cortex) for purposes of attentive spatial localization and orientation. Both systems interact together to define spatially invariant ORS recognition categories and to control visual search. From "3-D Vision and Figure–Ground Separation by Visual Cortex" by S. Grossberg, 1994, *Perception and Psychophysics, 55,* p. 58. Copyright by the Psychonomic Society, Inc. Reprinted by permission.

orientation (Goodale & Milner, 1992; Mishkin, Ungerleider, & Macko, 1983; Ungerleider & Mishkin, 1982).

Our visual search model incorporates properties of a visual object recognition system (ORS) that models aspects of neurobiological data showing how the inferotemporal cortex interacts with visual cortex, hippocampal formation, and pulvinar for purposes of attentive object search, learning, categorization, and recognition (Desimone, 1991, 1992; Desimone, Schein, Moran, & Ungerleider, 1985; Desimone & Ungerleider, 1989; Gochin, 1990; Gochin, Miller, Gross, & Gerstein, 1991; Harries & Perrett, 1991; Miller, Li, & Desimone, 1991; Mishkin & Appenzeller, 1987; Perrett, Mishkin, & Chitty, 1987; Spitzer, Desimone, & Moran, 1988). The Object Recognition System model is an Adaptive Resonance Theory (ART) network (Carpenter & Grossberg, 1991; Grossberg, 1980, 1987c). Recent neurophysiological experiments on object recognition in monkeys suggest that neurons in the inferotemporal cortex exhibit properties that are consistent with ART mechanisms. See Carpenter and Grossberg (1993) for a review. For our purposes, we use the facts that an ART recognition category can be used to read out a top-down expectation that primes consistent combinations of BCS boundary and FCS surface properties. Such a prime can be used to focus attention on expected cue combinations and to regulate a search for consistent or inconsistent conjunctive properties. A good enough match leads to a state of resonant attention, binding, learning, and recognition. A mismatch leads to a reset event that can further propel the search. The coarseness of the search can be regulated by a parameter called *vigilance* that computes how good a match is needed before search terminates.

A precursor of the present use of ART properties to explain visual search data is found in Grossberg (1978, Section 61). There an analysis is given of how *automatic* processing and *controlled* processing data from the visual search experiments of Schneider and Shiffrin (1976) may be explained by using ART mechanisms. Schneider and Shiffrin had identified automatic processing as a parallel search process and controlled processing as a serial search process. Grossberg (1978) described how both types of data may be explained by recursive operation of the same parallel matching and memory search mechanisms that are invoked herein. It is also suggested how the recognition categories and primed expectations that are learned in the Schneider and Shiffrin (1976) varied mapping (VM) and consistent mapping (CM) conditions differ. In particular, the CM condition enables the subject to learn visual chunks that control parallel matching against all memory set items, much as happens in the multi-item groupings discussed herein. Our search model augments these ART matching and search mechanisms with computationally precise properties of interacting boundaries, surfaces, and spatial maps.

Why are both object and spatial representations needed? At the risk of oversimplification, we can briefly hypothesize that object representations are relatively insensitive to variations in such variable object properties as their retinal position. A serious combinatorial explosion would occur if the brain needed to learn a separate representation for each object at every retinal position. In contrast, spatial maps do represent object positions. The interactions that link object and spatial representations are designed to enable desired objects to be recognized through spatially invariant recognition codes and to be spatially localized and engaged during behavior through spatial maps.

Our conception of how spatial maps work arose from models of Carpenter, Grossberg, and Lesher (1992, 1993) and Grossberg and Kuperstein (1986, 1989), which analyze aspects of neurobiological data about how the parietal cortex interacts with the visual cortex for purposes of spatial localization and orientation (Anderson, Essick, & Siegel, 1985; Fischer & Breitmeyer, 1987; Maylor & Hockey, 1985; Mountcastle, Anderson, & Motter, 1981; Rizzolati, Riggio, Dascola, & Umita, 1987; Wurtz, Goldberg, & Robinson, 1982). FACADE theory proposes that the BCS and FCS each interact reciprocally with the ORS to bind boundary properties (e.g., texture segregation) and surface properties (e.g., surface color and form) into categorical object representations. The FCS also interacts reciprocally with a Multiplexed Spatial Map to determine the spatial locations of particular combinations of surface properties (Figure 11). These "What" and "Where" processes are also envisaged to interact reciprocally with one another. Grossberg (1994) discusses these interactions in greater detail, including how 3-D boundary and surface representations are formed by using preattentive mechanisms, how they may interact with attentive object and spatial processes, and how search for moving targets may be controlled. This analysis incorporates modeling results about motion detection and segmentation by the magnocellular cortical processing stream (Grossberg & Mingolla, 1993; Grossberg & Rudd, 1989, 1992) and about the visual and attentive control of saccadic eye movements (Grossberg & Kuperstein, 1986, 1989).

A recapitulation of all these results herein would take us too far afield. The main purpose of this brief summary is to emphasize that the representations and operations of our search model were not invented to explain the search data discussed herein.

They were introduced to explain entirely different types of data. The search model hereby shows how to link search data, through the corresponding neural models, to the underlying processes of visual perception, object recognition, and spatial orientation that they reflect. For the remainder of this article, we describe our visual search model in computational terms and show how it can be used to quantitatively simulate data about visual search of static images. These computational rules are consistent with the neural model outlined earlier for visual, object, and spatial processing by visual, temporal, and parietal cortex. The rules thus have a plausible interpretation in terms of a large body of perceptual and neural data other than that on visual search that we simulated here.

Section 5. The SOS Algorithm: Spatial and Object Search

Although conceived as a heterarchy of neural networks with continuous and asynchronous dynamics, our algorithmic realization of visual search can be summarized as a four-step process (Figure 12). In Step 1, preattentive processing of the visual scene results in retinotopic registration of stimulus features. In Step 2, these retinotopic featural arrays support boundary segmentation and surface formation, which group the scene into separate candidate regions. During searches for known targets, a top-down priming signal can influence the organization of the search regions. This grouping step has been assumed by others to immediately and correctly define scenic objects, which in laboratory stimuli are individual target or distractor items. In Step 3, a candidate region is selected for further analysis. In a directed search, this step could be influenced by either bottom-up salience or top-down priming of target features selected by an ORS. For example, a salient emergent boundary segmentation in the BCS, as in the Humphreys et al. (1989) study, might bias BCS–ORS–FCS interactions; or a color prime, as in the Egeth, Virzi, and Garbart (1984) study, might amplify activation of a particular color–depth slab and thereby bias FCS–ORS–BCS interactions. Finally in Step 4, feature groupings within the selected candidate region must be compared with the stored target representation, which is assumed to be a categorical representation (or representations) within the ORS. A mismatch between all these feature groupings and the stored target representation causes a return to Step 3 for reset of the old region and selection of a new candidate region. A partial mismatch between the features in a multi-item candidate region and the stored target features may trigger a more vigilant search *within* the candidate region. This would cause a return to Step 2 to further segment the candidate region into subregions on the basis of a new featural dimension. If this recursive process does not yield a target match, then the entire candidate region is discarded and a new one processed. Search terminates when a match is found.

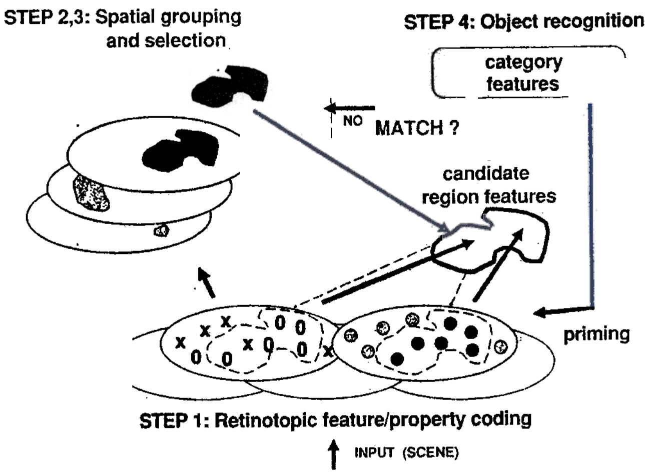


Figure 12. A visual search architecture. The scene is analyzed in parallel through the retinotopic registration of boundary segmentations and surface slabs. Candidate target regions are separated for recognition by the interaction of spatial grouping and object recognition processes.

Within this system, each step need only be partially completed for the next step to begin. That is, partial activation of one neural layer, which is not itself yet at equilibrium, may suffice to initiate processing at successive layers. In a response time minimization task, a speed–accuracy trade-off would determine the optimal confidence level at which each step would be complete enough for the next step to take place. The correct prediction of average search time or response time (RT) for a given scene requires the determination of a duration for each step and an algorithmic computation of time needed for scene grouping and search.

For many scenes, there may be no sharp distinction between segmentation, surface selection, and definition of a candidate region for search. In the computer simulations described later, processing times for these steps are lumped into a constant duration, which is added to search time for each candidate region that is chosen for recognition.

Our simulations instantiate three segmentation heuristics:

1. *Boundary and surface units influence grouping.* Conflicting groupings that are supported by separate featural dimensions are resolved in favor of a grouping supported by a single multiplexed featural dimension, for example, color or a prescribed color–depth combination. Grouping is biased to occur between items sharing the target feature value along that single dimension, much as a whole color–depth slab in the FCS can be primed.

2. *Spatial context influences grouping.* In simulations, this complex factor is simplified into the rule that featurally similar items can be grouped into the same candidate region if they can be connected by uninterrupted spatial paths whose width corresponds roughly to the diameter of items. This simplification of the spatial influences on grouping proves sufficient to account for a considerable amount of important search data. In general, we conceive this grouping as being performed by the full multiscale interactions of the BCS and FCS of FACADE theory, despite our simplifications.

3. *Featural contrast influences grouping.* In simulations, the probability that item groupings can become candidate regions is a function of stimulus saliency. For example, color saturation could bias a particular FCS color–depth slab, a particular BCS boundary segmentation, or both. High saliency may also allow textural elements or search items to perceptually pop into different perceived depth planes regardless of spatial arrangement, in a manner that is modeled in Grossberg (1994). Groupings that might otherwise be disallowed, because the path be-

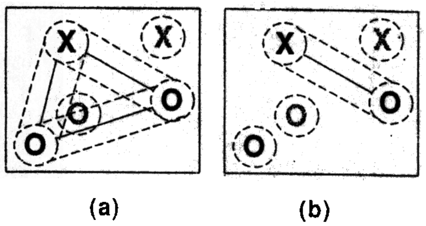


Figure 13. (a) The equations for the lines connecting all the items sharing target color are computed and stored. (b) If a distractor item of nontarget color is within a path of approximately item width that connects two target-color items, then that connection is disallowed and the grouping line is removed.

tween items that are alike on a single dimension (e.g., color = red) is interrupted by an item of different value on that featural dimension (e.g., color = blue) might nevertheless occur as, for example, when the contrast of red-to-background is much higher than the contrast of blue-to-background (see Figure 13).

To predict the mean search or RT for scenes of a given number of items requires that a duration be assigned to each of the four steps as well as the algorithmic computation of search time for any given display based on the principles stated earlier. The mean search time of the algorithm for target-present conjunctive scenes can be approximated by the *target-present search equation:*

$$\mathrm{RT} = R + (N + 1) \times (S + M). \qquad (1)$$

RT for target-absent scenes is given by the *target-absent search equation:*

$$\mathrm{RT} = R + N \times 2 \times (S + M). \qquad (2)$$

In Equations 1 and 2, R is the duration necessary to complete Step 1 (retinotopic feature registration), S is the duration necessary to separate a candidate region by Steps 2 and 3 (segmentation and selection), M is the duration necessary to match the candidate region with the target representation in Step 4, and N is the mean number of candidate regions into which the scenes containing a certain number of items, pseudo-randomly distributed, are initially segmented. Each multi-item candidate region must be recursively searched, resulting in a recursion factor of 2 in each equation for conjunctive scenes. More generally, this factor of 2 could take on some other value, as for triple-conjunction searches. Evaluation of this factor is not, however, as simple as counting the number of display dimensions (e.g., color, orientation, and motion) that are varied in a scene, as not all combinations of values on all dimensions may be included in a given scene construction. In Equation 1, this factor is cancelled because in target-present scenes only about half $(N + 1)/2$ of the candidate regions (N), on average, have to be evaluated before the target is found. Note that these equations yield only an approximation of search time for a given display because in our algorithm a candidate region including a single item need not be recursively searched or recognized. Thus, search times would be expected to be slightly shorter than the equations suggest. However, the simplified equations serve well to fit the data curves considered in the next sections.

Section 6. Simulation of Psychophysical Search Data

In this section, the SOS algorithm is used to determine RTs for a number of psychophysical search scenes. The model is capable of quantitatively simulating the search data reviewed earlier as well as additional findings using a *single choice of the parameters R, S, M,* and an algorithmic solution to N. Such an algorithmic solution for N on each display trial is not ad hoc because our theory proposes that the candidate regions that form the units of visual search are flexibly determined "on the fly" by the brain and vary for each scene even within the same display paradigm.

Simulation Methods

To compare the performance of the algorithm to experimental data, we simulated both the experimental scenes and the algorithm. Simulation of the experiments required a program to convert information in the methods section of an experimental report into a sequence of scene representations suitable for use as input to the simulated algorithm. Visual search experiments typically involve the random placement of various numbers of items across predefined potential scene locations. Whenever restrictions on the placement of items were specified in published reports, these were simulated.

Our implementation of a display simulator produced and stored x- and y-coordinates for each item contained in a trial as well as the value of each item along each feature dimension. In addition, the size of each item was recovered from the experiment description. For some of the experimental reports simulated, scene and item size were specified in subtended retinal angle and viewing distance. In this case, Cartesian coordinates and dimensions were recovered to allow comparison across experiments. The size of each item was approximated by recovering an item radius corresponding to the item width given in the experimental reports. This approximation sufficed because items were typically nearly square. Thus, although the form of each item was stored, the size of all experimental items was modeled as if they were circular. Finally, an ordinal value for the degree of featural contrast in the scene was entered by the programmer on the basis of reported color saturation. For all of the simulation points plotted in this section, 50 trials were simulated to recover average search times.

The SOS search algorithm was also simulated. For convenience, only form–color experiments were modeled, and color was always assumed to be the feature initially used for grouping. Altering these assumptions, for example, by grouping initially on form does not materially alter the results of our simulations. For each trial, the program outlined here was executed. Each scene was first segmented into a set of item groupings according to the following procedure:

1. The equations of the finite length lines connecting each item of target color to each other item of target color were computed and stored (Figure 13a). This step implements the notion that a target color can act as a prime for grouping.

2. Next, the distances between the centers of each of the items of distractor color and these lines were computed. If any item of nontarget color was between two items of target color and within a fixed proportion (P) of item radius away from the line grouping those items, then that line was deleted from the list of item grouping lines (Figure 13b).

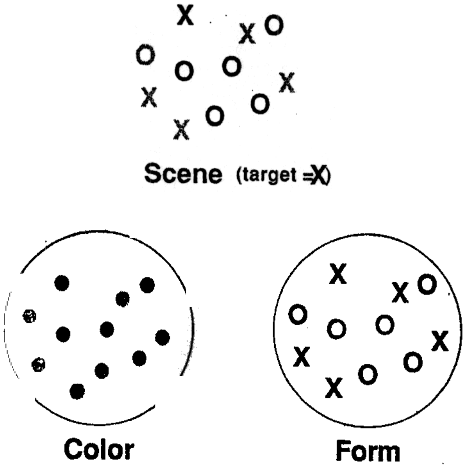


Figure 14. Step 1: The scene is analyzed through the retinotopic registration of boundary segmentations and surface slabs.

3. A percentage of these lines (C) inversely proportional to a nominal measure of the featural saliency, in this case color contrast, were also deleted at random from the line list.

4. The remaining set of lines was then used to create lists specifying item groups; that is, if a line connecting two items remained, those items were stored in the same item group.

This four-step segmentation procedure results in a set of N item groupings. It generates groupings of boundaries and colors that are consistent with how BCS/FCS representations would set up groupings for interaction with spatial and object mechanisms during search but does so by using far less computational time than would be required to run a full BCS/FCS simulation. Thus, this segmentation procedure should not be viewed as different from, but rather as merely cruder than, a full BCS/FCS implementation. After segmentation, the computation of search time proceeds as follows. An example is illustrated in Figures 14–18.

5. A constant duration (R) was added to the RT for the trial. This duration accounts for the time taken to register features (Figure 14).

6. An item group was selected at random from the list of unchecked groups. Note that this selection criterion could be replaced by one based on some measure of which group was currently winning a competition for selection without significantly distorting the reported results. Segmentation and selection time S was added to the RT for the trial. Parameter S accounts for the time needed to select a group of items from the background on the basis of a single feature (Figures 15 and 16).

7. The chosen group was compared with the target representation; that is, for each featural dimension, the target features were matched against the group features. The match time M was added to the RT for the trial. It accounts for the time needed to determine degree and nature of similarity between group and target representation (Figure 17).

8. The group was then recursively searched for target form (Figure 18). For example, if the entire group did not match the target on both dimensions but did match it on one dimension (e.g., color), then subgroups were formed and searched. Parameter S was again added to the RT for the trial. If this recursive segmentation yielded the target, parameter M was added to the RT for the trial and search was terminated. Otherwise, if segmentation yielded a nontarget form, M was added to the RT the trial, the group was marked as checked, and the search returned to Step 6.

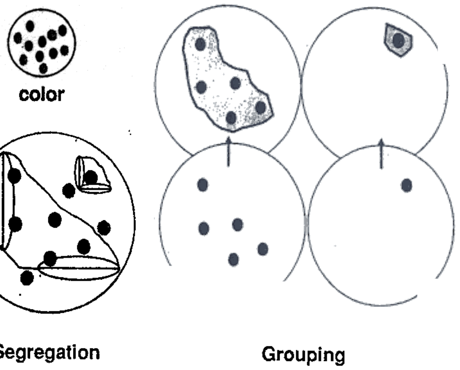


Figure 15. Step 2: The scene is segmented on the basis of a single feature dimension through processes of grouping and segregation.

Parameters

The parameter values S, M, R, the proportion of item radius used to model spatial influences on grouping (P), and the percentage of groupings disallowed as an inverse function of feature saliency (C) were selected to fit the data. This process of curve fitting was carried out in stages designed to explain an ever increasing set of data. For this reason it is discussed with reference to each piece of experimental data so as to best represent the actual procedure that was followed. Although parameter values were adjusted during development of the algorithm,

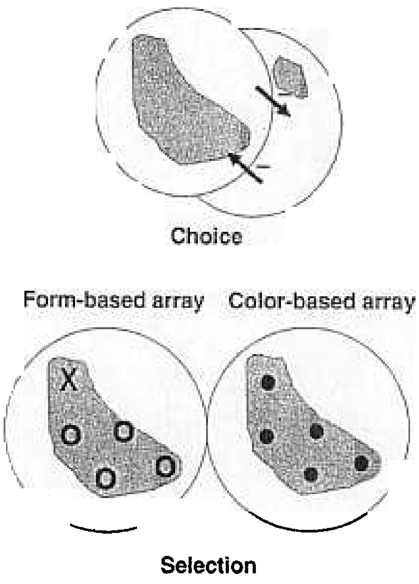


Figure 16. Step 3: A single candidate region is chosen across all feature maps for parallel conjunctive recognition.

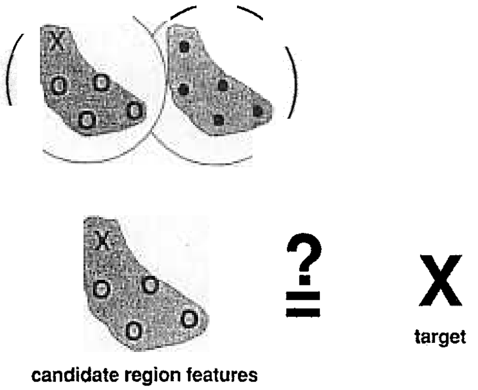


Figure 17. Step 4: The features within the candidate region are compared with the target representation through a limited capacity parallel process.

in the end *all the data were modeled by a single set of parameter values.*

Section 7. Simulation of Form–Color Conjunctive Search

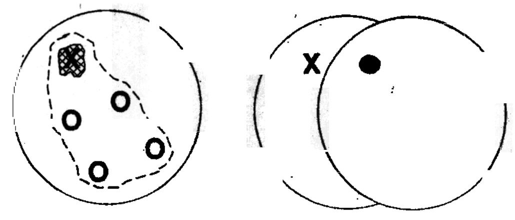
Treisman and Gelade (1980) conducted tachistoscope studies indicating serial conjunctive search. Wolfe et al. (1989) repeated this investigation in experiments using high-contrast scenes on a CRT and found fast conjunctive searches that could be explained by a parallel search process. The data from these fast form–color conjunctive searches may be quantitatively simulated by a two-step process. Initially, high color contrast supports grouping of target colored items into separate candidate regions or perceptual surfaces. These multi-item candidate groupings are then searched for the target form. Because there are fewer item groups than individual items, search times decrease. This process is essentially a recursive application of the original Feature Integration mechanisms with the exception that individual items would not necessarily become candidate groupings.

Both the Treisman and Gelade (1980) and the Wolfe et al. (1989) results are quantitatively simulated by the model. The scenes used to test the algorithm were generated by using the experimental description given in Experiment 7 of Wolfe et al. (1989). The viewing field was a square of length corresponding to 11.3° at a distance of 100 cm. Item dimensions corresponded to 0.85° square at this same distance and were randomly arranged in a slightly irregular 6×6 array. Color saturation was reduced, and the background color was changed to white to emulate the tachistoscope setup and reduce saliency.

Wolfe et al.'s (1989) data consist of essentially linear curves of varying slope. To model these data, factors determining both intercept or offset of the data lines and their slopes needed to be specified. The constant duration that offsets search among any number of distractors was captured by R, which includes the time to register retinotopic feature arrays as well as the time needed to generate a motor response. With $P = 1.0$, varying the percentage of groupings permitted as a function of stimulus saliency, in this case color contrast (C), resulted in different numbers of item groupings (N). The rate at which N increases with the number of items defines a slope that decreases with decreasing C. This property allows the data to be fit. The data lines were extrapolated to find an approximate y-intercept of $R = 540$ ms. Next, the target-present data slopes for both the high-contrast and the low-contrast data were fit. The data indicate a high-contrast slope of 6.1 ms/item and a low-contrast slope of 10.8 ms/item. Thus, the low-contrast slope is approximately 1.66 times steeper. When all the spatially allowable groupings were permitted in the high-contrast case ($C = 0\%$) and only 30% of the spatially allowable groupings were permitted in the low-contrast case ($C = 70\%$), then the ratio of the N slopes was found to be approximately 1.66. This ratio was checked by compiling average N values for various numbers of distractors. Table 1 shows N averaged over 50 trials for each number of scenic items.

Choice of parameter values for P and C fixed the ratio between slopes. The actual slopes of the data curves remained to be determined. By Equation 1, this requires setting a value for the sum $(S + M)$. In this case, 70 ms proved to model the data well. In all, this experiment constrained the parameters so that $R = 470$, $C = 70\%$ in the low-contrast case, $C = 0\%$ in the high-contrast case, and $(S + M) = 70$ ms. Parameters S and M were not individually constrained. It was found that the spatial path width parameter could be varied from $P = 1.0$ to $P = 1.5$ with little effect on the number of spatially allowable groupings (N).

Figure 19 shows the close match between simulation and experimental data. The target-absent slopes are just a result of doubling the target-present slopes, as in Equation 2. It should be noted that this simulation of data on fast form–color conjunctive scenes is consistent with recent neurophysiological data on



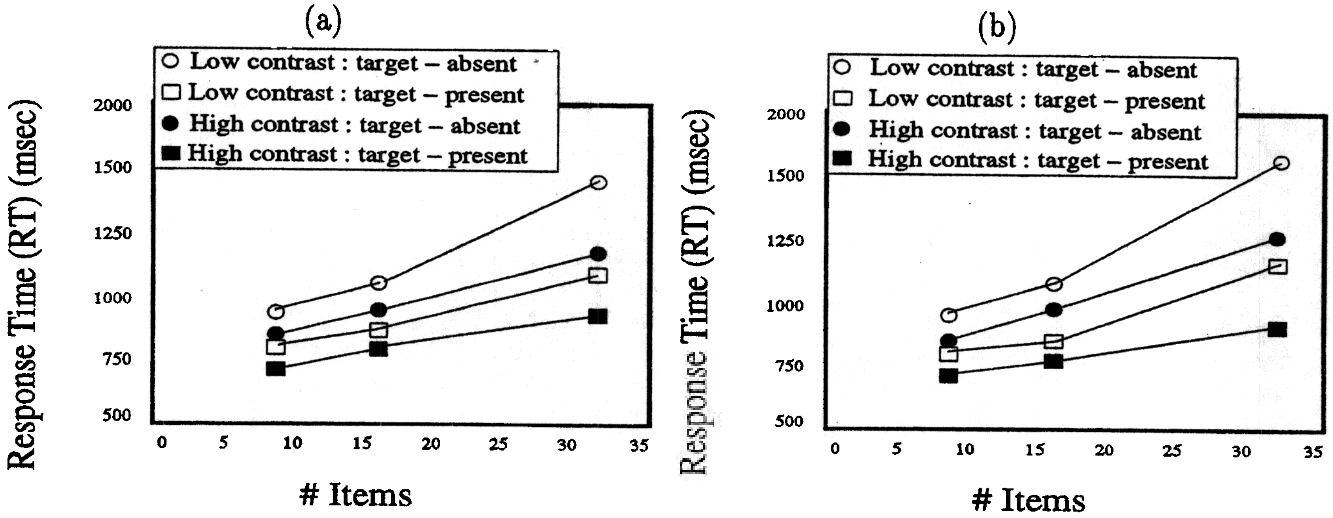
Segregation and Grouping Choice and Selection

Figure 18. Step 4a: If distractor features within the candidate region are too numerous to allow unambiguous determination of target presence, then the region is recursively segmented on the basis of the other target feature.

Table 1
Number of Candidate Groupings (N) Found for Scenes of Wolfe, Cave, and Franzel (1989)

	Number of items				
N	4	8	12	16	24
High contrast	1.14	1.26	2.14	2.74	4.66
Low contrast	1.14	1.95	2.6	3.36	5.8

Note. P (path width proportion) = 1–1.5; C (low-contrast) = 70%, and only 30% of the spatially allowable groupings were permitted; C (high-contrast) = 0%, and all spatially allowable groupings were permitted. A comparison of N values for the low- and high-contrast cases shows a relative slope ratio of approximately 1.66. See text for details.



Experiment
Redrawn from
Wolfe *et al.* (1989, Expt. 7)

Simulation

Figure 19. Increasing stimulus saliency yields faster conjunctive search. (a) From "Guided Search: An Alternative to the Feature Integration Model for Visual Search" by J. M. Wolfe, K. R. Cave, and S. L. Franzel, 1989, *Journal of Experimental Psychology: Human Perception and Performance, 15,* p. 426. Copyright 1989 by the American Psychological Association. Adapted by permission of the authors. (b) Model simulation.

monkey visual search (Chellazi et al., 1993). It can also be applied to data on fast conjunctive search in depth–motion and depth–color scenes.

Section 8. Color–Color Conjunctive Search

The data on color–color conjunctive search are also consistent with the model. It has been found that most scenes including color–color conjunctive items demand serial investigation (Wolfe & Friedman-Hill, 1992). In these scenes, segregation by color would tend to yield regions whose boundaries split right through pairings of squares that form items (Figure 20a). These regions could not be recursively searched because they would not contain full items. The grouping based on the first color would not lead to partial activation of the regions that include the second color. The spatial focus of attention would thus miss the conjunctive feature. Various exceptions to this finding support this hypothesis. In scenes of disk–annulus color–color items where the annulus is significantly larger than the disk, search times can be independent of the number of distractors (Wolfe & Friedman-Hill, 1992). In this case, the extent of target annulus color may be strong enough to support segregation of a multi-item candidate segmentation that does include the target disk colors, which can then be searched for target disk color (Figure 20b). This result has also been replicated by using houses with different wall and window colors. Although these data on color–color conjunctive search pose a serious problem for models like Guided Search and Feature Integration, they are consistent with the SOS model and illustrate yet again the importance of multi-item grouping in visual search.

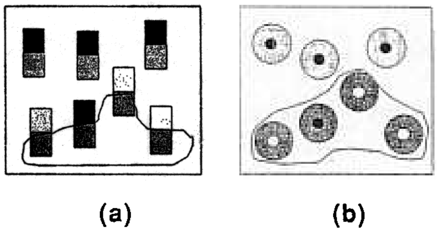


Figure 20. (a) Segmentation of a candidate region defined by a single target color yields a multi-half-item grouping that cannot be recursively searched for the other target color, thereby explaining steep search slopes. (b) Segmentation of a candidate region defined by a single target color yields a multi-item grouping that can be recursively searched, thereby explaining shallow search slopes.

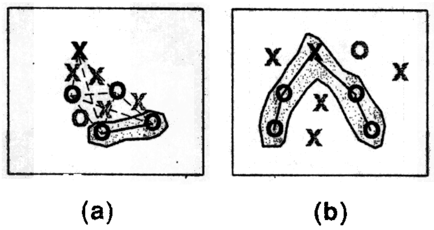


Figure 21. (a) In clumped scenes, the spatial grouping of items is difficult, and the scene is segmented into many candidate regions. (b) In spread-out scenes, spatial grouping is facilitated, and the scene can be segmented into very few candidate regions.

Section 9. Simulation of Clumped Versus Spread-Out Conjunctive Search

In the SOS model, multi-item groupings are formed only by items that can be connected by uninterrupted paths proportional to item width. In clumped scenes, the spatial segregation of multi-item target-color groups is hindered because it is less likely that there exists an unobstructed path between target-colored items. In spread-out scenes, there is a higher probability of clear paths and segregation is easier (Figure 21). This observation is capable of explaining data of Cohen and Ivry (1991) showing that conjunctive searches are faster in spread-out scenes.

Cohen and Ivry (1991, Experiment 3) found that in scenes in which conjunctive items are clumped together, search is slow and appears to proceed by serial processing, whereas in spread-out scenes search is much faster. The stimuli were presented on a MultiSync SD color monitor using yellow and blue Xs and Os. The items were displayed on two imaginary circles approximately 2.8° and 4.17° from the screen center. The character height of .46° was used to recover item radius. In the clumped condition, items were organized into groups separated by approximately .69°. In the spread-out condition, two adjacent letters were at least 1.37° apart. To reduce grouping effects, we restricted the runs of identical colors or letters along either circle to be 3 or less.

Cohen and Ivry (1991) reported target-present slopes of 18.2 ms/item and 5.2 ms/item, respectively, for the clumped and spread-out conditions. This results in a clump/spread-out slope ratio of 3.5. Because they made no attempt to lower the CRT color contrasts in the scene, the experiment was considered high-contrast, and all the spatially allowable groupings were permitted in our simulation of their data ($C = 0\%$). Despite the constraint on color runs, item groupings within and between the circles often occurred with $P = 1.0$. When P was increased to 1.5, the ratio between the clumped and spread-out slopes became 3.2. The simulation data used to check this slope ratio are reported in Table 2. Again, all N values were the result of averaging over 50 trials. As in the simulation of the Wolfe et al. (1989) data, R was set equal to 470 ms and $(S + M)$ was set equal to 70 ms. Figure 22 shows the close match between simulation and experimental data.

Increasing P beyond 1.5 altered N for the scenes used by Wolfe et al. (1989). Parameters P and C are not, however, brittle parameters because variations about their best values result in performance that degrades gradually and falls within the range of search slopes seen across different subjects. To illustrate this property, consider the limiting conditions as both P and C are increased. Grouping tends to disappear, and N approaches the number of items. The result is serial search among items as initially reported by Treisman and Gelade (1980). If P and C are reduced, grouping increases and N approaches one for all numbers of items. At this limit, the search slopes are flat.

Table 2
Number of Candidate Groupings (N) Found for the Scenes of Cohen and Ivry (1991)

	Number of items				
N	4	8	12	16	24
Spread out	1.60	1.08	1.58	1.94	2.60
Clumped	1.16	2.90	3.70	5.00	7.70

Note. P (path width proportion) = 1.5; C (high-contrast) = 0%, and all spatially allowable groupings permitted. A comparison of N values for the low- and high-contrast cases shows a relative slope ratio of approximately 3.2. See text for details.

Section 10. Simulation of Multi-Target and Single-Target Conjunctive Search

Minimal RT in conjunctive searches of scenes with two targets is faster than those with one target. Mordkoff et al. (1990, Experiment 3) reported results for six-item scenes. They used scenes viewed at a distance of 45 cm in which all items were located on a circle of radius 1.5 cm. Items were 1.1 cm tall and 0.7 cm wide. Each scene contained two Xs, two Os, and two Is, two colored red, two green, and two blue. Their result is significant because it violates the prediction of any model that hypothesizes a serial search among items.

This conclusion follows when one considers the following serial search scenario. The observer is confronted with six items and chooses one from the scene by some means, for example, randomly among all items, according to target likelihood computed by using evidence from all feature dimensions. This operation can be accomplished no faster than some time t. In a double-target scene, this fastest search time t would be more likely even if the means of item choice is random selection. However, the duration t would not be expected to vary as a function of the number of targets in the scene. Thus, the Mordkoff et al. (1990) data violate serial item search models.

These data can, however, be explained by a model that is based on segmentation and search among multi-item groupings. In scenes containing few items, separation of a target-color, multi-item grouping that contains one target would include a relatively high target-to-distractor feature ratio. In multi-target scenes, the target-to-distractor feature ratio within the multi-item grouping would be even higher (Figure 23). Within a real-time neural theory of object recognition such as ART, match time is faster when the input includes more evidence for target presence. See Grossberg and Stone (1986) for an analysis of data from lexical decision experiments where this is also true. Future research may help compare the predictions of this approach to multi-target tasks with those of hybrid race and coactivation models (Mordkoff & Yantis, 1993).

The match duration M can thus depend on the experimental display. For segmentations of few items that include the target, suppose that M is a decreasing function of the evidence for target presence within a multi-item grouping. In particular, M results from a limited capacity parallel process that selectively amplifies scenic data that match a prime while suppressing data that mismatch the prime. Such a matching process is an emergent property of suitably defined, competitive neural networks (Carpenter & Grossberg, 1991; Grossberg, 1980; Grossberg & Stone, 1986). Given system noise, the match time for a grouping of one target and one distractor has a mean of shorter duration than for a grouping of two targets.

The Mordkoff et al. (1990) data are given as cumulative probability distribution functions. To model these data, it is neces-

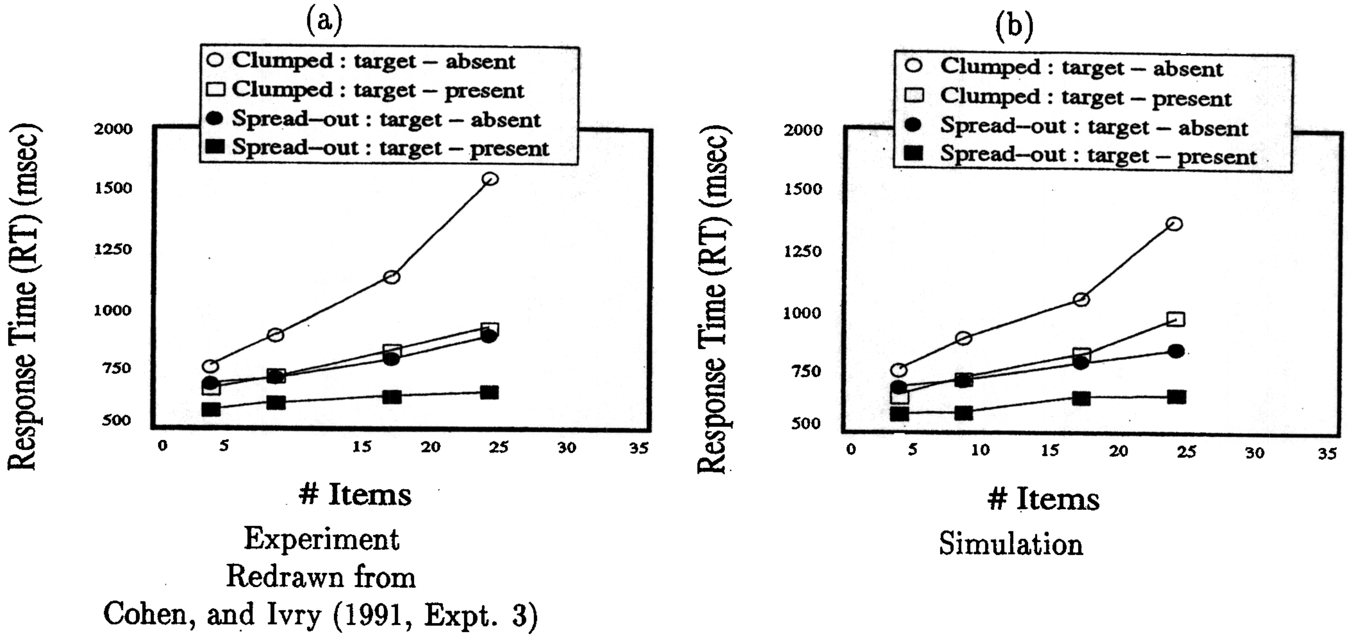


Figure 22. Spreading out items yields faster conjunctive search. (a) From "Density Effects in Conjunction Search: Evidence for a Coarse Location Mechanism of Feature Integration" by A. Cohen and R. B. Ivry, 1991, *Journal of Experimental Psychology: Human Perception and Performance, 17,* p. 898. Copyright 1991 by the American Psychological Association. Adapted by permission of the authors. (b) Model simulation.

sary to model RT as the result of a noisy or probabilistic system. RT, even for identical scenes, can be represented as a Gaussian distribution centered on some most likely mean. If we can show that the mean M for target-to-distractor ratio of 1:1 is 20 ms longer than the mean M for a target-to-distractor ratio of 2:0, then the data can be modeled.

In this simulation, P remained at 1.5 and it was assumed that the scenes were high-contrast, so all spatially allowable groupings were permitted ($C = 0\%$). The time constant R clearly had to be changed because our previous choice of R (470 ms) exceeded the search times for these scenes. A possible justification for reducing R from the value used to model the Wolfe et al. (1989) and the Cohen and Ivry (1991) data is that the Mordkoff et al. (1990) scene size was relatively small and the items were relatively large, thereby facilitating activation of the retinotopic feature arrays. The Gaussian distributions for the functions offset by R are given in Figure 24. They were integrated to derive the cumulative distribution functions used to generate the curves in Figure 25.

Note that the same effect could be the result of shorter segmentation time in double-target scenes. Although this possibility cannot be ruled out, variable recognition time as a function of evidence seems more likely at this time. These data do not serve to further constrain S or M or even their sum. To fix ideas, one permissible allotment of parameter values, which will be further motivated in the next section, is, as follows: $R = 300$ ms for this scene, $S = 40$ ms for all the scenes, $M(1{:}1) = 40$ ms, $M(2{:}0) = 20$ ms, and $M = 30$ ms for low target-to-distractor ratio segmentations that are recursively investigated.

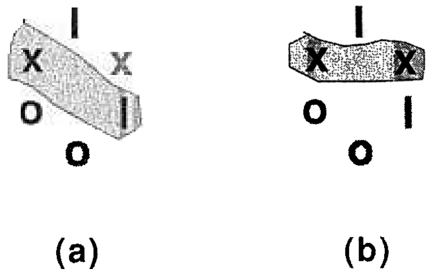


Figure 23. (a) A single-target scene yields a multi-item candidate region including one target and one distractor. This region includes some featural evidence for target presence and some contradictory evidence. (b) A double-target scene yields a multi-target candidate region in which evidence for target presence is doubly reinforced.

Section 11. Simulation of Feature Additivity Effects

Treisman and Sato (1990) found a strong correlation between ease of single-feature segregation and conjunctive search, thereby suggesting a role for multi-item grouping in search. However, they treated grouping as a special case or strategy and did not specify how segregation might be quantified and integrated into their theory. Instead, they focused on the feature inhibition hypothesis because of a piece of evidence that they took to argue against multi-item grouping and for simultaneous processing of multiple features.

This evidence is referred to as the *additivity of features effect.* A comparison of conjunctive search slopes indicates that each feature makes additive and independent contributions to slopes. As a result, the difference between a color–orientation target

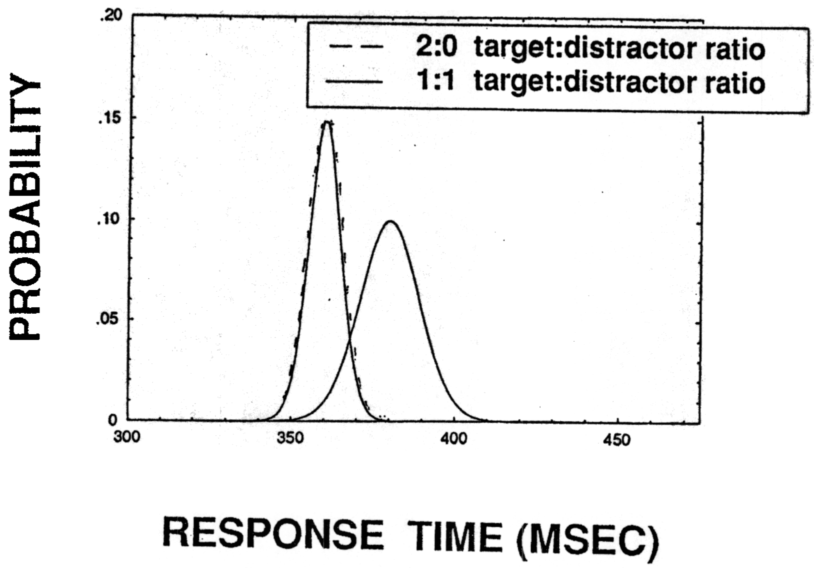


Figure 24. Gaussian distributions for match time M with one target and one distractor in the grouping are greater than those with two targets in the grouping.

search slope and a motion–orientation slope is the same as the difference between a color–size slope and a motion–size slope. This result only argues against multi-item grouping, however, if it is assumed that each candidate grouping is evaluated on the basis of a single feature. If search proceeded by separation of a group of items from the scene that shares one feature, and then separation of an item from that multi-item group is based on the other target feature, additivity would be expected.

The additivity of feature effects observed by Treisman and Sato (1990) can be simulated by the SOS model if parameter S is made into a function of feature dimension. In fact, their data constrain the relative values of the sum ($S + M$) for various combinations of feature dimensions. It is not clear from the data whether S or M is a function of the feature dimension. To fit the data, either S or M could have been made a function of feature dimension. If M were made a function of feature dimen-

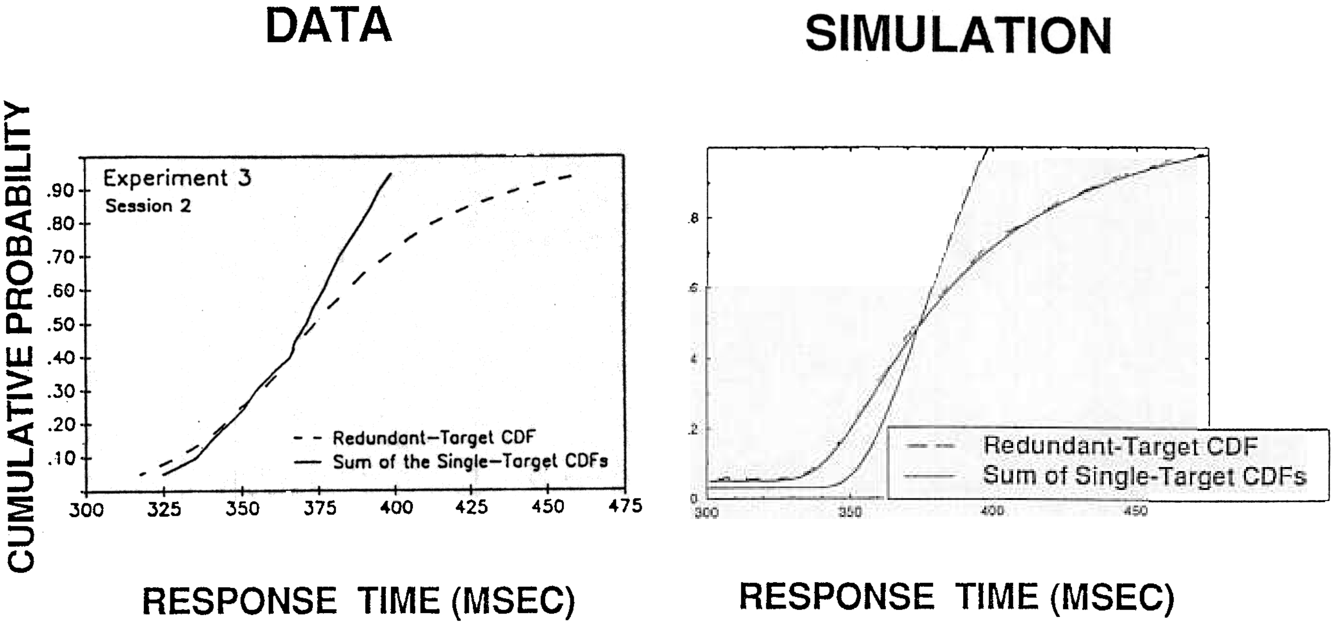


Figure 25. The probability functions shown in Figure 24 can be integrated to fit the Mordkoff, Yantis, and Egeth (1990) data. (a) From "Detecting Conjunctions of Color and Form in Parallel" by J. T. Mordkoff, S. Yantis, and H. E. Egeth, 1990, *Perception and Psychophysics, 5,* p. 165. Copyright 1990 by the Psychonomic Society, Inc. Adapted by permission. (b) Model simulation. CDF = Cumulative distribution function.

Table 3
Comparison of Experimental Data of Treisman and Sato (1990) on the Additivity of Features Effect and the Model Simulation

	Color		Size		Motion	
Feature	Actual	Predicted	Actual	Predicted	Actual	Predicted
Size	14.4	14.1	—	—	—	—
Motion	21.7	22.5	21.4	20.6	—	—
Orientation	29.5	29.0	27.5	28.0	36.4	36.5

Note. Values represent reaction time divided by number of items, expressed in milliseconds.

sion, however, changes in the parameter choices for the Mordkoff et al. (1990) simulation would have to be made. Our simulations accordingly used the following *S*(feature) values. The average of *S*(color) and *S*(orientation) was chosen to be 38 ms to make this simulation consistent with the previous simulations. In addition, *S*(color) = 10 ms, *S*(orientation) = 66 ms, *S*(motion) = 40 ms, and *S*(size) = 6 ms.

Parameter R was not relevant to this experiment because only slopes are compared. The previously defined values of match time $M = 30$ ms were used. The averages of values N (the number of candidate regions) for the Treisman and Sato (1990) experiment were computed as a function of the number of items and were generated for scenes with a scene size of 11.0° square, an item size of .8° square, and an intermediate level of feature saliency ($C = 50\%$) and were as follows: $N = 1$, for 4 items; $N = 2.2$, for 9 items; and $N = 4.5$, for 16 items. As in Tables 1 and 2, 50 trials were averaged for each number of items. The scene, item size, and saliency settings are compatible with Experiment 2 in Treisman and Sato (1990).

Table 3 compares the mean search rates (the average of 2 times the target-present slope and the target-absent slope) for conjunctive targets obtained by Treisman and Sato (1990) and those predicted by the model. As shown, the model is an excellent predictor of the additivity of feature effects. This is a significant result considering that Treisman and Sato argued against what they called the "segregation hypothesis" on the basis of the additivity of features effect. Because of the potential for recursive search in different dimensions in our theory, reaction times can show the effects of more than one dimension despite initial grouping on a single dimension.

The additivity of features effects can thus be explained by a model that operates by grouping multiple items that are distinguished by a single target feature, such as color or, more generally, a single multiplexed feature combination, such as color–depth, on an FCS surface representation. The data that Treisman and Sato (1990) were forced to treat as special cases can thereby be explained in a unified way. The model in a similar manner handles the data of Egeth et al. (1984), who found that search times for conjunctive targets wherein three items share target color are the same regardless of the number of distractors, as well as the Triesman and Sato (1990) finding that ease of segregation by a single feature dimension and search times were well correlated.

Section 12. Triple Conjunctive Search

Wolfe et al. (1989) reported search data for triple conjunctive targets that shared only one feature with each distractor. The slopes for size–color–orientation, triple conjunctive searches were 20% of the slopes for color–orientation, simple conjunctive searches. These data on the relative ease of triple conjunctive searches may, at first glance, seem to challenge the hypothesis that search is organized by item groupings that are distinguished by a single target feature. However, even in well-controlled experiments, such a search would predict faster triple conjunctive search for several reasons. First, the use of three versus two feature dimensions increases the likelihood that items can be grouped on the basis of a single feature dimension. Provided that any feature dimension could be used for initial grouping and that the spatial distribution of features is not controlled, the odds that spatially grouping of certain features occurs is higher in triple conjunctive scenes than in conjunctive scenes. Second, the requirement that the target shares only one feature with any distractor means that in any display, at least one target feature must be present in less than half of the items. Fewer items sharing one of the target features tends to decrease the number of groupings in that feature dimension.

These factors enable our model to account for the Wolfe et al. (1989) data. As Table 3 shows, size–color search slopes are less than half as steep as color–orientation search slopes. Because the triple conjunctive targets share only one feature, they could be found by simple size–color searches. Moreover, as discussed, the one shared feature restriction means that in any scene with an equitable feature distribution, each target feature will be present in only about one-third of the items. For example, in a 16-item, target-present scene, the most equitable distribution of target features among distractors yields a scene with 6 items sharing each target feature. Thus, on average, each target feature would be present in only two-thirds as many items as in a simple conjunctive search. Therefore, simply accounting for differences in search difficulty and the fact that the model need, at most, investigate all the items or item groupings defined by a single feature, the triple conjunctive slopes would be expected to be one-third as steep as the simple conjunctive slopes.

In addition, the visual system may not handle size independently of other features. Considerable evidence suggests that the visual cortex forms multiple boundary segmentations, each corresponding to a different range of relative depths from the observer (Grossberg, 1994). These segmentations realize a size-disparity correlation that tends to, at least partially, enable different segmentations to preferentially process different sizes. This boundary selectivity is passed along to multiple surface representations, whose filling-in of surface brightness, color, depth, and form is organized by the corresponding boundary segmentation. For this reason, items of different size may be automatically separated into different boundary segmentations and thus onto different surface representations, each of which implicitly represent a size–color conjunction. This type of separation, which is tied to data about the neural representations of visual percepts, provides an alternative to the unlikely possibility that a separate feature map encodes size, and indeed every searchable feature, independently of all other features.

Triple conjunctive search data are perhaps the strongest evidence for item selection based on the combined evidence from multiple featural dimensions. We hypothesize that such dimensions are the ones that are multiplexed on separate boundary and surface representations. In particular, spatially coincident discontinuities in multiple featural dimensions could yield more salient boundary segmentations. In this case, shorter segmentation and selection times for targets distinguished by two features would be expected. In fact, additive segmentation effects are indicated by recent work done by Rivest and Cavanagh (1991). They found that spatially coincident boundaries in multiple featural dimensions cooperated to define stronger and sharper contours than those defined by a single featural dimension. In a triple conjunctive search, two features may combine to speed object segmentation because these featural dimensions represent the same boundary or surface. In natural scenes in which environmental objects are typically distinguished from their backgrounds on the basis of several features, this facilitation effect would be advantageous without threatening the veridical perception of the objects.

Section 13. Conclusion

Much of the psychophysical data about search can be explained and quantitatively simulated by using a neural theory that suggests how spatial maps in the parietal cortex and object recognition categories in the inferotemporal cortex interact attentively with other brain regions to search representations in the visual cortex. These visual representations are parsed into multiple boundary segmentations and surface filling-in domains that enable the organization of groupings that may contain multiple items.

Previously, neural network models have been developed to analyze how the functional organization of each of these brain regions may give rise to emergent properties that match properties of behavioral data. These include models of 3-D boundary segmentation and surface representation by striate and extrastriate visual cortex (Grossberg, 1987b, 1994; Grossberg & Mingolla, 1985, 1987, 1993; Grossberg & Rudd, 1992; Grossberg & Todorović, 1988); models of attentive category learning, priming, search, and recognition by inferotemporal cortex, hippocampal formation, and pulvinar (Carpenter & Grossberg, 1993; Grossberg, 1987c); models of attentive spatial localization and orientation by the parietal cortex and its projections (Carpenter et al., 1992; Greve, Grossberg, Guenther, & Bullock, 1993; Grossberg, Guenther, Bullock, & Greve, 1993; Grossberg & Kuperstein, 1989); and models of temporal storage and recall of sequences of events, such as targets, in working memory by the frontal cortex (Boardman & Bullock, 1991; Bradski, Carpenter, & Grossberg, 1992). Unlike artificial neural network models such as back propagation, these biological neural networks incorporate neurobiologically plausible mechanisms and are capable of acting autonomously in real time.

Although it is not yet possible to synthesize all of these neural models into a unified neural architecture that is capable of searching a scene, we have here presented a search algorithm that incorporates many properties of these neural models. By using a fixed set of parameters, we have used this algorithm to quantitatively simulate many fundamental search data. The algorithmic SOS model also makes testable predictions that are currently under empirical investigation. Thus, whereas all the neural mechanisms of these cortical regions are not yet known, our model provides new constraints on how they need to work together to explain properties of visual search.

References

Aks, D. J., & Enns, J. T. (1992). Visual search for direction of shading is influenced by apparent depth. *Perception and Psychophysics, 52,* 63–74.

Anderson, R. A., Essick, G. K., & Siegel, R. M. (1985). Enclosing of spatial location by posterior parietal neurons. *Science, 230,* 456–458.

Arend, L. (1991). Surface colors, illumination and surface geometry: Intrinsic image models of human color perception. Submitted for publication.

Boardman, I., & Bullock, D. (1991). A neural network model of serial order recall from short-term memory. In *Proceedings of the International Joint Conference on Neural Networks* (Vol. 2, pp. 879–884). Piscataway, NJ: IEEE Service Center.

Bradski, G., Carpenter, G. A., & Grossberg, S. (1992). Working memory networks for learning temporal order with application to three-dimensional visual object recognition. *Neural Computation, 4,* 270–286.

Bravo, M., & Blake, R. (1990). Preattentive vision and perceptual groups. *Perception, 19,* 515–522.

Bundesen, C. (1990). A theory of visual attention. *Psychological Review, 97,* 523–547.

Carpenter, G. A., & Grossberg, S. (1991). *Pattern recognition by self-organizing neural networks.* Cambridge, MA: MIT Press.

Carpenter, G. A., & Grossberg, S. (1993). Normal and amnesic learning, recognition, and memory by a neural model of cortico-hippocampal interactions. *Trends in Neurosciences, 16,* 131–137.

Carpenter, G. A., Grossberg, S., & Lesher, G. W. (1992). A what-and-where neural network for invariant image preprocessing. In *Proceedings of the International Joint Conference on Neural Networks* (Vol. 3, pp. 303–308). Piscataway, NJ: IEEE Service Center.

Carpenter, G. A., Grossberg, S., & Lesher, G. W. (1993). *The what-and-where filter: A neural network for spatial mapping, object recognition, and image understanding.* Submitted for publication.

Chelazzi, L., Miller, E. K., Duncan, J., & Desimone, R. (1993). A neural basis for visual search in inferior temporal cortex. *Nature, 363,* 345–347.

Cohen, A., & Ivry, R. B. (1991). Density effects in conjunctive search: Evidence for a coarse location mechanism of feature integration. *Journal of Experimental Psychology: Human Perception and Performance, 17,* 891–901.

Desimone, R. (1991). Face-selective cells in the temporal cortex of monkeys. *Journal of Cognitive Neuroscience, 3,* 1–8.

Desimone, R. (1992). Neural circuits for visual attention in the primate brain. In G. A. Carpenter and S. Grossberg (Eds.), *Neural networks for vision and image processing* (pp. 343–364). Cambridge, MA: MIT Press.

Desimone, R., Schein, S. J., Moran, J., & Ungerleider, L. G. (1985). Contour, color, and shape analysis beyond the striate cortex. *Vision Research, 25,* 441–452.

Desimone, R., & Ungerleider, L. G. (1989). Neural mechanisms of visual processing in monkeys. In F. Boller & J. Grafman (Eds.), *Handbook of neuropsychology* (Vol. 2, pp. 267–299). Amsterdam: Elsevier.

DeYoe, E. A., & van Essen, D. (1988). Concurrent processing streams in monkey visual cortex. *Trends in Neuroscience, 11,* 219–226.

Duncan, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology: General, 113,* 501–517.

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

Egeth, H., 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.

Fischer, B., & Breitmeyer, B. (1987). Mechanisms of visual attention revealed by saccadic eye movements. *Neuropsychologia, 25,* 73–83.

Gochin, P. M. (1990, January 1). Pattern recognition in primate temporal cortex: But is it ART? In *International Joint Conference on Neural Networks* (Vol. 1, pp. 77–80). Hillsdale, NJ: Erlbaum.

Gochin, P. M., Miller, E. K., Gross, C. G., & Gerstein, G. L. (1991). Functional interactions among neurons in inferior temporal cortex of the awake macaque. *Experimental Brain Research, 84,* 505–516.

Goodale, M. A., & Milner, D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences, 15,* 20–25.

Greve, D., Grossberg, S., Guenther, F. H., & Bullock, D. (1993). Neural representations for sensory-motor control: I. Head-centered 3-D target positions from opponent eye commands. *Acta Psychologica, 82,* 115–138.

Grossberg, S. (1978). A theory of human memory: Self-organization and performance of sensory-motor codes, maps, and plans. In R. Rosen & F. Snell (Eds.), *Progress in theoretical biology* (pp. 233–374). San Diego, CA: Academic Press. (Reprinted in S. Grossberg [Ed.], 1982, *Studies of mind and brain,* pp. 500–639. Dordrecht, The Netherlands: Kluwer Academic.)

Grossberg, S. (1980). How does a brain build a cognitive code? *Psychological Review, 87,* 1–51.

Grossberg, S. (1987a). Cortical dynamics of three-dimensional form, color, and brightness perception: I. Monocular theory. *Perception and Psychophysics, 41,* 87–116.

Grossberg, S. (1987b). Cortical dynamics of three-dimensional form, color, and brightness perception: II. Binocular theory. *Perception and Psychophysics, 41,* 117–158.

Grossberg, S. (1987c). *The adaptive brain* (Vol. 1 & 2). Amsterdam: Elsevier/North-Holland.

Grossberg, S. (1992). 3-D vision and figure–ground separation by visual cortex. *Society for Neuroscience Abstracts, 18,* No. 313.4.

Grossberg, S. (1994). 3-D vision and figure–ground separation by visual cortex. *Perception and Psychophysics, 55,* 48–120.

Grossberg, S., Guenther, F. H., Bullock, D., & Greve, D. (1993). Neural representations for sensory-motor control: II. Learning a head-centered visuomotor representation of 3-D target position. *Neural Networks, 6,* 43–67.

Grossberg, S., & Kuperstein, M. (1986). *Neural dynamics of adaptive sensory-motor control: Ballistic eye movements.* Amsterdam: Elsevier/North-Holland.

Grossberg, S., & Kuperstein, M. (1989). *Neural dynamics of sensory-motor control* (expanded ed.). Elmsford, NY: Pergamon Press.

Grossberg, S., & Mingolla, E. (1985). Neural dynamics of perceptual grouping: Textures, boundaries, and emergent segmentations. *Perception and Psychophysics, 38,* 141–171.

Grossberg, S., & Mingolla, E. (1987). Neural dynamics of surface perception: Boundary webs, illuminants, and shape-from-shading. *Computer Vision, Graphics, and Image Processing, 37,* 116–165.

Grossberg, S., & Mingolla, E. (1993). Neural dynamics of motion perception: Direction fields, apertures, and resonant grouping. *Perception and Psychophysics, 53,* 243–278.

Grossberg, S., Mingolla, E., & Todorović, D. (1989). A neural network architecture for preattentive vision. *IEEE Transactions on Biomedical Engineering, 36,* 65–84.

Grossberg, S., & Rudd, M. (1989). A neural architecture for visual motion perception: Group and element apparent motion. *Neural Networks, 2,* 421–450.

Grossberg, S., & Rudd, M. (1992). Cortical dynamics of visual motion perception: Short-range and long-range apparent motion. *Psychological Review, 99,* 78–121.

Grossberg, S., & Stone, G. (1986). Neural dynamics of word recognition and recall: Attentional priming, learning, and resonance. *Psychological Review, 93,* 46–74.

Grossberg, S., & Todorović, D. (1988). Neural dynamics of 1-D and 2-D brightness perception: A unified model of classical and recent phenomena. *Perception and Psychophysics, 43,* 241–277.

Harries, M. H., & Perrett, D. I. (1991). Visual processing of faces in temporal cortex: Physiological evidence for a modular organization and possible anatomical correlates. *Journal of Cognitive Neuroscience, 3,* 9–24.

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

Humphreys, G. W., Quinlan, P. T., & Riddoch, M. J. (1989). Grouping processes in visual search: Effects with single- and combined-feature targets. *Journal of Experimental Psychology: General, 118,* 258–279.

Julesz, B. (1984). A brief outline of the text on theory of human vision. *Trends in Neurosciences, 7,* 41–45.

Kanizsa, G., & Luccio, R. (1987). Formation and categorization of visual objects: Hoffding's never confuted but always forgotten argument. *Gestalt Theory, 9,* 111–127.

Kleffner, D. A., & Ramachandran, V. S. (1992). On the perception of shape from shading. *Perception and Psychophysics, 52,* 18–36.

Maylor, E. A., & Hockey, R. (1985). Inhibitory components of externally controlled covert orienting in visual space. *Journal of Experimental Psychology: Human Perception and Performance, 11,* 777–787.

Miller, E. K., Li, L., & Desimone, R. (1991). A neural mechanism for working and recognition memory in inferior temporal cortex. *Science, 254,* 1377–1379.

Mishkin, M., & Appenzeller, T. (1987). The anatomy of memory. *Scientific American, 256,* 80–89.

Mishkin, M., Ungerleider, L. G., & Macko, K. A. (1983). Object vision and spatial vision: Two cortical pathways. *Trends in Neurosciences, 6,* 414–417.

Mordkoff, J. T., & Yantis, S. (1993). Dividing attention between color and shape: Evidence of coactivation. *Perception and Psychophysics, 53,* 357–366.

Mordkoff, J. T., Yantis, S., & Egeth, H. E. (1990). Detecting conjunctions of color and form in parallel. *Perception and Psychophysics, 48,* 157–168.

Mountcastle, V. B., Anderson, R. A., & Motter, B. C. (1981). The influence of attentive fixation upon the excitability of the light-sensitive neurons of the posterior parietal cortex. *Journal of Neuroscience, 1,* 1218–1235.

Nakayama, K., & 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.

Pashler, H. (1987). Detecting conjunctions of color and form: Reassessing the serial search hypothesis. *Perception and Psychophysics, 41,* 191–201.

Perrett, D. I., Mishkin, A. J., & Chitty, A. J. (1987). Visual neurons responsive to faces. *Trends in Neurosciences, 10,* 358–364.

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

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

Rivest, J., & Cavanagh, P. (1991). Cross media cooperation in localization of contours. *Investigative Opthalmology and Visual Science, 32,* 1024.

Rizzolati, G., Riggio, L., Dascola, I., & Umita, C. (1987). Reorienting attention across the horizontal and vertical meridians. *Neuropsychologia, 25,* 31–40.

Schneider, W., & Shiffrin, R. M. (1976). Automatic and controlled information processing in vision. In D. LaBarge & S. J. Samuels (Eds.), *Basic processes in reading: Perception and comprehension.* Hillsdale, NJ: Erlbaum.

Spitzer, H., Desimone, R., & Moran, J. (1988). Increased attention enhances both behavioral and neuronal performance. *Science, 240,* 338–340.

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

Treisman, A. (1985). Preattentive processing in vision. *Computer Vision, Graphics, and Image Processing, 31,* 156–177.

Treisman, A., & Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology, 16,* 97–136.

Treisman, A., & Gormican, S. (1988). Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review, 95,* 15–48.

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

Tsotsos, J. K. (1990). Analyzing vision at the complexity level. *Brain and Behavioral Sciences, 13,* 423–469.

Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems: Separation of appearance and location of objects. In D. L. Ingle, M. A. Goodale, & R. J. W. Mansfield (Eds.), *Analysis of visual behavior* (pp. 549–586). Cambridge, MA: MIT Press.

Wolfe, J. M. (1992). "Effortless" texture segmentation and "parallel" visual search are not the same thing [Research Note]. *Vision Research, 32,* 757–763.

Wolfe, J. M. (in press). Guided Search 2.0: A revised model of visual search. *Psychonomic Bulletin and Review.*

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

Wolfe, J. M., & Friedman-Hill, S. R. (1992). Part–whole relationships in visual search. *Investigative Opthalmology and Visual Science, 33,* 1355.

Wurtz, R. H., Goldberg, M. E., & Robinson, D. L. (1982, June). Brain mechanisms of visual attention. *Scientific American,* 124–135.

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

Received May 24, 1993
Revision received November 8, 1993
Accepted November 8, 1993 ■