The Parallel Guidance of Visual Attention

Jeremy M. Wolfe

In the fifth-act battle scene of Shakespeare's *Henry IV, Part 1,* rebel nobles are searching the battlefield at Shrewsbury for King Henry. The task is made more difficult by Henry's stratagem of disguising some of his soldiers in the royal colors. "The king," says one rebel, "hath many marching in his coats." Douglas, another rebel, replies:

> Now, by my sword, I will kill all his coats;
> I'll murder all his wardrobe, piece by piece,
> until I meet the king.

Shakespeare offers a good, if sanguinary, account of a real-world visual search task: a search for one target item in a field of distractors. Because of processing limitations, candidate targets must be processed in a serial, self-terminating manner. However, not all items in the field are reasonable candidate targets. The serial processing of unfortunate royal surrogates is made more efficient by processes that guide search toward likely kings and away from trees, rocks, sheep, and those soldiers fortunate enough not to be wearing the royal disguise.

Research with far less exciting visual search tasks in the laboratory suggests that this *guided search* is the normal mechanism used by individuals to find a desired target in a crowded visual world. Some basic attributes of the visual scene (e.g., colors, motion, orientation) can be extracted *in parallel* across the entire visual field or, at least, over a substantial piece of the visual field. Thus, in Figure 1a, all items can be processed simultaneously in a search for a white letter. Other processes (e.g., letter or object recognition) can operate over only a limited region of the visual field at any one time. These spatially limited resources must be deployed first at one location, then at another, in a *serial* manner.[1] Under some circumstances, all items are candidate targets, and that deployment is random. For example, in Figure 1b, a "T" is easily distinguished from an "L," but search for a "T" proceeds at random from item to item until the correct letter is found. In most cases, however, information from the parallel processors—the front end of the system—can be used to *guide* the deployment of spatially limited resources. In Figure 1c, only half of the items are candidate targets because a parallel processor for color can guide search to the white items in a search for a white "T," eliminating the black items.

As another example, consider a search for a red vertical target among green vertical and red horizontal distractors. This task involves a conjunction of color and orientation. Evidence suggests that we do not have parallel processors for conjunctions of features. However, this task can be accomplished efficiently by having a parallel "color" processor activate all "red" locations while an "orientation" processor activates all "vertical" locations. If these activations are summed, the greatest activation will be at the location of a red vertical item if it is present. A spatially limited process can be deployed to check that location for the presence of a red vertical item. Even if this process of activation is noisy, it will reduce the set of candidate targets from the set of all items to a subset. This is the essential idea behind the guided search model.[2]

In this review, the focus is on the properties of the parallel, *preattentive,* stage of guided search. What information is available for guiding the deployment of spatially limited processes? How is this preattentive representation of visual information related to other representations, such as those found in electrophysiological studies of visual cortex or those involved in texture segmentation?

Jeremy Wolfe is Associate Professor of Ophthalmology at Harvard Medical School and is affiliated with Brigham and Women's Hospital. He is a visiting Associate Professor at MIT and at Wellesley College. Wolfe received his undergraduate degree at Princeton University and his PhD at MIT. Address correspondence to Jeremy Wolfe, Center for Clinical Cataract Research, 221 Longwood Ave., Boston, MA 02115; e-mail: wolfe@psyche.mit.edu.

WHAT IS DOING THE GUIDING? THE NATURE OF THE PREATTENTIVE REPRESENTATION[3]

Parallel processing appears to be limited to a relatively small set of basic features. A partial list includes color, orientation, size, motion, stereoscopic depth, curvature, a variety of two-dimensional pictorial cues to three-dimensional space, and a variety of surface properties like luster (shininess).[4] Form is probably not a single basic feature. Rather, it seems to be composed of a small set of features, though the precise set is unknown and the featural status of form properties such as line termination and closure is still open to debate. Still, there is little debate that the list of preattentive, or parallel-processed, features is a limited list, and there is little evidence in the visual search literature that practice increases the number of such features.

Preattentive Features Are Not Early-Vision (V1) Features

About 10 years ago, it was tempting to think that the basic features

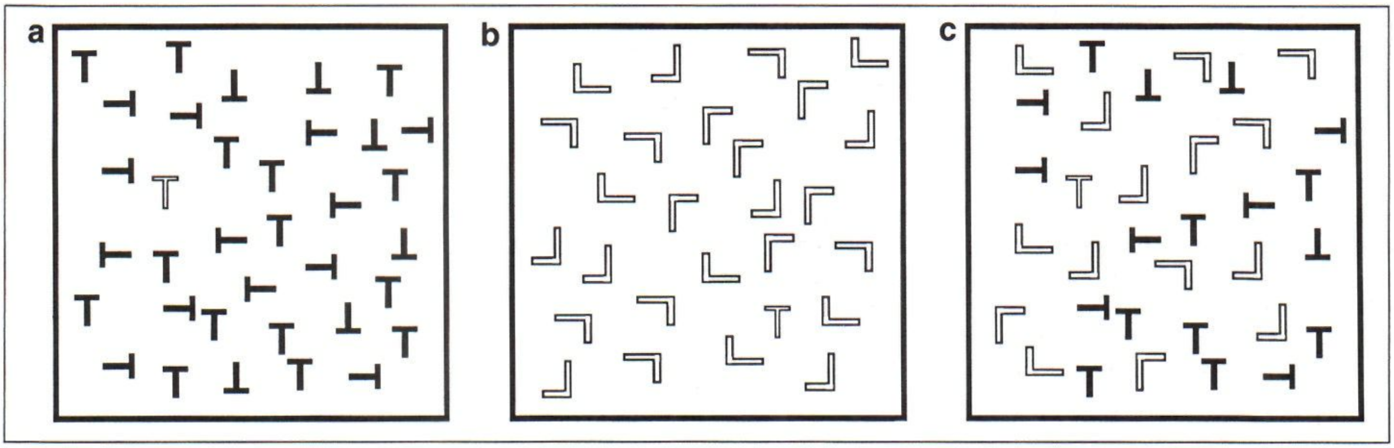

Fig. 1. The three broad classes of visual search tasks. In each case, the task is to find the white "T." (a) In parallel feature search, the time to find a target defined by a single feature does not increase as the number of distractors increases. (b) In serial search, items must be examined sequentially, and each additional distractor increases the search time. (c) In guided search, parallel processes restrict subsequent serial search to a subset of the items (here, the white items).

supporting parallel visual search[5] were the same as the basic features extracted in the first steps of visual cortical processing (e.g., color, size, orientation). Though the idea was attractive, there are at least two reasons why it is probably not correct: The lists of features are not identical, and features that do appear on both lists may have different properties.

The Two Lists Are Not the Same

Work in the past decade has revealed features that produce parallel visual search results but that have not been found to be tuning properties of single cells in the primary visual cortex (V1) of primates. Examples include perceived shape from shading cues, linear perspective cues to depth, and binocular luster (an appearance of shininess produced by unequal luminance in the two eyes).[6] In addition, some properties of V1 neurons (e.g., the ability to determine which eye has been stimulated) are not features in visual search,[7] suggesting that some V1 features are lost before the generation of the preattentive representation.

Cortical Features and Preattentive Features Have Different Properties

Even when an attribute like "color" or "orientation" can be studied with both cortical electrophysiology and the parallel visual search paradigm, it seems clear that often the same attribute name is being used for two quite different things. An example is illustrated in Figure 2a. There are four target items in the figure, each tilted 20° left of vertical. The distractors are tilted 20° to the right, 60° right, and 80° left. The task is very difficult even though the minimum difference in orientation between targets and distractors is 40°. Standard estimates of the orientation tuning[8] of cortical neurons are around 15° to 20°, so target and distractor lines must be stimulating very different populations of cortical neurons. Psychophysical orientation discriminations based on the outputs of these cells are on the order of 1°. Yet even on a homogeneous background of distractor items of one orientation, a target line cannot be found in parallel if the difference between target and distractor orientation is less than about 15°.[9]

Parallel processing of orientation in visual search is not merely cruder than cortical orientation processing.

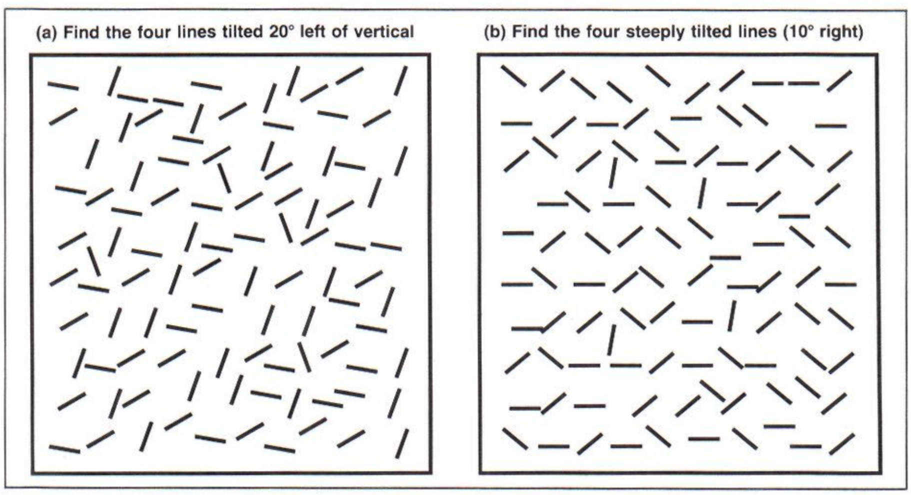

Fig. 2. Visual search for a specific orientation. (a) Search for lines tilted 20° left of vertical is difficult among heterogeneous distractors even though no distractor orientation is closer than 40° to the target orientation. (b) With the same angular relations between stimuli (all lines rotated 30° clockwise), the search becomes simple because the targets are *categorically* different from all distractors. Here, the targets (tilted 10° to the right) are the only "steep" items.

It is qualitatively different. This can be seen in Figure 2b: All the angular relationships between targets and distractors are the same as in Figure 2a, but all orientations have been rotated 30° clockwise. The target items are now oriented 10° to the right, and the distractors are now oriented 50° to the right, 50° to the left, and 90° (horizontal). More important, this rotation makes the targets the only "steep" lines in the field. Now they can be found easily. In a series of visual search experiments, my colleagues and I have shown that the parallel representation of orientation used in visual search is *categorical*. That is, when there are several orientations in a display, search is efficient only when the target is the sole member of an orientation category (e.g., steep, shallow, left, or right).[10]

A similar point could be made concerning V1 and preattentive processing of color information.[11] Also, visual cortex seems to have four to eight channels for size, but we find preattentive size processing to be limited to the categories "biggest" and "smallest."

The Preattentive Representation of Visual Information Is Sophisticated and Task-Specific

In the preceding section, I have argued that the early-vision features studied with cortical electrophysiology in V1 are different from the preattentive features studied with visual search. How are these two sets of features related? It seems likely that there is a hierarchical relationship between early cortical features and the preattentive features that can guide visual search. The latter are the product of further parallel processing of the former.[12] Some of that further processing is task-specific, and other visual tasks (e.g., texture segmentation) make different use of the output of early cortical processes. As evidence in support of these conclusions, it can be shown that the preattentive representation "knows" a lot, suggesting that it represents a relatively late stage in visual processing. At the same time, there is a lot that the representation does not "know," suggesting that it is using only a task-specific subset of the information available from earlier stages in processing.

The Preattentive Representation "Knows" a Lot

Studies of part–whole relations in guided search provide evidence of the sophistication of the processes that generate preattentive representations. Although it is possible to guide attention to conjunctions of two features (e.g., a red vertical item), guidance fails when the target is a conjunction of two examples of the same feature type. Thus, in Figure 3a, it is hard to find the item that is white and black among white-and-gray and gray-and-black distractors.[13] The same very inefficient search is found for orientation X orientation and size X size conjunctions. Recently, however, we have found a class of stimuli that produce efficient, guided searches for size X size and color X color conjunctions. An example is illustrated in Figure 3b, though the effect is more compelling in color. Here, the search is for a white house with black windows. The distractors are chosen so as to make this a color X color conjunction. This search and others like it suggest that within-feature searches are efficient when one color describes the *whole* object (the white house) while the other describes a *part* of the object (the black windows).

Part–whole color X color conjunctions allow for guided search. Part–part conjunctions (e.g., the house that is half red and half yellow) do not.[14] These results suggest that there is separate parallel processing for the colors of whole objects and the colors of parts, analogous to the separate processing of the color and orientation of items. Moreover, these results suggest that parallel, preattentive processing is capable of dividing the visual scene into items or objects with constituent parts—an indication of a substantial amount of prior processing in parallel.

Part–whole size X size conjunctions also yield efficient, guided

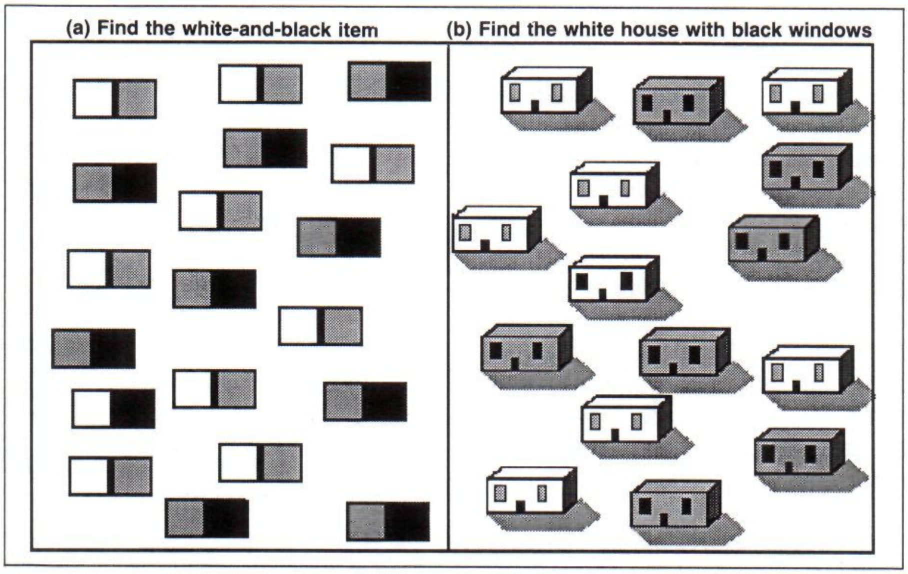

Fig. 3. Search for a conjunction of two colors. Search is very inefficient when the conjunction is between the colors of two parts of a target (a). However, search is much easier when the conjunction is between the color of the whole item and the color of one of its parts (b).

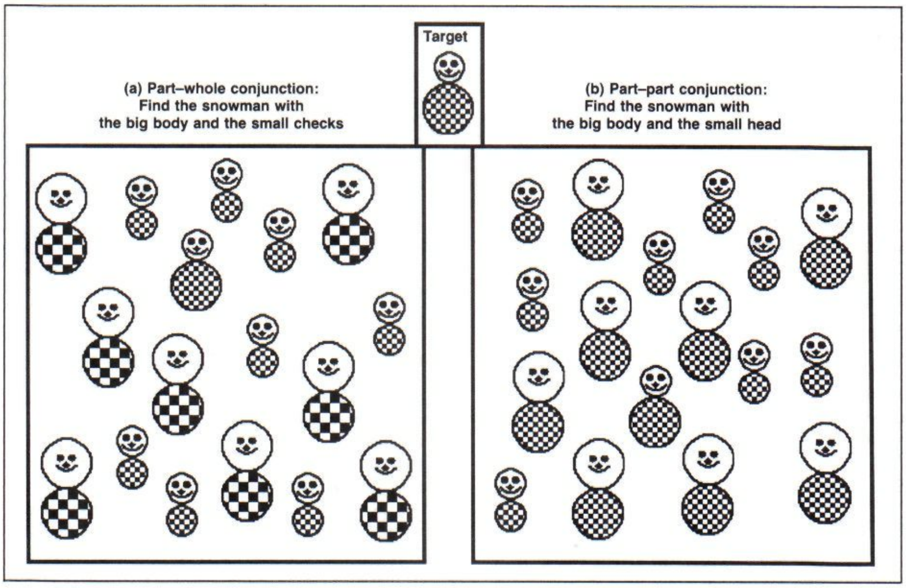

Fig. 4. Search for a conjunction of two sizes. Search is efficient if the conjunction is between the size of the whole item and the size of one of its parts (a). Conjunctions between the sizes of parts are hard to find (b).

search. Thus, in Figure 4a, it is relatively easy to find the snowman with the big body and little checks on his waistcoat among snowmen with big bodies and big checks and snowmen with little bodies and little checks. For contrast, Figure 4b shows a part–part conjunction with similar stimuli. It is hard to find the snowman with a big body and a small head among snowmen with big bodies and big heads and snowmen with small bodies and small heads. Interestingly, we have been unable to find any part–whole stimuli that support guided search for orientation X orientation stimuli. It is as if the color (or size) of a whole item can be represented separately from the color (or size) of part of the item, but the orientations of parts and wholes are not separately represented.[15]

There Is a Lot the Preattentive Representation Does Not "Know"

Although parallel processing of part–whole relationships is evidence that the preattentive representation "knows" a lot, it is also clear that much information is absent from the representation. It cannot differentiate between a "T" and an "L," for example (Fig. 1b). As noted, the representation does not code conjunctions directly, it cannot differentiate between lines that differ widely in orientation or patches that differ widely in color unless those differences cross categorical boundaries. Visual processing is not a single, unitary stream. Many different processing tasks occur at the same time, and visual search is guided by only a subset of the available information. This subset can be thought of as a representation created specifically for the task of guiding search.[16]

Visual Search and Texture Segmentation

The task-specific nature of the preattentive representation can be illustrated by comparison with the representation that supports "effortless" texture segmentation.[17] In Figure 5a, there is little or no evidence for effortless texture segmentation. The bulk of the texture is composed of black horizontal and white vertical lines, but one region contains black verticals and white horizontals. That region does not "pop out,"[18] or segment. However, it is possible to do a successful guided search for a black vertical or white horizontal target among distractors that are black horizontal and white vertical. By contrast, there is clear texture segmentation in Figure 5b. Here, the smaller region is composed of items that are vertical and oblique on a background of vertical-and-horizontal and horizontal-and-oblique items. As a visual search task, this is an orientation X orientation conjunction and is very inefficient. In fact, there is a lone vertical-and-oblique target in the lower right corner of the figure. As a visual search target, it does not pop out, even though a group of these items supports texture segmentation. The orientation X orientation items show texture segmentation because segmentation processes seem to integrate over multiple items, creating an impression of average orientation.[19] Such integration would be a

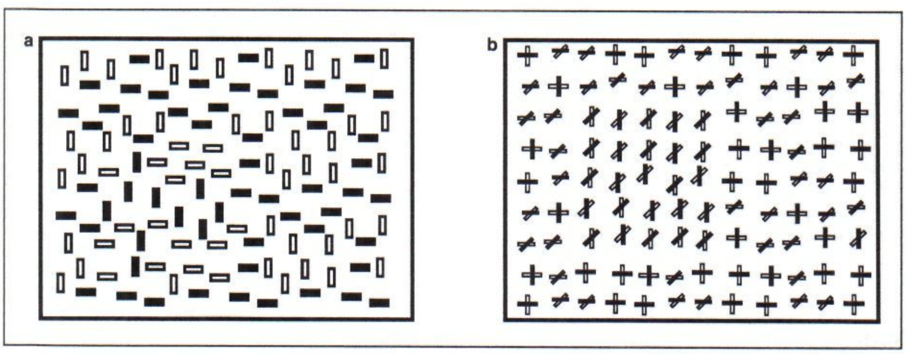

Fig. 5. Visual search versus texture segmentation. Conjunctions of color and orientation (a) support parallel search without supporting texture segmentation, whereas orientation X orientation conjunctions (b) support texture segmentation but not parallel search.

useless operation in visual search for a single item.

SUMMARY

In attempting to execute complex tasks like object recognition, the visual system is faced with significant limits on its processing capacity. Full processing of all locations in parallel is not possible. Some processes may operate on one item or, perhaps, on one limited region at a time. Deployment of the resources required for these tasks is under attentional control. To operate reasonably efficiently, that attention must be guided nonrandomly from item to item or location to location. To support guidance of attention, the visual system extracts information in parallel from visual input to create a representation intended for the specific purpose of guiding attention. That parallel, preattentive representation sacrifices a great deal of fine-grain information about attributes such as specific orientations and colors. It quickly parses input into featural categories, like "steep" and "shallow" in orientation. It is not, however, merely a crude or early representation of visual input. It is sensitive to such comparatively late perceptual properties as pictorial depth cues and part–whole relations. It is a representation well-suited to perform its appointed task—to quickly direct attention to a meaningful subset of items or locations in a visually rich environment.

Acknowledgments—I thank Anne Treisman, Kyle Cave, Judith Friend, Stacia Friedman-Hill, and two anonymous reviewers for their useful comments on drafts of this paper. The research was supported by the National Institutes of Health (Grant No. EY05087).

Notes

1. It may be possible to process clumps of items; e.g., H. Pashler, Detecting conjunctions of color and form: Reassessing the serial search hypothesis, *Perception and Psychophysics, 41,* 191–201 (1987).

2. Authors of reviews for *Current Directions* are asked to keep references to a minimum. The work described here builds on and is related to a large body of work from other labs. Interested readers should look at the bibliographies of the publications cited here for a more complete guide to this literature. Description of our work on guided search for conjunctions and the original formulation of the guided search model are found in J.M. Wolfe, K.R. Cave, and S.L. Franzel, Guided search: An alternative to the feature integration model for visual search, *Journal of Experimental Psychology: Human Perception and Performance, 15,* 419–433 (1989). Interested readers should certainly look at H.E. Egeth, R.A. Virzi, and H. Garbart, Searching for conjunctively defined targets, *Journal of Experimental Psychology: Human Perception and Performance, 10,* 32–39 (1984), as well as Anne Treisman's work. One of her recent articles on this topic is A. Treisman and S. Sato, Conjunction search revisited, *Journal of Experimental Psychology: Human Perception and Performance, 16,* 459–478 (1990).

3. The term *preattentive representation* is used here to refer exclusively to information that is extracted in parallel from the visual input and is available to guide the deployment of visual attention in visual search, hence the term preattentive. The term is used more broadly elsewhere. This topic is addressed at the end of this review.

4. A list, complete at the time it was created, can be found in A. Treisman, Properties, parts, and objects, in *Handbook of Human Perception and Performance,* K.R. Boff, L. Kaufmann, and J.P. Thomas, Eds. (Wiley, New York, 1986). See also J.T. Enns and R.A. Rensink, Preattentive recovery of three-dimensional orientation from line drawings, *Psychological Review, 98,* 335–351 (1991).

5. Common evidence for parallel processing in visual search is a pattern of reaction times (RTs) that are independent of the number of items displayed (set size). This yields RT × set size slopes near 0 ms/item. The classic pattern for a serial, self-terminating search is a slope of 20–30 ms/item on trials when a target is present and twice that, 40–60 ms/item, for target-absent trials. These conclusions must be taken with caution; see J.T. Townsend, Serial and parallel processing: Sometimes they look like Tweedledum and Tweedledee but they can (and should) be distinguished, *Psychological Science, 1,* 46–54 (1990). Guided searches have shallower slopes than serial searches, reflecting their status as serial searches through a subset of items. The smaller the subset, the shallower the slope. Most of the conclusions and assertions in this review are based on data from search experiments of this sort. The reader is directed to the cited publications for experimental details.

6. On shading, see V.S. Ramachandran, Perception of shape from shading, *Nature, 331,* 163–165 (1988). On linear perspective, see Enns and Rensink, note 4. On binocular luster, see J.M. Wolfe and S.L. Franzel, Binocularity and visual search, *Perception and Psychophysics, 43,* 81–93 (1988).

7. An example is eye-of-origin information, readily available in V1 but unavailable in visual search; Wolfe and Franzel, note 6.

8. The orientation tuning of a cell is a measure of the range of orientations to which that cell will respond with at least some fraction (often 50%) of its maximum firing rate. For a discussion, see L.A. Olzak and J.P. Thomas, Seeing spatial patterns, in *Handbook of Perception and Human Performance,* K.R. Boff, L. Kaufmann, and J.P. Thomas, Eds. (Wiley, New York, 1986).

9. D.H. Foster and P.A. Ward, Asymmetries in oriented-line detection indicate two orthogonal filters in early vision, *Proceedings of the Royal Society of London (B), 243,* 75–81 (1991).

10. J.M. Wolfe, S.R. Friedman-Hill, M.I. Stewart, and K.M. O'Connell, The role of categorization in visual search for orientation, *Journal of Experimental Psychology: Human Perception and Performance, 18,* 34–49 (1992). This article gives references to some standard psychophysical data on cortical orientation processing. Further wrinkles on preattentive orientation processing are found in J.M. Wolfe and S.R. Friedman-Hill, On the role of symmetry in visual search, *Psychological Science, 3,* 194–198 (1992).

11. M. D'Zmura, Color in visual search, *Vision Research, 31,* 951–966 (1991); A.L. Nagy and R.R. Sanchez, Critical color differences determined with a visual search task, *Journal of the Optical Society of America A, 7,* 1209–1217 (1990).

12. See, for instance, P. Cavanaugh, M. Arguin, and A. Treisman, Effect of surface medium on visual search for orientation and size features, *Journal of Experimental Psychology: Human Perception and Performance, 16,* 479–492 (1990). This is an elegant article showing how preattentive orientation features can be created from variations in features such as color and texture.

13. The actual experiments with these color X color conjunctions were done with more colorful stimuli (e.g., target: red and green; distractors: red and blue, blue and green); J.M. Wolfe, K.P. Yu, M.I. Stewart, A.D. Shorter, S.R. Friedman-Hill, and K.R. Cave, Limitations on the parallel guidance of visual search: Color X color and orientation X orientation conjunctions, *Journal of Experimental Psychology: Human Perception and Performance, 16,* 879–892 (1990).

14. This is not just a complex conjunction of size and color. In controlled experiments, we have varied the size of items so that the whole red of a small item is no bigger than the yellow part of some other item in the same display. Guided search is still possible.

15. This difference between size and color on one side and orientation on the other is by no means arbitrary. Consider that, in general, the color of a part is not constrained by the color of the whole, whereas the orientation of a part may be *defined* by its relationship to the primary axis whole. See, e.g., D. Marr and H.K. Nishahara, Representation and recognition of the spatial organization of three-dimensional shapes, *Proceedings of the Royal Society of London (B), 200,* 269–294 (1977).

16. This idea is couched in conditional language for a reason. It is probably a mistake to think that the preattentive representation exists as a neurally distinct locus in the brain. In the absence of clear evidence, it seems more plausible to imagine that it is a virtual map of the visual world, generated for the purposes of guiding visual attention but instantiated in neurons that are doing other tasks at the same time. This is analogous to the multiple functions of a retinal ganglion cell that functions as part of a color system and a contour system at the same time.

17. We run into a terminological problem here because *preattentive* has been used to refer to both parallel visual search and texture segmentation; see, e.g., B. Julesz and J.R. Bergen, Textons, the fundamental elements in preattentive vision and perceptions of textures, *Bell System Technical Journal, 62,* 1619–1646 (1983). Here, as discussed in note 3, I use the term to refer exclusively to the representation that is used to guide the deployment of visual attention. The matter is discussed more extensively in J.M. Wolfe, "Effortless" texture segmentation and "parallel" visual search are not two measures of the same thing, *Vision Research, 32,* 757–763 (1992).

18. Treisman sometimes uses pop-out as a technical term in her feature integration theory. In the context of guided search, pop-out is a description of the phenomenological experience that occurs when visual attention is grabbed by a particularly salient stimulus.

19. Seen in Figure 6 of Wolfe, note 17.