# Getting in to Guided Search

Jeremy M Wolfe (1,2)

Todd S Horowitz (1,2)

Evan M Palmer (1,2)

Kristin O Michod (3)

Michael J Van Wert (1)

1. Visual Attention Lab, Brigham and Woman's Hospital, 64 Sidney St., Suite 170 Cambridge, MA 02139

2. Department of Ophthalmology, Harvard Medical School, Boston, MA

3. Department of Psychology, University of Delaware, Newark, DE

Mailing Address
Jeremy Wolfe
Visual Attention Lab
Brigham and Woman's Hospital
64 Sidney St., Suite 170
Cambridge, MA 02139
Wolfe@search.bwh.harvard.edu

The world delivers more input than the visual system can handle. In response, the visual system has developed a variety of ways to limit the amount of information that it processes at once. For example, our foveated retina transmits the most detailed information from only a very small part of the visual field at any one moment. Since we cannot see the entire world at once, we must move our eyes in order to get that level of detail from a different portion of the visual field. That retinal image still contains more information than the system can process. Visual selective attention is a set of mechanisms to further restrict processing to a subset of the input (e.g. a single location or object) at one time. As a consequence, if we want to know if a specific object is present, we will often need to search for it, even if it is easily visible.

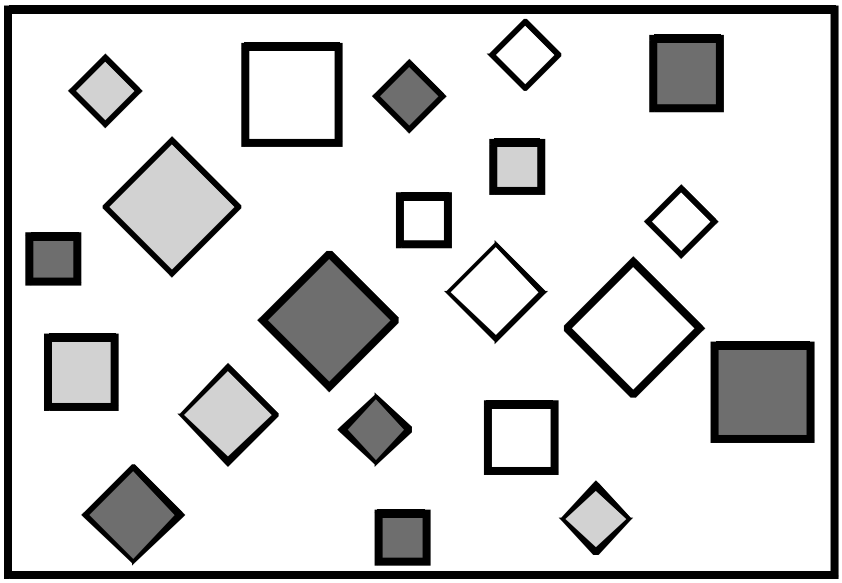


**Figure One: A simple visual search stimulus**

Thus, in Figure One, the medium size, light gray diamond is perfectly visible but not immediately discovered until you search and direct your attention to it.

A vast body of empirical information on visual search has accumulated over the past quarter century (for reviews see Kinchla, 1992; Pashler, 1998; Sanders & Donk, 1996; Wolfe, 1998) and numerous models have been developed to account for portions of the data (e.g. Cave, 1999; Deco, Pollatos, & Zihl, 2002; Grossberg, Mingolla, & Ross, 1994; Hamker, 2006; Herd & O'Reilly, 2005; Hubner, 2001; Parkhurst, Law, & Niebur, 2002; Thornton & Gilden, 2007). Our model, Guided Search, is a model of human visual search that has been in development for almost 20 years ( Wolfe, 1994, 2007; Wolfe, Cave, & Franzel, 1989; Wolfe & Gancarz, 1996). It is currently in its fourth incarnation, Guided Search 4.0 (GS4). GS4 accounts for reaction time (RT) and error data in a range of visual search tasks as well as or better than other models. Still, one might think that, after 20 years, it is time for a model to either explain the data or quietly retire. Of course, as with any active field of science, the 'problem' lies in the data. Some aspects of the data have always proven difficult to model (e.g. target absent trials, Chun & Wolfe, 1996; Cousineau & Shiffrin, 2004; Hong, 2005; Zenger & Fahle, 1997). In other cases, new data provide fresh challenges for a model.

This chapter will describe several lines of research that modify our understanding of the dynamics of guidance in Guided Search. The results are not specific to Guided Search but provide constraints for any model of search.

The "guided" part of Guided Search can be illustrated in Figure One. When you searched for that the medium size, light gray diamond, you might intuitively feel that you did not search at random. You were more likely to attend to items of the correct color, shape, or

size. These basic visual attributes 'guide' the deployment of selective attention (Wolfe & Horowitz, 2004). This chapter is concerned with the temporal dynamics of this guidance. The data reviewed below makes three points about the time course of guidance.

1) It takes several hundred msec for guidance to become effective after the guiding information is available.
2) Curiously, it seems to take a similar amount of time for guidance to become effective each time a new stimulus is presented, even if the guidance is the same as on the previous trial. That is, if you search another version of Figure One for another medium, gray diamond; the guidance settings from the first search would not be immediately available to the second search.
3) More generally, the earliest moments of an extended visual search are different from later moments. The processes of attentional selection evolve over the course of a search.
4) Different aspects of search evolve at different rates. Here we describe the example of guidance to scene-based properties like surface type (e.g. Is there a target on top of a block?). This sort of guidance appears to take much longer to develop on each trial than guidance to simpler properties like color or orientation.

**1. A quick outline of Guided Search 4.0**

Guided Search is an effort to explain why some targets take longer to find in a visual search than others. A red dot among green dots will be found quickly and the number of green dots will not have much impact on the RT. A search for a specific letter among a

variety of different letters will take longer and, even if the letters are big enough to make eye movements unnecessary, each additional distractor letter will add something like 20-30 msec to the average RT. Real world search tasks exaggerate these effects. A fire engine, racing through the parking lot, will be found rapidly independent of the number of cars, while the proverbial needle in a haystack will take much more time, even if the needle is visible, and that time will scale with the size of the haystack.

In her highly influential Feature Integration Theory (FIT; Treisman, 1988; Treisman & Gelade, 1980), Treisman argued that the efficient searches for red dots, fire engines and so forth, were those searches where the target was defined by a single, salient, basic feature like red color or leftward motion. Such searches could proceed in parallel across the whole search display. If target identification required the *binding* together of features as in the letter search, then attention would need to be directed to each item in turn in a serial manner. The model proposed an initial, "preattentive," parallel feature processing stage and a subsequent, "attentive," serial binding stage (c.f. Broadbent, 1958; Hoffman, 1979; Neisser, 1967).

Guided Search borrowed this two-stage architecture and added the core principle that information from the first stage could guide attentional deployments in the second stage. The core example is a conjunction search for a target defined by the conjunction of two features; for example, a small blue square among big blue and small yellow squares. Since identification of a small blue item requires the binding of color and size, it requires attention. Nevertheless, search for such stimuli is quite efficient (Egeth, Virzi, & Garbart,

1984; McLeod, Driver, & Crisp, 1988; Nakayama & Silverman, 1986; Wolfe, 1998; Wolfe, Cave, & Franzel, 1989). GS proposes that the first-stage color processing would guide attention toward blue items while, at the same time, first-stage size processing would guide attention toward small items.

More precisely, attention is deployed to the most active location in an "activation map." Activation is based on a weighted average of bottom-up salience and top-down guidance. A limited set of basic feature dimensions contribute to the activation map (reviewed in Wolfe & Horowitz, 2004). Bottom-up, stimulus-driven salience is based on local differences in a dimension. Thus, a vertical line surrounded by horizontals is salient. The same line surround by lines tilted 10 deg off vertical is not salient (Julesz, 1984; Moraglia, 1989; Nothdurft, 1993). There are numerous sophisticated models of the computation of bottom-up salience (Hamker, 2004; Itti & Koch, 2000; Krummenacher, Muller, & Heller, 2001; Li, 2002; Parkhurst, Law, & Niebur, 2002; Zhaoping & Koene, 2007). They are broadly similar though they differ importantly in, for example, their proposed neural substrates and the degree of modularity of processing of different dimensions (size, orientation, etc). Top-down, user-driven guidance is more complex. We believe it to be based on categorical representations of features (e.g. "steep" and "shallow," not 12 and 75 deg; Wolfe, Friedman-Hill, Stewart, & O'Connell, 1992). GS proposes that a target object would be described in a very limited vocabulary of these categorical attributes. Other models have different, usually richer, descriptions of the top-down information (Hamker, 2006; Hochstein & Ahissar, 2002; Navalpakkam, Rebesco,

& Itti, 2004; Torralba, Oliva, Castelhano, & Henderson, 2006). This topic awaits more data.

Also open for debate is exactly how the system sets weights to use some signals and ignore others (e.g. from irrelevant singletons, Bacon & Egeth, 1994; Lamy & Egeth, 2003; Lamy & Tsal, 1999; Rauschenberger, 2003; Theeuwes, 1991; Yantis, 1993). Some of this weight-setting seems to be essentially a priming effect in which the last stimulus configures the search for the next (e.g. DiLollo, Kawahara, Zuvic, & Visser, 2001; Hillstrom, 2000; Kristjansson, Wang, & Nakayama, 2002; Olivers & Humphreys, 2003) but, as discussed below, that can't be the whole story (Muller, Reimann, & Krummenacher, 2003; Wolfe, Butcher, Lee, & Hyle, 2003).

Returning to the example of a search for a small blue item, dimensional weights would be adjusted so that size and color information made more of a contribution to the activation map than other dimensions like motion and orientation. Feature weights would be adjusted to favor blue and small over yellow and big. It does not seem to be possible to set the weight on bottom-up saliency to zero. In this case, yellow items near blue and big near small all produce salience signals that act as noise in this particular task. Combined with noise within the visual system, this will degrade the activation map so that the small blue target is not necessarily the most active item. Search will be biased toward the target, but not perfectly, resulting in search that is moderately efficient, falling between the easiest feature searches and the hardest inefficient searches where no feature guidance is possible. (Illustrated, in black and white, in Figure Two).

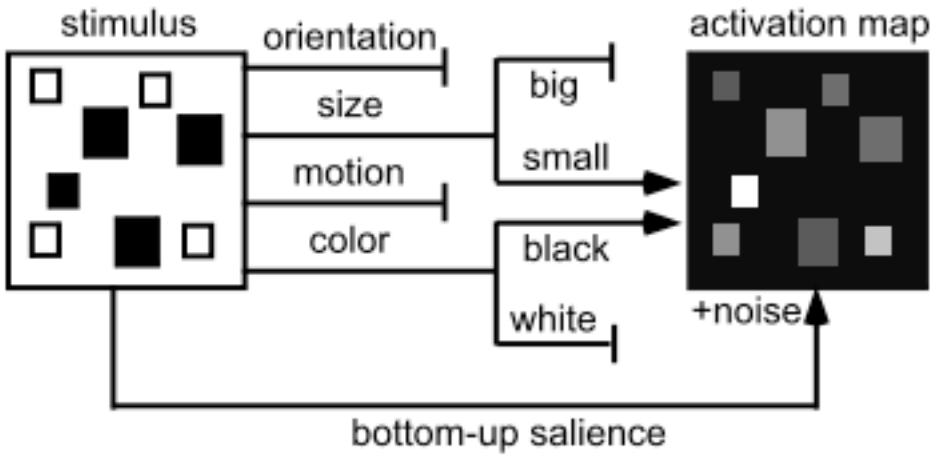



**Figure Two: In a search for small, black items, the activation map guiding attention would be based on inputs from Size and Color. Weights for other dimensions (e.g. Motion and Orientation) would be reduced. Within the dimension, weights for Black and Small would be increased while other features (e.g. White and Big) would be decreased. Noise and a mandatory contribution from bottom-up salience complete the inputs to the activation map.**

In earlier versions of GS, it was proposed that attention was directed to the most active locus in the activation map. If that did not prove to be the target, attention would move to the next most active, and so on, until the target was found or all (or almost all) items had been rejected. This simple serial search model is inadequate for several reasons. First, as will be discussed in the second part of this chapter, subsequent research has shown that observers do not keep track of rejected distractors in a way that would permit this sort of sampling without replacement (Horowitz & Wolfe, 1998, 2001, 2003, 2005). Second, depending on your assumptions about memory in search, simple items are being processed at a rate of one object every 20-40 msec in standard search tasks. However, the lower bound of estimates for the time required to identify a single item is on the order of

100 msec (Thorpe, Fize, & Marlot, 1996). Accordingly, GS4 models object recognition as an asynchronous diffusion process (Ratcliff, 1978; Ratcliff, 2006). Specifically, as shown in Figure 3, when an item is selected, information begins accumulating. If that information reaches a target threshold, a target-present response is generated. If that information reaches a distractor threshold, the item is rejected as a possible target. Identification of one object need not end before the process begins for the next item. Attention can be deployed to the next item while the prior item is still in the process of being identified. At that point, two items would be in the identification process at the same time. The maximum number of items that can be in the identification stage at the same time is a parameter of the model. A number on the order of 4 seems to work well in current simulations (and, perhaps coincidentally, fits nicely with capacity estimates for visual working memory - Cowan, 2001; Luck & Vogel, 1997).

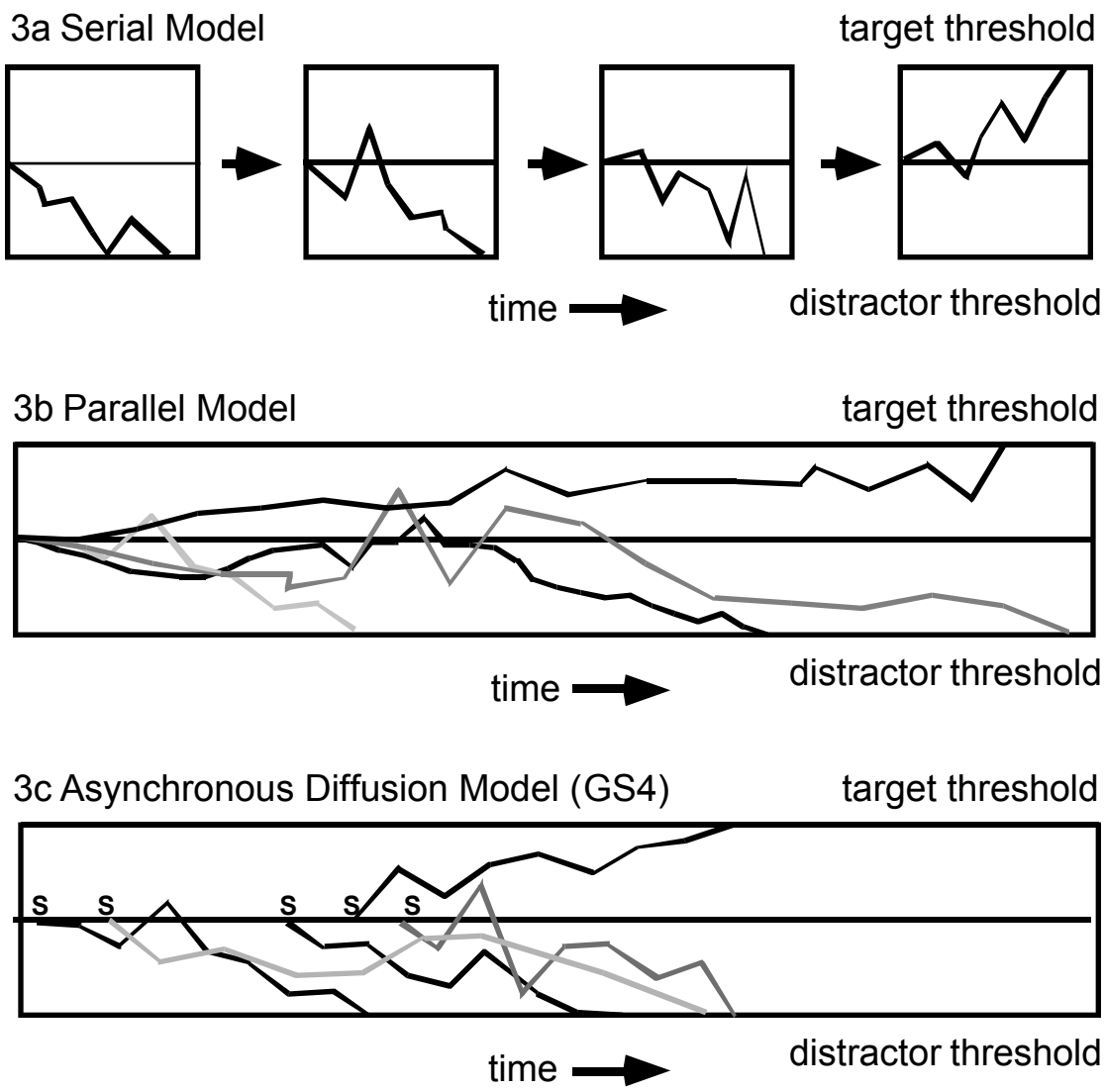



**Figure Three: Using a diffusion process to model the categorization of an item as target or distractor, this figure cartoons serial and parallel accounts of visual search as well as the hybrid, GS4 account. Each "S" in 3c marks the selection of a new item. In 3a and 3b, all selections occur at the left-hand axis.**

The asynchronous diffusion aspects of GS4 blur the distinction between serial and parallel models of visual search. If the diffuser can handle only one item at a time, you have a strict serial model (Fig 3a). If all items can begin diffusing at the same time, you

have a parallel model (3b). If the rate of diffusion varies inversely with the number of items that are diffusing, you have a limited-capacity parallel model. GS4 is a limited-capacity serial/parallel hybrid model (3c). It places a limit on the number of items that can be diffusing toward recognition at any moment. Therefore, once the set size is above that limit, there must be some serial aspect to the process of search. However, multiple items can be in the process of recognition at the same time, giving the model a parallel aspect as well. More details can be found in Wolfe (2007).

One final introductory note about the architecture of GS4: In the spirit of Feature Integration Theory, GS4 holds that object identification requires selective attention of the object in order to permit binding. If you can't recognize an object before you attend to it, what do you <u>see</u> at that object's location before attention is directed there? This topic has been the subject of a great deal of debate. Positions range from the claim that if you don't attend to something you don't see it (Mack & Rock, 1998; Mack, Tang, Tuna, & Kahn, 1992; Noe, Pessoa, & Thompson, 2000; O'Regan & Noe, 2001) to claims that objects (Thorpe, Fize, & Marlot, 1996; VanRullen & Thorpe, 2001) or whole scenes ( Li, VanRullen, Koch, & Perona, 2002) can be recognized without attention, or at least without much attention. It is always a methodological challenge to claim that a task is being done "without attention." This mirrors the classic debate between early selection (Broadbent, 1958) and late selection models (Deutsch & Deutsch, 1963; reviewed in Pashler, 1998).

Rather like its response to the serial/parallel debate, GS4 wants to have it both ways on the early/late selection debate. The debate centers on the position of a bottleneck on a pathway from eye to conscious perception. Suppose, however, that there are two pathways through the visual system, a common enough idea (Goodale, 1996; Held, 1970; Ungerleider & Mishkin, 1982). GS4 proposes that one pathway gives rise to object recognition. The bottleneck of selective attention lies in this pathway making it the embodiment of an early selection model. The second pathway is non-selective, not bottlenecked, but severely limited in its capabilities. It is capable of quickly registering image statistics (Ariely, 2001; Chong & Treisman, 2003) and global scene properties that might give rise to an understanding of the layout of a scene or even its semantic category (e.g. "beach scene", Oliva & Torralba, 2001; Torralba, Oliva, Castelhano, & Henderson, 2006). From the vantage point of conscious visual perception, it would be this non-selective pathway that fills the visual field with some sort of perceptual "stuff." Note that the non-selective pathway is not a magic end run around the limitations of selective attention. Each pathway is severely limited: the selective pathway in its capacity and the non-selective pathway in its abilities.

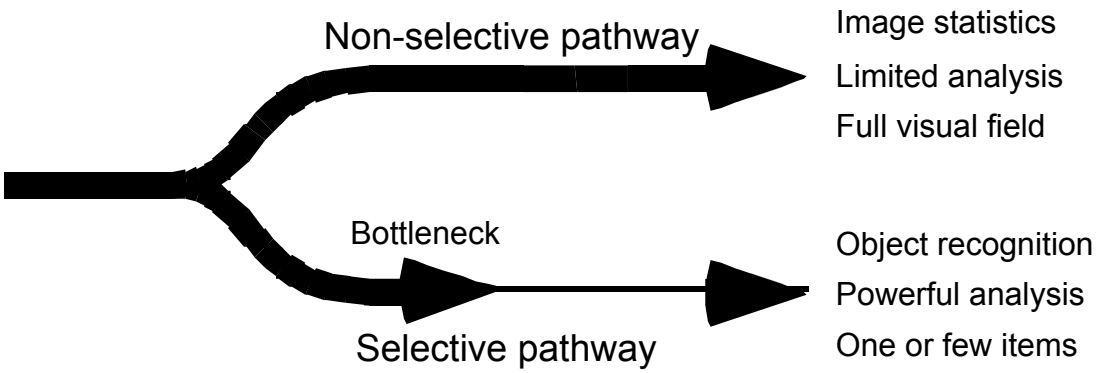



**Figure Four: The large-scale architecture of GS4 differentiates between a selective pathway that does object recognition on a very limited subset of the stimulus at any**

**one time and a non-selective pathway that does a very limited analysis of the entire scene in parallel.**

**2. How long does it take to "change your mind"?**

Consider that basic conjunction search for a small blue target, described in the previous section. This search will be reasonably efficient because observers can use the information about the color and size of the target to guide the deployment of attention. What is the time course of the guidance? Returning yet again to Figure One, you can look at the array without a task, in an unguided way. If you are then asked to look for a small, dark square, you can reconfigure your visual system to guide your attention to those targets. We are interested in the temporal dynamics of that act of getting in to Guided Search. Here we describe four sets of experiments that reveal the dynamics of the start of guidance. The first set asks how long it takes to change your mind, switching from one search goal to another.

Consider a search task in which the observer did not know the identity of the target because it could change on each trial. We asked observers to search for the unique item in an array composed of rectangles that could be red or green, big or small, vertical or horizontal (Wolfe, Horowitz , Kenner, Hyle, & Vasan, 2004). On each trial, one pair of attributes defined the unique target (e.g. BIG GREEN – CAPS denote target attributes), but the participant was not informed of the target attributes before the search display was presented. The two types of distractors on that trial each shared one target attribute. In

this case, those would be BIG red and small GREEN distractors. On this trial, the orientation would be irrelevant and uniform.

In the absence of information about target identity, one might do this task in a number of ways. One might look for an oddball in a set defined by color or size. One might figure out that multiple BIG red and small GREEN distractors mandate either a BIG GREEN or a small red target. Whatever our observers were doing, it took about 1250 msec on average to find the target in this "Uninformed" condition. Unsurprisingly, if the target did not change over trials, observers learned what to look for and they could do the task more quickly. Thus, when the observers searched for BIG GREEN or some other fixed target on every trial in a "Blocked" condition, mean RT was about 600 msec.

In the critical conditions, observers were cued about target identity at some time prior to the appearance of the search array. In these conditions, a cue was presented at fixation, some time prior to the onset. The cue could be a copy of the actual target or the words describing the target.

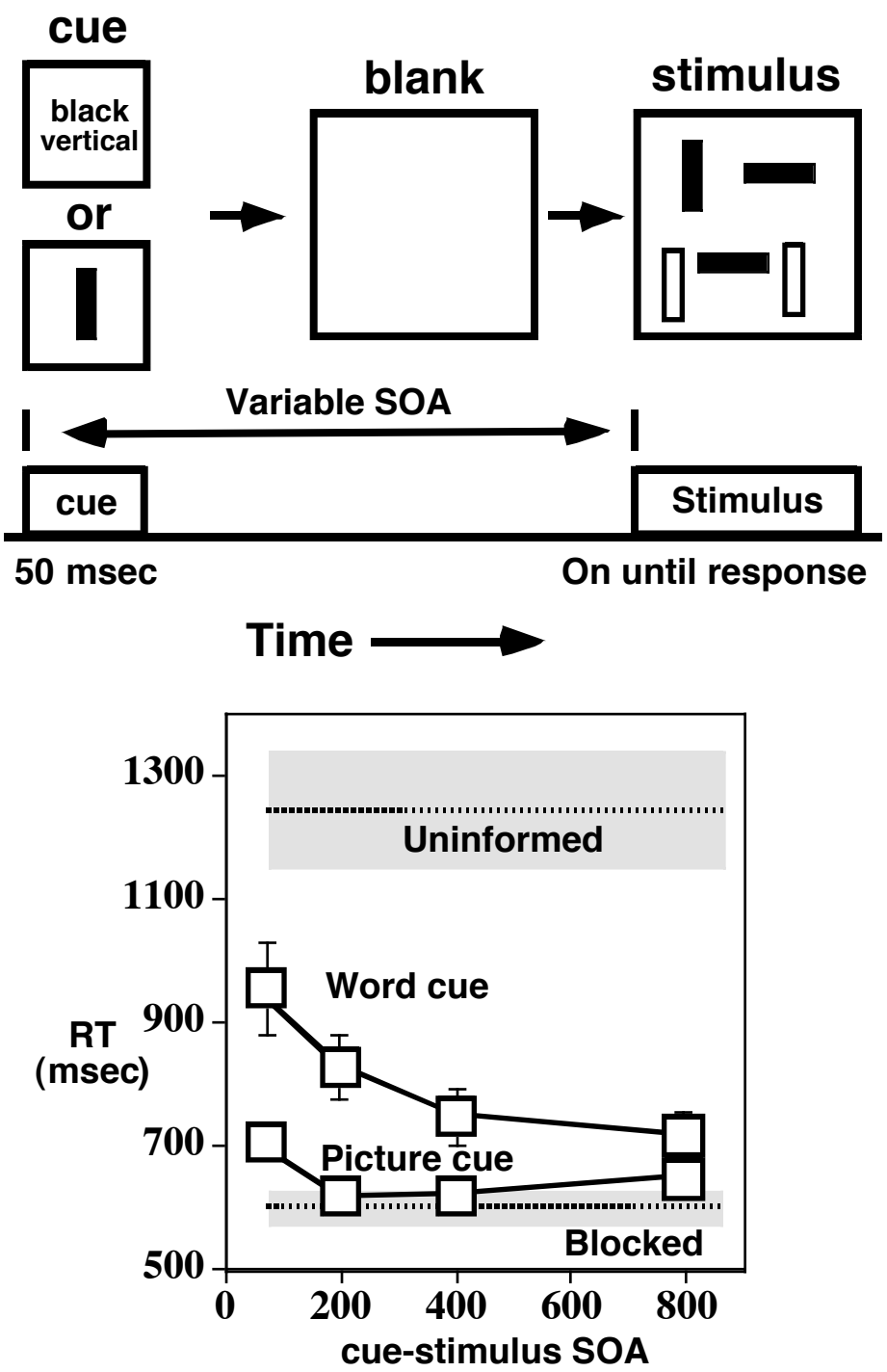



**Figure Five: The design and sample results from the "advance warning" experiments. Dashed lines represent the Blocked and Uninformed baseline conditions. Gray regions are 95% confidence limits. Data figure redrawn from Wolfe, Horowitz, Kenner, Hyle, & Vasan (2004). How fast can you change your mind? The speed of top-down guidance in visual**

**search. *Vision Research, 44*(12), 1411-1426.**

Figure Five shows the basic result. When the cue was physically identical to the target, 200 msec was adequate advance warning to make the cued RT indistinguishable from the blocked RT. It is not surprising that it took longer for the word cues to become fully effective. After all, the words needed to be read. More interesting was the fact that word cues were never as good as picture cues. Even after hundreds of trials, writing "BIG GREEN" produced guidance that was less effective than showing the observer the big green item. That penalty went away when target type repeated by chance. Thus, even if word cues were used, if the cue and the target were BIG GREEN for two trials in a row, then the RT for the second trial was as fast, on average, as the picture-cued trials. This is consistent with the idea that seeing the target on one trial primes the observer to find it on the next (priming of pop-out, Hillstrom, 2000; Kristjansson, 2006; Lamy, Bar-Anan, Egeth, & Carmel, 2006; Maljkovic & Martini, 2005; Maljkovic & Nakayama, 1994). It suggests that the picture cue acted not only as top-down information but also as a prime.

These data indicate that it takes a minimum of about 200 msec to change the setting of the top-down guiding weights so that attention can be guided to items of the correct color, size, or orientation. In those 200 msec, the weights get to the same state that they are in on trials in the Blocked condition, where the weights can remain more or less fixed over many trials. One might think that this represents the time to get guidance going. However, that is not quite right, as the next set of experiments will show.

**3. Even consistently mapped guidance takes time to get started**

The preceding experiment measures the time to start guidance only if we assume that guidance is fully in place at the start of a trial under blocked conditions when the target identity is consistent and known. This would imply a model in which guidance was something like a filter, eliminating (or, at least, attenuating) items with the wrong attributes.

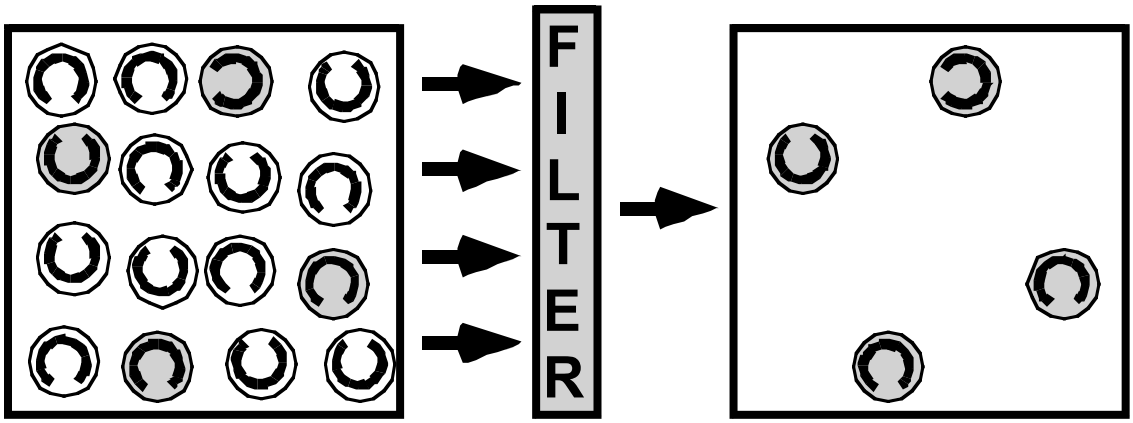



**Figure Six: Guidance as a filter**

Figure Six illustrates this point with stimuli from a new set of experiments. The observer was faced with a set of Cs in four possible orientations. All but one opened up or down. The target, present on each trial, opened to the left or right, and observers were asked to report its orientation. Each C was placed on a colored disk. Suppose that observers knew that the target was always on a gray disk. If the top-down guidance to gray acted like a filter, then search for the target should be like search through the set of four gray items. If no guidance were available, this would be a search through the 16 Cs on the left side of the figure.

To examine the temporal dynamics of guidance, the appearance of the colored disks was offset in time from the appearance of the Cs as illustrated at the top of Figure Seven.

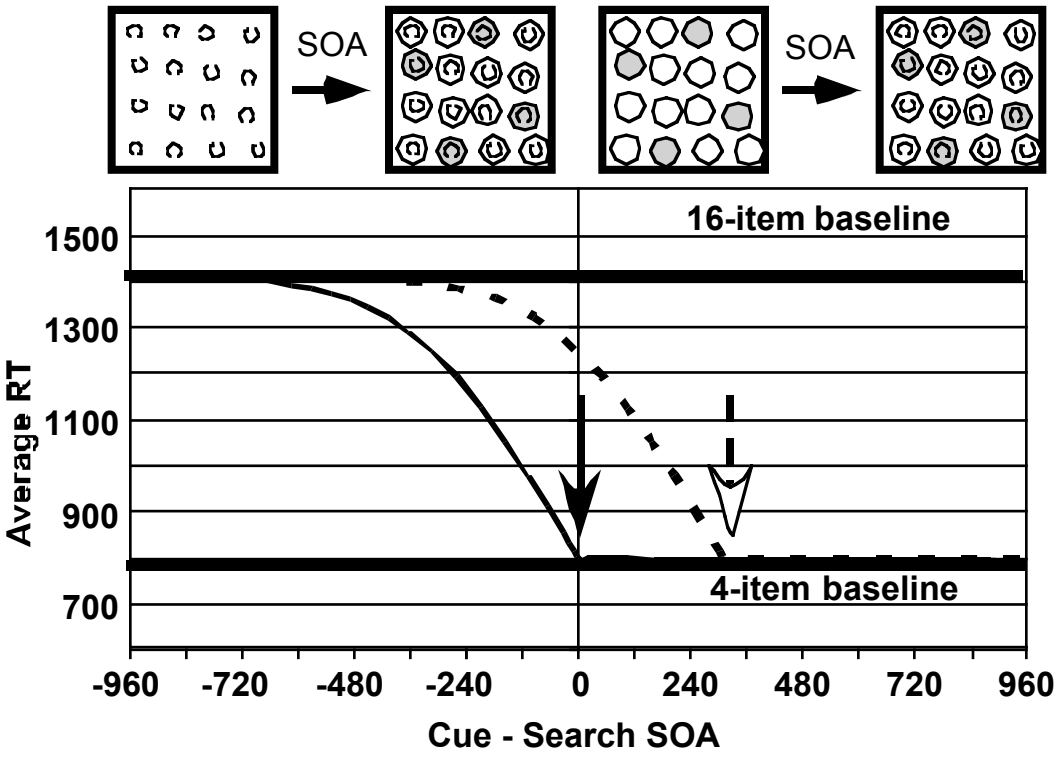



**Figure Seven: Hypothetical outcomes for the time-to-guide experiments. Solid line assumes that color guidance is available as soon as the color is available. Dashed line assumes that guidance begins about 300 msec after the color becomes available.**

Consider a simple, two-state model. At any moment, observers are either searching through all 16 items in an unguided manner, or they are restricting their search to the four items of the target color. For simplicity, assume that there is a sharp transition between those two states. Suppose that the Cs appear 400 msec before the color cue (SOA = -400 in Figure Seven). When the Cs appear, observers must search through 16 items because there is no guiding information. After 400 msec, the color information appears. Once it

becomes effective, this becomes a search through four items. The RT, therefore, is a mixture distribution of some purely unguided searches, when the observer finds the target before the color ever appears, and some that benefit from eventual guidance. As the SOA becomes increasingly negative, there is a greater chance that the search will finish before the color becomes available. At the longest negative SOAs, RTs should approximate the 16-item baseline, the time required to find a target when there is no color guidance. The four-item baseline is the RT for an unguided search through a set of just four items. Figure Seven shows the predictions of such a model for the case where guidance starts as soon as the guiding information is presented (solid line) and a second case where guidance starts some time after the onset of the guiding information (dashed line). In this case, that delay is arbitrarily set to about 300 msec. Note that the declining function hits the 4-item baseline RT at the SOA corresponding with the delay in the onset of guidance. This is marked by the arrows at 0 msec for the solid line and 300 for the dashed line.

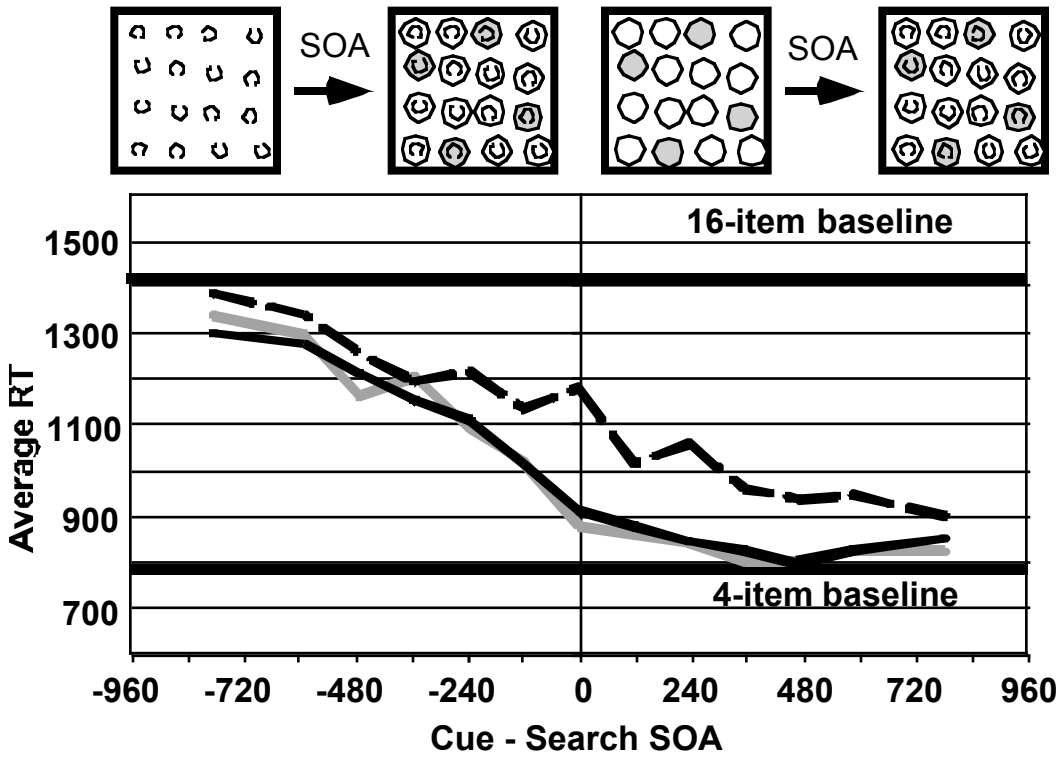


**Figure Eight: Results for one set of time-to-guide experiments. Black line shows data from Blocked trials. Gray line shows data for Mixed trials with consistent mapping of target and distractor colors. The dashed line shows data for Mixed trials with inconsistent mapping.**

The data in Figure Eight show the average of the median RTs for 15 observers. Each observer was tested for 30 trials at each of 13 SOAs. Each observer also completed 27 trials of unguided search with set sizes of 4 and 16 items to establish the baselines, plotted as horizontal lines at their median value. A full account of these experiments is presented in Palmer, Van Wert, Horowitz & Wolfe, in preparation.

The data plotted with a dashed line are most analogous to the data from the previous "advance warning" experiments. On these trials, observers did not know from trial to trial which color set would contain the target. Moreover, a target color on one trial could be a distractor color on another (inconsistent mapping Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977). Clearly, guidance is not fully effective in this condition until many hundreds of msec after the onset of the color. Why is guidance so delayed? In this experiment, observers had to figure out the target color for themselves. They were told that there would be 12 items of one color and 4 of another and that the target was always in the smaller set. When a target color on one trial could become a distractor color on the next, substantial penalties, akin to task switching costs, are seen (e.g. Mayr & Kliegl, 2003; Wylie & Allport, 2000).

The data plotted in gray show the results when observers do not know the color of the target on this trial but do know that the set of target colors and the set of distractor colors are distinct from each other (consistent mapping). This speeds the onset of fully effective guidance but, critically, at an SOA of 0 msec, performance is significantly worse than the 4 item baseline and has not reached asymptote. This is the sign that guidance is not fully effective at the onset of the guiding colors.

The most important data for the present argument are those plotted in black. They show the results for the blocked condition where observers knew that the 4 item subset containing the target would always be the same color. The 12 distractors also preserved the same color for the entire block. It is obvious that this Blocked condition produced the same results as the Consistent Mapping condition. Even when observers knew that the target was always, for example, in the pink subset, they were not able to begin fully effective guidance at the moment of color onset. These data cannot pinpoint the precise time when guidance is fully available, but clearly it takes at least 200 msec. We can firmly reject the hypothesis that guidance is fully available from stimulus onset even under conditions when an observer can maintain the same 'guiding principles' for an entire block of hundreds of trials.

The data shown in Figure Eight came from a study using relatively desaturated colors, and the actual search stimuli (i.e. the Cs) were black, but placed on colored placeholders that carried the guiding information. However, we have replicated this result several times, including experiments using saturated hues and experiments where the Cs

themselves carry the guiding color. In all cases, the basic result is the same: at SOA 0, guidance is not fully active. We can reject the model of Figure Six. Guidance is not a simple filter that, once established, can sit athwart the visual pathway, ready and waiting for the next stimulus. Instead, guidance must be re-established on every trial and that takes several hundred msec.

**4. Where, in theory, is the guiding representation?**

In Treisman's original formulation, there was a 'preattentive stage' that could process a limited set of basic features in parallel. It was followed by an attentive stage that accomplished the binding of features into object representations in a serial manner, one object at a time (Treisman & Gelade, 1980). Guided Search (Wolfe, Cave, & Franzel, 1989) borrowed this essentially linear architecture and added the notion that the preattentive stage could guide the deployment of attention in the attentive stage. Subsequent work has revealed problems with this architecture. Attention seems to be guided by a specific abstraction of what were the traditional "preattentive features." For example, guidance by orientation seems to be based on a categorical representation of orientation. Guidance is more effective if the target is categorically unique: the only steep, shallow, left- or right-tilted item (Hodsoll & Humphreys, 2005; Wolfe, Friedman-Hill, Stewart, & O'Connell, 1992). Moreover, in guidance by orientation, a 90 deg representation provides the greatest orientation contrast. Guidance seems to be quite insensitive to the 180 deg difference between upright and inverted, even for clearly polar objects (Wolfe, Klempen, & Shulman, 1999). These findings pose a problem because, if preattentive information is a coarse categorization of the input, how do later stages

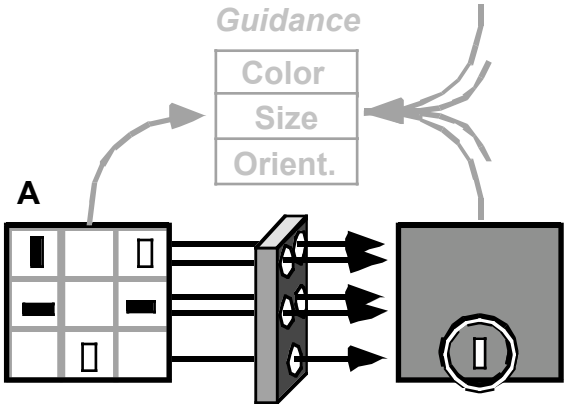
recover information? After all, the end user of all this visual processing can determine which end is up and can tell the difference between 20 deg and 30 deg "steep" orientations.

The point is perhaps made more clearly in the case of preattentive processing of intersection type (Bergen & Julesz, 1983; Gurnsey & Browse, 1989; Julesz & Krose, 1988; Wolfe & DiMase, 2003). The distinction between a "T" junction and a "+" intersection can be used by early visual processes - for example, to make initial inferences about occlusion of one object by another (e.g. Rensink & Enns, 1995; Yantis, 1995). Later processes also have access to this distinction. After all, we can easily distinguish between an uppercase T and a lowercase t on the basis of the intersection type. Nevertheless, visual search for + junctions among T junctions or vice versa is very inefficient (Wolfe & DiMase, 2003). In a linear scheme, it is awkward to explain how intersection information can be available before and after guidance, but is not available to guide.

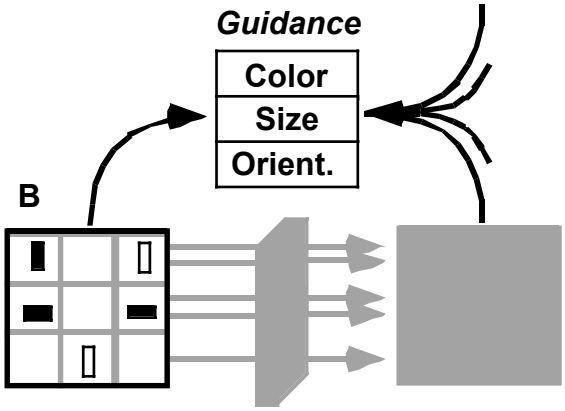
In the latest versions of Guided Search (GS4), we have argued for a "guiding representation" that is not in the main pathway from visual input to object recognition (Wolfe, 2007; Wolfe & Horowitz, 2004). Instead, we propose that guidance sits (figuratively) to one side of the pathway, controlling the bottleneck between early visual processes and later object recognition processes. The guiding representation is abstracted

from the visual input but cannot be directly perceived, nor is it one of the building blocks of later, perceptual representations. In this view, it is a separate control device.

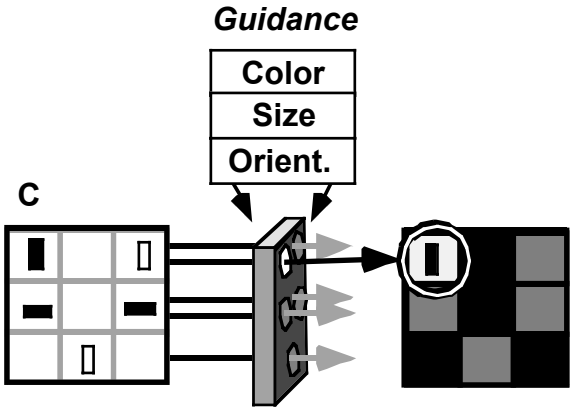
The findings concerning the timing of guidance fit into this framework as shown in the cartoon in Figure Nine.




At stimulus onset, search is unguided
and an item is selected at random.




Guiding information accumulates




Subsequently selection of an item
is guided by basic feature information

**Figure Nine: Guided Search 4 proposes that guidance is based on a representation that lies off the main path from early vision to object recognition. In this view, the deployments of attention (marked by the circle) in a search would initially be unguided (9a). Guiding information would accumulate over time (9b) and subsequent deployments would be guided (9c).**

Imagine that the observer is looking for a black vertical line. The cartoon in Figure 9a shows an essentially unguided, initial, feed-forward sweep of activity after stimulus onset. At the same time, information about the stimulus is being fed to the guiding representation (9b). This might be feed-forward information from early visual stages or feed-back information from some later stage. Guided Search is currently agnostic on this point; hence, the various input arrows with unspecified origins. As shown in Figure 9c, after some period of time the guiding representation has enough information to determine where "black" and "vertical" items might be. At that point, guidance can be used to constrain selection and it becomes more likely that attention will be directed to a black vertical target. This view is broadly consistent with the notion of guidance as "reentrant process" (Di Lollo, Enns, & Rensink, 2000; DiLollo, Kawahara, Zuvic, & Visser, 2001; Lamme & Roelfsema, 2000) and models like the Reverse Hierarchy Theory (Ahissar & Hochstein, 1997; Hochstein & Ahissar, 2002).

**5. Are the first moments of standard guided searches 'unguided'?**

The account cartooned in Figure Nine makes the prediction that visual search is initially unguided, with available guidance developing as search progresses. The experiments

presented thus far point in that direction but some converging evidence would be helpful. The results of the blocked condition of the time-to-guide experiments show that, at SOA 0, guidance to four items is not exactly the same as search through four items, presented without other distractors. However, that task differs from a standard search task in its dissociation of the guiding color information from the rest of the stimulus. In this section, we describe evidence for a delayed onset of full guidance in standard visual search.

The logic of these experiments is very simple. In standard visual search experiments, set size is varied to allow estimation of the slope of the RT x set size function. Since that function is assumed to be linear, three or four set sizes are typically considered adequate. However, if guidance takes time to develop, searches through small set sizes should be relatively unguided and should produce relatively steep slopes compared to larger set sizes. In other words, the RT x set size function should be curved instead of linear, with large set sizes having shorter RTs than would be predicted by linear extrapolation from small set size RTs. This should be especially true for "guided" search tasks like conjunction search and less evident for "unguided" search tasks such as a spatial configuration search for a T among Ls or 2 among 5s. In the following experiments, we tested observers on both conjunction search and spatial configuration search using eight set sizes spanning a large range and sampling densely at the smallest set sizes.

In the first experiment, set sizes were mixed within blocks. The conjunction search consisted of search for a red vertical bar among green vertical and red horizontal bars, and the spatial configuration search consisted of search for a T among Ls (“TvL”), where

both Ts and Ls could appear in any of four 90° rotations. The upper left panel in Figure Ten shows the mean RTs for correct target-present trials for the conjunction and TvL tasks. The accompanying regression lines were computed on only set sizes 1-4. This figure makes three points. First, the initial slopes are very similar: over the first four items, conjunction search is just as inefficient as spatial configuration search. Second, RTs for the larger set sizes were systematically faster than would be predicted by linear extrapolation from the smaller set sizes. Third, this compressive non-linearity is greater for the conjunction data than the TvL data. These three points are reinforced in the lower left panel of Figure Ten, which plots the target present slopes for the two search tasks for small and large set size ranges. Note that we are using this approach only to illustrate that the functions are compressive. We are not making any claim that the functions are bi-linear, since any set of four set sizes would be roughly linear.

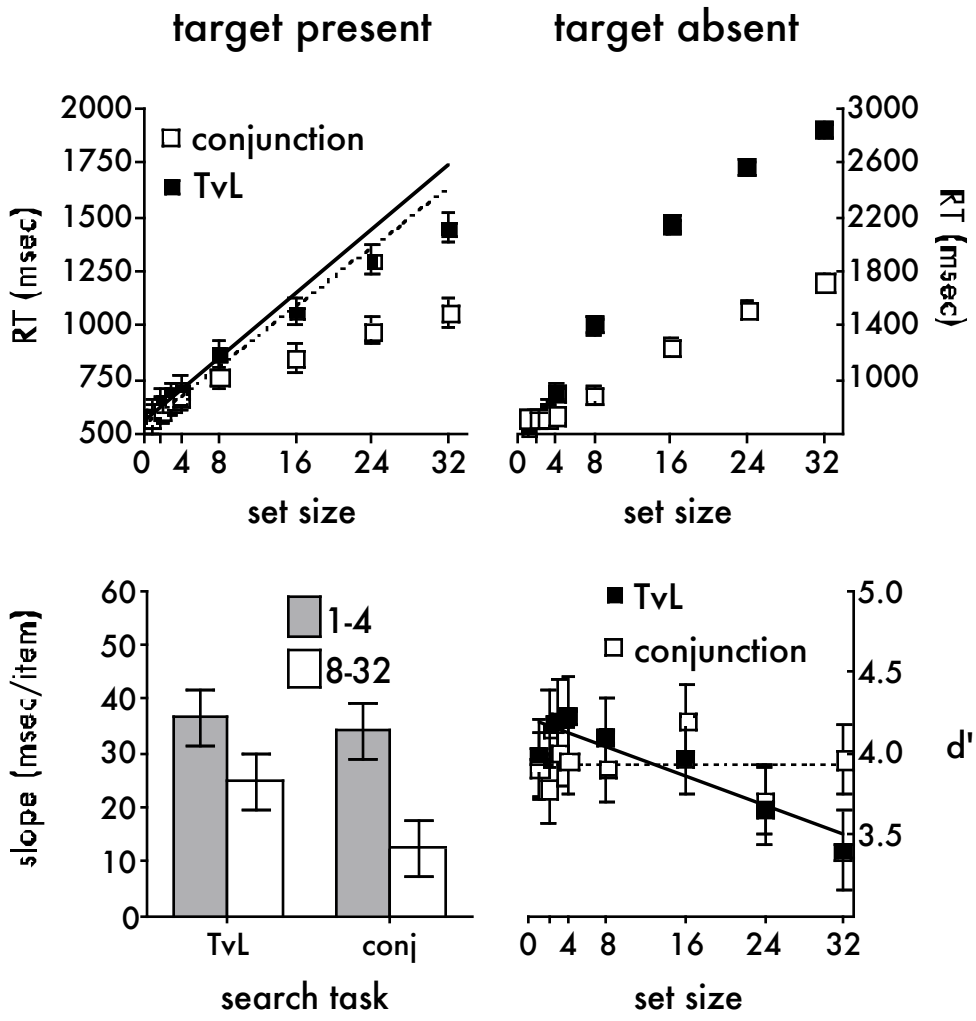



**Figure Ten: Compressively non-linear RT x set size functions in visual search. Upper left panel plots the target-present RT x set size functions for the TvL (solid symbols) and the conjunction (open symbols) search tasks. Regression lines are based on set sizes 1-4 for TvL (solid line) and conjunction (dashed line). Upper right panel plots the target-absent RT x set size functions. Lower left panel shows target present slopes for set sizes 1-4 (shaded bars) and 8-32 (open bars) as a function of task. Lower right panel shows d' as a function of set size for TvL (solid symbols) and conjunction (open symbols) tasks. Error bars in all figures except the d' figure are**

**within-subjects 95% confidence intervals based on the interaction shown. In the d' figure, confidence intervals are based on the simple effect of set size for that task.**

These data are consistent with the hypothesis that guidance takes some time to become effective. For the fast searches through small set sizes, the "guided" conjunction search is no more efficient than the presumably unguided TvL task. The tasks diverge at larger set sizes; after, we would argue, color and orientation have had a chance to effectively guide the selection of items.

Can we explain these results as a product of speed/accuracy tradeoffs? Remember that RTs for the two search tasks were comparable at the small set sizes, diverging only at the larger set sizes. Thus, a speed/accuracy tradeoff account must predict that errors at large set sizes should be substantially greater for the conjunction task than for the TvL task. If we convert those errors into d' measures, a speed accuracy tradeoff would be reflected in a decline in d' as set size increases. The lower right panel of Figure Ten shows that the TvL task shows more of a decline than the conjunction task. In fact, d' did not vary with set size for the conjunction task, so the compressive non-linearity in those data cannot be explained by a speed-accuracy tradeoff. The smaller non-linearity in the TvL data, however, can be at least partially accounted for by a speed-accuracy tradeoff. This analysis suggests that the difference between TvL and conjunction search is actually underestimated in these data: if we could hold TvL accuracy constant across set sizes, we would have seen even less compression.

A plausible account of the data from this experiment would make three points.

1) When examined with sufficient resolution, RT x set size functions from standard search tasks are not necessarily linear. This point, by itself, is worthy of note because it has been assumed that the functions are basically linear and most models produce linear functions.
2) At least part of the non-linearity in the TvL data is likely due to a speed-accuracy tradeoff. Modeling this sort of non-linearity would require correctly specifying the "quitting rule" for visual search. When do observers give up on unsuccessful searches? This has been a hard problem in the modeling of search.
3) The non-linearity in the conjunction search data does not appear to be readily explained as a speed-accuracy tradeoff, since d' remains constant across set sizes. Search efficiency for conjunction and TvL tasks are initially very similar but diverge at larger set sizes. This is consistent with the idea that the initial stages of the search are relatively unguided but become guided as time goes on.

The slope at low set sizes would be consistent with a rate of processing of 35-70 msec per item (depending on your modeling assumptions). If we propose that search becomes increasingly guided after 4-5 items have been examined, this would produce an estimate of about 200-300 msec for the onset of guidance, comparable to the estimates from the previous experiments in this chapter.

**6. A replication and some complications**

Thus far, the data presented here fit with the cartoon in Figure Nine. The initial phase of a search appears to be unguided, with guidance gradually developing to full strength over

the course of several hundred msec. In this section, we wish to consider two matters that complicate this account. First, in a second experiment, described below, we again sampled set sizes densely in conjunction and spatial configuration searches. We successfully replicated the decelerating function for conjunction search. However, compared to the TvL task of the previous experiment, we obtained more marked non-linearity in the nominally "unguided" spatial configuration search. This cannot be attributed to a simple speed-accuracy tradeoff. The second complication is that the estimates of the time required to get guidance going seem long. In some physiological studies, reentrant processes have latencies on the order of 200 msec, but others have shorter latencies (Lamme & Roelfsema, 2000). Why does it seem to take so long to start simple guidance by a feature like a color?

The second experiment was motivated by our concern that the shape of the RT x set size functions might be strongly influenced by the mixture of set sizes. The average set size was 11.25. Perhaps observers quit searching through the longer set size displays because they had set a quitting threshold based on the smaller set sizes. Accordingly, in the second experiment, the set size variable was blocked. We maintained the same conjunction search task for comparison purposes, but employed a different spatial configuration task, search for a digital 2 among digital 5s ("2v5").

Figure Eleven shows the data in same format as Figure Ten. The conjunction search data replicated the previous experiment in every aspect. RTs at large set sizes were substantially faster than predicted by linear extrapolation from the small set sizes (open

symbols in the upper left panel of Figure Eleven). The slopes for search through 1-4 items were again around 35-40 msec/item, but the slopes for the larger set sizes were substantially shallower (lower left panel of Figure Eleven). This cannot be explained by a speed-accuracy tradeoff, since d' was constant across set sizes (open symbols in the lower right panel of Figure Eleven. This close replication suggests that search strategies were not affected by whether the set sizes were mixed or blocked.

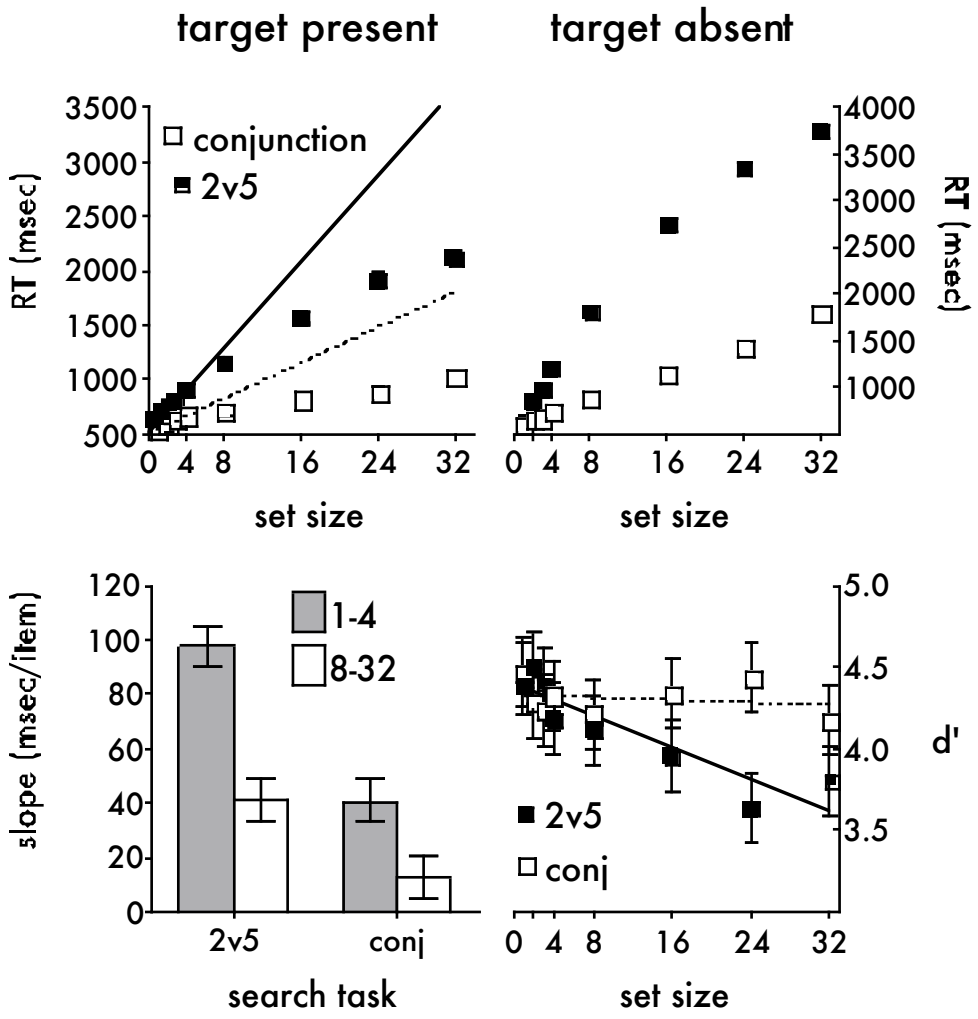



**Figure Eleven: Compressively non-linear RT x set size functions in visual search (blocked set size version). Upper left panel plots the target present RT x set size functions for the 2v5 (solid symbols) and the conjunction (open symbols) search**

**tasks. Regression lines are based on set sizes 1-4 for 2v5 (solid line) and conjunction (dashed line). Upper right panel plots the target absent RT x set size functions. Lower left panel shows target present slopes for set sizes 1-4 (shaded bars) and 8-32 (open bars) as a function of task. Lower right panel shows d' as a function of set size for 2v5 (solid symbols) and conjunction (open symbols) tasks. Error bars in all figures except the d' figure are within-subjects 95% confidence intervals based on the interaction shown. In the d' figure, confidence intervals are based on the simple effect of set size for that task.**

However, the data for the 2v5 task were different than what we observed for the TvL task in the first experiment. The 2v5 task was substantially less efficient than the conjunction task for small set sizes and its RT x set size function was markedly compressive (filled symbols in the upper left hand panel of Figure Eleven), more so than TvL search (lower left hand panel of Figure Ten). This is not the result we would predict for unguided spatial configuration search. Some of the non-linearity in the 2v5 task can be explained due to speed-accuracy tradeoffs, since d' again declined as a function of set size. However, while d' drops off with set size at roughly the same rate in the two experiments, the compressive non-linearity is much greater for the 2v5 task. Some additional explanation is required.

The current architecture of Guided Search contains a possible source for a non-linearity not associated with guidance. Recall from Figure Three, that GS4 models search as an asynchronous diffusion process. Information about each item, as it is selected, diffuses

toward a target or distractor boundary. If the item is rejected as a distractor, it is removed from the diffuser and a new item is selected. The parameters of this process are discussed elsewhere (Wolfe, 2007). One important parameter for present purposes is the capacity of the diffuser. How many items can be diffusing at the same time? A reasonable estimate might be about 4, the proposed capacity for visual short-term memory (Cowan, 2001; Luck & Vogel, 1997). For present purposes, the exact number is not important, what is important is that this architecture predicts differences between the beginning of a search and its later stages.

Suppose that capacity is four. What happens at the first moment of selection? One possibility is that four items are selected at the same time. Subsequent items are selected only as each of the initially selected items is dropped from the diffuser. Thus, there is a transition from parallel selection of multiple items to a steady state selection at a rate related to the slope in unguided search tasks (on the order of one every 50 msec). If the set size is small, most or all of the items in the display might be selected in this initial phase. If four items are selected as efficiently as one, why does the RT increase quite steeply with small set sizes? One possibility is that more resources can be devoted to processing each item if the diffuser is not loaded to capacity. Thus, one might imagine that information in the diffuser would accumulate four times as fast when one item was in the diffuser than when four items were in the diffuser.

Suppose that only one item can be selected at a time – a strict serial bottleneck, even if up to four could be in the diffuser at one time. In that case, the beginning of each search

would be characterized by a steady increase in the number of items, up to the capacity of the diffuser. If the rate of accumulation of information about an item is proportional to the number of items in the diffuser, then the rate of accumulation for the first selected item would slow as other items were selected into the diffuser.

The point of this delineation of possibilities is to show that many variants of the GS4 architecture are characterized by an early period of search that is different from steady-state search and that will evolve into the steady-state if the search is reasonably prolonged. For those searches that have very small set sizes, the entire search will have characteristics of the early period of search. Simulations of these various GS4 configurations tend to produce decelerating, non-linear RT x set size functions. If capacity is about four and if the throughput rate is about 50 msec per item, the transition from the early stage of search to the steady state will take about 200 msec, broadly consistent with the data. At the present time, the data do not constrain the model to the point where it is possible to firmly specify a capacity or to determine if the initial act of selection involves one or several items. All of these options propose some sort of transition over the course of an extended period of search.

If curvilinear functions fall out of the GS4 architecture, is there any need to propose that guidance takes time to develop? The time-to-guide experiments (Figures Seven & Eight) suggest that the answer is yes. If guidance were in place at the beginning of a trial, like the filter of Figure Six, then the initial loading of the diffuser would only include items of the correct color in those experiments. If that were the case, then the RT for SOA 0

should be the same as the RT for a set size of four. In either case, the diffuser should be loaded with a set of four items that must include the target. Apparently this is not the case. Fully effective guidance develops after stimulus onset.

**7. Guidance by surface type**

As a final example of the evolution of search process during a single search, consider the problem of looking for people in a real scene. In some sense, this must be a guided search. We look for people in places where people might be – on horizontal surfaces that will support them, for example, and not in less plausible places like midair or on the wall (Torralba, Oliva, Castelhano, & Henderson, 2006). This sort of guidance is not the feature guidance of a standard conjunction search (Wolfe, Cave, & Franzel, 1989). Nor is it the location guidance of spatial cuing experiments (Posner & Cohen, 1984). Some of our recent data suggests that it is a form of guidance with its own very slow time course. Perhaps this guidance is slow because it involves feedback from the "non-selective" pathway illustrated in Figure Four.

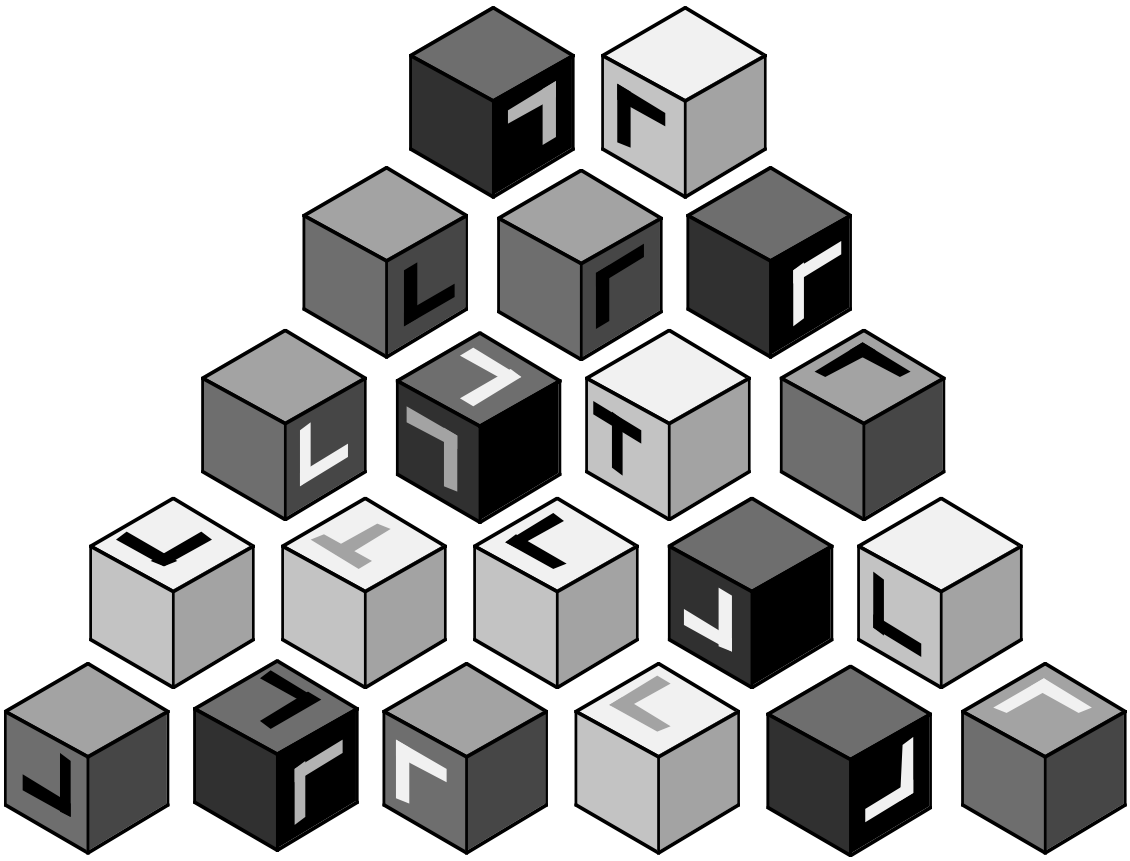


**Figure Twelve: Stimuli for search by surface type. Observers could be asked to look for a "T" (Look for a T of any color on any surface), a colored T (Look for the black T) or a T on a particular surface (Look for the "T" on the top of a cube).**

Figure Twelve is a grayscale illustration of the stimuli we used. Actual stimuli were vividly colored. Observers looked for a T among Ls. A single T was present on half the trials. The other half of the trials contained only Ls. Five conditions were tested. In the Unguided condition, observers searched for a T among Ls. In the Blocked Color condition, observers searched for a T of a specific color (red, blue or yellow) among Ls of various colors. In the Mixed Color condition, the guiding color was changed on each trial. A word cue was given in 1000 msec advance of each trial to specify the color. In the

Blocked Surface condition, observers searched for a T on a specific type of cube surface (top, left, or right) for 300 trials in a row. In the Mixed Surface condition, the guiding surface was changed on each trial. A word cue was given in advance of each trial to specify the surface. In all of the guided searches, the guiding property defined a subset of one third of the items. That is, if the target was a red T, one third of the Ls were also red. If the target was a T on top of a cube, one third of the Ls were on top of cubes.

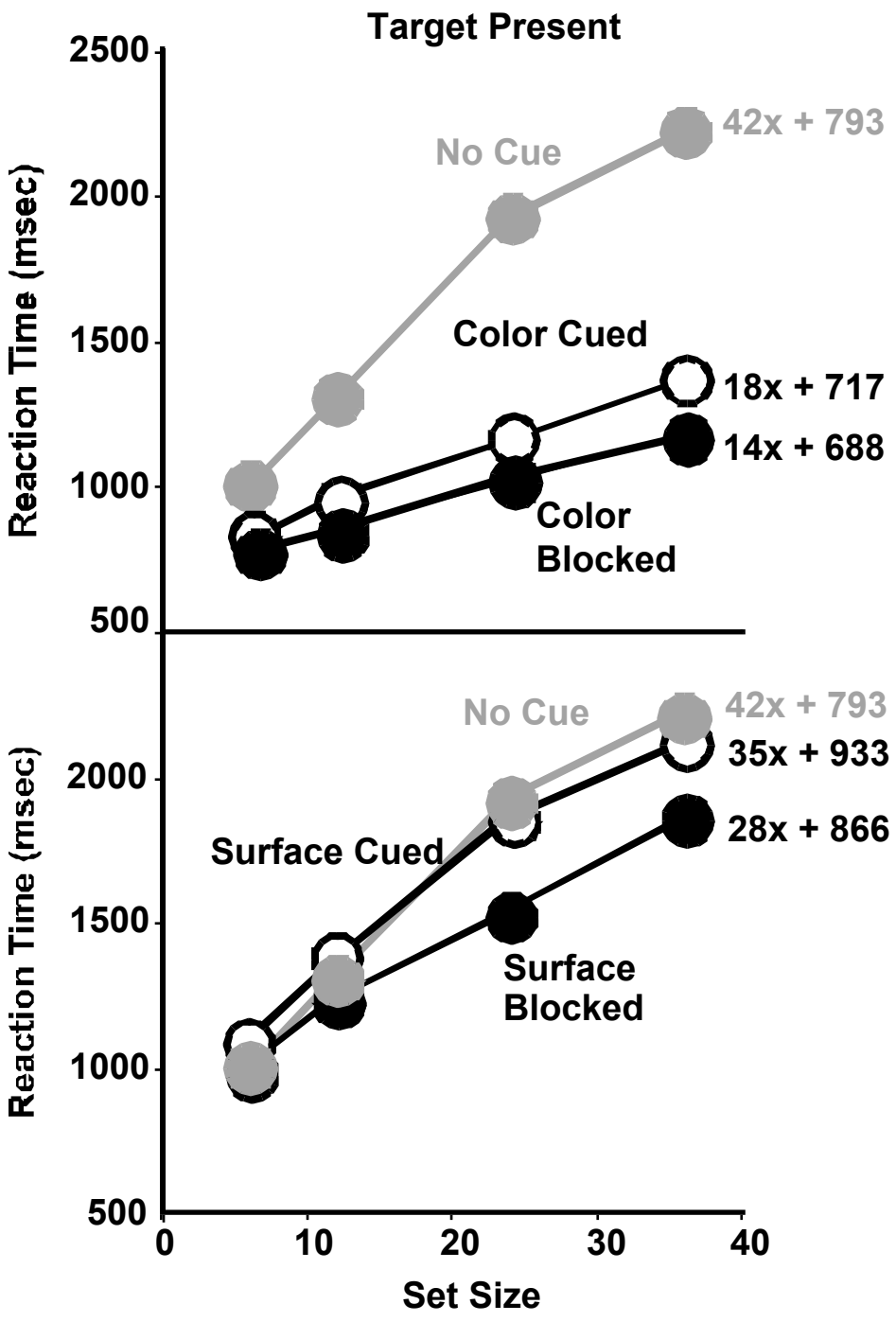



**Figure Thirteen: Mean target-present RT data for searches guided by color and surface information. Unguided baseline data are shown in gray.**

The results for target present trials are shown in Figure Thirteen. Color guidance replicated standard guided search results. One third of the items were of the guiding color. Consequently, slopes dropped by a factor of three relative to the Unguided condition. Results for guidance-by-surface showed much less evidence for guidance. Indeed, when the surface was changed on each trial (Surface Cued), there was no significant guidance for target present trials.

Why did this experiment "fail"? It is intuitively clear that guidance of this sort *must* exist. If you are told that your child's missing sock is *somewhere* on the floor, you will be able to guide attention to the floor and away from the walls, ceilings, and shelves of his room. One possibility is that guidance to surface properties is a different form of guidance, one that takes a long time to develop. Figure Thirteen shows a hint of this in the Surface Blocked conditions. Here the RTs are the same as the Unguided RTs for smaller set sizes. They deviate for the larger set sizes.

This hypothesis is supported more clearly by the target absent data shown in Figure Fourteen. When surface was the guiding property, RTs for the blocked and cued versions of the task were very similar to the Unguided baseline for set sizes of 6 and 12. They deviate at set sizes 24 and 36. This cannot be attributed to a simple speed-accuracy tradeoff.

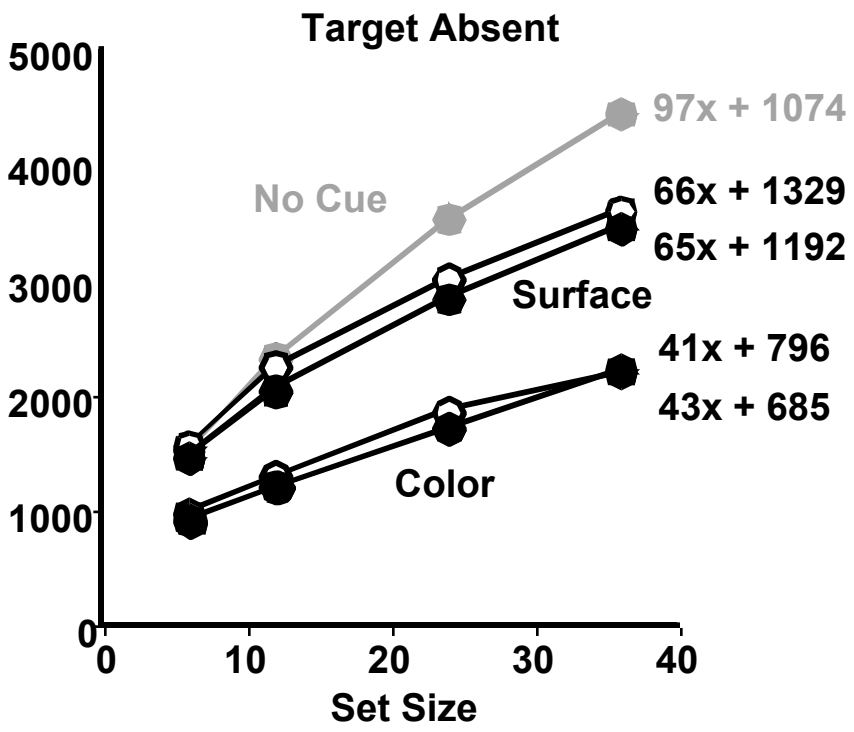



**Figure Fourteen: Target-absent data for search using the stimuli illustrated in Figure Twelve. Filled symbols indicate blocked conditions. Open symbols indicated that the guiding feature or surface was specified by cue on each trial.**

Work on guidance by scene properties is in its infancy. For the present, we offer as a working hypothesis the idea that guidance by surface is similar to guidance by classical features like color. However, it involves reentrant processes operating on a much slower time scale. It would be unsurprising to find that the nervous system might use the same trick in two different ways. Examples abound. If we stay within the realm of visual attention, we find that illusory conjunctions occur for basic features. An observer might report seeing red vertical in a display containing red horizontal and green vertical stimuli (Treisman & Schmidt, 1982) and, at a slower time scale, one can see illusory words when letters are incorrectly combined (Treisman & Souther, 1986).

**8. Conclusions**

Twenty-five years ago, models of visual search had a linear feel to them; a series of boxes, connected by arrows that flowed from input to perception. Models of vision and visual physiology tended to have similar architectures. The output of retinal receptive fields was used to build simple cells that built complex cells and so on until you reached cells that responded to faces and yellow Volkswagens (Barlow, 1985 ; Weisstein, 1973). It is clear that this was an oversimplification. For example, we now know that there are more fibers feeding back from cortex to LGN than feeding forward (Cudeiro & Sillito, 2006). Current theories recognize that understanding the feed-forward sweep of information from eye to brain probably only accounts for the first 100 msec or so of visual experience (Serre, Oliva, & Poggio, 2007). After that, reentrant feedback mechanisms will tend to alter and complicate any feed-forward story. Steady-state visual perception – the visual experience of an observer looking for an extended period of time at a scene – is different from the initial experience of the first moments of input.

In this chapter, we have seen evidence for this temporal evolution of the processes of vision in the context of the guidance of visual search. Even if you know what you are looking for and even if you have been looking for the same thing on trial after trial, visual search at the moment of stimulus onset is not as effectively guided as it will be a couple of hundred msec later. On the initial, feed-forward sweep, guidance does not appear to be fully engaged. Only with time are those aids to selection used to their full effect.

**Bibliography**

Ahissar, M., & Hochstein, S. (1997). Task difficulty and visual hierarchy: Counter-streams in sensory processing and perceptual learning. *Nature, 387*(22 May 1997), 401-406.

Ariely, D. (2001). Seeing Sets: Representation by statistical properties. *Psychological Science, 12*(2), 157-162.

Bacon, W. F., & Egeth, H. E. (1994). Overriding stimulus-driven attentional capture. *Perception and Psychophysics, 55*(5), 485-496.

Barlow, H. B. (1985 ). The twelfth Bartlett memorial lecture: The role of single neurons in the psychology of perception. *QJEP, 37A*, p121-145.

Bergen, J. R., & Julesz, B. (1983). Rapid discrimination of visual patterns. *IEEE Trans on Systems, Man, and Cybernetics, SMC-13*, 857-863.

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

Cave, K. (1999). The FeatureGate model of visual selection. *Psychological Research, 62*(2-3), 182-194.

Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision Res, 43*(4), 393-404.

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

Cousineau, D., & Shiffrin, R. M. (2004). Termination of a visual search with large display size effects. *Spat Vis, 17*(4-5), 327-352.

Cowan, N. (2001). The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behav Brain Sci, 24*(1), 87-114; discussion 114-185.

Cudeiro, J., & Sillito, A. M. (2006). Looking back: corticothalamic feedback and early visual processing. *Trends Neurosci, 29*(6), 298-306.

Deco, G., Pollatos, O., & Zihl, J. (2002). The time course of selective visual attention: Theory and experiments. *Vision Reseach, 42*(27), 2925-2945.

Deutsch, J. A., & Deutsch, D. (1963). Attention: Some theoretical considerations. *Psychological Review, 70*, 80-90.

Di Lollo, V., Enns, J. T., & Rensink, R. A. (2000). Competition for consciousness among visual events: the psychophysics of reentrant visual pathways. *Journal of Experimental Psychology: General, 129*(3), 481-507.

DiLollo, V., Kawahara, J.-i., Zuvic, S. M., & Visser, T. A. W. (2001). The preattentive emperor has no clothes: a dynamic redressing. *J Exp Psychol Gen, 130*(3), 479-492.

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

Goodale, M. A. (1996). One visual experience, many visual systems. In T. Inui & McClelland (Eds.), *Attention and Performance XVI* (Vol. 16, pp. 369-393). Cambridge, MA: MIT Press.

Grossberg, S., Mingolla, E., & Ross, W. D. (1994). A neural theory of attentive visual search: Interactions of boundary, surface, spatial and object representations. *Psychological Review, 101*(3), 470-489.

Gurnsey, R., & Browse, R. A. (1989). Asymmetries in visual texture discrimination. *Spat Vis, 4*(1), 31-44.

Hamker, F. (2006). Modeling feature-based attention as an active top-down inference process *Biosystems, 86*(1-3), 91-99.

Hamker, F. H. (2004). A dynamic model of how feature cues guide spatial attention. *Vision Res, 44*(5), 501-521.

Held, R. (1970). Two modes of processing spatially distributed visual stimuli. In F. O. Schmitt (Ed.), *The Neurosciences: Second study program.* (pp. 317-324). New York: Rockefeller U,.

Herd, S., & O'Reilly, R. (2005). Serial search from a parallel model. *Vision Res, 45*(24), 2987-2992.

Hillstrom, A. P. (2000). Repetition effects in visual search. *Perception and Psychophysics, 62*(4), 800-817.

Hochstein, S., & Ahissar, M. (2002). View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron, 36*, 791-804.

Hodsoll, J. P., & Humphreys, G. W. (2005). The effect of target foreknowledge on visual search for categorically separable orientation targets. *Vision Res, 45*(18), 2346-2351.

Hoffman, J. E. (1979). A two-stage model of visual search. *Perception and Psychophysics, 25*, 319-327.

Hong, S.-K. (2005). Human stopping strategies in multiple-target search. *International Journal of Industrial Ergonomics, 35*, 1-12.

Horowitz, T. S., & Wolfe, J. M. (1998). Visual search has no memory. *Nature, 394*(Aug 6), 575-577.

Horowitz, T. S., & Wolfe, J. M. (2001). Search for multiple targets: Remember the targets, forget the search. *Perception and Psychophysics, 63*(2), 272-285.

Horowitz, T. S., & Wolfe, J. M. (2003). Memory for rejected distractors in visual search? *Visual Cognition, 10*(3), 257-298.

Horowitz, T. S., & Wolfe, J. M. (2005). Visual Search: The role of memory for rejected distractors. In L. Itti, G. Rees & J. Tsotsos (Eds.), *Neurobiology of attention* (pp. 264-268). San Diego, CA: Academic Press / Elsevier.

Hubner, R. (2001). A formal version of the Guided Search (GS2) model. *Percept Psychophys, 63*(6), 945-951.

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Res, 40*(10-12), 1489-1506.

Julesz, B. (1984). A brief outline of the texton theory of human vision. *Trends in Neuroscience, 7*(Feb), 41-45.

Julesz, B., & Krose, B. (1988). Features and spatial filters. *Nature, 333*, 302-303.

Kinchla, R. A. (1992). Attention. *Annual Review of Psychology, 43*, 711-742.

Kristjansson, A. (2006). Simultaneous priming along multiple feature dimensions in a visual search task. *Vision Res, 46*(16), 2554-2570.

Kristjansson, A., Wang, D., & Nakayama, K. (2002). The role of priming in conjunctive visual search. *Cognition, 85*(1), 37-52.

Krummenacher, J., Muller, H. J., & Heller, D. (2001). Visual search for dimensionally redundant pop-out targets: Evidence for parallel-coactive processing of dimensions. *Perception and Psychophysics, 63*(5), 901-917.

Lamme, V. A., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends Neurosci, 23*(11), 571-579.

Lamy, D., Bar-Anan, Y., Egeth, H. E., & Carmel, T. (2006). Effects of top-down guidance and singleton priming on visual search. *Psych Bulletin & Review, 13*(2), 287-293.

Lamy, D., & Egeth, H. E. (2003). Attentional capture in singleton-detection and feature-search modes. *J Exp Psychol Hum Percept Perform, 29*(5), 1003-1020.

Lamy, D., & Tsal, Y. (1999). A salient distractor does not disrupt conjunction search. *Psychonomic Bulletin and Review, 6*(1), 93-98.

Li, F. F., VanRullen, R., Koch, C., & Perona, P. (2002). Rapid natural scene categorization in the near absence of attention. *Proc Natl Acad Sci U S A, 99*(14), 9596-9601.

Li, Z. (2002). A salience map in primary visual cortex. *Trends Cogn Sci, 6*(1), 9-16.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390*(20 Nov), 279-281.

Mack, A., & Rock, I. (1998). Inattentional blindness: Perception without attention. In R. D. Wright (Ed.), *Visual Attention* (Vol. 8, pp. 55-76). New York: Oxford.

Mack, A., Tang, B., Tuna, R., & Kahn, S. (1992). Perceptual organization and attention. *Cognitive Psychology, 24*, 475-501.

Maljkovic, V., & Martini, P. (2005). Implicit short-term memory and event frequency effects in visual search. *Vision Res, 45*(21), 2831-2846.

Maljkovic, V., & Nakayama, K. (1994). Priming of popout: I. Role of features. *Memory & Cognition, 22*(6), 657-672.

Mayr, U., & Kliegl, R. (2003). Differential effects of cue changes and task changes on task-set selection costs. *J Exp Psychol Learn Mem Cogn, 29*(3), 362-372.

McLeod, P., Driver, J., & Crisp, J. (1988). Visual search for conjunctions of movement and form is parallel. *Nature, 332*, 154-155.

Moraglia, G. (1989). Display organization and the detection of horizontal lines segments. *Perception and Psychophysics, 45*, 265-272.

Muller, H. J., Reimann, B., & Krummenacher, J. (2003). Visual search for singleton feature targets across dimensions: Stimulus- and expectancy-driven effects in dimensional weighting. *J Exp Psychol Hum Percept Perform, 29*(5), 1021-1035.

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

Navalpakkam, V., Rebesco, J., & Itti, L. (2004). Modeling the influence of knowledge of the target and distractors on visual search. *J of Vision, VSS04*(abs E123).

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

Noe, A., Pessoa, L., & Thompson, E. (2000). Beyond the grand illusion: What change blindness really teaches us about vision. *Visual Cognition, 7*(1), 93-106.

Nothdurft, H.-C. (1993). Saliency effects across dimensions in visual search. *Vision Research, 33*(5/6), 839-844.

O'Regan, J. K., & Noe, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*.

Oliva, A., & Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision, 42*(3), 145-175.

Olivers, C. N. L., & Humphreys, G. W. (2003). Attentional guidance by salient features depends on intertrial contingencies. *J. Exp. Psychol: Human Perception and Performance, 29*(3), 650-657.

Parkhurst, D., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision Res, 42*(1), 107-123.

Pashler, H. (1998). *The Psychology of Attention.* Cambridge, MA: MIT Press.

Pashler, H. E. (1998). *Attention*. Hove, East Sussex, UK: Psychology Press Ltd.

Posner, M. I., & Cohen, Y. (1984). Components of attention. In H. Bouma & D. G. Bouwhuis (Eds.), *Attention and Performance X* (pp. 55-66). Hillside, NJ: Erlbaum.

Ratcliff, R. (1978). A theory of memory retrieval. *Psych. Preview, 85*(2), 59-108.

Ratcliff, R. (2006). Modeling response signal and response time data. *Cognit Psychol, 53*(3), 195-237.

Rauschenberger, R. (2003). Attentional capture by auto- and allo-cues. *Psychon Bull Rev, 10*(4), 814-842.

Rensink, R. A., & Enns, J. T. (1995). Pre-emption effects in visual search: evidence for low-level grouping. *Psychological Review, 102*(1), 101-130.

Sanders, A. F., & Donk, M. (1996). Visual search. In O. Neumann & A. F. Sanders (Eds.), *Handbook of perception and action, Vol. 3: Attention.* (Vol. 3, pp. 43-77). London: Academic Press.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychol. Rev., 84*, 1-66.

Serre, T., Oliva, A., & Poggio, T. (2007). A feedforward architecture accounts for rapid categorization. *Proc Natl Acad Sci U S A, 104*(15), 6424-6429.

Shiffrin, M. R., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psych. Review, 84*, p127-190.

Theeuwes, J. (1991). Exogenous and edongenous control of attention: The effect of visual onsets and offsets. *Perception and Psyhcophysics, 49*(1), 83-90.

Thornton, T. L., & Gilden, D. L. (2007). Parallel and serial processes in visual search. *Psychol Rev, 114*(1), 71-103.

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature, 381*(6 June), 520-552.

Torralba, A., Oliva, A., Castelhano, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: The role of global features on object search. *Psychological Review, 113*(4), 766-786.

Treisman, A. (1988). Features and objects: The 14th Bartlett memorial lecture. *Quart. J. Exp. Psychol., 40A*, 201-237.

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

Treisman, A., & Souther, J. (1986). Illusory words: The roles of attention and of top-down constraints in conjoining letters to form words. *Journal of Experimental Psychology: Human Perception and Performance, 12*, 3- 17.

Treisman, A. M., & Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cognitive Psych., 14*, 107-141.

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

VanRullen, R., & Thorpe, S. J. (2001). Is it a bird? Is it a plane? Ultra-rapid visual categorisation of natural and artifactual objects. *Perception, 30*(6), 655-668.

Weisstein, N. (1973). Beyond the yellow Volkswagen detector and the grandmother cell: A general strategy for the exploration of operations in human pattern recognition. In R. L. Solso (Ed.), *Contemporary issues in cognitive psychology: The Loyola Symposium*. (pp. 17-51). Washington, DC: Winston/Wiley.

Wolfe, J., Horowitz , T., Kenner, N. M., Hyle, M., & Vasan, N. (2004). How fast can you change your mind? The speed of top-down guidance in visual search. *Vision Research, 44*(12), 1411-1426.

Wolfe, J. M. (1994). Guided Search 2.0: A revised model of visual search. *Psychonomic Bulletin and Review, 1*(2), 202-238.

Wolfe, J. M. (1998). Visual search. In H. Pashler (Ed.), *Attention* (pp. 13-74). Hove, East Sussex, UK: Psychology Press Ltd.

Wolfe, J. M. (1998). What do 1,000,000 trials tell us about visual search? *Psychological Science, 9*(1), 33-39.

Wolfe, J. M. (2007). Guided Search 4.0: Current Progress with a model of visual search. In W. Gray (Ed.), *Integrated Models of Cognitive Systems* (pp. 99-119). New York: Oxford.

Wolfe, J. M., Butcher, S. J., Lee, C., & Hyle, M. (2003). Changing your mind: On the contributions of top-down and bottom-up guidance in visual search for feature singletons. *J Exp Psychol: Human Perception and Performance, 29*(2), 483-502.

Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided Search: An alternative to the Feature Integration model for visual search. *J. Exp. Psychol. - Human Perception and Perf., 15*, 419-433.

Wolfe, J. M., & DiMase, J. S. (2003). Do intersections serve as basic features in visual search? *Perception, 32*(6), 645-656.

Wolfe, J. M., Friedman-Hill, S. R., Stewart, M. I., & O'Connell, K. M. (1992). The role of categorization in visual search for orientation. *J. Exp. Psychol: Human Perception and Performance, 18*(1), 34-49.

Wolfe, J. M., & Gancarz, G. (1996). Guided Search 3.0: A model of visual search catches up with Jay Enoch 40 years later. In V. Lakshminarayanan (Ed.), *Basic and Clinical Applications of Vision Science* (pp. 189-192). Dordrecht, Netherlands: Kluwer Academic.

Wolfe, J. M., & Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience, 5*(6), 495-501.

Wolfe, J. M., Klempen, N. L., & Shulman, E. P. (1999). Which End is Up? Two representations of orientation in visual search. *Vision Research, 39*(12), 2075-2086.

Wylie, G., & Allport, A. (2000). Task switching and the measurement of "switch costs". *Psychol Res, 63*(3-4), 212-233.

Yantis, S. (1993). Stimulus-driven attentional capture. *Current Directions in Psychological Science, 2*(5), 156-161.

Yantis, S. (1995). Perceived continuity of occluded visual objects. *Psychological Science, 6*(3), 182-186.

Zenger, B., & Fahle, M. (1997). Missed targets are more frequent than false alarms: A model for error rates in visual search. *J. Experimental Psychology: Human Perception and Performance, 23*(6), 1783-1791.

Zhaoping, L., & Koene, A. R. (2007). Feature-specific interactions in salience from combined feature contrasts: Evidence for a bottom–up saliency map in V1. *J. of Vision, 7*(7), Article 6, Pages 1-14.