THE DEPLOYMENT OF VISUAL ATTENTION: TWO SURPRISES

Jeremy M Wolfe

Brigham and Women's Hospital and Harvard Medical School
221 Longwood Ave., Boston, MA 02115 USA
wolfe@search.bwh.harvard.edu

1. SUMMARY

The visual system is not capable of processing of all aspects of a scene in parallel. While some visual information can be extracted from all locations at once, other processes, including object recognition, are severely limited in their capacity. Selective attention is used to limit the operation of these limited-capacity processes to one (or, perhaps, a few) objects at a time. Searching for a target in a scene, therefore, requires deployment of attention from one candidate target to the next until the target is found or the search is abandoned. Common-sense suggests that distractor objects that have been rejected as targets are marked in some fashion to prevent redeployment of attention to non-target items. Introspection suggests that sustained attention to a scene builds up a perception of that scene in which more and more objects are simultaneously recognized.

Neither common-sense nor introspection are correct in this case. Evidence suggests that covert attention is deployed at random among candidate targets without regard to the prior history of the search. Rejected distractors are not marked during a search. Prior to the arrival of attention, visual features are loosely bundled into objects. Attention is required to bind features into a recognizable object. For an object to be recognized, there must be a link between a visual representation and a representation in memory. Our data suggest that only one such link can be maintained at one moment in time. Hence, counter to introspection, only one object is recognized at one time. These surprising limits on our abilities may be based on a trade off speed for apparent efficiency.

Keywords: Vision, visual attention, visual search, guided search, memory, object recognition, human experimental psychology

2. INTRODUCTION

Faced with a new scene, we immediately see something. However, we do not immediately perceive everything. Thus, you might emerge from customs at the airport to be faced with a crowd of faces, one of whom should be the friend who has come to pick you up. It is not possible to simultaneously process all of the faces (not to mention the other objects in the scene) to the point of recognition. As a result, you need to search. Search from face to face in an apparently serial manner (29; 38) will either lead you to your friend or will lead you to the bus and to a reassessment of the nature of friendship.

Two aspects of the course and consequence of such a search are the topics of this paper. First, it seems reasonable to assume that, if you deploy attention to a face and determine that it is not your friend, that you will somehow mark that face so as to avoid revisiting it. Second, even if you do not recognize multiple objects when first confronted with a new scene, it seems intuitively clear that, after prolonged search, the visual scene will contain multiple, simultaneously recognized objects. The purpose of this paper is to demonstrate that neither of these reasonable hypotheses is actually true. In a field of items that are equivalent in their ability to attract attention, attention appears to be deployed at random with no regard to the prior history of deployments. When attention is deployed to an item, it becomes possible to recognize that item. However, when attention is redeployed away from the item, the item is no longer actively recognized. It may be remembered, just as an item that is out of sight is remembered. But our data indicate that simultaneous recognition of multiple objects does not occur.

This paper is organized into four sections. In the first section, we review some of the basics of laboratory visual search experiments. Next, we discuss the evidence that the deployment of attention is more anarchic than commonsense would predict. A third section considers the visual consequences of attention. Finally, the implications of these results will be discussed.

3. VISUAL SEARCH IN THE LABORATORY

3.1. Introduction to Search Methods

In a standard laboratory visual search experiment, observers search for a target item among a number of distractor items. In a typical version, the target would be present on 50% of the trials. The total number of items (the "set size") would be varied. The dependent measures are the "reaction time"(RT) - the amount of time required to press a key to indicate the
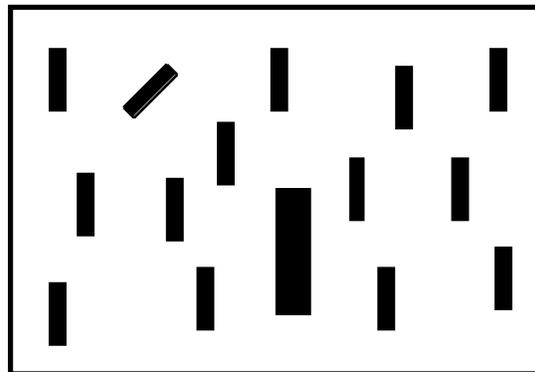

Figure One: Highly efficient search. Targets defined by salient basic features can be found, independent of the number of distractors. Here targets are defined by size and orientation.

presence or absence of a target - and the accuracy of that response. Most of the results presented here will be RT data. The measure of greatest interest is the slope of the function relating RT to set size. This is a measure of the efficiency of search. The most efficient searches have slopes near zero, suggesting that all items can be processed at the same time, without capacity limits. Examples are shown in Figure 1.

The most efficient searches are searches for targets defined by a basic feature among homogeneous distractors (e.g. red among green, big among small, etc.) The set of basic features for visual search contains obvious candidates like color (e.g. 2; 10), size(e.g 4), and orientation (e.g 3; 26). It also contains less obvious features like lustre (61) and a variety of depth cues (14). The full list contains perhaps a dozen features (reviewed in 57).

The presence of an attribute is easier to detect than its absence. This leads to so-called "search asymmetries" (50) where the search for A among B produces a steeper slope than a search for B among A. An example is shown in Figure 2.
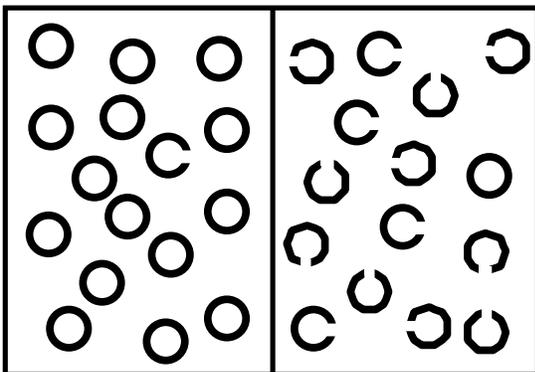

Figure Two: It is easier to find the presence of a feature (here the line terminators in the "C") than it is to find the absence of a feature. (After Treisman).

At the other end of the continuum of search tasks are inefficient searches. With slopes of about 20-30 msec/item on target present trials and about twice that on target absent trials. It is, of course, possible to have search tasks with arbitrarily steep slopes. One source of steep slopes is a need to fixate items. If the items cannot be classified as distractor or target without fixating each item, then search slopes will come to
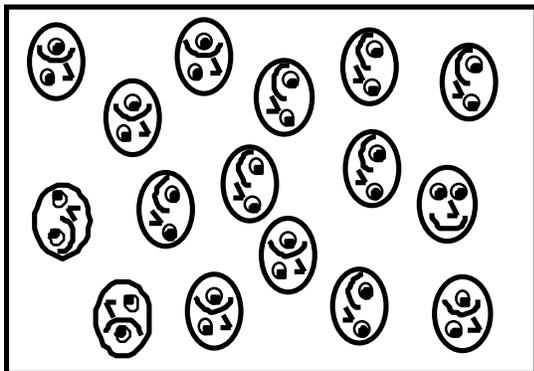

Figure Three: Even ecologically significant stimuli like faces produce inefficient search if they do not differ from distractors in basic features.

reflect the speed of eye movements (2-4 fixations per second) and will yield slopes of greater than 100 msec/ item. In experiments, like those described here, that are concerned with the covert deployment of attention, care must be taken to assure that eye movements are not required. It is less important, in most cases, to require rigorous fixation since the pattern of RTs appears to be essentially the same whether eye movements are permitted or not (66).

The class of inefficient searches includes all those for which basic feature information is of no use. This includes searches for easily identifiable objects like faces and animals where identification is based on the relationship of features to one another rather than to the mere presence of a defining feature. Our data indicate that the shape of an object is not a basic feature for visual search. If local features like line termination are controlled, search for one shape among other, quite different shapes is inefficient (59).

3.2. Conjunctions and Guided Search

Most natural searches are neither feature searches nor random searches among preattentively equivalent items. Most searches involve targets that, while they are not defined by a single unique feature, are defined, at least in part, by basic feature information. Thus, the hunt for your friend at the airport requires a search but it is a search through a subset of visible objects. Little time will be spent examining suitcases and car rental signs (13).

Laboratory search experiments have concentrated on the less natural case of *conjunction search*. In a typical conjunction search, targets are defined by the presence of two features (e.g. a *black vertical* target) among a mix of distractors that have one or the other of these features (e.g. white *vertical* and *black* horizontal distractors).
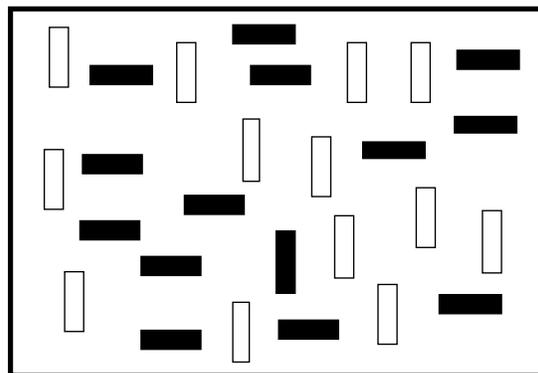

Figure Four: Conjunction search. Find the black vertical item.

Work in the 1970's and early 80's seemed to show that conjunction searches were uniformly inefficient (48). These and other data led to Treisman's very influential proposal that searches could be divided into two categories: Feature searches that could be performed in parallel and all other searches that required serial, item by item, inefficient search. This hypothesis was one of the central propositions of Treisman's original formulation of her "Feature Integration Theory" (48). However, subsequent research revealed that conjunction search could be quite efficient (e.g. 9; 24; 28; 34; 49; 60; 67). At first, it appeared that these efficient conjunctions searches might represent specific exceptions to the general rule of inefficient conjunction search (23; 27).

However, it has become increasingly clear that search for any conjunction of basic features can be efficient if the features are salient enough (see discussions in 56; 58). Indeed, there are several published reports of conjunction searches that yield search efficiencies that are indistinguishable from those produced by basic features (e.g. 40; 52; 55).

In retrospect, this is not a surprise. As the earlier example should have made clear, it is intuitively obvious that attention is somehow guided to likely targets. The Guided Search model makes the claim that this guidance comes from preattentive feature information (6; 56; 60; 62). That is, Guided Search holds that no preattentive process has explicit information about conjunctions. However, to continue with the example from Figure Four, a color processor can guide attention toward black items while an orientation processor can guide attention toward vertical items. The combination of these sources of guidance will tend to guide attention toward items that are both black and vertical (see Figure Five).
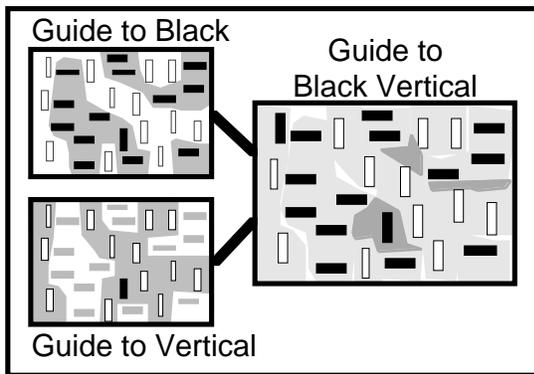

Figure Five: The core idea of Guided Search is that basic feature information can be used to guide attention to targets defined by more than one feature.

Revisions of Feature Integration Theory incorporate feature guidance (46; 49) as do some other models (e.g. 51). On the other hand, there are models, notable Duncan and Humphreys' (11) Similarity model that propose explicit preattentive processing of conjunctions.

3.3. The Myth of Two Classes of Search Tasks

The influence of the 1980 version of Feature Integration Theory has been long and wide. An unintended consequence has been the wide-spread assumption that there are two types of visual search, "serial" and "parallel" and that specific tasks can be placed in one of these two categories on the basis of the slope of the RT x set size function.

In fact, as should be clear from the preceding discussion, search tasks yield a continuum of slopes from efficient to inefficient with no value dividing these slopes into two principled groups. To illustrate this point, we pooled 2000+ search slopes from a range of different feature, conjunction, and letter searches. The distribution of slopes is shown in Figure Six.
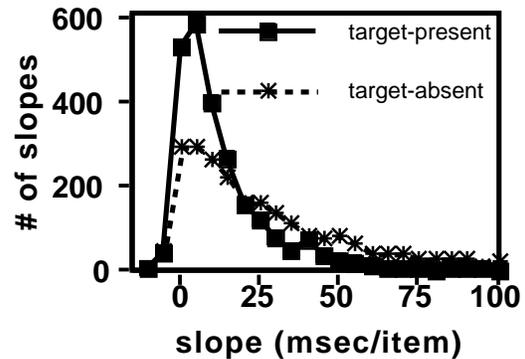

Figure Six: The distribution of 2000+ search slopes showing that there is no obvious division of tasks into search classes on the basis of slope alone (redrawn from 58)

The purpose of this exercise is not to argue that all search tasks are drawn from the same distribution. If we sort the slopes by the type of search task, it is clear that different types of task produce different distributions of slopes. Figure Seven shows the target present slopes of Figure Six broken into three broad classes of search: feature searches, conjunction searches, and searches such as a search for a "T" among "L"s that have traditionally served as benchmark "serial" tasks.

The distributions are clearly different. Thus, search slope can be predicted (albeit imprecisely) from a knowledge of the search task. It is the reverse that does not work. It is not possible to place a dividing line at, say, 10 msec/item and declare searches on one side to be qualitatively different from searches on the other.
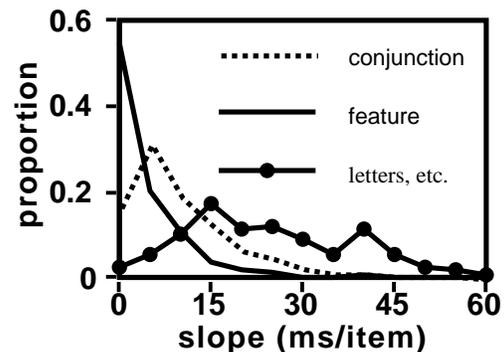

Figure Seven: Distribution of target-present slopes divided by type of task. (redrawn from 58)

There are a number of ways to understand this continuum of search slopes. In the context of the Guided Search model, all searches involve preattentive guidance of the deployment of spatial attention. For the tasks described here, the prime source of variation lies in the effectiveness of that guidance. In the most efficient feature searches, guidance is sufficient to direct attention to the target before it is deployed to any distractors. In an inefficient search such as a search for a T among Ls, guidance still limits search to the Ts and Ls. Attention is not directed to blank space or away from the search display. However, within that set of letters, there is no further guidance and search proceeds at random. Conjunction

tasks represent an intermediate case in which preattentive feature information guides attention but guides it imperfectly so that some distractors attract attention and search slopes are intermediate. In this framework, it is important to understand the rules for deployment of attention. That topic is addressed in the next section.

4. THE DEPLOYMENT OF ATTENTION: THE FIRST SURPRISE

4.1. The standard models

There are two broad classes of models of the deployment of attention. The preceding discussion has assumed a *serial* model in which attention is deployed from item to item. Alternatively, a limited-capacity resource could be allocated to multiple items *in parallel.* Guided Search generally assumes a serial model. However, in principle, preattentive processing could guide the allocation of a distributed resource rather than guiding the deployment of an item-sized attentional "spotlight". Both classes of model can predict the patterns of RTs seen in search experiments(43-45). Intermediate positions are possible. Several models propose a serial deployment of attention, not from item to item, but from one group of items to the next (e.g. 15; 31). In fact, the dichotomy between serial and parallel models may have been overstated. Consider a conveyor belt. Items may be loaded on and off the belt in series but multiple items are on the belt in parallel (see also 16; for a more extensive discussion of this idea see 25).

A hallmark of virtually all of these models of attentional deployment has been the assumption that information accumulates during the course of a trial. In serial models, this takes the form of the assumption that rejected distractors are inhibited or marked in some way so that attention is not re-deployed to previously rejected items (e.g. 1; 20; 42). Phenomena like inhibition of return (IOR) have been invoked as plausible mechanisms of distractor marking (32; 33) though efforts to find evidence for IOR in visual search have had a checkered career (18-20; 65).

In parallel models, within-trial 'memory' generally takes the form of a local accumulation of evidence over the course of a trial (in the manner of 35). Thus, in a search for a T among Ls, information about the T-ness or L-ness of each item would accumulate over time until one item was confirmed as a T or all items were confirmed as Ls. Our recent data violate the predictions of this core assumption about the deployment of attention.

4.2. The Experiments of Horowitz and Wolfe (17)

To test the hypothesis that information accumulates during the course of a visual search trial, we compared a fairly standard search with a condition designed to minimize the accumulation of information. In the first experiment, the task was a standard T among Ls search. Both Ts and Ls could appear, randomly, in any of four orientations: 0, 90, 180 and 270 deg. As usual, the subject's task was to report as quickly as possible whether or not the target letter was present in the display. Targets were present on 50% of trials. The set sizes were 8, 12, or 16. Letters subtended 1 deg at the 57 cm viewing distance.

There were two stimulus conditions in the experiments: Dynamic and Static. The Static condition was a variation on a standard visual search experiment. The stimulus presentation

consisted of 20 cycles of an 83 msec presentation of the search display and a 24 msec mask composed of all of the line segments that could go into the Ts and Ls. The total stimulus duration, therefore, was 2220 msec..
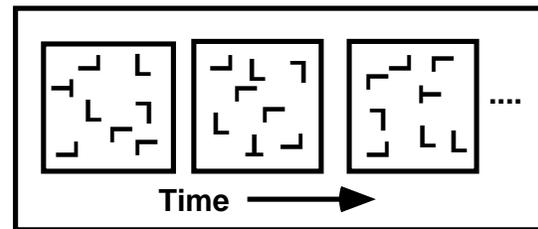

Figure Eight: The dynamic search condition. The same elements are plotted in each frame but their positions are changed randomly.

In the Dynamic conditions (shown above), the stimuli were randomly relocated every 111 msec. This did not involve any sort of coherent motion of stimuli. In this version of the experiment, a Dynamic trial consisted of five cycles of four independent frames of 83 msec duration with the 24 msec masks in between. Suppose that the trial was a target present trial with a "T" and eleven "L"s. Each of the four frames would present those twelve items in new random positions. If necessary, Ss could respond after the 2220 msec stimulus display. In practice, RTs of this length accounted for less than 2% of the data.

The Dynamic condition was intended to make any marking of rejected distractors irrelevant. If search involves serial selection of items, then the Dynamic condition should force selection *with replacement* from the set of items on the screen (That is, a given distractor might be checked more than once). The standard serial view of the Static condition has been that it involves selection *without replacement* (A given item would <u>not</u> be checked more than once.). In a standard serial, self-terminating search, the observer must sample an average of half of the items on target-present trials. Modeling shows that the average number of samples in the Dynamic case equals the set size. This does not mean that each item in the display is sampled. In sampling with replacement, some items may be sampled multiple times. It follows that Dynamic target present slopes should be twice as steep as the Static target present slopes, if there is marking of rejected distractors in the Static condition.

A second version of this experiment was run without the masks. In this case, the Static condition is truly static. Nine subjects were tested for 200 trials in each condition, randomly distributed over 3 set sizes.

Figure Nine shows the RT and errors as a function of set size for Experiment One. Results for Exp. 2 are comparable. The slopes for the Dynamic condition were not twice the slopes of from the static condition - falsifying the prediction of the standard serial model. Target-present slopes in static and dynamic conditions did not differ significantly in either version of the experiment. (Exp. 1: t(8)=.13, p <.50, Exp. 2: t(8)= 1.52, p>.15). Note in Figure Nine that target-absent slopes are actually *shallower* for the Dynamic case than for the Static case. While the Dynamic mean RTs do appear to be longer than the Static, that RT cost is reliable only in Experiment 2 (F(1,8)=18.81, p < .005). We suspect that the increased mean RTs reflect subjects' decreased confidence in their responses. Consider a subject who *believes* she has found a target. In the Static case, the physical stimulus is still

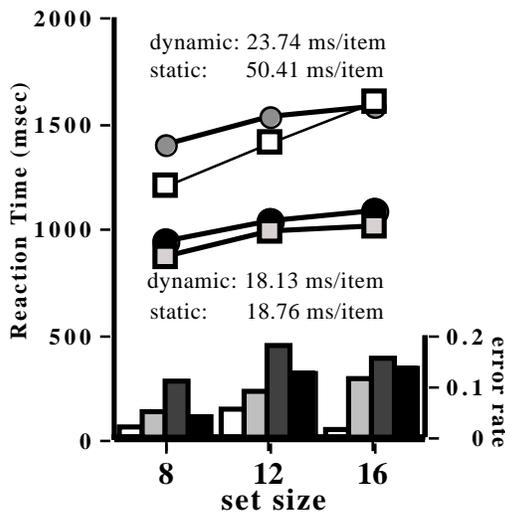available for confirmation, while in the Dynamic case, it is not.

Figure Nine: Mean RT data for dynamic and static conditions of the first experiment (with masks). Upper curves are target absent. Lower are target present. Note that dynamic, target present slopes are very similar to static slopes. Bars give error rates in the following order: Static false alarms, static misses, dynamic false alarms, dynamic misses.

These results would be uninteresting if subjects, in the dynamic condition, could direct attention to one location and simply wait for the target to appear in that location. However, the position of the target was constrained in order to thwart any such strategy. In Experiment One, the target only appeared at one of four locations (one in each of the four independent frames). Here a "sit and wait" strategy would lead to failure on 93.75% of target present trials. In Experiment 2, the target changed location on every trial but remained at one of four eccentricities (again, chosen at random from trial to trial). In this case, a "sit and wait" strategy would fail on 75% of trials.

These data would have been a fairly straight-forward, if surprising, refutation of the predictions of the standard accounts of the marking of rejected distractors were it not for the error rates. Subjects make more errors in the Dynamic condition than in the Static condition. This is not surprising. Stimuli are more degraded in the Dynamic condition and, as noted in connection with the RT difference, subjects can continue to attend to a location and confirm the existence of a target in the Static condition but not in the Dynamic condition. That said, the error rates complicate the analysis of the result because of the likelihood of a speed-accuracy tradeoff. Given the more frequent errors in the Dynamic case and given the increase in those errors with set size, we must assume that the slopes in the Dynamic case are underestimates of the "true" slope. Could that "true" Dynamic slope be twice the "true" Static slope and, thus, consistent with marking of rejected distractors in the Static condition? In an effort to answer this question, we conducted a replication of the experiment with a design intended to reduce the error rates.

4.3. Experiment Three: Another Version

In this third experiment, we eliminated the option to respond "no" by having subjects respond to target identity, rather than target presence. A target letter "E" or "N" was present on each trial, embedded in distractors selected from the remaining letters of the alphabet (except for "I" and "J"). Subjects identified the target letter. Again, we compared Static and Dynamic conditions. Methods were similar to the experiments described above. Since subjects would always know that a target was present, we reasoned that they would be less inclined to abandon a difficult search with a guess. This should lower errors.

Our results showed that, once again, the slopes were statistically indistinguishable with the Dynamic slope of 29.5 ms/item being slightly shallower than the Static slope of 34.67 ms/item. The effort to reduce errors worked. Error rates were substantially lower in this experiment (5.6% overall for the Dynamic condition, 2.8% for the Static). Nevertheless, there are still twice as many errors in the Dynamic condition. Is this difference sufficient to mask a true 2:1 relationship between Dynamic and Static slopes? The point is arguable but we think that it is implausible to propose that a relatively few errors could, in effect, cut the Dynamic slope in half. It is possible, for example, to calculate the missing RTs that would be needed to double the Dynamic slope. The details of this error correction analysis are given on our website (search.bwh.harvard.edu). In brief, in order to double the Dynamic slope, one would need to assume that all errors come from trials where the reaction time should have been much longer than almost any of the correct RTs in the actual data. As a different approach, we can look at the results only for the subjects with the smallest differences between Dynamic and Static error rates. In this subset of the data, we still find that Dynamic and Static slopes are essentially the same.

4.4. Memory-free search?

How should these results be interpreted? Recall the predictions of the standard, serial, self-terminating search model. If we assume that rejected distractors are marked in the Static case and that they cannot be marked in the Dynamic case, then the target present slopes in the Dynamic case should be twice those in the Static case. The experiments yield Static and Dynamic slopes that are indistinguishable from each other. These data appear to falsify the hypothesis that rejected distractors are marked in the Static condition and not in the Dynamic condition. Given the distractors *could not* be marked in the Dynamic condition, it would seem to follow that they were not marked in the Static condition either. That is, it would appear that items are sampled from the display *with replacement* in both the Dynamic and Static cases. We have dubbed this the *memory-free search* hypothesis.

The memory-free hypothesis only applies to covert deployments of attention and not, for example, to overt eye movements. It is possible that previously fixated locations are marked in visual search (19). Covert attention and overt eye movements are usually linked (e.g. 21). Attention can be deployed at a faster rate than can the eyes. Nevertheless, some memory for prior fixation might be all the memory needed in real-world visual search. It is also important to note that the memory-free hypothesis proposes a lack of memory for rejected distractors. It does not propose a lack of memory for accepted targets. Targets must be remembered, once they are found, otherwise it would be impossible to perform repeated

searches through the same display (e.g. Where are those two kids of mine?). The act of rejecting a distractor is different than the act of accepting a target. Perhaps it is the act of coding targets into memory that produces the attentional blink (8; 36; 37).

4.5. Examining the effect of trial length

Beyond simple speed-accuracy trade-offs discussed above, there is another way for Dynamic and Static conditions to produces the same target present slopes even if search is memory-free in the Dynamic and memory-based in the Static condition. Alex Backer (personal communication) noted that the theoretical distribution of RTs is uniform and finite in the memory-based case while it has an exponential, potentially infinite upper tale. That is, suppose that a display contains ten items. In an accurate memory-based search, the observer never searches through more than ten items. In an memory-free search, however, the subject could search forever. Very long searches will be very rare, but they should occur in theory.

In practice, long RTs are less likely. After a certain point, observers will tend to give up and guess. Of more specific relevance to these experiments, Backer noted that we used 20 frames of 100 msec each. If subjects did not find a target during the 2000 msec of stimulus exposure, they would have to guess. As a consequence, RTs that would have been significantly longer than 2000 msec would have been removed from the RT distribution. Under one set of assumptions, it happens that the loss of these long RTs would be enough to reduce the theoretical slope of a memory-free Dynamic search to the slope of a hypothetical, memory-based Static search.

More generally, Backer's analysis predicts that slopes in the Dynamic condition should be strongly influenced by the duration of the stimulus display. Slopes in the Static condition are only influenced at short display durations. As a consequence, this analysis predicts that Dynamic search slopes will be shallower than Static at short durations and longer at long durations with a fairly narrow range of durations producing roughly equal slopes in the two conditions.
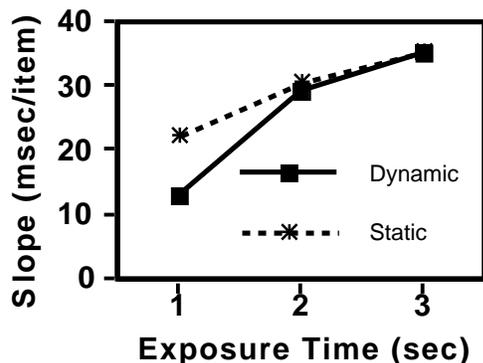

Figure Ten: Slope as a function of exposure duration of Dynamic and Static search displays. Note that the slopes converge as the duration gets longer.

In order to assess the possibility that we had inadvertently stumbled on the point of equality, we tested subjects at display durations of 1, 2, and 3 seconds. The task was the "E or N?"

task described above. Methods were similar to those described for that experiment.

Figure Ten shows the results of this experiment. At the shortest duration, in partial support of Backer's hypothesis, the slopes for the Dynamic case are somewhat shallower than the slopes for the Static case. The effect is smaller than predicted but is in the predicted direction. Recall, however, that Backer's hypothesis predicts that the slopes for the Dynamic case should rise quite dramatically. In fact, as the duration gets longer, the slopes for the Static and Dynamic conditions appear to converge. There is no evidence that Dynamic slopes rise to twice the Static slopes even when the stimulus is presented for 3 seconds.

4.6. Implications of Memory-free Search

The title of this paper refers to "two surprises". The possibility of memory-less search is the first of these surprises. Before turning to the second, it is worth considering some of the implications of memory-less search for our understanding of the deployment of attention.

1) At the most basic level, memory-less changes our view of the deployment of attention. We had thought it was relatively orderly. Perhaps order is expensive and perhaps reality is more anarchic, based on a simple, rapid strategy that avoids the overhead of tagging checked locations.

2) If rejected distractors are entirely unmarked, models like Guided Search would develop a problem with perseveration. Attentional deployment is biased toward the fovea (5; 64). The standard account allows attention to work its way toward a peripheral target by rejecting and marking more central distractors and then moving outward. If there is no such marking, why doesn't attention get stuck at the fovea or on the brightest or the most salient stimulus? One possibility is that there is some limited memory, perhaps a memory for the positions of the last one or two distractors. It is unclear that limited memory of this sort would have been detected in the experiments reported here. Incomplete memory has been suggested in other search contexts (e.g. 1).

3) The rate of attentional deployment in the standard models is estimated by doubling the target present slope. Thus, the standard slopes of 20-30 msec for inefficient search, implies a rate of one item every 40-60 msec. If search is memory-free, the rate is estimated directly from the target present slope, making it twice as rapid. There are investigators who have theoretical and empirical difficulty with serial selection at a rate of 40-60 msec/item because they think that attentional deployment requires several much slower steps (e.g. 12; 53). A rate of 20-30 msec/item would be even more challenging.

4) Parallel models of attention would also be disturbed by this memory-free finding. In a standard parallel model, information accumulates at each location about the likelihood of target presence. The Dynamic condition renders this accumulation function, if it were available, irrelevant. How then is it possible to search with the same efficiency in Dynamic and Static cases? These results would seem to require a parallel model that analyzes multiple, independent snapshots of the search display.

5. POST-ATTENTIVE VISION:
   THE SECOND SURPRISE

5.1. The Roles of Selective Attention in Object Recognition.

Earlier in this paper, it was asserted that deployment of attention to an object is a prerequisite to the recognition of that object. Why should that be the case? Selective attention serves two roles in the perception of objects. First, attention is required for the proper binding of features in objects. Prior to the arrival of attention, features of an object are not well bound to each other (47). As an illustration, see Figure Eleven.
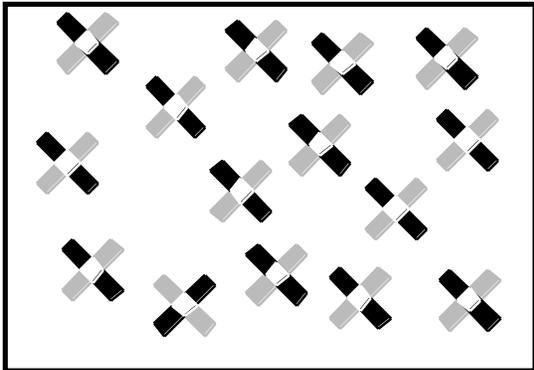

Figure Eleven: Find the black line, tilted to the right.

This is a conjunction search, logically similar to the color X orientation search shown in Figure 4. In that case, guidance from preattentive color and orientation information could lead to efficient search. Here, however, guidance fails because each "X" is treated as an object with the features "black" and "gray" and "left" and "right". Prior to the arrival of attention, the relationship of features to each other within an object is unclear. The features are "bundled" with the object but they are not "bound" (59).

In its second role in object recognition, attention controls traffic through a tight bottleneck between the visual representation of an object and its representation in memory. Recognition of a visual object requires three things. First, there must be a visual object to see and recognize. Second, there must be a representation of that object in memory. Otherwise, the observer cannot know the identity of the object. the observer would be agnosic. Finally, there must be a link between the visual and memorial representations. This notion of a link is critical. An observer might be seeing a cow and thinking of a car. We would not want this observer to 'recognize' the cow as a car. Hence, it is not enough for the two representations to coexist in time. They must be linked.

5.2. Post-attentive vision and Repeated Search

We have found that the number of links that can be maintained at any one time is very small - perhaps as small as one. The prime evidence for this conclusion comes from experiments using a "Repeated Search" paradigm in which observers search multiple times through the same set of stimuli. This is illustrated in Figure Twelve. The capital letters remain present throughout a series of N repeated searches. They do not flicker. They are not masked in anyway. Only the letter at the center changes, indicating the target for the current

search. Thus, in Figure Twelve, the observer searches first for the letter 'f', next for a 'b', and so on.
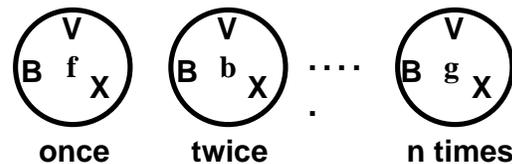

Figure Twelve: The Repeated Search paradigm. Observers, search over and over through the same, unchanging display. In this case, the display is the letters "B", "V", and "X".

We know from prior experience that the first search through these letters will be inefficient. It appears that observers must search from item to item until they find the target or, in the example shown here, until they are convinced that an "f" is not present. Search is inefficient because each letter is recognized only when attention is directed to it.

The critical question in Repeated Search concerns the fate of the effects of attention on an object after attention has been directed elsewhere. If attention allows the binding of features and the linking of visual to memorial representations, does that binding and linking survive when attention departs? The Repeated Search paradigm provides a way to answer this question. If binding and linking survive, then multiple links will be built connecting vision and memory. Eventually, all items in the display will be recognized at the same time. If the observer is then asked about an element in the display, that request will activate the node in memory. That node in memory will be linked to the visual stimulus and the observer should be able to respond, "yes", without a search. That is, RT should no longer depend on set size because the other items in the display should be irrelevant. If, on the other hand, links do not accumulate, then an inefficient search will be required each time a new target probe is presented.

5.2.1.   Methods

We have performed repeated search experiments with a wide range of stimuli including letters (as shown in Fig. 12), novel objects, and 'real' objects. Details can be found in Wolfe et al. (63). Here, we will illustrate the basic result with an experiment that used conjunction stimuli of the sort shown in Figure 13.
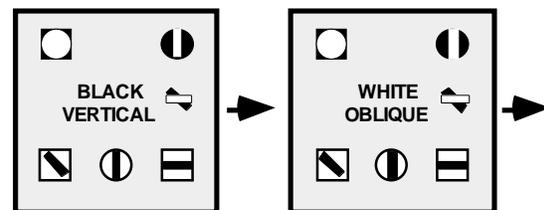

Figure Thirteen: A Repeated Search task. Observers look for the target defined by the words at the center of the display. The surrounding search array does not change.

The actual stimuli were conjunctions of color and form/orientation. Conjunction search of this sort, with variable targets and many types of distractors, is inefficient - at least on the first trial. In this experiment, observer's searched through the same display five times. One hundred sets of five trials were run at each of two set sizes, allowing us to compute slopes of the RT x set size function for each repetition.

In addition to the Repeated Search condition, an Unrepeated Search condition was run. In this case, the items changed on each trial. This condition provides a baseline for comparison. No links can be built up over repetitions in this case because no stimuli are repeated.

5.2.2.   Results

Figure Fourteen shows the results for this experiment. The upper panel shows mean RTs as a function of repetition. The lower panel shows the slopes of the RT x set size functions.
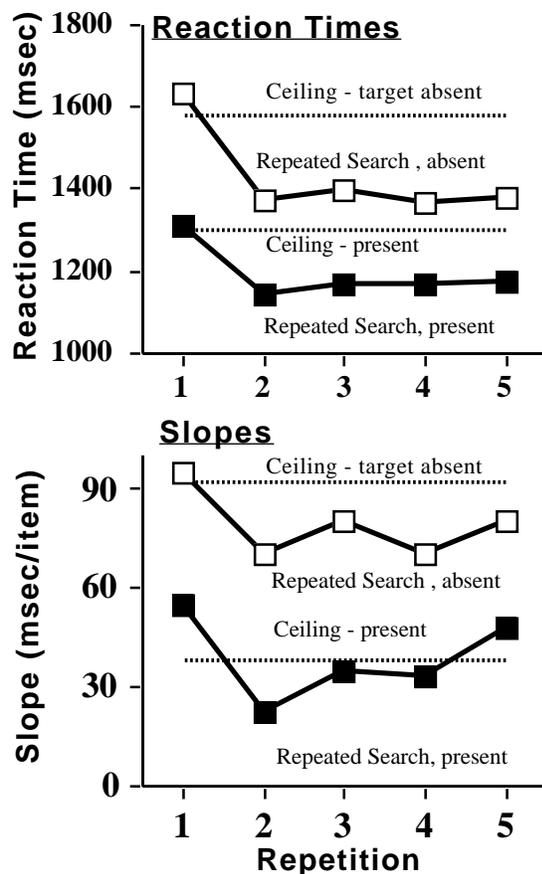

Figure Fourteen: RT and slope results of Repeated Search for conjunction targets compared to Unrepeated search for the same type of targets.

Note that, for the control "Ceiling" condition, repetition is meaningless. The stimuli are new on each trial. Accordingly, the single mean RT and slope values are plotted as straight lines, constant across the repetition variable that is of interest in the Repeated Search condition. In Repeated Search, there is some apparent improvement in the RTs from the first to the second repetition of the stimuli. However, other experiments in this series show that to be an effect of the masking of the probe words at the center by the surrounding visual stimuli (63). That masking is present on all trials in the Ceiling condition and, accordingly, the Ceiling RTs are very similar to the repetition one, Repeated Search RTs.

Turning to the slopes, we again see a hint of an improvement, mostly on the target absent trials. However, there are two important points to be made. First, even after any improvement, the search remains very inefficient. There is no hint that repeated search through these stimuli has produced the efficient search predicted if multiple items are simultaneously recognized - simultaneously linking their visual and memorial representations. Second, the target present slopes are essentially the same in the Repeated Search and Ceiling conditions, indicating that repeated search through the stimulus did not lead to the development of any representation that could facilitate search.

5.2.3.   Discussion

We have repeated this basic finding with letters and objects; always obtaining the same general pattern of results. If search is inefficient on first exposure to a stimulus, it remains inefficient after repeated searches through that stimulus. In many cases, there is no significant change in the slope of RT x set size functions or in error rates. (Wolfe, et al., 1999). Concerned that five repetitions might be too few, we had subjects search 350(!) times through the same sets of three or five letters. Even in this extreme case, search efficiency did not improve in the Repeated Search condition.

6.   GENERAL DISCUSSION

6.1. The Role of Memory in Visual Search

Two findings have been highlighted in this paper. First, the results from the Dynamic Search experiments indicate that rejected distractors are not marked during the course of a visual search. Second, the work with Repeated Search shows that search for ever changing targets does not become more efficient with repeated search through the same display. This can sound like some sort of 'attentional stupidity' or like a denial of any role for memory in visual search. Such a position would be not only counter-intuitive but wrong. Starting with the dynamics of a single search, while subjects may not keep track of rejected distractors, they must keep track of accepted targets. That, after all, is the purpose of the search. Brad Gibson and his colleagues (personal communication) have illustrated this point in a simple extension of our work. They had subjects discriminate between displays containing one or two targets. The displays could be either static or dynamic. The static case was easy. The dynamic case was virtually impossible. In the static case, subjects could find and retain the first target and then proceed to search for the second. In the dynamic case, this was impossible (given that the targets were identical. With two different targets, the results would be different.) Our claim that "visual search has no memory" is a claim of amnesia for the course of the search, not for its consequences.

The Repeated Search, post-attentive experiments are open to similar misinterpretation. It would be foolish to deny that subjects learn and remember something about the displays in repeated search tasks. After multiple searches through one display, the contents of that display are committed, at least, to some short term memory. Indeed, we compared performance on the Repeated Search tasks to performance on memory search tasks. For example, in the experiment where subjects searched through the same letters 350 times, we also included a memory search condition in which they committed letters to memory and then searched that memory 350 times. Efficiency (slope) and RT were actually somewhat faster in the absence of the visual stimulus though errors were somewhat higher.

The conclusion is that the presence of the visual stimulus conveys no benefit in Repeated Search performance.

As in the Dynamic Search experiments, this does *not* mean that subjects do not learn the locations of targets. Once you learn that bathroom is around the corner to the left, you do not have to search randomly for it. In a search paradigm, Chun and Jiang (7) have shown that subjects can learn the layout of meaningless search displays if they are repeated. That learning seems to be implicit. That is, they behave as if they remember the displays, even in the absence of any ability to explicitly recognize them among novel displays.

Our results do not deny an ability to remember displays. They merely show that the physical presence of the display does not allow a short-cut around the limited-capacity of search through that memory.

6.2. Why does visual search have no memory? Implications for artificial search mechanisms.

If one were building a search device from the ground up, one might think that it should be constructed with characteristics other than those described here for human visual search. Why not build a search mechanism that marked rejected distractors and, thus, gained efficiency over a mechanism that did not? Why not build a visual system that could have multiple links between visual and memorial representations of objects? Of course, answers to such questions are speculative. However, when Nature picks an apparently inferior way to perform a task, we may guess that the superior method was too expensive.

In the case of the marking of rejected distractors, we know that there are mechanisms of inhibition that serve to keep attention away from previously attended items(20). The most prominent of these is "inhibition of return" (IOR - 33; 39). Another apparently different mechanism has been dubbed "visual marking" (41; 54). Why should visual search use these mechanisms to avoid resampling of rejected distractors? In this case, the cost may be time. Distractors in visual search are being rejected at a rate of about 30-50 Hz (20 - 30 msec/item). These inhibitory mechanisms seem to require an order of magnitude more time (e.g. 22). By the time that this sort of inhibition could be applied, search might well be over.

Interestingly, the time course of inhibition is similar to the time of saccadic eye movements (3-4Hz). Klein and MacInnes (19) have new evidence that IOR might aid search, not by marking rejected distractors but by preventing the eyes from returning to previously fixated locations. One can imagine covert deployments of attention working cooperatively with slower, overt movements of the eyes. The eyes go to a location. Attention randomly samples 6 - 10 objects, probably in the neighborhood of fixation. This sampling is done without marking distractors but when the eyes move again, IOR prevents the same location from being the target of another eye movement. In longer searches, this could act to limit the amount of resampling of rejected distractors.

The cost of multiple links between vision and memory seems qualitatively different. It may be very hard to prevent 'cross talk' if multiple links are present. If the scene contains a car and cow and memory contains representations of a car and a cow, it is important not to attempt to drive the cow or milk the car. Selective attention may be the price we pay for accurate recognition. Kevin O'Regan (30) suggests that we can afford to pay this cost because the world serves as its own memory.

Ignoring the odd case of laboratory displays with randomly changing items, the world is a fairly stable place. A cow and a car, if present at one instant, are likely to be present at the next. Even if they move, they move on trajectories that are predictable in the short term. Thus, rather than simultaneously recognizing multiple objects, we can maintain a single link from vision to memory, secure in the knowledge that we can use visual search to quickly reacquire an object if we need it. At 30-50 objects/sec we can afford to do a lot of selection.

7. CONCLUSION

There may be other ways to build a search mechanism. Perhaps slow deployment of attention would work if combined with an ability to simultaneously recognize multiple objects. However, humans and, we presume, other animals have done well with a fast but sloppy selection mechanism and a narrow channel between vision and memory.

1 Arani, T., Karwan, M. H., & Drury, C. G. (1984). A variable-memory model of visual search. *Human Factors, 26*(6), 631-639.

2 Bauer, B., Jolicœur, P., & Cowan, W. B. (1996). Visual search for colour targets that are or are not linearly-separable from distractors. *Vision Research, 36*(10), 1439-1466.

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

4 Bilsky, A. A., & Wolfe, J. M. (1995). Part-whole information is useful in size X size but not in orientation X orientation conjunction searches. *Perception and Psychophysics, 57*(6), 749-760.

5 Carrasco, M., & Frieder, K. S. (1997). Cortical mgnification neutralizes the eccentricity effect in visual search. *Vision Research, 37*(1), 63-82.

6 Cave, K. R., & Wolfe, J. M. (1990). Modeling the role of parallel processing in visual search. *Cognitive Psychology, 22*, 225-271.

7 Chun, M., & Jiang, Y. (1998). Contextual cuing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology, 36*, 28-71.

8 Chun, M. M., & Potter, M. C. (1995). A two-stage model for multiple target detection in RSVP. *Journal of Experimental Psychology : Human Perception & Performance, 21*(1), 109-127.

9 Cohen, A., & Ivry, R. B. (1991). Density effects in conjunction search: Evidence for coarse location mechanism of feature integration. *J. Exp. Psychol. Human Perception and Performance, 17*(4), 891-901.

10 D'Zmura, M. (1991). Color in visual search. *Vision Research, 31*(6), 951-966.

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

12 Duncan, J., Ward, R., & Shapiro, K. (1994). Direct measurement of attention dwell time in human vision. *Nature, 369*(26 May), 313-314.

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

14 Enns, J. T., & Rensink, R. A. (1990). Scene based properties influence visual search. *Science, 247*, 721-723.

15 Grossberg, S., Mingolla, E., & Ross, W. D. (1994). A neural theory of attentive visual search: Interactions of

boundary, surface, spatialand object representations. *Psychological Review, 101*(3), 470-489.

16 Harris, J. R., Shaw, M. L., & Bates, M. (1979). Visual search in multicharacter arrays with and without gaps. *Perception and Psychophysics, 26*(1), 69-84.

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

18 Klein, R. (1988). Inhibitory tagging system facilitates visual search. *Nature, 334*, 430-431.

19 Klein, R. M., & MacInnes, W. J. (1999). Inhibition of return is a foraging facilitator in visual search. *Psychological Science, in press*(July).

20 Klein, R. M., & Taylor, T. L. (1994). Categories of cognitive inhibition with reference to attention. In D. Dagenbach & T. H. Carr (Eds.), *Inhibitory processes in attention, memory, and language.* . New York: Academic Press.

21 Kowler, E., Anderson, E., Dosher, B., & Blaser, E. (1995). The role of attention in the programming of saccades. *Vision Research, 35*(13), 1897-1916.

22 Mackeben, M., & Nakayama, K. (1987). Sustained and transient aspects of extra- foveal visual attention. *Invest. Ophthalmol. Vis. Sci. (suppl), 28*, 361.

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

24 McLeod, P., Driver, J., Dienes, Z., & Crisp, J. (1991). Filtering by movement in visual search. *J. Experimental Psychology - Human Perception and Performance, 17*(1), 55-64.

25 Moore, C. M., & Wolfe, J. M. (2000). Getting beyond the serial/parallel debate in visual search: A hybrid approach. In K. Shapiro (Ed.), *The Limits of Attention: Temporal Constraints on Human Information Processing* . Oxford: Oxford U. Press.

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

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

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

29 Nothdurft, H. C. (1993). Faces and facial expression do not pop-out. *Perception, 22*, 1287-1298.

30 O'Regan, K. (1992). Solving the 'real' mysteries of visual perception. The world as an outside memory. *Canadian J. of Psychology, 46*, 461-488.

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

32 Posner, M. I. (1980). Orienting of attention. *Quart. J. Exp. Psychol., 32*, 3-25.

33 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.

34 Quinlan, P. T., & Humphreys, G. W. (1987). Visual search for targets defined by combinations of color, shape, and size: An examination of the task constraints on feature and conjunction searches. *Perception and Psychophysics, 41*, 455- 472.

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

36 Raymond, J. E., Shapiro, K. L., & Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task: An attentional blink? *J. Experimental Psychology: Human Perception and Performance, 18*(3), 849-860.

37 Shapiro, K. L. (1994). The attentional blink: The brain's eyeblink. *Current Directions in Psychological Science, 3*(3), 86-89.

38 Suzuki, S., & Cavanagh, P. (1995). Facial organization blocks access to low-level features: An object inferiority effect. *Journal of Experimental Psychology: Human Perception and Performance, 21*(4), 901-913.

39 Taylor, T. L., & Klein, R. M. (1999). On the causes and effects of inhibition of return. *Psychonomics Bulletin and Review, 5*(4), 625-643.

40 Theeuwes, J., & Kooi, J. L. (1994). Parallel search for a conjunction of shape and contrast polarity. *Vision Research, 34*(22), 3013-3016.

41 Theeuwes, J., Kramer, A. F., & Atchley, P. (1998). Visual marking of old objects. *Psychonomic Bulletin and Review, 5*(1), 130-134.

42 Tipper, S. P., Weaver, B., & Watson, F. L. (1996). Inhibition of return to successively cued spatial locations: Commentary on Pratt and Abrams (1995). *Journal of Experimental Psychology: Human Perception & Performance, 22*(5), 1289-1293.

43 Townsend, J. T. (1971). A note on the identification of parallel and serial processes. *Perception and Psychophysics, 10*, 161-163.

44 Townsend, J. T. (1976). Serial and within-stage independent parallel model equivalence on the minimum completion time. *J. Mathematical Psychology, 14*, 219-239.

45 Townsend, J. T. (1990). Serial and parallel processing: Sometimes they look like Tweedledum and Tweedledee but they can (and should) be distinguished. *Psychological Science, 1*, 46-54.

46 Treisman, A. (1993). The perception of features and objects. In A. Baddeley & L. Weiskrantz (Eds.), *Attention: Selection, awareness, and control.* (pp. 5-35). Oxford: Clarendon Press.

47 Treisman, A. (1996). The binding problem. *Current Opinion in Neurobiology, 6*, 171-178.

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

49 Treisman, A., & Sato, S. (1990). Conjunction search revisited. *J. Exp. Psychol: Human Perception and Performance, 16*(3), 459-478.

50 Treisman, A., & Souther, J. (1985). Search asymmetry: A diagnostic for preattentive processing of seperable features. *J. Exp. Psychol. - General, 114*, 285-310.

51 Tsotsos, J. K., Culhane, S. N., Wai, W. Y. K., Lai, Y., Davis, N., & Nuflo, F. (1995). Modeling visual attention via selective tuning. *Artificial Intelligence, 78*, 507-545.

52 von der Heydt, R., & Dursteler, M. R. (1993). Visual search: Monkeys detect conjunctions as fast as features. *Investigative Ophthalmology and Visual Science, 34*(4), 1288.

53 Ward, R., Duncan, J., & Shapiro, K. (1996). The slow time-course of visual attention. *Cognitive Psychology, 30*(1), 79-109.

54 Watson, D. G., & Humphreys, G. W. (1997). Visual marking: Prioritizing selection for new objects by top-down attentional inhibition of old objects. *Psychological Review, 104*(1), 90-122.

55 Wolfe, J. M. (1992). "Effortless" texture segmentation and "parallel" visual search are *not* the same thing. *Vision Research, 32*(4), 757-763.

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

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

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

59 Wolfe, J. M., & Bennett, S. C. (1997). Preattentive Object Files: Shapeless bundles of basic features. *Vision Research, 37*(1), 25-44.

60 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.

61 Wolfe, J. M., & Franzel, S. L. (1988). Binocularity and visual search. *Perception and Psychophysics, 44*, 81-93.

62 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.

63 Wolfe, J. M., Klempen, N., & Dahlen, K. (1999). Post-attentive vision. *Experimental Psychology: Human Perception and Performance, in press*.

64 Wolfe, J. M., O'Neill, P. E., & Bennett, S. C. (1998). Why are there eccentricity effects in visual search? *Perception and Psychophysics, 60*(1), 140-156.

65 Wolfe, J. M., & Pokorny, C. W. (1990). Inhibitory tagging in visual search: A failure to replicate. *Perception and Psychophysics, 48*, 357-362.

66 Zelinsky, G. J., & Sheinberg, D. L. (1997). Eye movements during parallel / serial visual search. *J. Experimental Psychology: Human Perception and Performance, 23*(1), 244-262.

67 Zohary, E., & Hochstein, S. (1989). How serial is serial processing in vision? *Perception, 18*, 191-200.