Mechanisms of Visual Search

by

Seth Andrew Herd

B.A., Earlham College 1997

M.S., University of Colorado, 2001

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Department of Psychology and Center for Neuroscience

2005

UMI Number: 3190399

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

This thesis entitled:
Mechanisms of Visual Search
written by Seth Andrew Herd
has been approved for the Department of Psychology and Center for Neuroscience

Randall C. O'Reilly

Yuko Munakata

Michael Mozer

Date

The final copy of this thesis has been examined by the signatories, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.

Herd, Seth Andrew (Ph.D., Psychology and Neuroscience)

Mechanisms of Visual Search

Thesis directed by Dr. Randall C. O'Reilly

This dissertation addresses the strategies employed for visual search tasks, and the mechanisms underlying them. The empirical portion of the work addresses the hypothesis that many search tasks are performed by inspecting a display by areas, and moving attention among these areas when there are too many items to be processed at once. The experiments investigate the previously unaddressed question of whether the number of items inspected in parallel varies with the difficulty of the task, or whether a fixed number of items are searched at once, due to central processing limitations. The results indicate that the number of items inspected with each attentional fixation does vary with the difficulty of the task. However, this variance appears to depend on individual strategies, rather than being uniform among observers. In addition, there is evidence that the number of items searched in parallel depends on display size as well as stimulus types, with larger displays being searched relatively more efficiently. The second portion of the work seeks to develop a comprehensive theory of visual search, integrating existing theories that address particular types of search. The conclusion of a review of relevant literature is that there are several distinct strategies that are employed for particular types of searches. These strategies include emphasizing a subset of the items using top-down featural attention, searching in parallel by using a broad spatial attention, and searching using random or nearly random eye movements. The conclusions drawn from the broad literature review is partially consistent with current theories based on behavioral data, but differs in that the number of items searched at once is not definite, but rather the chance of locating a target falls off smoothly with distance from the current center of gaze, and with the spacing of stimuli. The last portion of the work seeks to identify the neural mechanisms that give rise to these strategies of search and their successes and failures. A previously unrecognized link is drawn between search patterns

and the size of receptive fields of neurons responsive to target/distractor differences.

Acknowledgements

This dissertation would have not been possible without the support of many people over the course of my time as a graduate student. First and foremost, I want to thank my advisor Dr. Randall O'Reilly for being both a good mentor and a good friend. Randy has given me the freedom to pursue my own interests, and the guidance to ensure that that pursuit was productive. Other faculty members here have also provided valuable advice and served as role models in various respects. In particular, Yuko Munakata has served as an example of combining enthusiasm for big questions with careful work on and thought about the small pieces that make them up. All of my friends have offered valuable support, but a few also gave specific help on this dissertation. Erika Nyhus provided support, advice, proofreading and editing, time as an observer for the experiments, assistance with statistics, and more support. Beth Mulligan generously gave invaluable help in coding the statistical tests for Experiment 1. Erica Wohldman, Richard Busby, and James Kohl read and commented on previous drafts, and served as observers for the experiments. My thanks go to everyone who made this project and the wonderful learning experience I've had at this school possible.

Contents

Chapter

List of Tables

Table

List of Figures

Figure

Chapter 1

Introduction

This dissertation addresses the question of how visual search (VS) works. The topic of visual search is special in that so much evidence is available; it has been explored thoroughly by both behavioral means, and using neuroscientific approaches that provide a great deal of evidence on the brain mechanisms that underlie it. The project undertaken here is to use the constraints provided by this wealth of evidence to arrive at a more complete answer to the question than any offered by existing theories. Despite the wealth of evidence available, no existing theory seeks to provide a general explanation of visual search. The current project seeks to provide such a general framework based on known brain mechanisms.

The first chapter addresses the question by empirical means. The current hypothesis of serial-parallel search is tested by several different experiments. These experiments seek to further test the theory, and refine it in ways that existing experiments have not. The serial-parallel search hypothesis (Pashler 1987; Wolfe, Michod, and Horowitz, submitted) proposes that visual search proceeds by identifying a limited number of items in parallel, then moving to another group of items if the target is not found.

The experiments seek to test this theory using converging behavioral methods. In addition, novel statistical means are used to test whether the number of items that are identified varies with the type of targets and distractors that are the subject of search. The results of these experiments provide mixed support for the theory. It is found that the number of items processed in parallel does seem to vary for different searches. However, some of the experiments

do not match the predictions of the theory. Based on the experimental evidence, I conclude that this theory is an incomplete description of search, even within similar types of searches.

In the next three chapters I take a broad view of the question, and bring to bear a great deal of existing evidence. The second chapter reviews existing theories of search, and behavioral evidence. This evidence reveals existing theories are all inadequate to explain the full spectrum of visual search tasks, and that substantially different mechanisms are employed in different situations. The third chapter reviews converging evidence from data about brain function, and finds that the known mechanisms of brain function help explain why visual search has particular successes and limitations in different situations.

The fourth chapter explicates a theory intended to bring together the behavioral and neuroscientific evidence into an explanation of visual search that is more comprehensive than existing theories. It seeks to reconcile different theories into one, identifying theories that are not needed to explain the evidence, and establishing the relationship between the different theories that are necessary parts of a complete understanding. Two neural network models are detailed, each of which provides evidence that one of the emergent principles of brain function on which the theory is based hold true, at least for particular artificial systems based on known brain function.

Because of the broad focus of this work, there are some inevitable limitations. Some of the evidence on which the theory is based is not reviewed extensively here, but rather accepted based on previous analysis. An effort is made to do this only where the conclusions are relatively uncontroversial within the group of experts on that topic, as revealed by a lack of dissenting opinions in the literature. Some more subtle effects within visual search findings are not explained completely, although an effort is made to make the theory presented here compatible with possible explanations. The theory is not instantiated either mathematically, or as a simulation. Therefore it does not offer quantitative predictions. The theory is not designed to be easily falsifiable, but an attempt is made to identify which aspects of it require further testing,

and which are relatively certain given existing evidence.

Because the subject matter is rather limited, it is worth questioning the worth of this work in relation to the larger projects of cognitive psychology and neuroscience. A complete understanding of visual search has practical as well as theoretical value. The practical value of understanding visual search lies in improving the searches that people engage in every day, from search for links on a web page, to searches for items in an airport X-ray display. The theoretical value of a more complete understanding of search includes a better understanding of the mechanisms of attention in the brain; the understanding arrived at here has implications for mechanisms of attention both across modalities, and for higher level attention to a particular task. In addition, the theory developed here is mechanistic, making it appropriate for guiding the construction of artificial systems for conducting searches. Such artificial systems also have both obvious practical value and theoretical value, in allowing more complete simulations of the visual system, that more fully incorporate realistic mechanisms of attention and representation.

1.1 Background

In order to situate the current work, it is necessary to introduce some existing work on and theories of visual search. The review proceeds roughly as a history of research on visual search.

In the typical visual search paradigm, a variable number of items are displayed at once on a computer monitor. The observer decides as quickly as possible whether a particular target object is among them, and indicates this decision with a key press. Typically a maximum of one target per display is used, the observer knows in advance precisely what this target looks like, and the target is present on 1/2 of the trials. All the results discussed below use such a paradigm unless otherwise stated.

1.1.1 Basic Results: Feature vs Conjunction Searches

Early visual search results seemed to paint a clear and informative picture. It seemed that targets defined by a single feature, such as a horizontal bar among vertical bars, or a red item among green items, were found quickly, and the speed did not seem to depend on the number of distractors in the display. Such searches were termed "parallel" searches since the fact that more items were searched just as quickly means they must all be processed at the same time, (or popout search, since the target's identity need not be known in advance, but rather the target seems to "pop out" of the background). Targets defined by a conjunction of features, (for instance, a red vertical bar among red horizontal and green vertical bars) however, were found more slowly when more distractors were present. Therefore a red X (conjunction of red color and X shape) would be found more quickly when presented among five green X's and five red O's than the same target among ten of each distractor.

The average amount of extra time per added distractor was referred to as the **search slope**. It was found that search slope when the target was absent was roughly twice that when the target was correctly judged as present (a 2:1 absent-present slope ratio-see figure 1.1). This is exactly what one would see if observers did search for conjunctions one at a time, and find the target, if one is present, after searching through half of the distractors on average. On the average target-present trial, only half the distractor items will be identified. Therefore, each extra distractor item adds half as much time as it takes to identify it. On target absent trials, every item will be identified to ensure that no distractor is present; therefore every item added to the display adds as much time as it takes to identify it.

1.1.2 Feature Integration Theory (FIT)

Treisman and Gelade (1980) synthesized these findings into a Feature Integration Theory (FIT) stating that attention is necessary to bind together individual features and allow recognition of objects composed of different features. In this theory, attention could be deployed to

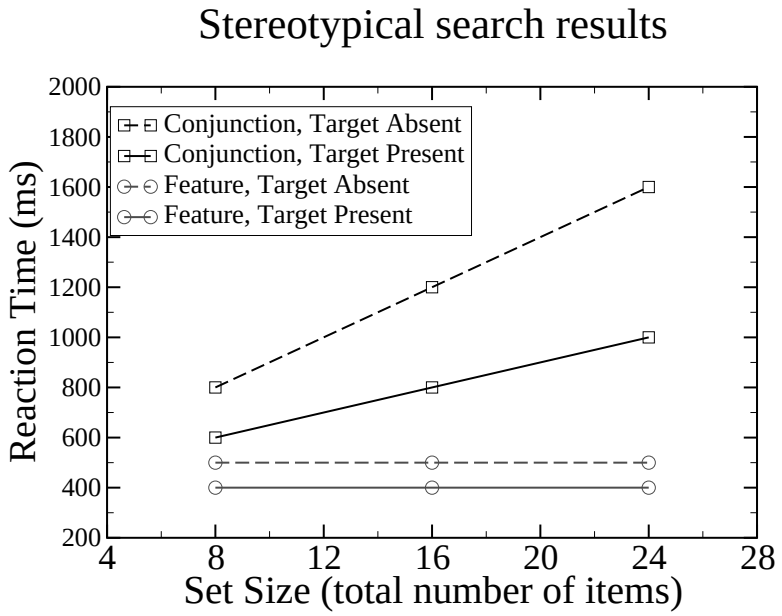

Figure 1.1: A caricature of early findings for conjunction and feature search. Note that feature searches take no extra time with added distractors, and so must be performed in parallel, whereas conjunction searches take extra time with added distractors. Also note the 2:1 absent:present ratio of search slopes. This ratio is consistent with an item-by item self-terminating search, in which only 1/2 the items are searched on average before the target is found and the search ends. On target absent trials, every item must be identified before search ends; thus on average every extra item adds twice as much time to the search on target absent trials as it does on target present trials

only one location at a time, therefore the term “serial” for searches combining features. This dichotomy between “attentive” searches (serial, attention devoted to each item in turn) and “preattentive” searches (parallel, all items processed at once) established a measure of visual system representation that could be explored experimentally. Features that could be located in parallel were assumed to be represented early in the visual system. This mapping was potentially productive, allowing exploration of the structure of the visual system using behavioral measures.

This account of visual search has since been considerably complicated. During the time since FIT was introduced, the idea that visual search findings have a simple mapping onto the physiology of the visual system has disappeared as further studies have complicated the basic findings. More recent studies, provide evidence that a slightly different but still relatively direct mapping does exist, and one goal of the dissertation is to explore that possibility. One aspect of the theory offered here is that features leading to rapid search are represented by many neurons with small spatial receptive fields (RFs) that encompass only one stimulus. This idea is discussed further in the second chapter on theories of visual search. Before going into the details of this mapping and the evidence for it, however, it will be useful to review the complications to the above view of visual search that have arisen since its conception.

1.1.3 Complications in Visual Search findings

1.1.3.1 Parallel Models of Search

The first difficulty with Feature Integration theory and its claim of serial search for conjunctions is that there is another plausible explanation of the evidence. It has long been known that limited capacity parallel models could produce the same basic pattern of data (Townsend, 1971). In these models, object identification proceeds simultaneously for all items, but the rate of identification is slowed by the number of items present, due to the limited capacity of the process. Several neural network models have shown such results (Ward & McClelland, 1989;

Palmer, 1995; Deco & Zihl, 2001). Such models can very naturally reproduce the constant search slope; my own model of parallel search showed a nearly constant slope on a range of set sizes of four items to 49 items, without parameter adjustment (Herd & O'Reilly, In press).

It has been argued (Wolfe, 1998a) that the 2:1 absent-present slope ratio is explained more parsimoniously by a serial process. It is a necessary feature of a process that stops when a target is located or when all objects have been examined. Each extra object will be examined on only half of target present trials, so the average search time rises by half as much as the target absent trial in which it will be inspected every time. This logic seems to have been compelling for many researchers, but it has recently been called into question by more recent findings, and parallel models of search are prominent in the literature (Chelazzi, 1999).

1.1.3.2 Efficient Search for Conjunctions

The first major empirical challenge to feature integration theory is the fact that some conjunctions targets display a near zero search slope. This highly efficient search indicates that they must be located in parallel without being visited serially by attention as FIT proposed. A number of demonstrations have been made of these effects, showing that feature conjunctions between dimensions, for example, shape and color (Nakayama & Silverman, 1986; Cohen, 1993), contrast polarity and shape (Theeuwes & Kooi, 1994) or form and motion (von Muehlenen & Mueller, 2000), could be discriminated without additional time for additional distractors.

1.1.3.3 Continuum of Search Efficiencies

The second major change in the data is that there now seems to be no strict dichotomy between "parallel" and "serial" searches. Using a wider variety of stimili, search slopes vary from quite small, very close to zero, to quite high, on the order of several hundred ms per additional item for a difficult search. There is no evidence for any threshold dividing efficient from inefficient search (Wolfe, 1998b). It seems that searches can be more accurately described

in terms of a continuum of efficiency (although there is more subtle evidence for a qualitative shift from parallel to serial processing; this is the focus of the empirical work presented here). I will hereafter follow Wolfe (1998a) in referring to searches as ranging from very efficient to very inefficient, and reserve the terms serial and parallel for theories about the mechanism of their performance.

1.1.3.4 Slope Ratios are Not Always 2:1 in Inefficient Search

A 2:1 ratio of target present to target absent slopes seems highly indicative of a serial self terminating search (that is, one in which search continues until the target is found or all objects have been inspected). Finding this ratio in many search tasks made serial search seem more plausible than a parallel search, as discussed above. However, it has long been known that some searches deviate from this ratio, showing closer to 1:1 ratios at small set sizes, (Pashler, 1987). Some ratios are larger than 2:1, for instance, a challenging 2 vs 5 search showed a ratio of 3.6:1. (Wolfe, Michod, & Horowitz, submitted). In a meta analysis of many inefficient searches, the average slope ratio was 2.4:1 (Wolfe, 1998b).

1.1.3.5 Visual Search Processes have Little Memory

A 2:1 ratio absent-present search slope ratio is consistent with a serial self-terminating search only if the search process never revisits a checked location. It has recently been shown using several approaches that search processes do not perfectly track the locations they have already visited (e.g. Horowitz & Wolfe, 1998, 2003). It now seems that only about four total locations are remembered (McCarley, Wang, Kramer, Irwin, & Peterson, 2003; Snyder & Kingstone, 2000).

It is possible for observers to circumvent the limitations of memory for searched items by using a systematic progression of search. For instance, searching every item in a given quadrants of a display, and moving among quadrants in a memorized order, so that the number

of items to remember within each sub-region is always small. However, eye tracking studies have never reported evidence of such systematic search; observers instead seem to sample areas of the display at random with their eye movements (Scialfa & Joffe, 1998).

If search processes cannot accurately track a great number of locations, search cannot be truly serial by item and self terminating with set sizes over four or five; once search has visited more items than can be accurately remembered, there is no way to know when every item has been visited. Therefore a 2:1 ratio cannot be evidence of an item-by-item self-terminating search, as was once thought.

1.1.3.6 Eye Movements

Given all of these complications to the original logic for a serial process, it would seem that parallel theories of search are equally as plausible as serial theories. There is one major reason that most researchers believe that many searches are performed serially: observers clearly move their eyes during challenging searches. This fact has long been noted informally, and has now been quantitatively reported in many eye tracking studies. This fact makes fully parallel models of search fairly implausible for the wide range of tasks (including standard conjunctions search) for which observers are known to make eye movements.

It can be argued that eye movements are merely gathering information for an internal representation that is searched in parallel. However, search cannot be uniformly parallel for several reasons. First, most visual areas are retinotopic, that is, their informational content depends on the current orientation of the eyes, and changes when the eyes move. Thus the information processing in most visual areas changes dramatically whenever the eyes move. Unless these areas are somehow irrelevant to the real work of visual search, search with eye movements cannot be entirely performed in parallel. Because retinotopic areas make up the great bulk of brain regions devoted to visual processing, it is highly unlikely that the bulk of processing for search is carried out in parallel across eye movements.

Second, it is known that visual perception decreases dramatically with distance from the center of gaze (Liversedge & Findlay, 2000). The same is true of neural resources devoted to each unit of space, and receptive field sizes also increase dramatically with distance from the fovea. Because visual search displays occupy a much larger space than can be covered by the fovea, the direction of a gaze must be important to the information in any brain representation. So even if search is taking place at the level of a retinally invariant representation, eye movements (saccades) are very likely to be critical for determining how much information is available about each item. If information is added serially to parallel computations, those computations clearly have a large serial component.

But if eye tracking studies nearly eliminate many parallel models of search, they do the same for serial models in which attention visits each item. Most of these studies also show that much less than half of the items are fixated on a target present trial. Brown and Gilchrist (2000) showed about 3 fixations on average in searching a display of 16 heterogenous objects for a color/shape conjunction. Williams and Reingold (2001) showed around 5.5 average fixations to search 24 items for a triple conjunction of shape, orientation, and color. Scialfa and Joffe (1998) recorded almost four fixations on average to search among around 80 items for a conjunction of orientation and contrast polarity (white vs black lines on a gray background).

1.1.4 Guided Search

In order to deal with the above complications of conjunction search findings, a modification of FIT called Guided Search (GS) was proposed by Wolfe, Cave, and Franzel (1989). The addition of this theory is the idea that the same mechanisms that allow parallel fast detection of easy targets (“basic features” in the terminology of FIT and GS) can help guide attention to more difficult (“conjunctive”) targets. The proposed mechanism involves “preattentive feature maps” that track where in the visual field certain features occur. If participants are able to form spatial maps of both features involved in the conjunction, directing attention using a sum of the

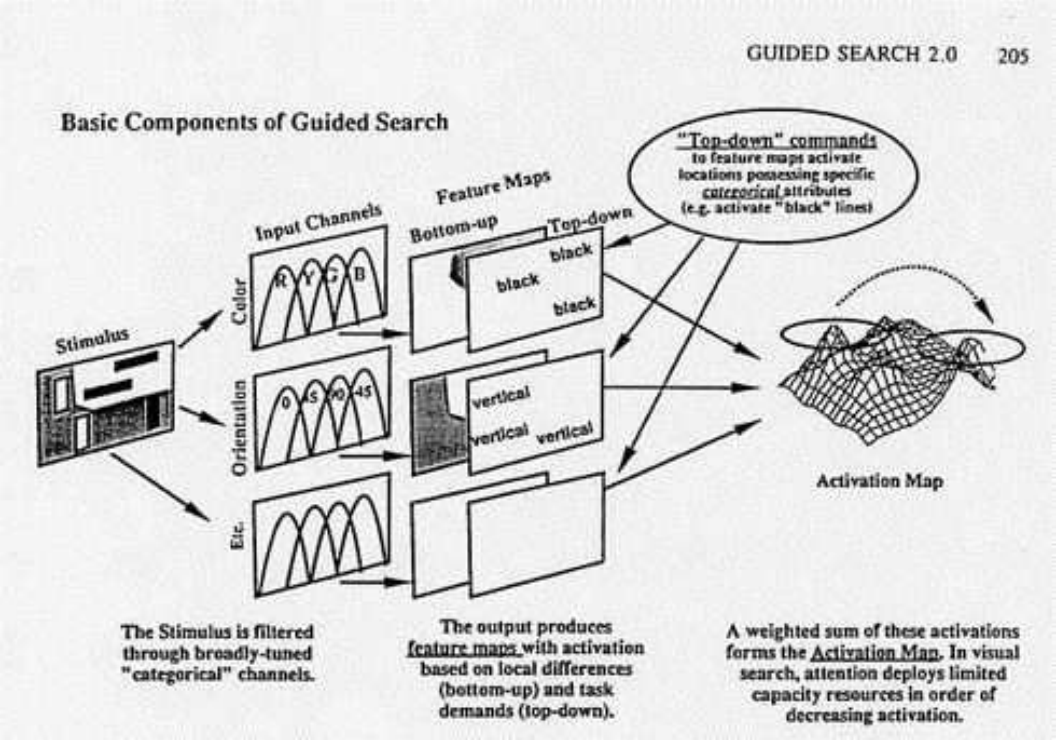

Figure 1.2: Guidance mechanisms proposed by Guided Search 2.0, (Wolfe 1994). Note that there are separate maps for different feature types, and that the guidance stage is computed once, prior to serial attentional fixations. Reproduced from Wolfe (1994)

two maps will lead to a focusing of attention on the target without any need to focus it serially on any of the distractors. Summing activities of a 'horizontal' and 'red' feature map would have the strongest activation at the target location, since it will be the only location activated in both of the two maps (see figure 1.2). With such a mechanism, the features of a conjunction do not need to be "bound" into an object in order to direct attention to the correct location

Activities in each feature map, and therefore in the "activation map" that sums them, are controlled by both top-down and bottom-up criteria. The bottom-up rules are that items gain activation according to mismatch with neighbors. Both space and similarity are taken into account, so that an item closely surrounded by very different items would produce a high activation in that feature map, while an object with only identical items around it produces no activation. Top-down activation is given according to task demands, and selects one feature to produce higher activations in its particular feature map.

The issue of a continuum of search efficiencies is also addressed by GS. Each additional distractor with some similarity to the target will increase the likelihood that the first attentional fixation will miss the target. This happens because every nontarget has a finite chance of being more active than the target on the feature map through random noise (Wolfe, 1994). Because more easily discriminated features lead to more reliable guidance of search, the effective search slope is lower for some targets. A target defined by highly discriminable features will be located on the first attentional fixation. A target defined by less discriminable features might take many fixations on average to locate, since its relatively weak target activation on the feature maps is more likely to be exceeded by at least one of the distractors along with random noise. Each distractor adds more time on average in this case, since each one has a better chance of exceeding the target's activation level.

Guided Search has more trouble explaining the 2:1 search ratio, where it exists for conjunction searches. Because search is guided, attention should visit the target well before half the objects on average are examined. This would lead to a ratio greater than 2:1 if search continued

through all items in the target absent case. Guided Search version 2 (Wolfe, 1994) postulates that search is terminated before all items are searched, using an adaptive criteria that becomes more conservative when targets are missed. The precise way that criteria shifts gives rise to the 2:1 ratio. As Wolfe himself has commented (Wolfe, 1998a), this mechanism does not have the pleasingly parsimonious explanation of the 2:1 ratio that serial self-terminating models have.

However, more recent evidence shows that search does not have a 2:1 ratio for many easily discriminable conjunctions, as discussed above. The theory does predict the 2:1 ratio when guidance is not possible. GS collapses to a serial, self-terminating search when guidance is not possible. Search is guided only when the features defining an object are easily discriminable to the guiding processes ("preattentive" processes in the terminology of GS). The computational model proposed for guided search 2.0 contains thresholds for feature difference, so that features below a certain discriminability produce no activation at all on the feature maps. Therefore the 2:1 slope ratio arises naturally in unguided searches. However, the model must use a very rapid search rate of about 50 ms per item searched; this rate is questionable in light of neuroscience evidence, reviewed later.

Guided Search remains an attractive explanation for many types of search. It has gone through four distinct versions, and is still easily the most cited theory of visual search. However, it does not account for all visual search data, as its author has taken pains to point out (e.g Wolfe, 1998b, 1998a; Horowitz & Wolfe, 1998).

1.1.5 The Speed of Attention

One challenge to the Guided Search model comes from recent physiological methods. Such measures indicate that voluntary shifts of attention seem to take at least 100 ms and likely around 150 ms (Woodman & Luck, 1999, 2003; Chelazzi, Miller, Duncan, & Desimone, 1993). These estimates are far longer than the 20 ms or so per item that is often taken to indicate serial search, and longer than the 50 ms per item posited by guided search. A variety of studies with

fine time discrimination are reviewed by Ward (2001). These studies include behavioral studies involving stimulus onset asynchronies (SOA) intended to estimate the time course of attention, as well as Event Related Potential (ERP) studies and single cell studies. He concludes that there is no direct evidence (or compelling indirect argument) that attention can shift faster than 100 ms.

Because the findings are not in complete agreement, I will briefly review the tasks. Two methods of cuing attentional shifts give very different estimates of how fast attention can move. Higher speeds (just under 100ms) of attentional shifts have been estimated from tasks using exogenous shifts, when an object appears to guide attention to a new location, as in the Posner cuing task (Posner, 1980). Attention seems to move even faster if the first item disappears as the cuing object appears, as fast as 33 ms (Mackeben & Nakayama, 1993). But in visual search tasks attention is not guided by onsets, but must move either volitionally or automatically among objects that are present before and after attention visits them.

Another type of estimate measures speed of endogenous attentional shifts, those that must be internally generated rather than externally cued by an object onset. These generally produce very high estimates of attentional dwell time, 500 ms for a complete attentional shift, and as much as 300 for a shift to begin (Ward, Duncan, & Shapiro, 1996).

It should be noted that the more direct ERP measures of Woodman and Luck (1999, 2003) suggest that attention can move a good deal faster in a visual search paradigm, perhaps 100-150 ms for a shift. However, because ERP data are averages of many trials, they cannot show the precise timing of individual trials. The data do indicate that there is on average a detectable shift at 100 ms, meaning that individual trials could be faster or slower. The more precise single-cell recordings in monkeys indicate 200 ms to complete a shift to a target after display onset, but this could be slowed by the time needed to differentiate the target.

To reconcile these findings with theories of search involving fast attentional movements, one might hypothesize that attention can move faster automatically than volitionally. All of the

above studies involved exogenous cues to redirect attention, whereas a visual search display can allow attention to shift without higher-level guidance. For instance, guidance in the GS theory is hypothesized to be set up at the beginning of a trial, and to then work at a relatively low level, so that it is not volitional in the sense of being directed by consciousness or less controversially, by working memory.

Wolfe, Alvarez, and Horowitz (2000) produced an estimate of automatic attention of 84 ms per shift. This estimate is larger than the 50ms assumption of GS2, but it is likely still too fast. In this study the "volitional attention" condition consisted of quickly shifting displays ("frames"), with a different letter appearing at each of eight constant positions in a circle. The observer's task was to shift attention by one position at each frame shift, and report whether a target had appeared at the attended location. The target appeared only once. A stairstep procedure for frame time was used, and it showed that observers were about 70% accurate at about 274 ms for each frame rate. The authors concluded that volitional shifts of attention required about 274 ms. This finding was contrasted to a frame rate of 84 ms per frame for the same level of accuracy in the "anarchic attention" condition, for an estimate of 84 ms for a non-voluntary shift of attention. However, this conclusion seems highly dubious. In this condition a target appeared on every frame, in a random position. It was assumed that attention must move at random among the locations every 84 ms to accurately detect the target. However, if observers could simultaneously attend to even two possible locations at once, they could produce the observed accuracy even with imperfect target identification. Because attending to more than one location is highly plausible (and indeed has since been proposed by one of the authors, Wolfe et al., submitted), the methodology of this study is highly unconvincing.

A more recent study from the same group used a refined variant of this method (Horowitz, Holcombe, Wolfe, Arsenio, & DiMase, 2004). This time there was no "automatic" condition, but instead a "tracking" condition. In this condition, attention was guided by an apparent motion: twelve circles appeared and disappeared in alternate frames, so that six circles were visible

at a time, and the percept was of these six circles moving either clockwise or counterclockwise. The task was to track a given (illusorily) moving circle, and to report what letter briefly appeared after a random number of time steps. The letter was followed by a mask. The volitional condition was identical, except that all twelve circles appeared on every frame, and instead a tone sounded to signal this attentional shift.

This study removed the major confounds in the earlier work (Wolfe et al., 2000), but still found large differences between the conditions. This time the volitional condition produced an estimate of total attentional movement time of around 300 ms, while the tracking condition produced estimates of 150-200ms for an accuracy of 66% (when the letter was presented for 150 ms before masking).

It is clear that estimates of automatic attention vary with the difficulty of the discrimination used to measure attention; using frame times of 107 ms still allowed 40% accuracy. However, it seems that accurate target discrimination in visual search requires high accuracy, although the discrimination is usually much easier due to use of only a few items in each task.

It is also debatable whether attention in a visual search is accurately modeled by either the exogenous (visually cued) process or the tracking process. In a visual search task, neither an onset nor apparent motion exists to cue attention. Therefore that the volitional condition offers a better analog to the standard visual search paradigm. It could be argued that the necessity of directing attention to a particular stimulus dramatically slows attentional shifts (Wolfe et al., 2000). But this task involves no interpretation of a cue (which as the authors note would surely take time) but rather a memorized and practiced shift, which seems more similar to the low level guidance theorized for attentional shifts in the GS model and related models.

But regardless of whether the time for attentional movements is more similar to endogenous or exogenous shifts, it is clear that the evidence is heavily against shifts even as fast as the 50 ms posited for the Guided Search model. It seems that the more recent study (Horowitz et al., 2004) has brought the authors to the same conclusion, since the most recent version of

Guided Search now posits a limited parallel identification process (Wolfe et al., submitted).

1.1.6 Serial-Parallel Theories of Search

A modified model is consistent with both the findings of 2:1 search slopes in many paradigms, and with a relatively slow movement of attention. If a small group of items is identified during each attentional movement, attention need visit only a small number of areas in each display, few enough that the limited memory capacity of search can track them all. Search can under this condition be serial self-terminating, explaining the common 2:1 absent-present slope ratio. And if, say, four items are identified during each attentional fixation, fixation times would be four times as long as those calculated by assuming an item-by-item search.

I am terming such a search serial-parallel. In this type of theory, a parallel search takes place during each attentional fixation, and this search is repeated serially. I independently arrived at a theory of serial-parallel search for non-temporal reasons, based on considerations of known structure and function in the visual system at a neural level, as discussed in chapter two, sections 3. However, the idea has long been present in the visual search literature, although it has remained at a relatively low profile— the idea is mentioned and glossed over in the two most recent major reviews of the visual search literature (Wolfe, 1998a; Chelazzi, 1999).

Most general theories of visual attention include the possibility of attention to a limited area including more than one item, notably the zoom lens model of Eriksen and St James (1986). Some such models have even proposed that such a strategy could be common in visual search (Bundesen, 1998). In these theories the number of items attended to varies according to the requirements of the task.

Pashler (1987) proposed a serial-parallel model of visual search, based on findings that small set sizes display absent-present slope ratios near one. Zohary and Hochstein (1989) proposed a similar model, based on a different experimental manipulation, with the difference that time to process each set of items (“clump”) is constant rather than dependent on the number,

as Pashler proposed. Duncan and Humphreys (1989) proposed a similar model in which four items are identified in parallel, although they did not propose that these items should be located together in a clump. A number of more recent studies of visual search have reached similar conclusions (Palmer, Verghese, & Pavel, 2000; McElree & Carrasco, 1999). A notable addition to these theories is the most recent update of the guided search model, GS 4.0 (Wolfe et al., submitted). This model proposes that four items are identified in parallel, with no requirement for clumping.

1.1.6.1 Number of Items Identified in Parallel

The above theories all vary in the number of items proposed for identification. More general theories propose that the number should vary, according to particular task demands. All the models of visuals search have tentatively hypothesized a specific number of stimuli that can be identified in parallel, and the estimates range from sixteen (Zohary & Hochstein, 1989) to eight (Pashler, 1987) to four (Duncan & Humphreys, 1989; Wolfe et al., submitted). The estimate of Wolfe et al. (submitted) seems the most certain; the graphs shown in Pashler (1987) generally show an intermediate slope size between set sizes 4 and 8 (figures 1.3 and 1.4). In addition, the slope estimates of that study seem to be driven more by an increase in target absent slopes beyond set size 4 rather than the decrease in positive slopes at larger values.

One intriguing possibility is that a constant number of items identified in parallel could be linked to central cognitive limitations. Duncan and Humphreys (1989) explicitly based their hypothesis of four items on the idea that only four items can be held in visual working memory. Cowan (2001) has recently proposed a general capacity of four items in working memory. If visual recognition involves the entry of items into working memory hypothesized by Duncan and Humphreys (1989) and in the Theory of Visual Attention (which also posits a four item capacity Bundesen, 1998, 1990), this limitation could account for the limit on parallel visual search capacity.

This idea is backed by experimental evidence showing that short-term memory can hold information about three or four objects, regardless of the number of features composing each object (Vogel, Woodman, & Luck, 2001; Lee & Chun, 2001; Luck & Vogel, 1997), although this claim has been disputed (Davis & Holmes, 2005; Olsson & Poom, 2005; Alverez & Cavanagh, 2004). Other literature has proposed that the approximately four item limits on visual processes for subitization (fast counting) and tracking are linked (Trick & Pylyshyn, 1994; vanMarle & Scholl, 2003). Another point in favor of a four item constant capacity is the fact that Wolfe et al. (submitted) tested three types of stimuli that produced fairly different search slopes, yet found a similar discontinuity around 4 items for each type.

Given the findings of a capacity of around four objects in three separate types of visual tasks, and in three variations of visual search tasks, it seems possible that there could a general capacity for four objects on average for visual processing. If this is the case, we would expect that the capacity of parallel object identification is also four objects, as suggested by (Wolfe et al., submitted) on purely empirical grounds. It is striking and surprising that the three different types of stimuli used by Wolfe et al (submitted) all seem to show nonlinearities in the search function at the same point. Set sizes between 4 and 8 were not employed, so firm conclusions about exact break points cannot be drawn. However, visual inspection of their graphs shows that, at set size 8, reaction times for all types of stimuli fall a similar distance below the point predicted by extrapolating the slope of set sizes 1-4 (two of these are shown in figure 1.5).

The theory of visual search proposed in chapter four makes the opposite prediction of the supposition that working memory capacity limits the number of items searched in parallel. If the theory proposed here holds true, we would find that the number of items identified in parallel changes depending on the discriminability of distractors from targets. Theories based on Signal Detection (Palmer et al., 2000; Geisler & Chou, 1995) also predict such changes with set size, based on hypothesized limitations in perceptual processes. These theories will be reviewed in the next chapter, section 2.2.1.

1.2 Experiments

These experiments are designed to provide further evidence of serial-parallel search, and to refine theories of such search. The primary hypothesis these experiments test is that the number of items searched in parallel varies inversely with the discriminability of targets from distractors, that is, relatively easier stimuli leads to processing of more items during each attentional fixation;. To do this, these experiments involve a wider range of stimuli and task conditions than used in previous work.

One question left open by previous work is whether serial attention to limited areas is the product of eye movements. All of the existing studies involved eye movements, and so it could be that their results indicate that displays under some minimum size do not invoke eye movements. The proposed studies investigate this by including conditions that do and do not allow eye movements.

The challenge was inferring when observers shift attention without the use of eye-tracking equipment. This issue is approached directly, by including a condition in which observers refrain from eye movements, and indirectly, using a promising method of inference from search slopes. This second method is slightly involved, and is explained below.

1.3 Experiment 1

Experiment 1 follows the experimental logic of Pashler (1987) and Wolfe et al. (submitted) for identifying the number of items processed at once during visual search. The logic is based on search slope ratios, as mentioned above in section 1.1.1. Pashler used standard conjunction tasks (red and green, Os and Ns, Ts or Ls in different experiments). He found that small set sizes (below 8 items) produced slope ratios of around 1:1, and that they had slope ratios that were significantly less than 2.0. At set sizes larger than 8, he found slope ratios around 2:1 (see figure 1.3.

These findings are consistent with a model in which up to 8 items at a time are identified

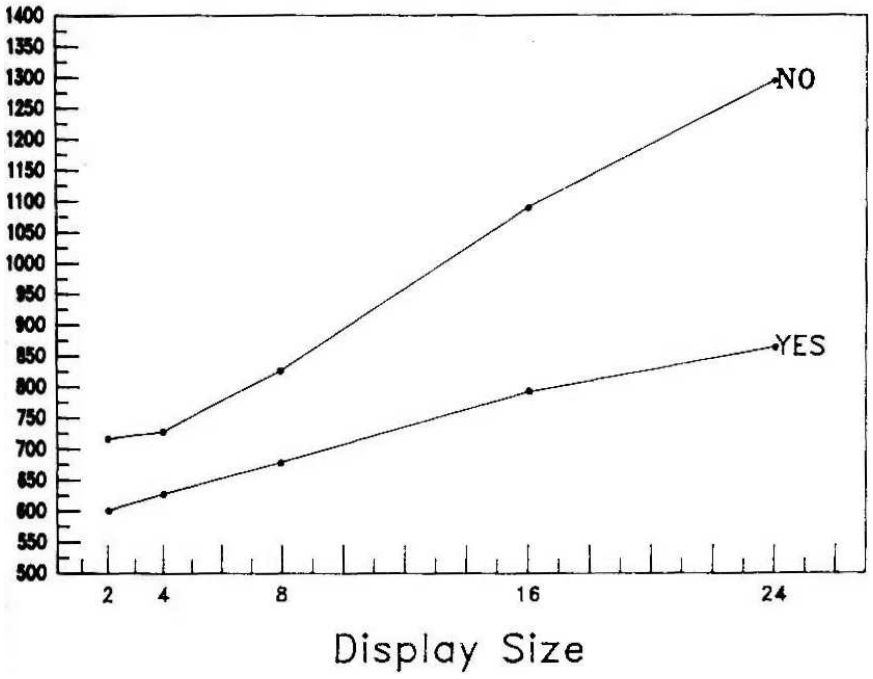

Figure 1.3: Data from Pashler (1987), Experiment 2b. A conjunction search for a red O among red Ts and green Os. Note the lines are approximately parallel prior to 4 items and divergent thereafter. Reproduced from Pashler (1987)

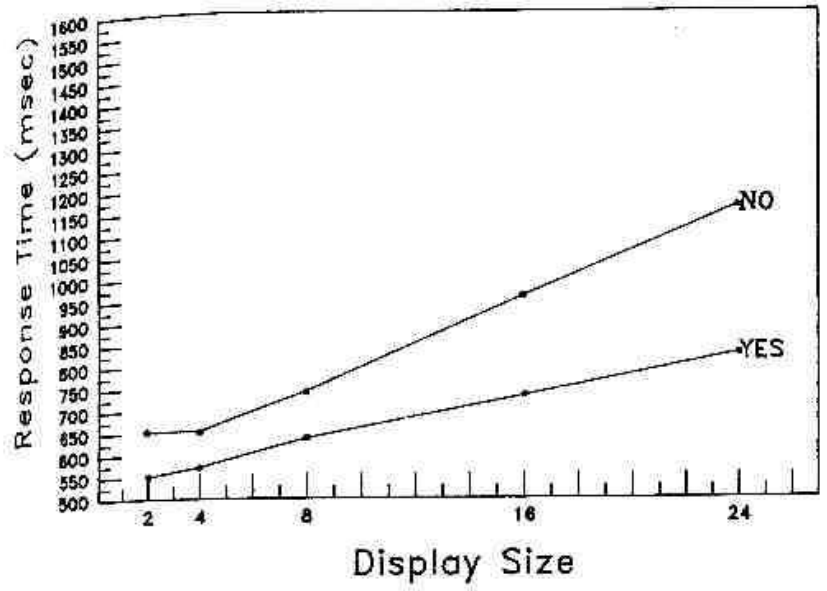

Figure 1.4: Data from Pashler (1987), Experiment 3b. A conjunction search for a red O among red Ns and green Os. Note the approximately parallel lines prior to 8 items and the divergent line after. Reproduced from Pashler (1987)

in parallel, with the rate of identification slowing down as more objects are added. At set sizes of 8 and below, all items add to search times for both conditions, since they are all processed. Above set size 8, each added item contributes only half as much time on average in the target present condition, since the target will be found earlier on half the trials. In the target absent condition, every group will be scanned, so each item contributes its full extra time on average. The search slope ratio thus changes from 1:1 to 2:1 above 8 items.

If the number of items processed in a single attentional fixation varies either between trials or between individual observers, we would see an intermediate search slope ratio, between 1:1 and 2:1, over the range of variability. This is the case in Pashler's (1987) Experiment 3b, which used a larger number of trials, providing greater reliability (see figure 1.4). The slope ratio between set sizes 4 and 8 is about 4:3 in these data.

Wolfe et al. (submitted) arrived at a similar conclusion using slightly different methods. They focused primarily on the target-present slopes, noting that slopes were significantly greater in the lower set sizes (see figure 1.5). This is consistent with the idea that on average only every other group beyond the first must be attentionally fixated before the target is found, when it is present, so that groups beyond the first contribute only half as much time to search on average. They also showed that slope ratios were larger (closer to 1:1) in the small set sizes (below 4). They used several sets of stimuli, including a standard color-orientation search (red verticals among blue vertical and red horizontal lines), a rotated T among rotated Ls, and a digital 2 among digital 5s, a mirror image reversal.

Based on their data, they proposed a very similar model to that of Pashler (1987). The only difference in their proposed model is that in the model of Wolfe et al. (submitted) there is no requirement for items to be spatially clumped to be identified together, or for a group of items to enter and leave the identification process together. However, this difference does not affect the predicted search slopes. The main difference is therefore the number of items identified in parallel. Pashler's (1987) model proposed approximately 8 items; Wolfe et al (submitted)

proposed four items.

Experiment 1 uses methods similar to those of Pashler (1987) and Wolfe et al. (submitted) to replicate and refine their findings. A greater variety of stimuli were used, to test the hypothesis that the number of items identified in parallel will vary inversely with the difficulty of target-distractor identification. The stimulus sets were chosen so that none was a very efficient search, but so that each type varied from the others in difficulty. Three stimulus sets were used. The easiest was finding an E among Fs (hereafter the EvF task); slightly harder was locating a rotated T among rotated Ls (very similar to one task used in Wolfe et al., submitted; hereafter this task is referred to as the TvL task), and the hardest search was for a rotated complex shape that differed from distractors only in that it lacked one line (hereafter the shape task). Figure 1.6 shows an example display of each type.

The primary question this experiment addressed was whether the number of items identified in parallel during each attentional fixation varies for different types of stimuli. A secondary question was whether this number would also vary with strategy. A manipulation of the ordering of set sizes was included in an attempt to manipulate strategy. The reasoning was that observers might be slow to switch strategies. For instance, observers might be more likely to maintain the strategy of a mostly parallel search that worked well at set sizes 2 and 4 when they reached set size 6, and more likely to maintain the strategy of quick attentional shifts that worked on set sizes 16 and 8 when they reached set size 6 after seeing larger set sizes. If this were the case, the number of items identified in parallel would differ across orderings of set sizes.

1.3.1 Method

1.3.1.1 Participants

Participants were recruited from the subject pool of introductory psychology students, and the paid subject pool. Sixty-four people participated, all with normal or corrected-to-normal vision. Twenty-six of these participated for credit, while 37 were paid $10 for approximately

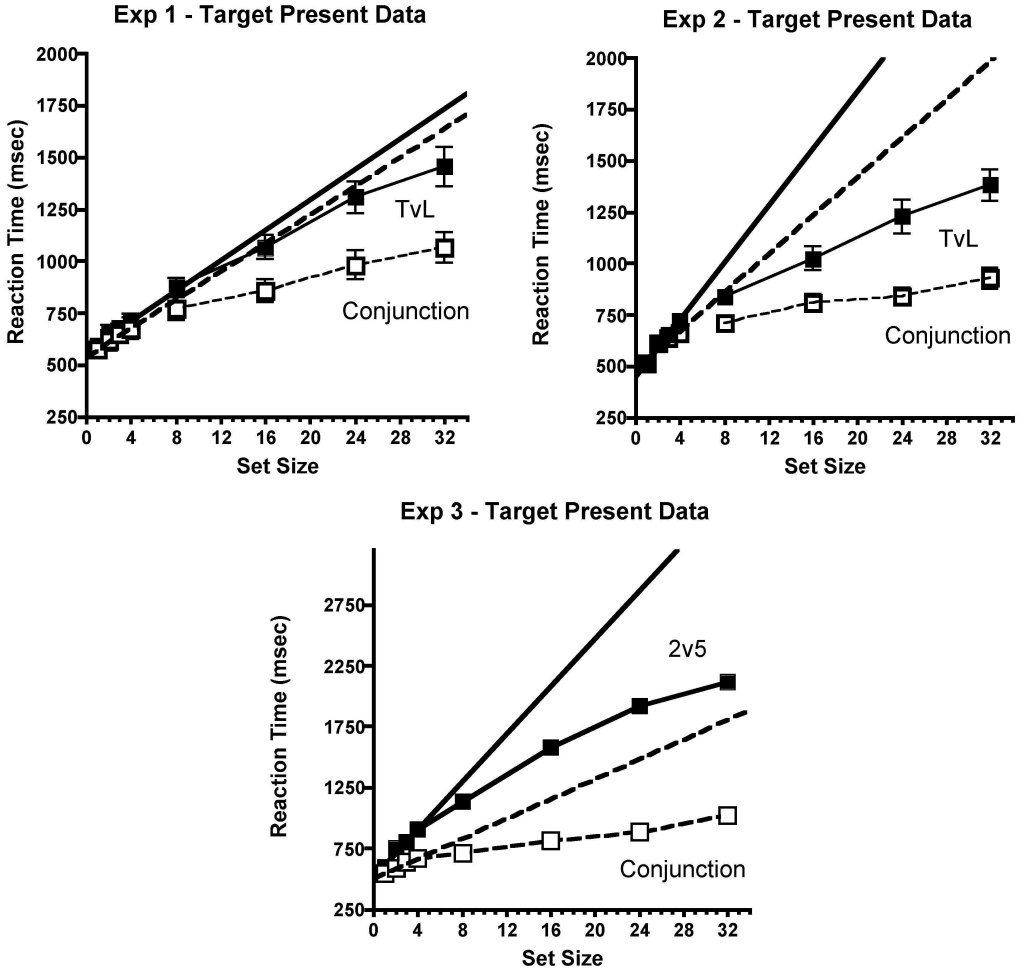

Figure 1.5: Target present results for each of the 3 experiments of Wolfe et al (submitted). White boxes are orientation X color search; black boxes are T vs L search in exps 1 & 2, 2 vs 5 search in expt. 3. Straight lines are extrapolated slope of set size 1-4, solid for the T vs L or 2 vs 5 task, dashed for conjunction. Reproduced from Wolfe et al (submitted).

45 minute long sessions. Two subjects were graduate students in the department who participated as observers in Experiment 5; they were unfamiliar with the hypotheses being tested. All observers were between approximately 18 and 30 years of age. Data from one observer (from the unpaid subject pool) was excluded from the analysis due to extremely fast reaction times and low accuracies.

1.3.1.2 Apparatus and stimuli

Stimuli were generated using the xcss software developed by Randy O'Reilly for similar purposes. Stimuli were be presented and responses recorded using the E-prime software package, run on Windows 2000 computers with VGA monitors. Searches were performed with three stimulus sets. The easiest was finding an E among Fs; slightly harder was locating a rotated T among rotated Ls (very similar to one task used in Wolfe et al. (submitted)), and the hardest search was for a rotated complex shape that differed from distractors only in that it lacked one line. Figure 1.6 shows an example display of each type.

Stimuli were located on an invisible 7 X 7 grid subtending about 22 degrees on both axes. A fixation cross always occupied the center position of the grid. The position of each item was varied by up to about .75 degrees of "jitter", random vertical and horizontal offset. This jitter was included, as in most visual search tasks, to prevent easy grouping of distractors. In all three tasks, stimuli subtended about 1.25 degrees by 2.5 degrees of visual arc for the EvF and TvL stimuli. The Shapes stimuli subtended about 2.5 by 2.5 degrees square. Line width for all stimuli was about .25 degrees of arc. The above arc measures are calculated based on the viewing distance of about 45 cm, but no chin rest was used to maintain a constant distance. Stimuli were white, presented on a dark background. When targets were present in a display their position was random, but constrained to fall outside the 8 positions surrounding fixation. This was intended to decrease variation in reaction times due to eccentricity differences.

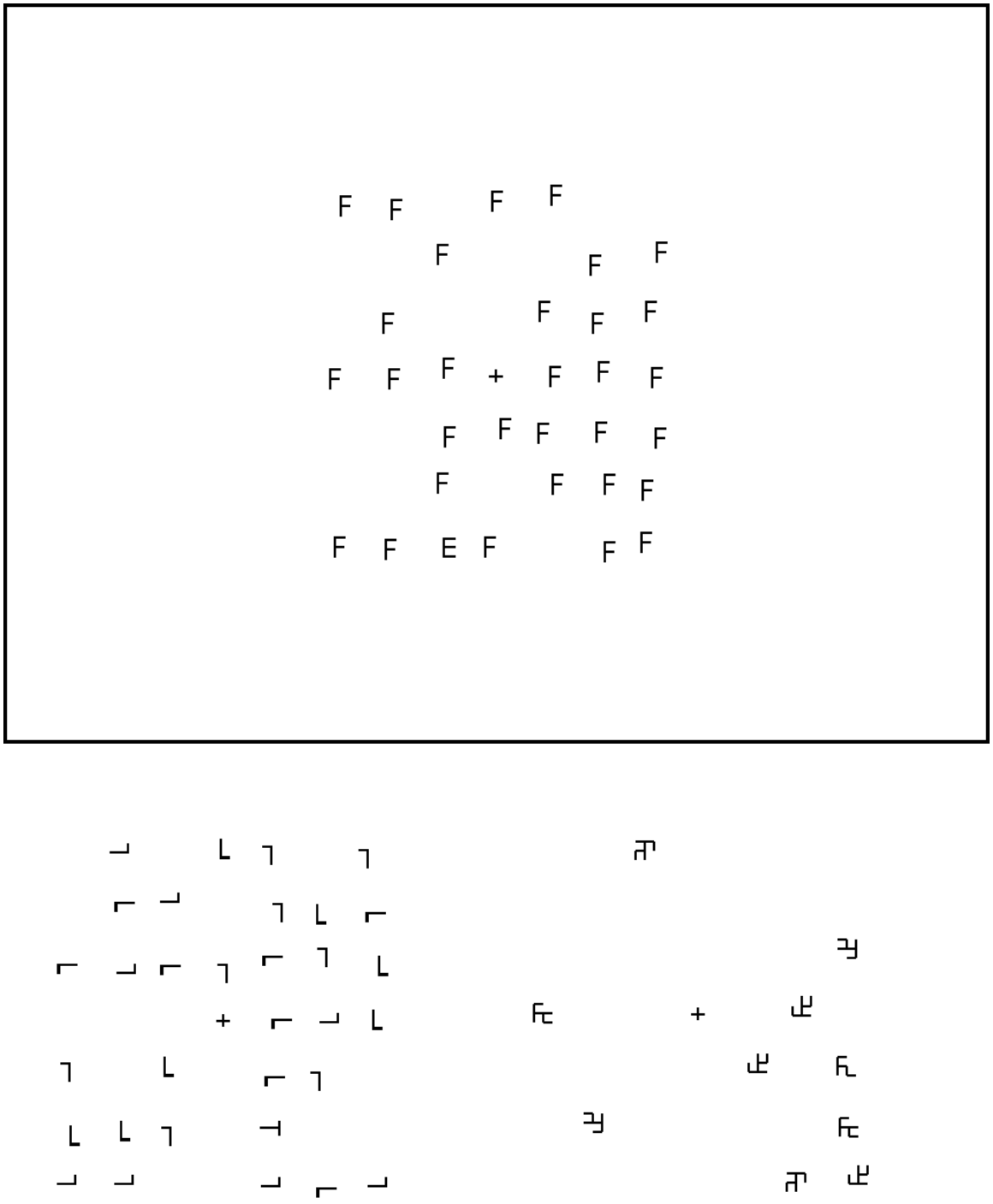

Figure 1.6: Displays from experiments 1, 2, and 3. Colors are reversed; displays were shown as white characters on a black background.All displays shown are the largest set size in that condition, with a target present. The black outline around the EF display shows the size of the 12 inch monitor; no such border was visible in the experiments.

1.3.1.3 Design and Procedure

The design was a 2 (target present or absent) X 6 (set sizes) X 3 (stimulus types). The set sizes used for both the EvF stimuli and the TvL stimuli were 2, 4, 6, 8, 16, and 32. Those for the shape stimuli were 1, 2, 3, 4, 6, and 10. These sizes were chosen to sample a large range of set sizes, but to provide greater resolution in the smaller ranges, where the discontinuities in search slopes were hypothesized. The shape stimuli were assigned smaller set sizes because they took much longer to search; large set sizes would consume a good deal of time for each display.

Each search task asked observers to make a forced choice judgment as to whether the display contained a single target, or only distractors. On each trial, the fixation cross was presented as a warning for one second. The search display then appeared, and remained present until a response was made, or 5 seconds elapsed, at which point the message 'please respond faster' was presented, and the trial was discarded. Responses were made by pressing the "1" key on the number pad (right side) of a standard keyboard to indicate that the target was present, and the "2" key, also on the number pad, to indicate it was not present. Observers were asked to make responses "as quickly as possible without making errors." Feedback was presented after each trial. If the response was correct, the reaction time for that trial was given. The feedback remained visible for 1.5 seconds, then the next trial began. A rest period was given between each block.

Each type of stimuli was presented within a block. The order of blocks was randomized, so that the order of different search types varied at random. In the main blocks, set sizes were organized into sub-blocks, to allow a consistent strategy to develop, as in Wolfe et al, (submitted). The order of set sizes was held constant, in an ascending or descending order, which was roughly counterbalanced across participants. This order was chosen to influence strategy choices; we reasoned that if strategies are different for large displays than for small displays, the choice of strategy might be determined by whether the participant had seen larger or smaller displays previous to any particular one. Each sub-block consisted of 10 target-present and 10

target absent trials, in random order. There were six set sizes for each stimulus type, for a total of 120 trials within each stimulus type block, for a total of 460 trials across all three blocks. Each block was preceded by 12 practice trials. In these trials, each set size and target present/absent condition was sampled once, in a random order.

1.3.2 Results

The results are similar to those usually reported in this type of paradigm; we focus on two patterns in the data that are not usually analyzed. First, there is an indication of the nonlinearities reported by previous experiments that densely sampled the lower set sizes (Pashler, 1987; Wolfe et al., submitted). The slopes between set size 1 and 2 for the Shapes stimuli show the clearest evidence; they are parallel for that segment, and afterward adopt a slope ratio around 2:1.[1] The EvF and TvL slopes do not show such clear evidence of a nonlinearity. The target present slopes for both decrease from set size 4 to 6, steepen again, and decrease in the last part of the graph. One might imagine break points at several spots on these graphs, or merely an exponential or otherwise relaxing function of set size, with some amount of noise. Because these graphs are averages of many subjects, they may obscure individual differences that happen at different points; the data analysis focuses on revealing systematic tendencies in the results from individual observers by more subtle means than simple averages.

The general decrease in slopes at large set sizes may well be the result of a speed accuracy tradeoff. The second unusual aspect of these data is that they showed somewhat higher error rate than most search findings. There several possible contributing factors. The first is the low number of total trials with each stimulus type. The amount of practice was relatively low compared to most studies. The second factor is the inclusion of individual trial reaction times with feedback; these could well have encouraged observers to “compete” against their own times, emphasizing speed at the expense of accuracy. The last factor is the lack of auditory

[1] Slope calculations were not done as best fit lines to multiple set sizes, since there were no a priori hypotheses about precise change points for slopes; the slope ratios for each individual line set average about 2:1, and each set is individually significantly different from 1 at all line segments above set size 3.

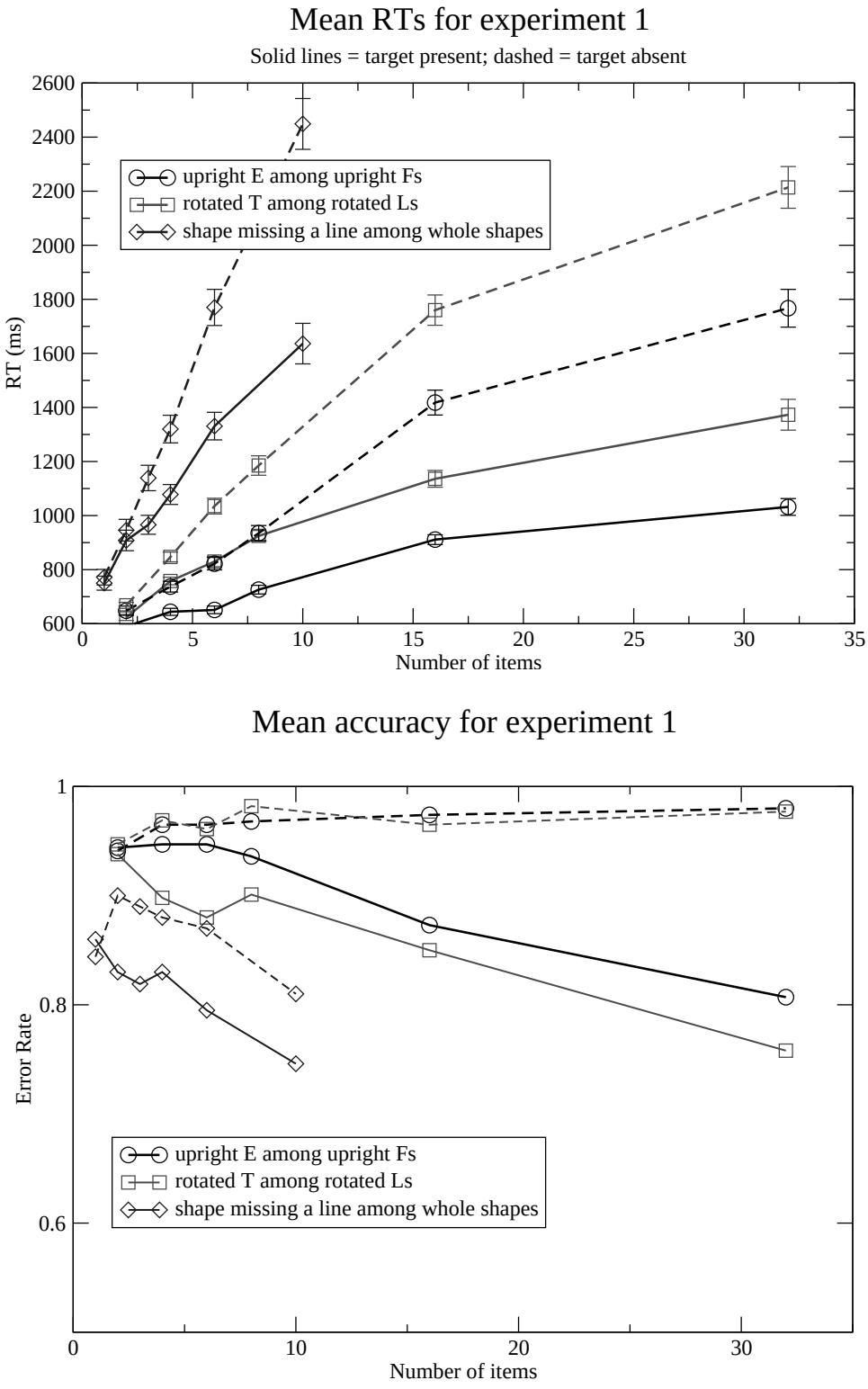

Figure 1.7: Data from Experiment 1. Reaction times are harmonic means (reverse log transformation of the means of log transformed reaction times). Error bars are standard error of the mean, between subjects.

feedback. Some visual search experiments report use of a buzzing tone as negative feedback on error trials; this device could be effective in emphasizing accuracy. The issue of a possible speed/accuracy tradeoff is addressed in the analysis below.

The data analysis proceeded as follows. Error trials were eliminated from the data set; the error rates were analyzed separately. Reaction times were log-transformed, to correct for the asymmetric distribution usually seen in human reaction time data and produce a roughly normal distribution (Ulrich & Miller, 1993). This correction is similar to looking at median rather than mean reaction times. Data were analyzed for outliers by constructing a model of each stimulus type, set size, and present/absent condition for each participant, and constructing models to test whether any individual data point significantly altered the model. Those data points that had a significant effect were eliminated as outliers. The significance value was adjusted for multiple comparisons, with a Bonferroni correction (Judd & McClelland, 1989) for 460 individual tests, the maximum number of tests performed for each participant, one for each individual data point. Therefore if the data were truly normally distributed (within participant and condition), only one data point would be mistakenly rejected out of every 20 participants on average, due to the standard alpha value before correction of .05. This conservative criteria eliminated only about 50 data points out of all 63 participants, but eliminated all trials below 200 ms, a common truncation point as an estimate of an absolute minimum decision time. This style of outlier analysis is more appropriate to an analysis of conditions with widely varying means than is the truncation method (dropping all RTs above and below a certain level). Truncation has a differential effect on conditions with large and small means, and so may bias the results of analyses comparing these conditions (Ratcliff, 1993).

Once the outliers were removed, the remaining data were averaged to produce a single value for each participant in each condition. This procedure produced 36 data points for each participant, one for each combination of stimulus type, set size, and target present/absent condition. These points were then reverse-log transformed, to recover the between-conditions

variability. There were 6 such harmonic mean data points per participant for each stimulus type and absent/present condition.

To find the most likely break point in each observer's data, five separate bi-linear models were fitted to the target-present[2] reaction times for each of the possible 5 break points for each observer's data [3] There was substantial variation in the model with the best fit across different observers, and all possible break point appeared more than once.

First the break points used by the one best-fitting model for each subject and stimulus type was tested for systematic variation with block order. This was a between subjects variable; about half (33) of observers had seen all blocks presented in ascending order; about half (30) had seen them presented in descending order. These tests showed no significant difference; the effect of order on breakpoints in the EvF task produced p = .22; for the TvL task p = .14; and in the Shapes task p = .69. These differences neither confirm nor disprove the hypothesis that breakpoints would vary with strategy differences induced by block order effects. Because this was not the central hypothesis being tested, it is not considered further.

The more central hypothesis was the comparison of break points across different stimulus types. The break points taken as the input for a set of nonparametric tests. A Wilcoxon rank sum test was used, since it is unlikely that the break points we found are normally distributed. The test was performed on the difference calculated for each participant between breakpoints for each pair of stimulus sets.

There were significant differences between best break points for each of the three stimulus types. The difference between breakpoints for the EvF stimuli and the TvL stimuli had

[2] The slope ratios analyzed by Pashler (1987) turned out to be poor candidates for such model fitting. Any ratio is not in general normally distributed (Judd & McClelland, 1989) and these in particular were poorly behaved due to the relatively small number of trials for each person in each condition. Therefore an analysis more similar to that of Wolfe et al (submitted) was performed, with the focus on target present trials only. This approach also avoids some possible difficulties with assuming a self-terminating model for target-absent reaction times; a number of authors (e.g. Wolfe and Chun 1995) have suggested that the termination criteria for search varies in a way that is not straightforward.

[3] Each of these models fitted a line of one slope to the lower range of set sizes, and a line of independently varying slope to the higher set sizes. Five variants of the model were run, with all the possible distinctions between low and high set sizes allowed by the six set sizes examined. The models with the break point between smallest or largest set size are not truly bi-linear, but linear with an extra point predictor for the highest or lowest set sizes.

Variable	Mean	Median	SD	P value
EvF breakpoints	12.2	16	5.4	
TvL breakpoints	10.1	8	5.9	
Shapes breakpoints	4.4	6	1.82	
EvF - TvL	2.0	0	7.3	.042
EvL - Shapes	7.6	10	5.7	$< .0001$
TvF - Shapes	5.7	5	6.8	$< .0001$

Table 1.1: Summary breakpoints and differences between break points in Experiment 1

.

a mean of 1.96, meaning that the best fitting breakpoint was on average at a set size of about 2 larger for the EvF stimuli than for the TvL stimuli. This difference was significant, S(63) = 138.5, p = .042. The difference between breakpoints for the EvF stimuli and the Shapes stimuli had a mean of 7.66, meaning that the best fitting breakpoint was at a set size of about 8 larger for the EvF stimuli than for the Shapes stimuli. This difference was significant, S(63) = 797.5, $p < .0001$. The difference between breakpoints for the TvL stimuli and the Shapes stimuli had a mean of 5.66, meaning that the best fitting breakpoint was at a set size of about 6 larger for the TvL stimuli than for the Shapes stimuli. This difference was significant, S(63) = 590.5, $p < .0001$. The results of these tests, and the values calculated for the breakpoints themselves, are presented in table 1.1

It must be noted that the comparisons between breakpoints for either the EvF or TvL stimuli and those for the Shapes stimuli are questionable. Because different set sizes were used, it is likely that even pure noise in the data would result in a significant difference on this test. However, the comparison between the EvF and TvL stimuli does not have this problem. The finding of significant differences indicates that the two data sets have a substantially different shape.

One outstanding question is whether the reaction times in this task are better fit by a bi-linear function, or are more truly logarithmic in form. Therefore I performed the same test on the data under an exponential transform. The linear fit models thus became equivalent to performing a bi-logarithmic fit to the data, fitting a logarithmic curve, but with two separate

log constants. The Wilcoxon paired rank-sum test produced a value of $S(63) = 147.5$, $p < .03$. The differences between the TvL and EvF conditions and the Shapes condition were still highly significant, although again this test is unreliable. Interestingly, this transformation changed the best fitting break points for only a single observer, and only in the EvF and TvL conditions; all other best fits were the same. The means and standard deviations of break points thus remained very similar to the above analysis.

It is possible that the differences in best break points across stimulus conditions was caused by different speed/accuracy tradeoffs, rather than truly different underlying search functions. To address this possibility, the analysis was performed again, corrected for effects of accuracy. Accuracy was included as a covariate in the model for break points within each subject's data. The accuracy level for that observer, stimulus set, set size, and present/absent condition was the covariate for the harmonic mean of reaction times in the same condition.

This correction had a similar effect to correcting for speed/accuracy tradeoff by dividing RT by accuracy, as suggested by Townsend and Ashby (1983), (reported in Wolfe et al., submitted), and employed by Wolfe et al. (submitted). It performs the same function of adding a correction for accuracy, but it gives that correction a variable coefficient fitted individually for each participant. The analysis proceeded as did the original analysis, using the best fitting bilinear break point for each subject in each stimulus type as input to a Wilcoxon signed rank test. The results of this test for differences between the EvF condition and the TvL condition were significant, $M(63)= 271.5$, $p = .004$. The mean difference between conditions increased to 2.57. The differences between EvF and TvL conditions and the Shapes condition remained highly significant but of dubious validity. The significance levels and means obtained are reported in table 1.2.

Because correcting for accuracy actually improves the significance level of the test, it seems highly likely that the observed differences are not the result of a speed-accuracy tradeoff, but are rather obscured by it. The correction for accuracy reduced the average best break point

Variable	Mean	Median	SD	P value
EvF breakpoints	10.3	8	5.4	-
TvL breakpoints	7.7	6	4.7	-
Shapes breakpoints	3.37	3	1.52	-
EvF - TvL	2.6	2	6.8	.004
EvF - Shapes	6.9	5	5.5	$< .0001$
TvL - Shapes	4.4	4	4.9	$< .0001$

Table 1.2: Summary breakpoints and differences between break points after in Experiment 1 correcting for accuracy levels

.

for all conditions; it was reduced by an average of nearly two per participant in the EvF task, nearly three per participant in the TvL task, and slightly more than one per participant in the Shapes task. Because accuracy falls substantially at larger set sizes, the estimates of break points after controlling for accuracy are likely to be more accurate than those from the initial analysis. The estimates for the TvL stimuli of a shift at about set size 6 (really anywhere between 4 and 6, since no intermediate set sizes were tested) is consistent with the conclusions of Wolfe et al. (submitted) of a break point between their set sizes of 4 and 8. However, the standard deviations of these estimates remain high; this could be due to noise in the data, or to actual variance in strategies among observers.

1.3.3 Discussion

The results confirmed the hypothesis at least in part. The findings of significant differences between best break points indicates that there is more to the data than the linear search function that is usually discussed. Because there are significant differences there are systematic variations from linearity or logarithmicity. The average best break points occur in the predicted direction across different stimuli. If the functions were a relaxing function other than a logarithm, such as a fractional power function ($sqrt(x)$), the best break points should differ in the opposite direction across stimulus types.

The difference in best break points between conditions is consistent with the interpreta-

tion offered by Pashler (1987) and (Wolfe et al., submitted). The lower portion of each curve has a steeper slope, a pattern consistent with a shift from parallel to serial search at some intermediate set size. This "serial-parallel" hypothesis proposes that search proceeds in parallel over a limited number of items, and switches between groups of items serially. At small set sizes, observers use attention spread to all items. With more items, observers tend to restrict attention to a subset of the items.

Previous work has raised the question of whether the number of items processed in parallel during each attentional fixation varies with the type of item. However, this is the first study to present emperical evidence of such variance, as well as the first I am aware of to deal with the theoretical implications of the distinction. The current findings suggest that the set size at which observers tend to switch to a strategy of multiple attentional fixations varies with the type of stimuli used. The point at which the slope shifts is systematically higher in the easier EvF condition than in the more difficult TvL condition.[4]

The best fit break points are higher when the stimuli are more difficult to process (as measured by reaction times at a given set size). This pattern is consistent with observers tending to search more items in parallel with the easier stimuli, before switching to a serial search strategy at relatively larger set sizes. The issue of strategy shifts has been little discussed in previous work on visual search, although it makes sense of otherwise contradictory findings; see chapters 2-4. I suggest that observers tend to process small display sizes in a single attentional fixation, and switch this strategy to attentionally fixating only part of the display when set sizes become larger.

This interpretation is independent of whether or not observers use an attentional shift to verify the presence of a target. The time to execute this shift would be a constant addition to total search time, and therefore would not affect search slopes.

This is the first study to test the idea that the number of items effectively searched dur-

[4] Although the estimates of break point are much lower still in the much harder shapes condition, the difference in set sizes does not allow for a test of reliability; therefore I focus on the difference between EvF and TvL conditions.

ing each attentional fixation varies with difficulty of discriminating targets from distractors. However, the conclusions are supported by a variety of other evidence. This hypothesis was made based on theoretical work described in chapters 2-4, combining previous behavioral results with known neurophysiological evidence. The conclusion drawn there is not that search operates over a fixed number of items on each attentional fixation. Rather it seems that all items are processed, but with decreasing efficiency away from the point of fixation, due largely to crowding from other stimuli.

The available evidence regarding eye movements supports this conclusion. A recent set of detailed studies recorded many eye movements made by monkeys performing standard visual search tasks. Although it was not the focus of their work, the results reported by Motter and Belky (1998b) indicate that their monkey subjects effectively searched more items on each attentional fixation when the search was relatively easy as measured by total reaction times at a given set size.

The primary focus of Motter and Belky's study was on the effect of stimulus spacing on the probability of target capture at different retinal eccentricities. They showed that normalizing the retinal distance by the average nearest neighbor distance (ANND) produced a constant probability of target capture across set sizes. These results indicate that stimulus crowding controls the ease of target location for any given eccentricity and stimulus set. These results are consistent with the conclusion drawn in chapters 3 and 4: limitations in visual search performance are largely caused by unclear neural representations when more than one stimulus item is within the receptive field of the relevant neurons.

Motter and Belky (1998b) showed that normalizing retinal eccentricity into units of average average nearest neighbor distance produced a curve that was constant across set sizes, but not across stimulus sets. Different stimulus sets produced curves with the same shape but quite different locations. Thus the number of items effectively searched during each fixation varied with the type of stimuli.

These findings reinforce the conclusions of the present study. Previous theories of parallel search must be expanded to account for the finding that different numbers of items are effectively processed with different stimulus sets. The current findings show that the number of items searched in parallel is not a product of the capacity of visual working memory, as has previously been suggested (e.g Duncan & Humphreys, 1989), since that capacity should be roughly constant across stimulus sets, rather than varying with the relationship of targets to distractors.

One important disclaimer is that these results do not uniquely support a "high threshold" theory in the terms of signal detection theory. In such a theory, the number of items processed on each attentional fixation is determinate, with each item either being identified with certainty or not identified at all. Indeed, in light of Experiment 2 and the results of Motter and Belky (1998b), it seems more likely that each stimulus is processd partially, with some processing occuring for every item in the display.

The search functions seem to have break points as predicted by serial-parallel theories of search. However, the wide variance in best fit break points suggests that the break points are not consistent across observers. Although many observers showed best fit break points at set size 16 (as suggested by the mean RT data, figure (1.7), this was true of less than half of the observers even in the EvF condition, and the remainder were widely distributed. Because the goodness-of-fit was not compared across observers, some of this variance may be due merely to noise. However, it seems likely that the wide variance may also be caused by strategy differences between observers, tasks, and order. Since observers had practice with one stimulus type before seeing the other, they may have in some cases adopted a different strategy after practice. A strategy of delaying eye movements may allow more items to be processed in the first attentional fixation, as discussed in section 2.4. This difference would produce a higher break point on average. This is consistent with the subjective impression of strategy choice in the task; one can either maintain fixation and respond based on the target location coming to attention, or

eye movements can be made to potential targets as quickly as possible. The issue of possible strategy differences is discussed more fully in Chapter 4.

While it seems likely that the break points observed are the result of long initial fixations, it is also possible for a parallel model to produce such varying breakpoints. It could be supposed that the operation of such a model becomes less time consuming with additional information in a nonlinear fashion; extra information slows a decision, but beyond a certain amount of extra information, the rate of slowdown becomes more modest.

While such a model does not seem a priori likely, nonlinearities in the human information processing system abound. One test of this alternate hypothesis is performed in Experiment 5.

Because these current results are best viewed in light of a more complete review of the literature and the issues involved, a full discussion is left to the section on theory (4.2.2).

1.4 Experiment 2

Experiment 2 was designed to provide converging evidence for the estimates of number of stimuli perceived in parallel given by Experiment 1. Brief stimulus presentations were used, to prevent eye movements, and possibly attentional movements. The reasoning was as follows.

If observers generally choose to move their attention (and/or eyes) when more than X number of objects is present, and X varies depending on the discriminability of targets from distractors (as tested by Experiment 1), then the number of objects that can be discriminated in a short time should vary in the same order. That is, if observers move their eyes after identifying four T vs L stimuli, but after eight E vs F stimuli, then we would expect that they can in fact identify at least eight of the latter and four of the former in a single attentional fixation.

The method of Experiment 2 was to present stimuli briefly, so as to allow only a single attentional fixation. Observing the conditions at which accuracy drops below ceiling should then identify the limits of processing within a single attentional fixation, and these should icorrelate with the set sizes at which observers choose to use attentional shifts, as estimated in Experi-

ment 1. This method is attractively simple, but it rests on the as-yet-untested hypothesis that a brief presentation will prevent attentional shifts. In fact this interpretation is novel; prior experiments that used a brief presentation were interpreted in terms of serial search performed on the iconic memory of the image (e.g. Bergen & Julesz, 1983). However, more recent evidence indicates that attention probably moves a good deal slower, around 150 ms for a voluntary shift, as discussed in section 1.1.5.

Each set of stimuli were therefore presented for only 100 ms, too short for any voluntary movement of attention, and very likely too short for any helpful involuntary movement. After an additional 100 ms, a mask appeared. The total time of 200ms is likely long enough for a single attentional movement, and it is clear that an iconic memory can be attended after physical onset (Sperling, 1956, Bergen & Julesz, 1983). However, it seems unlikely that more useful information can be obtained after a first attentional fixation has promoted some information to higher processing. It is possible that attending to one portion of iconic memory will act to disrupt the remainder. It seems likely that only one fully effective attentional fixation is possible, but the interpretation of this experiment will not rest on that possibility. Rather this method will be cross-verified with that used in experiment one.

1.4.1 Methods

1.4.1.1 Participants

A subset of 26 participants for Experiment 1 (those from the introductory psychology subject pool, rather than paid subjects) were also tested on Experiment 2. Experiment 2 was presented separately, and always after Experiment 1.

1.4.1.2 Stimuli

The stimuli employed were identical to those from Experiment 1. However, only three set sizes were used, since piloting indicated that performance on larger set sizes was near floor

performance. The set sizes tested for the E vs F stimuli were 2, 8, and 16. The set sizes tested for the T vs L stimuli were 2, 4, and 8 items. The set sizes tested for the shape stimuli were 1, 2, and 3 items. Masks were constructed from the same stimuli, placing one target and four distractor items at the same spot that had been occupied by each stimulus, each with a small (approx 1/4 degree) random jitter on both X and Y coordinates. The resulting masks are shown in figure 1.8.

1.4.1.3 Procedure

On each trial, the fixation cross appeared for 1000 ms, before the stimuli appeared. They remained visible for approx 100 ms (the software used does not have millisecond timing capabilities, but times were accurate to about 10 ms). After an additional 100 ms, the mask display appeared. The mask remained visible until a response was made. Feedback was given as in Experiment 1, but without information on the response time.

Stimuli were blocked as in Experiment 1; each type of stimuli was isolated to a block, and trials of each set size were isolated to sub-blocks. Ten trials each of target-present and target-absent displays were randomly ordered within each sub-block. A rest period was given between each block, as well as between Experiment 1 and Experiment 2.

1.4.2 Results

Results are shown in figure 1.4.2. Accuracy falls off dramatically with set size, and the Shapes stimulus set shows very low accuracy relative to the other two stimulus types, in line with the longer reaction times for the Shapes set in Experiment 1. The separation between lines could show one of two effects. The first is systematic response biases toward target present responses; an optimal decision making strategy would place the decision criterion so that an equal number of target present and target absent response are made (if target detection and rejection have equal utility). The second possibility is that identification is actually better for

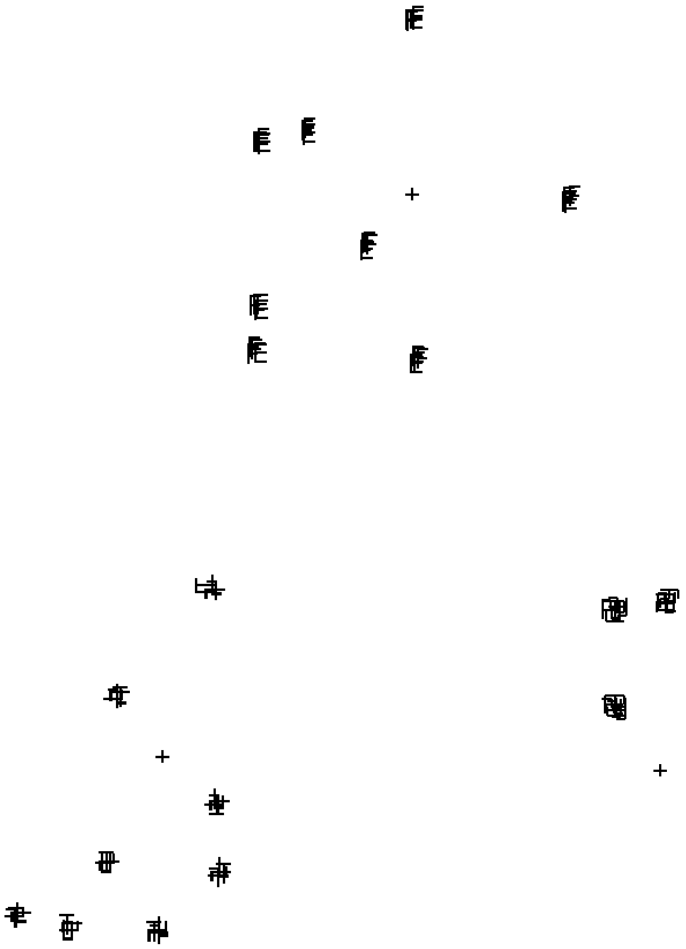

Figure 1.8: Mask displays from Experiment 2. The displays were shown in white on black. Masks appeared 200 ms after the onset of the display, including 100 ms of the search display, and 100 ms of a blank screen.

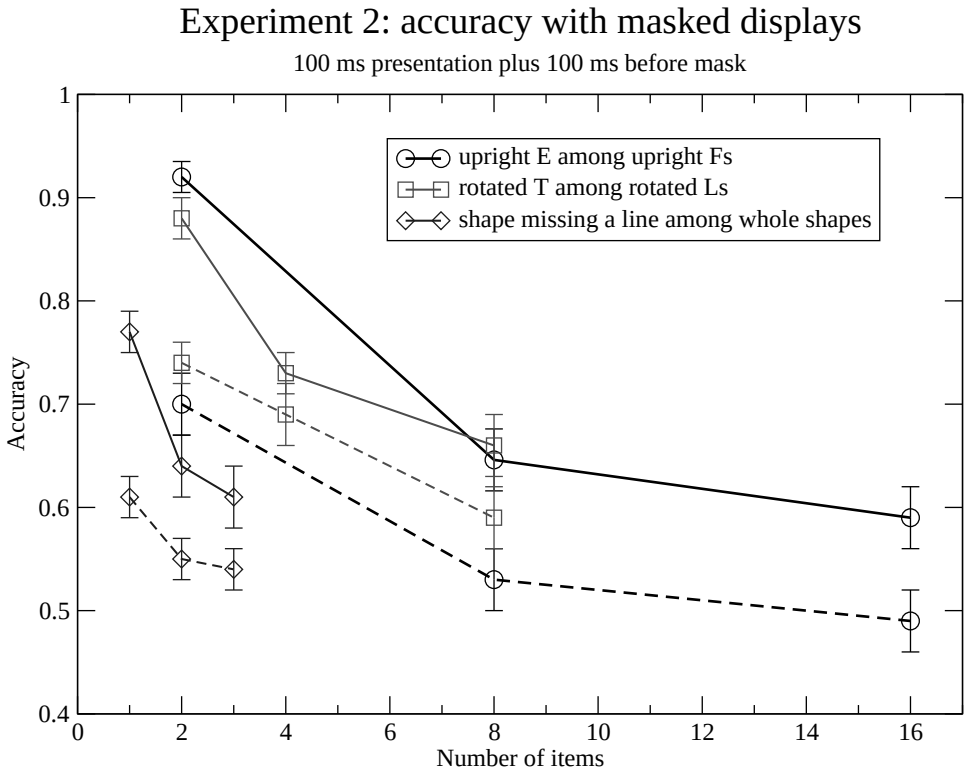

Figure 1.9: Results of Experiment 2. Each point is accuracy at a given set size, with targets scattered at random. Solid lines are target present trials; dashed lines are target absent trials.

Set Size	1	2	3	4	8	16
EvF # processed	-	1.26	-	-	1.44	1.64
EvF prob. each	-	.63	-	-	.18	.12
TvL # processed	-	1.22	-	1.72	2.08	-
TvL prob. each	-	.61	-	.43	.26	-
Shapes # processed	.39	.36	.45	-	-	-
Shapes prob. each	.39	.18	.15	-	-	-

Table 1.3: Summary of identification estimates as a number of items identified and probability of identifying each item

.

target present trials; judgments of target presence are more certain, whereas the target absent response is always a guess.

The results are analyzed as follows. The number of stimuli perceived on average, assuming (probably incorrectly) that identification is discrete, that is, it is performed perfectly accurately on some stimuli, and is not performed at all for others. Second, the chance of perceiving each item is calculated, as a measure of the extent to which item processing happens in parallel.

The first estimate is constructed by taking the average of accuracy across present and absent conditions (to control for response biases), and normalizing for the range between chance, .5, and 1, perfect identification of each stimulus. At a set size of four, and a hit rate of .8 for target present, .7 for target absent, and a set size of 2, we would estimate that (.8+.7) = .75 - .5(chance) = .25 * 2 (normalizing) = .5 chance of identifying each item or .5 * 4 (number of items) = 2 items identified if we assume a set number of items identified.

For instance, using this method, we would estimate that the number of Ts and Ls perceived at set size 4 is ((.74+.69/2)-.5)*2*4=1.72, The first part of this calculation represents the chance of each individual item being correctly identified; in the example above, this is ((.74+.69/2)-.5)*2 = .43. So each T or L item has a .43 chance of being identified at set size 4. The complete set of such estimates is shown in table 1.3.

1.4.3 Discussion

This set of estimates above is inconsistent with both a serial process acting item-by item, and with a capacity unlimited parallel process with a high threshold perceptual state (one in which each item is either identified accurately or not at all, with no uncertain states). A serial process acting item-by-item should show a constant number of items perceived if a constant time per item is assumed; if the time is greater with more crowding, the item estimates should actually go down instead of up with increasing set size. An unlimited capacity parallel process with no perceptual uncertainty should show a constant chance of identifying each item. Instead the estimates show an increasing number of items effectively searched as set sizes increased, while the effective chance of identifying each item grows smaller as set size becomes larger.

This pattern of results is could be the product of a serial process acting over several stimuli at once in parallel, (as discussed in section 1.1.6 in the introduction and further in section 4.2.2). The results could also be a product of a purely parallel process limited either by perceptual uncertainty (see section 2.2.1, or by negative interactions between multiple perceptual process (see section 3.2.2).

In any of these cases, this paradigm does not yield estimates for number of items perceived in parallel that are consistent with the results of Wolfe et al. (submitted), or with the estimate from Experiment 1 of about 6 for the TvL stimuli. If only about two T among L stimuli are distinguished in parallel, their results should show a change in linear slope between set size two and four. They clearly do not; see figure 1.5.

One explanation of these data is that an attentional fixation usually lasts longer than 200 ms, and the perceptual process normally in place are thus limited by the lack of time. This is reasonable given the 200-300ms range for first saccade latencies in eye tracking studies (see section 2.4). This long attentional fixation hypothesis is in sharp distinction to the more common fast serial model (see section /refsect.speedatt).

There is an additional difficulty with these data. They seem to indicate that the EvF

stimuli are perceived less accurately than the TvL stimuli. This is at odds with the finding of higher break points on average in Experiment 1; it is also at odds with the longer latencies of search for T among Ls (although this could accounted for if we assumed that attention moved much more quickly among the EvF stimuli than the TvL stimuli, there is no reason to suspect such large differences in the speed of attention).

The deficit in judging target presence in the EvF stimuli is largely in making correct target absent judgments. This suggests an alternate hypothesis. The subjective impression in performing this task is that the mask resembles the E stimuli, so that the impression is always that an E is present. Inspecting the masks in figure 1.8 reveals that those for the EvF stimuli reveal the lower bar of the E character. In this case the mask seems to perform an effective backward masking of the stimuli, since it contains features relevant to the decision task. This could happen if decision processes are unable to ignore information delivered after the search display ends and the mask display appears; in this case, the information from the mask display indicates that a target is present at every location.

Masks for the other characters do not reveal such clear discriminative features, nor do they seem to produce the subjective impression that the target was present on every trial. However, it seems possible that some amount of backward masking is at play in these situations, as well.

These results indicate that the choice of masking stimuli is crucial in visual search paradigms, since they can also be regarded as discrimination tasks that are effected by backward masking. Clearly the use of features important to the discrimination task harms task performance. However, masking is ineffective if the mask shares no features of interest with the task. The use of a mask can disrupt processing that is already completed, rather than simply halting it.

A more useful SOA paradigm for relating fixed duration search displays with persistent displays would be the use of successive displays, each with items in different positions, and

each with the possibility of containing the target. This situation is more closely analogous to that produced by a series of eye movements during normal search.

In light of the difficulty of EvF stimuli in this paradigm, and the severe mismatch with the data from Experiment 1 and the previous estimates for TvL stimuli (Wolfe et al., submitted), it seems unwise to take seriously the estimates of number of items identified in parallel that are produced here. There are two likely explanations for the relatively low estimates of items identified in parallel that would be obtained. The first is that normal attentional fixations last longer; this is likely part of the explanation, since fixation durations are usually longer than 200 ms. The second explanation is backward masking from distractors. The accidental inclusion of a discriminating feature in one of the stimulus masks highlights the effect that choice of mask can have on the outcome of brief presentation displays.

However, factoring out the likely effect of that particular mask, the stimuli are in the same order of difficulty as the ordering revealed by the break analysis of Experiment 1. The shape stimuli are much more difficult to make judgments on than either TvL or EvF, while the EvF stimuli are likely slightly easier than the TvL (there is clearly a response bias toward target present responses in this case; however, it seems likely that a less closely related mask would allow better overall performance).

The results of Experiment 2 are thus in broad agreement with those of Experiment 1, but suggest that attentional fixations are usually longer than 200ms, and/or that masking severely harms performance. They may also suggest that an eye movement made to check a potential target's identity is crucial. This possibility is investigated in the Experiment 5. However, it can also be argued that using brief displays elicits search strategies quite distinct from those elicited with persistent displays, essentially changing the task from a search task to a decision task with complex information. This argument is made in section 4.2.1.

1.5 Experiments 3 & 4

Experiments 3 and 4 were designed as an independent test of the hypothesis that visual searches with these stimuli generally proceed by a serial-parallel search by locations. These experiments employ multiple targets (12 targets in each target-present display in Experiment 3; 8 in Experiment 4). These targets are either grouped together, or spread throughout the display at random. If attention moves from region to region in a display with multiple targets, it will encounter at least one target much faster on average if they are spread throughout the display than if they are localized to one region. For instance, if attention rests on 1/4 of the display at once, and each portion of the display has a target within it, we would expect reaction times to be approximately twice as fast on average as the situation in which targets are present in only one quadrant of the display (since the quadrant with targets will be visited after 2.5 attentional fixations instead of 1 when targets are in all quadrants).

If, on the other hand, attention visits locations singly, we would expect that the two displays will be searched in the same amount of time. This would also be the case if the number of items inspected in each fixation is small (searching two items at a time at random would gain nearly no advantage from targets spread throughout the display).

It is also possible that these searches are guided (although the Guided Search model considers the difference between Ts and Ls to not be a guiding feature, Wolfe & DiMase, 2003). If this is the case, grouping many targets together could even speed reaction times, by producing a large guidance signal from one location of the guidance map.

1.5.1 Methods

1.5.1.1 Participants

Participants consisted of a subset of participants from Experiment 1. 8 people participated in Experiment 3; these were participants from the introductory subject pool. 22 people

participated in Experiment 4; these participants were also a subset of those that participated in Experiment 1, but this subset were all paid and recruited through the psychology department's paid subjects website. These experiments were always run after Experiment 1 (and Experiment 2 for observers in Experiment 3), so participants had practice with the stimuli.

1.5.2 Stimuli

The design of experiments 3 and 4 used many more distractors and multiple targets. In Experiment 3, each display consisted of a filled 19 by 11 grid was presented, subtending about 35 degrees by 22 degrees of visual angle, and containing 208 items (the center spot was reserved for a fixation cross, although we did not ask that observers maintain fixation during the trial). In Experiment 3, the targets were rotated Ts, and the distractors were rotated Ls, identical to those used in the other experiments. Twelve targets were presented in the target present conditions, which again composed half of all trials. Targets were constrained to appear within four spaces of the display edge, to reduce the number of trials in which targets were located on the first fixation.

In Experiment 4, targets and distractors were the shape stimuli from Experiment 1 and 2. Displays consisted of a filled grid 9 by 7 spaces for 62 items, with the center space reserved for the fixation cross. This display subtended about 25 by 20 degrees of visual angle. 8 targets were presented on each target present trial, constrained to appear within two spaces of the edge. In both experiments, individual items were offset by a random jitter factor of up to about 1/2 degree.

In both experiments, targets appeared in two arrangements, either randomly placed or grouped. In the ungrouped condition, targets were placed at random, with the constraint that targets could only appear within four spaces of the display edge . In the grouped condition, all targets had to fall within two horizontal and two vertical locations of a central target. Examples of the displays used (with colors reversed for printing) are shown in figure 1.10

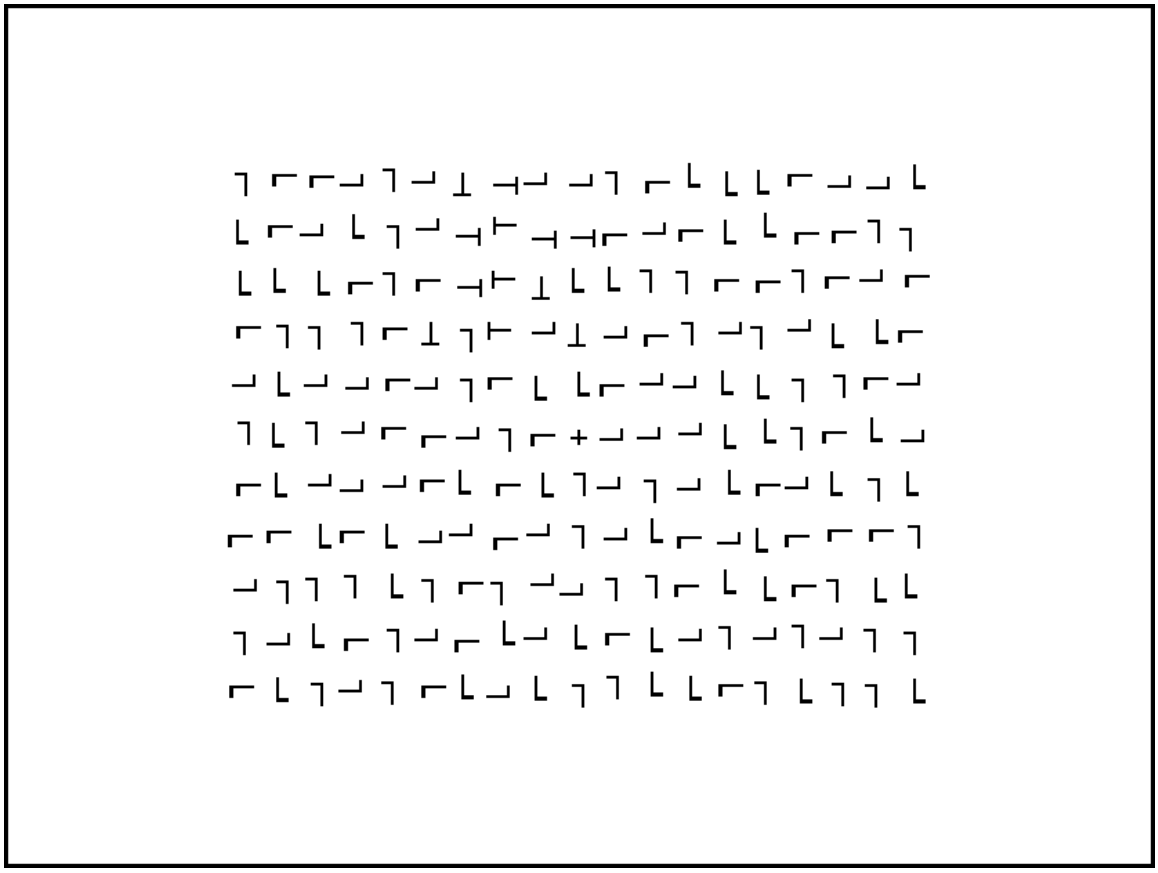

Figure 1.10: Displays from Experiments 3 and 4. The top display is from a grouped target present condition of Experiment 3 with the outline showing the size of the monitor; the bottom display is from an grouped target present condition from Experiment 4.

1.5.2.1 Procedure

All trials took place within one block. Ten grouped target present displays and ten ungrouped target present displays were randomly intermixed with twenty target absent trials. There were also eight practice trials. Display presentation was as for Experiment 1; a fixation cross appeared for 1000 ms, followed by the search display, which remained present until a response was made. Feedback was given, which included reaction times when the response was correct. The feedback display remained for 1.5 s, then the next trial began. There was a 5 second maximum time in Experiment 3; Experiment 4 eliminated this limitation to allow for the more difficult search.

Observers were not informed that some target present trials would have grouped targets. They were correctly informed that targets would be present on half the displays.

1.5.3 Results

Reaction time results were analyzed only for correct target present trials; accuracy results were analyzed only for target present trials. In Experiment 3 the mean RT for the ungrouped condition was 1638 ms; the mean of the grouped condition was 1662 ms. The standard error of the mean difference was 118 ms. A paired T test indicated that the difference in reaction times was not close to significance;, $t(8) = .20997$; $p = .84$. The mean error rate for the ungrouped condition of Experiment 3 was .91; the mean error rate in the grouped condition was .76. A paired T test showed a significant difference in error rate, $t(8) = 4.9$, $p = .0012$. Standard error of the mean difference was .029.

In Experiment 4, the mean RT for the ungrouped condition was 1844 ms; the mean of the grouped condition was 2145 ms. The standard error of the mean difference was 179 ms. A paired T test indicated that the difference in reaction times was marginally significant, $t(22) = 1.682$; $p = .1$. The mean error rate for the ungrouped condition was .94; the mean error rate in the grouped condition was .84. A paired T test showed a significant difference in error rate,

$t(22) = 3.2$, $p = .0046$. Standard error of the mean difference was .03.

1.5.4 Discussion

Experiment 3 revealed that reaction times were little affected by the difference between grouped and ungrouped targets. However, the accuracy between conditions was significantly different. The results of Experiment 4 are similar, but suggest that there are probably differences in reaction times, although these are likely small differences.

The results of both experiments are consistent with the hypothesis that displays are searched by areas. However, they also suggest that observers used a strategy of sampling only a few areas of the display, since this is adequate to find a target when it is present and scattered at random. No observers report noticing that targets were grouped in Experiment 4, although a few observers in Experiment 3 commented on the grouped targets.

The lack of reaction time differences in Experiment 3 also suggests that some amount of guidance was present. Grouped targets subjectively seem to form a different “texture”, and it is possible that some observers focused on this as the target of search. Some observers commented on the fact that targets were grouped when asked about strategy. A split between these strategies would explain the similar reaction times on correct trials between the grouped and ungrouped conditions.

In Experiment 4, some observers reported using a strategy of searching the perimeter of the display, or of a line-by-line scan as in reading (with the more difficult display of Experiment 4). However, most observers reported a pattern of search that divided the display approximately by quadrants. Because this pattern would include the targets on most grouped trials, it is likely that targets were missed even when they were fairly close to fixation. This would also help account for the fact that reaction times were not nearly as different as a search by quadrants or even sub-quadrants would allow.

While these results are consistent with the hypothesis of a search by areas, there are sev-

eral alternate hypotheses. Some parallel models could predict the accuracy difference, although by supposing that additional targets within a region are detected by a large scale detector, and that this detector's accuracy is affected sublinearly by the number of targets within its scope; that is, the accuracy of many detectors working on individual targets could exceed that of a single detector with many targets within its scope. In addition, completely serial models might suppose that attention moves to items near the current fixation in preference to moving further across the display. This would also account for the results.

Therefore the results of these experiments suggest a serial-parallel search by area, but do not eliminate other explanations convincingly. The same could be said of all the previous experiments as well. Because the results of the experiments discussed so far have many features that are suggestive of particular explanations of search, but none that are conclusive, they are better interpreted within a broader review of the relevant literature. That review is performed in the first two sections of the second chapter, and the further interpretation takes place in the concluding chapter on theory.

1.6 Experiment 5

Experiment 5 serves two purposes. The first is to establish that the pattern of results from Experiment 1 is not due to averaging over a large number of observers. The second is to compare the pattern of results of searching these displays with eye movements to the pattern produced without eye movements.

In order to obtain more reliable data for single subjects, only four observers were tested, with many trials for each one. Two observers (one of them the author) repeated the same paradigm from experiment one, 8 times each. Two other observers repeated only the Ts and Ls condition from Experiment 1, ten times each, and did so without making eye movements.

1.6.0.1 Participants

Participants were graduate students in the cognitive psychology program. They were paid approximately $13/hour. Participants (other than the author) were not aware of the hypotheses being tested.

1.6.0.2 Stimuli and Procedure

Stimuli and procedures were identical to those of Experiment 1, with four exceptions. Trials were self-paced, with feedback screens remaining visible until a key was pressed. Second, participants in the no-eye-movement condition did not see the shape condition blocks. Third, trials were not limited to five seconds, to allow for longer search times. Fourth, observers in the no-eye-movements condition were instructed to indicate eye movements by pressing the zero key, either instead of making a response, or at the feedback screen. These trials were eliminated from the analysis. About 3% of trials were eliminated by this procedure.

1.6.1 Results

No inferential tests were performed on these data; instead, the relatively small uncertainty of each data point allows an interpretation based on a the relative Shapes of search functions with and without eye movements. Both similarities and differences between the two types of search are evident. Target present slopes and Shapes are similar with and without eye movements. However, target absent trials are substantially different; slopes are much larger when observers maintain fixation.

Results are displayed in figure 1.11 for easy comparison between the results using eye movements and for no eye movements. Results are shown full size for easier discrimination of small set sizes in the following figures. Reaction times are harmonic means as discussed in the results section of Experiment 1; error bars are SEM taken from RT distributions before log transformation, and so are not true standard errors of the means graphed, but only approxima-

tions.

1.6.2 Discussion

The data from the observers who searched with eye movements is highly similar to the group data from Experiment 1. This provides evidence that the shapes of the curves observed there are not solely the product of group averaged data.

The data also reveal some similarities and differences between search with and without eye movements. Because the differences are striking, it seems clear that eye movements were routinely used in the searches of Experiment 1. The use of eye movements is highly likely to at least introduce a large serial component to otherwise parallel processing. The logic underlying this claim is discussed in section 1.1.3.6 in the introduction. If a large serial component is present in the searches, it seems likely that the results of Experiment 1 are due to that component, rather than to nonlinear interactions within a parallel model.

First, the target-present trials at low set sizes for the TvL condition are strikingly similar. The two slopes for search with eye movements lie between the two without, indicating no substantial difference between conditions. Thus eye movements do not seem to affect the pattern of search results in this range. It has been reported that the presence or absence of eye movements does not constrain the pattern of RT data when display items are relatively large and spacing is wide (Klein & Farrell, 1989). However, other aspects of these data call this conclusion into question.

The target absent slopes are substantially larger without eye movements, even at small set sizes. Eye movements may make it easier to accurately conclude that no targets are present in one of two ways. First, they may allow potential target locations to be checked more quickly than covert attention. This hypothesis is just the opposite of the supposition of Wolfe et al. (2000) that “attention is fast but volition [and eye movements are] slow” (title). Second, it may be not the movement rate of attention but rather the time taken to discern potential targets from

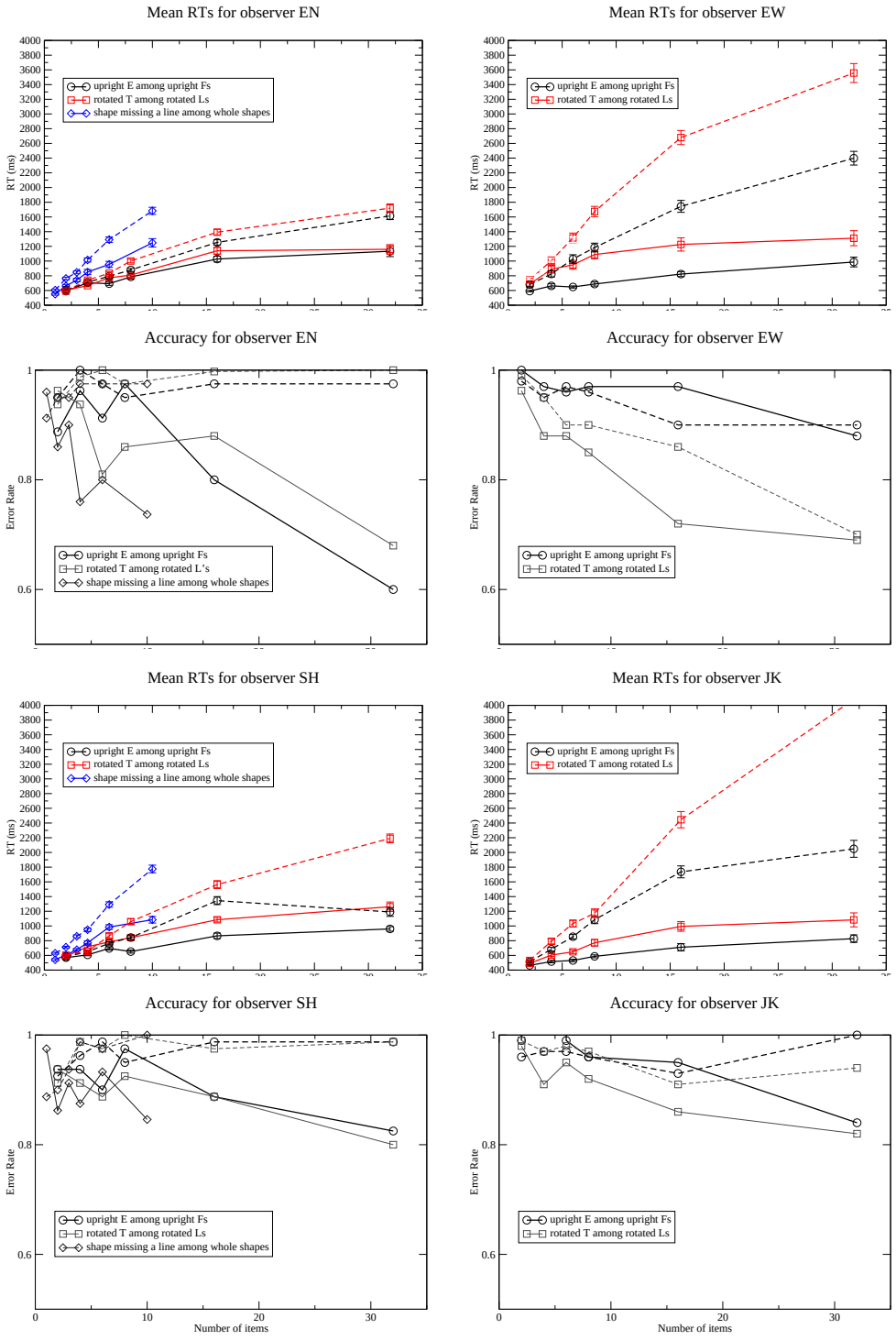

Figure 1.11: Data from Experiment 5. Each response time graph is paired with the associated accuracy data directly below. The two graphs shown on the left (observers EN and SH) are for search including eye movements; the graphs on the right (observers EW and JK) are for search without eye movements. Note the dramatically longer reaction times for large target-absent displays at large set sizes when eye movements are not made.

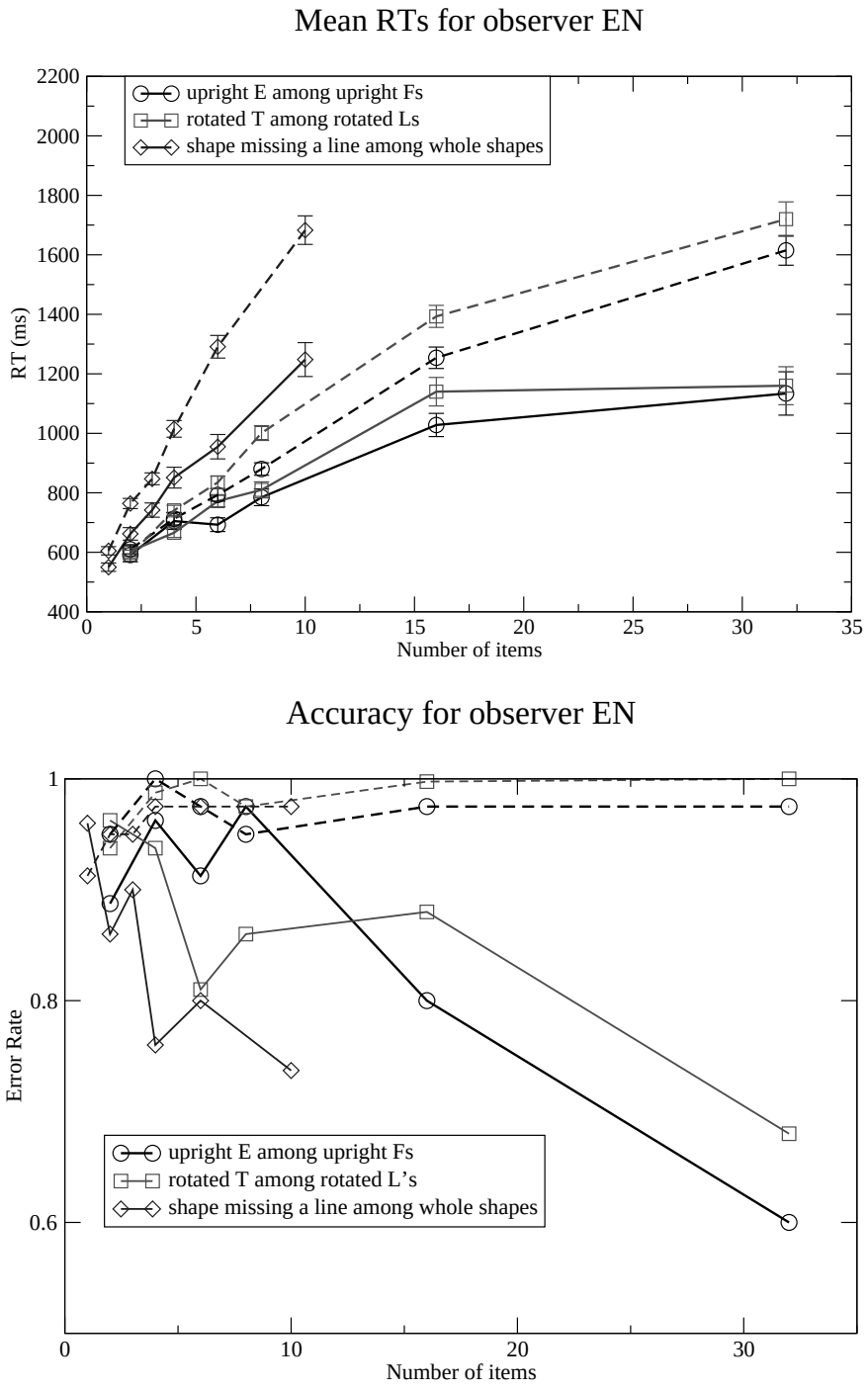

Figure 1.12: Data from Experiment 5, observer EN, who searched with eye movements. Axes are identical to those in the following graphs to allow easier comparison across observers.

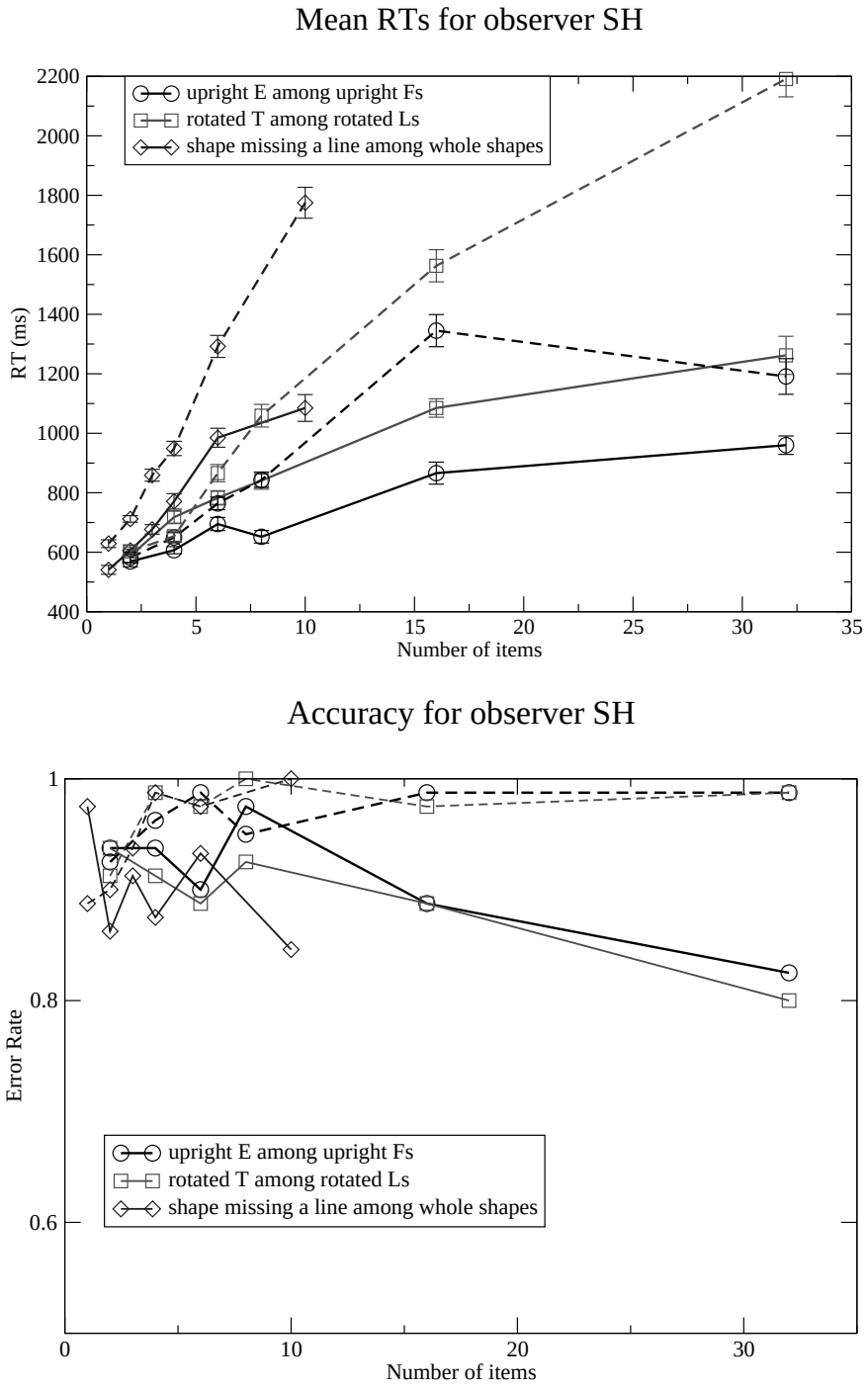

Figure 1.13: Data from Experiment 5, observer SH, who searched with eye movements. Axes are identical to those in the previous and following graphs to allow easier comparison across observers

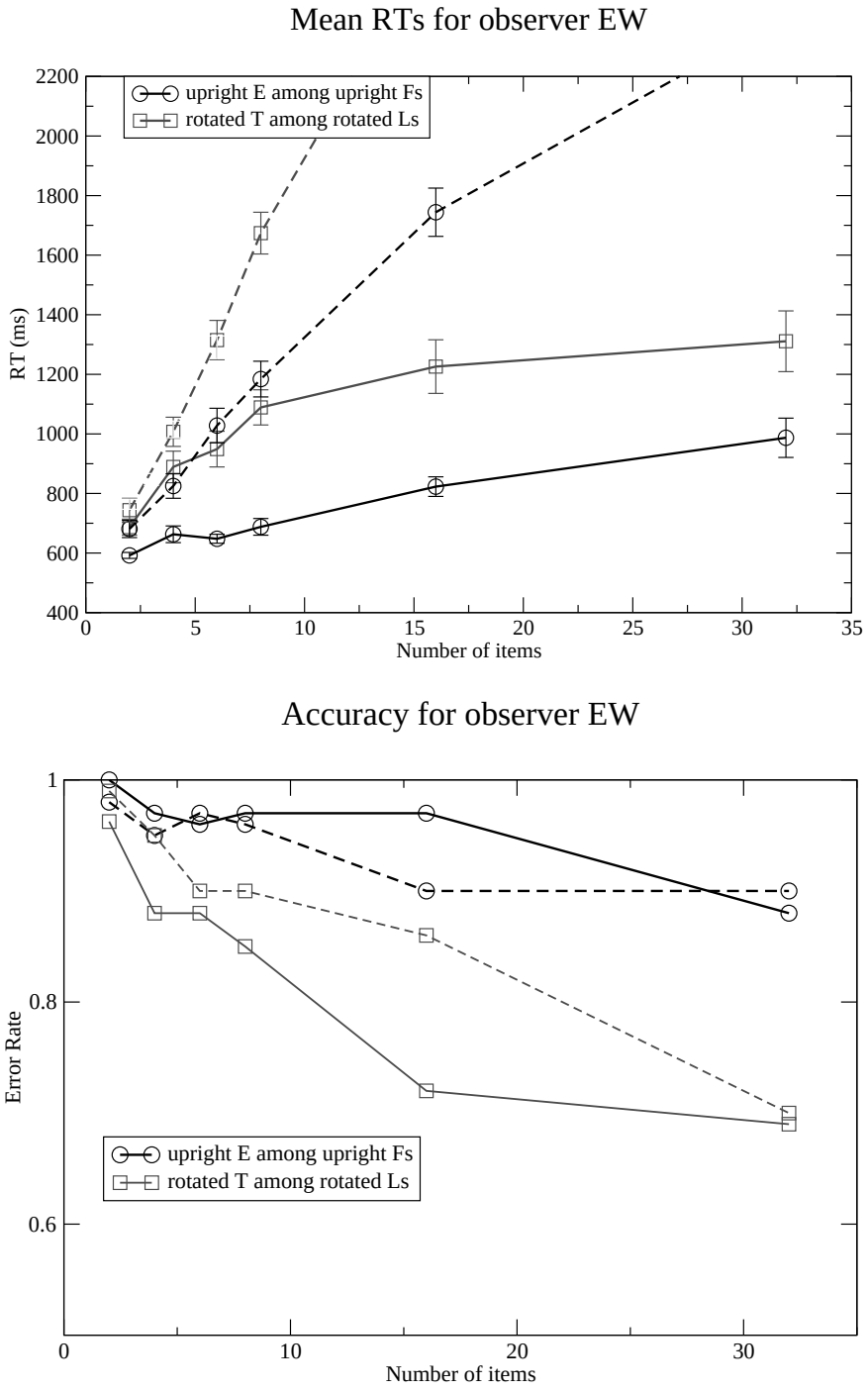

Figure 1.14: Data from Experiment 5, observer EW, who searched without eye movements. Large reaction times are not graphed to allow easier comparison of small set sizes across observers

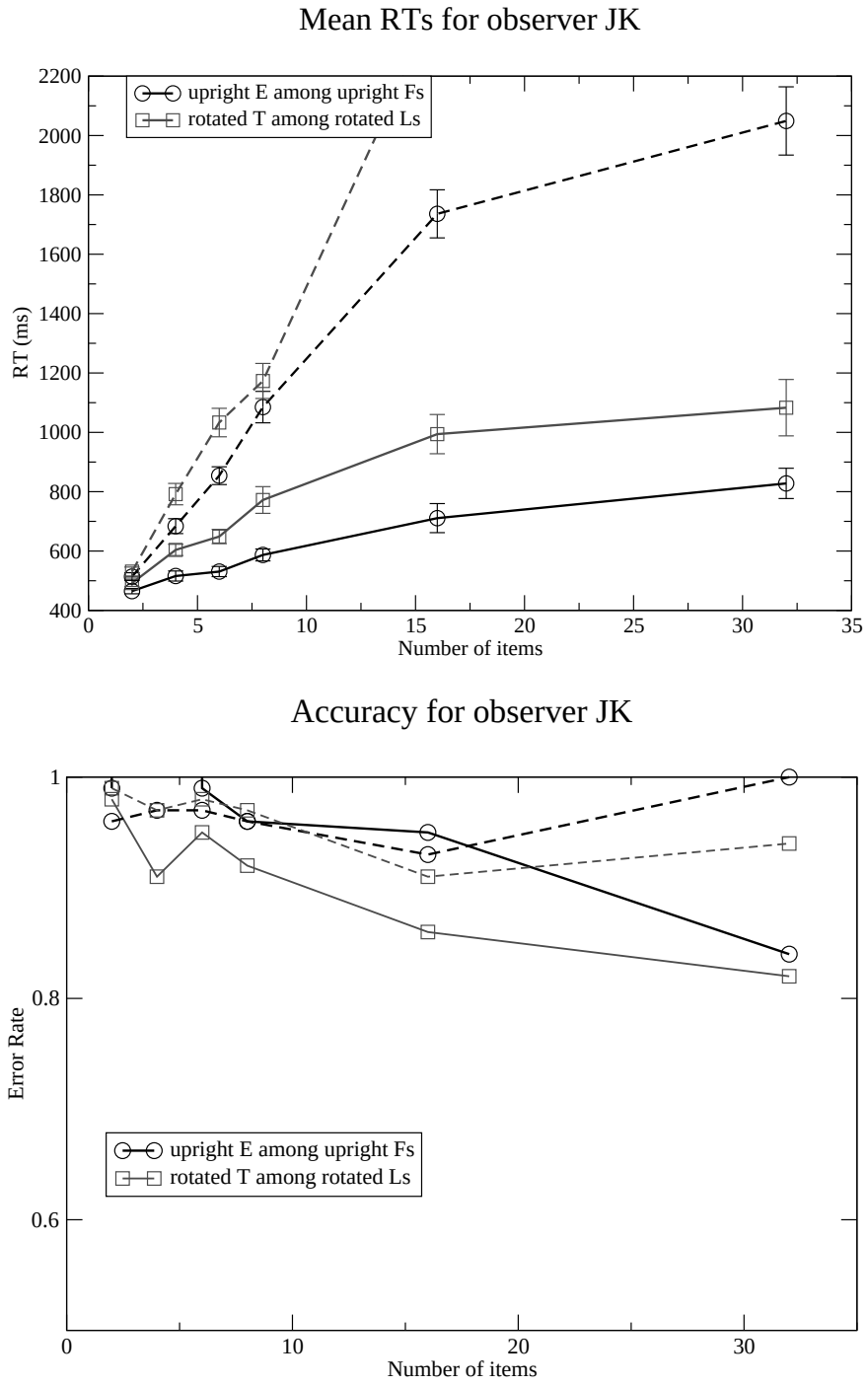

Figure 1.15: Data from Experiment 5, observer JK, who searched without eye movements. Large reaction times are not graphed to allow easier comparison of small set sizes across observers

distractors using covert attention. The possible reason for this is the large size of receptive fields responsive to differences as subtle as T vs L or E vs F. This is discussed in the review of receptive field sizes 3.2.4, and in the concluding theory section, 4.2.1.

Finally, these data suggest that some target present searches may actually be conducted more efficiently without eye movements. Although these data are obviously only preliminary, the search slopes for target present trials in the EvF condition appear more linear than those in the condition with eye movements. The prevention of eye movements may have lead to a strategy of parallel search; those two observers gave subjective reports that do not include movements of attention prior to attention settling on the target. The data suggest that, for locating the target when it is present in the EvF stimuli, this strategy may be more effective than the serial movements of attention that result from eye movements.

This raises the possibility that search is performed by default strategies even in situations in which other strategies are more effective. There is also evidence for this in the study of Scialfa and Joffe (1998) in which two observers practiced conjunction searches with eye tracking; after a number of sessions, both observers stopped making eye movements and reaction times improved while accuracy remained constant. It is possible that these strategies do not require practice, but are usually discovered through practice. If this is so, training with specific instructions could prove quite effective for some types of search.

The results of Experiment 5 suggest some possible strengths and weaknesses of different strategies. These speculations require more experimentation, and are not pursued further here. However, the results clearly indicate that eye movements were used in the task of Experiment 1, and shows that they are especially important in verifying that no target is present.

1.7 Experiment 6

This experiment explores a question completely different from that of the preceeding experiments. It is best understood in relation to the section on efficient search of a subset in the

second chapter, section 2.3.2. It is presented here for consistency, but the reader is advised to skip ahead until the context of the experiment is explicated.

Experiment 6 eliminates an important possible explanation for the findings of efficient search of a color-defined subset of items. The previous experiments demonstrated that searching is efficient for a conjunction defined by a known color, and an orientation different from that of the other items of that color (Friedman-Hill & Wolfe, 1995) (See figure 2.2 for an example of the displays used). There are two possible ways this efficient search could have been performed.

In the previous experiments, it was possible to deduce the identity of the target item without finding it, by looking at the orientation of the non-target-colored items. Search could then proceed in the ordinary fashion, with a target template exactly specifying the target's features. The finding that subset search takes about 150ms longer than a standard popout search , and longer than a standard color-orientation conjunction search for the same set size, allows time for this process to occur.

The more interesting explanation is that top-down emphasis of the colored subset allows the item of unlike orientation to "pop out" of the display without any knowledge of that orientation. If this is the case, these findings demonstrate the existence of a distinct strategy of search that may be usefully applied to other conjunction searches as well; top-down attention is used to attend only to items sharing a relevant and easily discriminable feature. This possibility is also interesting because it would provide clear evidence that at least 8 items can be selected as a group.

Experiment 6 tests these two alternate hypotheses. It uses a paradigm similar to that in the studies of Friedman-Hill and Wolfe (1995), but instead of using a uniform angle of the irrelevant color set, it uses random angles. This manipulation makes it impossible to infer the angle of the target stimuli, and use that angle to guide search. Instead the item must perceptually pop out from the similarly colored items once that set is attended to. If top-down attention can eliminate some items from a perceptual popout effect, the mechanisms underlying popout must

be different and more flexible than has previously been thought.

1.7.1 Method

1.7.1.1 Participants

The 12 observers in this study were a subset of those from Experiment 1. All observers in this experiment were paid, and recruited through the psychology department's paid subject pool. All had previously performed Experiment 1, so they had practice with search tasks, although with different stimuli.

1.7.1.2 Stimuli

Stimuli consisted of colored straight lines of various orientations. The lines were about 2.5 degrees in length and about .25 degrees in width. They were colored red and blue, and were highly saturated, with each color consisting of the maximum RGB setting for the computer monitor for that color. Precise hue and luminance information is not available; informally the colors were highly saturated and very bright.

Items were arranged in the same 7 by 7 grid used in Experiment 1, composing about 22 by 22 degrees of visual arc. The center position of the display was always occupied by a white fixation cross. Each items position was varied by a random jitter factor of about .74 degrees on each dimension. Each display consisted of an equal number of red and blue lines. The red lines were always the target color, and they all shared the same orientation, except for the target when it was present; that angle was chosen at random for each trial. When the target was present, its orientation was chosen at random, but constrained to be at least 30 degrees different from that of the other red lines. The orientation of each blue line was random, and unconstrained by any relation to any other line.

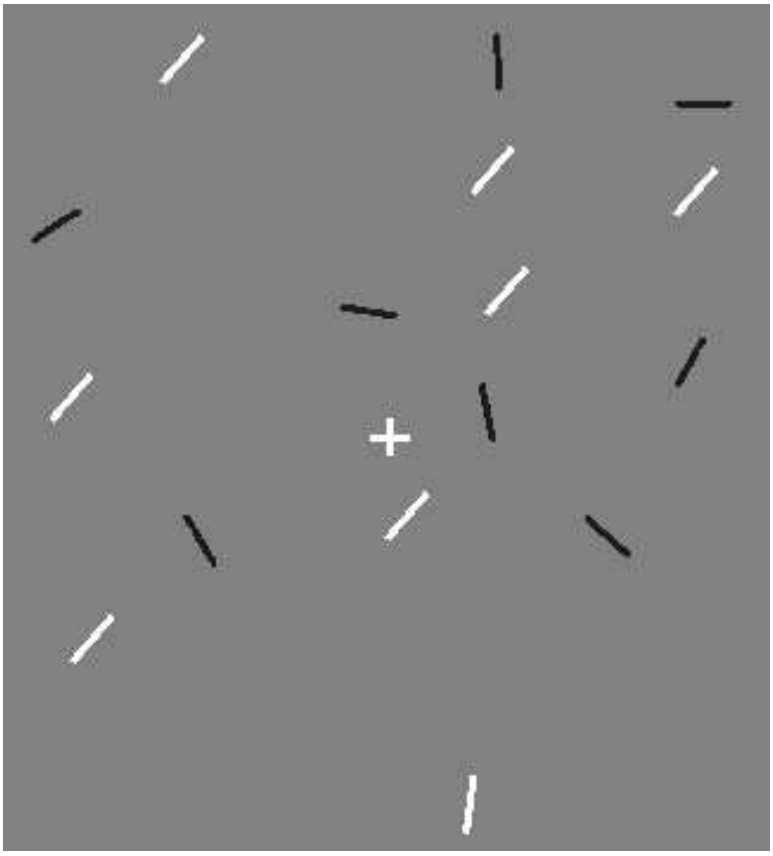

Figure 1.16: Display from Experiment 6. Target present condition with set size 16. Lines shown in white were red; lines shown in black were blue. The background was black, and the fixation cross was shown in white.

1.7.1.3 Procedure

Each trial began with presentation of a fixation cross. After 1000 ms, the search display appeared, and the cross remained. The search display remained visible until a response was made, or 5 seconds elapsed. If the 5 second time limit was reached, a screen appeared asking the observer to respond faster. Otherwise, a feedback screen appeared which included the reaction time on correct trials. This feedback screen remained visible until the observer pushed a button to start the next trial.

Five set sizes were used; these were 4, 8, 16, 24, and 32. Stimuli were presented in one block. Each set size was presented in a separate sub-block, as in the previous experiments, but the order of these sub-blocks was random, rather than ascending or descending. Each sub-block consisted of 20 target present and 20 target absent trials, for a total of 200 trials. 20 practice trials preceded the main block. Practice trials consisted of two target present and two target absent trials of each set size, and the order was random across both set size and present/absent condition.

1.7.2 Results

Results are shown in figure 1.17. Compare to the results of Friedman-Hill and Wolfe (1995), shown in figure 1.18. Reaction times increase with set size, but for small set sizes, target present and target absent reaction times are similar, a pattern that indicates parallel search at small set sizes. However, the higher error rates made in target present trials at all set sizes larger than four suggests a response bias toward the target absent response.

1.7.3 Discussion

The results of Experiment 6 are nearly identical to those of Friedman-Hill and Wolfe, (1995), despite superficial differences in the displays used. These results demonstrate that observers do not use the nontarget line orientations to identify the target orientation, and then

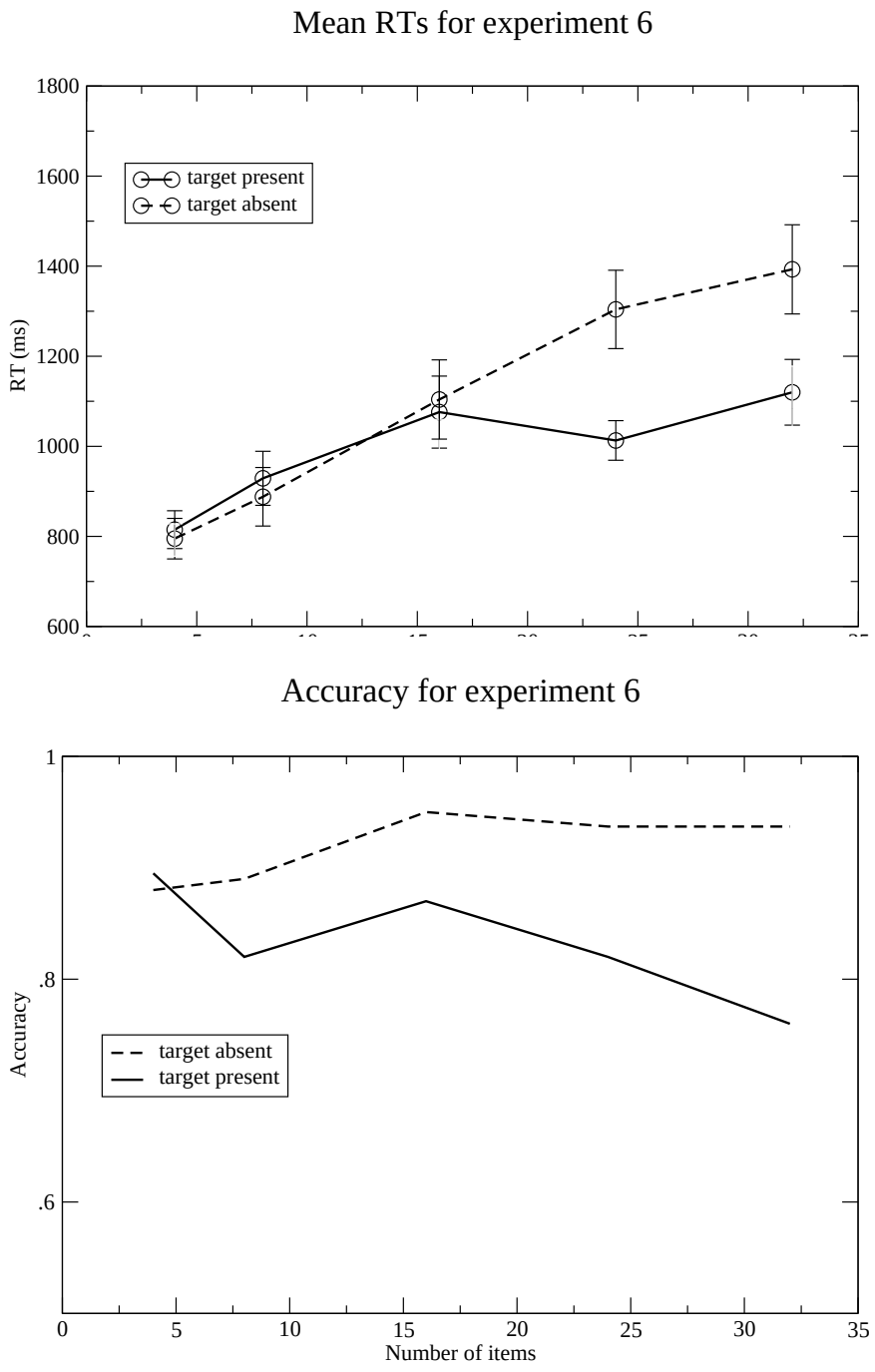

Figure 1.17: Data from Experiment 6. Subset search data at various times. Note that target present trials take as long as target absent trials for smaller set sizes.

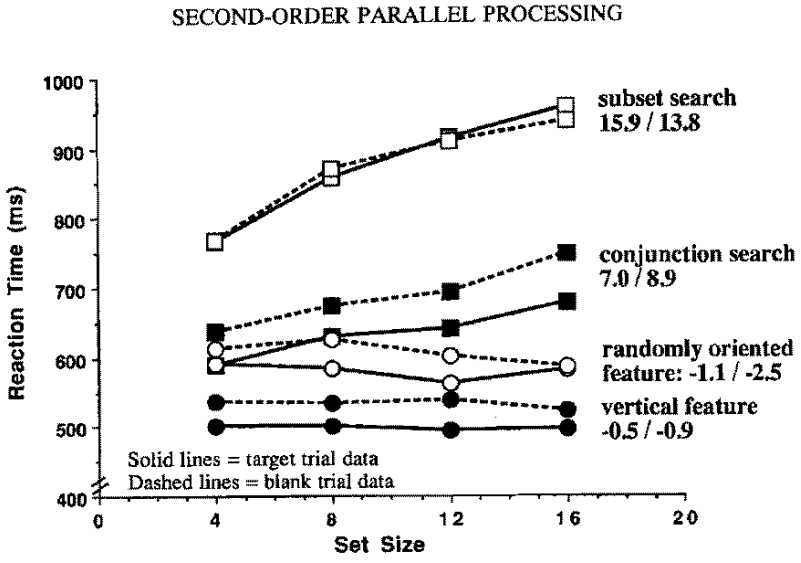

Figure 1.18: Data from Experiment 1 of Friedman-Hill and Wolfe (1995). The relevant line is labeled “subset search”. The comparison graph for conjunction search suggests a different mechanism. The subset search is unique in that target present trials are as fast as target absent trials for small set sizes.

search for the specific target, since this strategy was impossible with the randomly oriented distractor lines. Instead obervers must be able to tell the difference between a uniformly oriented subset and one with a single line of unlike orientation. This could happen because the target pops out once the relevant colored subset is selected (as discussed in section 2.3.2.1), or by keying response off of a different global neural signature that differs when there is an unlike oriented item in the selected subset (see section 2.3.2.3).

The results indicate that selecting the relevant subset takes time; the search takes about 150-300 ms longer than similar conjunction searches reported in Wolfe et al. (1989) and Friedman-Hill and Wolfe (1995). The mean reaction times of present and absent trials appear to be the same, a rarity in search tasks (other exceptions are discussed in section 2.3.3. Although the confidence intervals obtained in the current experiment are not sufficient to conclude that present and absent search times are truly the same (confidence intervals (about twice the SEM error bars shown in figure 1.17) for each individual data point span 200-400 ms). However, Friedman-Hill and Wolfe (1995) also reported identical mean reaction times between present and absent trials at small set sizes. The replication of this finding in two studies suggests that the mean reaction times at small set sizes are likely to be similar although probably not identical. The mean RTs for target-present and -absent seem similar at smaller display sizes, but become quite different at larger set sizes. This pattern of results would suggest that the search is parallel for a range of display sizes, and either becomes serial, or becomes more difficult so that it requires more time to make an accurate decision on whether or not a target is present at larger set sizes.

These results show a distinct mechanism of search different from that employed for searches in which the target identity is known in advance. These results are important to a complete theory of visual search because they demonstrate that many objects can be analyzed as a group simultaneously (up to 8 objects with a like number of a different color as indicated in this experiment; and any combination of 16 objects total are indicated in Experiment 2 of Friedman-Hill and Wolfe 1995). This grouping mechanism takes extra time, and seems to function serially

(see section 2.3.2), implying a top down attentional source. Although this mechanism is equally applicable to a standard conjunction search, observers do not seem to employ this mechanism in that case; see section 2.3.3 for evidence that this is the case. Because the same mechanism is not evident in every situation in which it applies, and because it seems to operate in serial, the obligatory, low level, and automatic grouping process proposed by grouping theories of search are not a likely explanation.

These findings are also important because of their relevance to the mechanisms by which popout occurs. A standard explanation of popout effects is that they are based on low level visual processes. see the section on popout effects (Appendix A) in the appendix for a discussion of these mechanisms and an alternate explanation.

For a more complete discussion of the implications of these results for theories of visual search, see section 2.3.2.

Chapter 2

Theories and Behavioral Evidence

2.1 Overview- Method and Goals

How does visual search work? The previous chapter followed the standard approach to the question, asking "how does visual search work in these particular experimental paradigms?". The following chapters attempt a more complete answer. This integrative review is based on the many partial answers provided by existing theories and behavioral evidence, and constrained by known visual system architecture and function.

In this section, I approach the question of how visual search works in its broader sense. I attempt to deal with the full range of existing theories of visual search. The goal is to identify not only which theories are correct, but under what circumstances each applies. Using this analysis of existing theories, I synthesize a theory at a broader scale, one that identifies both the algorithmic nature of visual search under different conditions, and the neural mechanisms that underly the full range of search strategies.

This review fills a gap in the current literature. Existing theories and work almost universally ask restricted questions about search ("How does search work in these situations" or "What is one mechanism contributing to search performance"). Work on integrating across types of search paradigms, theories, and neuroscience evidence will better situate future theoretical and empirical projects, by giving a progress report on what conclusions can be drawn from existing evidence, and what issues remain to be resolved.

2.1.1 Structure of Review and Theory

The remainder of the dissertation is structured as follows. This chapter presents a brief overview of the theory presented fully in chapter 4. The first section of this chapter then lays out the major types of theory that have been employed to explain visual search results. The second section is a review of some particularly relevant experimental findings, and an evaluation of the additional assumptions needed for each type of theory to explain evidence on all types of visual search. This review shows that no existing theory is capable of explaining the full spectrum of visual search findings, and so establishes the need for a new theory. In addition, it establishes an outline for that theory, by showing what additional assumptions are needed to allow existing theories to account for the full range of data.

In Chapter 3, the major neuroscience evidence is reviewed, and related to the behavioral evidence. Finally the conclusions reached are summed up in a neural level theory in chapter 4. This theory also address the algorithmic level strategies that arise from the proposed system, and I discuss which strategies are likely to be employed for particular types of search tasks. This theory is referred to simply as a neural theory of visual search, and abbreviated NTVS.

2.1.2 Outline of Conclusions

The central conclusion of this theoretical review is that no existing theory of visual search can account for the full range of findings, so different mechanisms must be at work in different experimental paradigms. The most important mechanisms underlying visual search are the following:

- **Neural response properties in the ventral object recognition stream.** Sections 3.2.1 and 3.2.2 describe a previously unrecognized link between behavioral target detection, and clear and abundant target representation at the neural level. The search situations that allow such representation are determined by the representational structure of the ventral stream, as discussed in section 3

- **Guidance of attention** by item features, as emphasized by Guided Search (Wolfe et al., 1989; Wolfe, 1994). However, the prevalence of such guidance is brought into question by data reviewed in section 2.4

- **Grouping by item features**, as emphasized by Spatial Object Search (SOS, Grossberg, Mingolla, & Ross, 1994). However, the cases that provide strong evidence for search by grouping, reviewed in section 2.3.2, require a substantially slower and more effortful mechanism than that proposed by existing theories. The presence of grouping mechanisms in standard conjunction search is unlikely but not impossible in the light of evidence reviewed in section 2.3.3.

- **A conspicuity zone** or visual lobe allowing effective recognition only near the center of gaze, (Motter & Belky, 1998b; Palmer et al., 2000). The contribution of such theories is discussed in sections 2.2.1 and 2.4. While these theories are useful, the evidence reviewed in sections 2.4.2.1 2.4.3 shows that the nature of the conspicuity zone is not an area within which target detection can occur, but rather a steady decrease in the probability of target detection with distance from the center of gaze, item crowding, and target/distractor similarity.

- **Signal Detection Theory (SDT)**, a set of considerations about making decisions based on multiple noisy perceptual signals. While these considerations are useful, more assumptions are required to form a mature theory of search. It is shown that some of the assumptions adopted by current theories (e.g. Palmer et al., 2000; Verghese, 2003) are not adequate to explain the behavioral data. In particular, the assumption of independent item representations, and a discrete conspicuity zone are shown to be unlikely in light of the evidence reviewed in sections 3.2.2 and 2.4.2.1 respectively.

The theories mentioned above explain a great deal about visual search when combined as outlined in the concluding chapter. I attempt to outline the domain within which each thoery

applies, since this is not made explicit by existing work. However, I also show that even within its domain of applicibility, existing evidence requires some modification of each theory.

NTVS proposes an account of search based at the mechanistic level, but extending to the algorithmic level that most theories address. Existing data on the function of the visual system is applied to explain the successes and failures of human visual search in a variety of situations. The evidence for these systems and their function is reviewed in chapter 3, and the systems are summarized at the start of chapter 4. In this chapter, the theory is discussed only at the algorithmic level. The mechanisms described in chapters 3 and 4 function to produce four distinct modes or strategies of search:

- **Parallel search without eye movements.** This strategy is employed when the search is easy or when the display is not present long enough for effective eye movements. The considerations of SDT (2.2.1) and the structure of the ventral visual object recognition system (3) explain search under these conditions. One important and previously overlooked factor in the success of parallel search is whether or not representations of a target are ambiguous due to stimulus crowding within the larger receptive fields of higher-level neurons (see section 3.2.2. This mode of search is discussed fully in section 4.2.1.

- **Serial search with eye movements.** This strategy is the one most commonly employed. Each fixation is a parallel search, with the same considerations outlined above. The usual mode of search is therefor a serial-parallel search, as discussed in Chapter 1. When decision processes have some evidence for target presence, an eye movement is used to foveate a target and verify its identity. The probability of identifying (and locating) a target falls off with the following factors: distance from the fixation point; similarity of target to distractors; and item crowding. The neural mechanisms giving rise to this function are presented in section 3 and summarized in section 4.2.1. New empirical evidence for this search behavior is presented in Experiment 1 of chapter

1, and existing evidence is reviewed in the introduction to chapter 1, and in section 2.4.2.1. This hypothesis offers an explanation of the common finding of 2:1 present-absent search ratios: search is serial-self terminating, but occurs over few enough areas for spatial memory to accurately track them all. In some instances, spatial attention is used to enhance processing of stimuli near fixation, at the cost of those farther away, as suggested by the zoom lens metaphor for spatial attention. The theory remains agnostic on both the guidance of search by combined features, as outlined in the Guided Search theory (Wolfe, 1994), and on the contribution of grouping of items, as outlined in the Spatial Object Search of Grossberg et al. (1994). Neither theory is necessary to explain known search findings. The neural mechanisms reviewed in chapter 3, section 3 and summarized in chapter 4 suggest that both of these strategies are possible, but possibly slower than a relatively unguided search on relatively ungrouped items. This mode of search is discussed fully in section 4.2.2.

- **Serial search with covert attention.** This strategy is generally used only when eye movements are prevented in a given experimental paradigm. Spatial attention acts to disambiguate representations of items in high level neurons with large receptive fields, since eye movements cannot place a potential target in the fovea where small RFs allow unambiguous representations. In all other respects, the same mechanisms and considerations are present in serial search with and without eye movements. This mode of search is discussed fully in section 4.2.3.

- **Serial search by grouping.** This strategy is used relatively rarely, in circumstances outlined in section 2.3.2. Attention is used in discrete steps, with attention first to a certain feature, and then spatial attention to areas containing that feature. Representaions of unattended items are attenuated by spatial attention to a relevant group. The target can then “pop out” of an attended group once it is selected, if the attended group produces a substantially different high-level representation when it does or does not

contains a target. This mode of search is discussed more fully in section 4.2.4.

The hypothesized mechanisms underlying the above strategies are based on a review of relevant neuroscience evidence in the third chapter. There is now solid evidence for attention to features and locations (as well as objects, although object attention plays little role in either experimental or real world searches). But accounting for search results requires taking into account two additional properties of the neural systems involved. One is the particular nature of feature and object representations in the ventral stream of the visual system, discussed in sections 3.2.1 - 3.2.4. The second is the mechanisms underlying attention, discussed in section 3.3.

2.2 Major Theories of Search

The review of behavioral evidence is organized in relation to existing theories. I begin by laying out the three classes of theory that are commonly used to explain visual search. It is shown that each of these types of theories can explain search most parsimoniously by adopting some assumptions of the other types of theory. The additional assumptions needed for each theory to explain the basic findings laid out in the introduction is discussed in the following sections that detail each type of theory, while the assumptions necessary to explain some particularly constraining findings are discussed in the section on behavioral evidence that follows.

2.2.1 Signal Detection Theories of Visual Search

One class of theories attempts to account for visual search results using basic considerations about making decisions based on imperfect perceptual processes. These considerations are codified in Signal Detection Theory (SDT, Harvey, 2004, Unpublished; Verghese, 2003).

The basic insight is that it becomes harder to accurately detect a target when there are more items present that might be targets (distractors in visual search), even with perfectly parallel perception of those items. If one assumes noisy perceptual processing, any one of the

distractors could result in an output from the sensory systems that resembles that produced by a target. To avoid dramatically increasing false alarms, decision processes must adopt a stricter criterion for making a target present response when more distractors are present. This results in only an insignificant number of missed targets when the discrimination between targets and distractors is easy (as in classical "feature search" conditions), but a dramatic increase in missed targets and therefore a decrease in accuracy in more difficult conditions, such as a more difficult feature search; see figure 2.1

Simple signal detection accounts do an excellent job of explaining visual search accuracy data when the following conditions are met: presentations are brief and of constant duration, items are widely spaced (to allow independent processing of each), and items only differ along a single dimension (Palmer et al., 2000). Most experimental work on visual search from a SDT perspective therefore focuses on these conditions. If these conditions are not met, SDT requires additional assumptions. Each of the above conditions are violated in most visual search tasks, so additional assumptions are necessary to explain a range of findings. Exactly what additional assumptions are needed is addressed below.

In addition to the need for additional assumptions, the response time pattern for visual search tasks is of great interest, and SDT does not directly consider time. Most real world searches do not allow only a brief glance, but rather allow the observer to gather information over an extended period of time. Under these conditions, perceptual processes are not very noisy: objects can be perceptually discriminated with very high accuracy given sufficient time. The interesting question becomes one of how speed and accuracy are traded off. Any predictions of speed require assumptions outside of SDT. However, fairly basic assumptions allow the predictions of SDT models to be brought to bear on reaction time search tasks. Diffusion process models apply the basic insights of SDT to reaction times by making the reasonable assumption that information accumulates over time in a roughly linear fashion. In the treatment that follows, I will make the more generous assumption that lower accuracy predictions

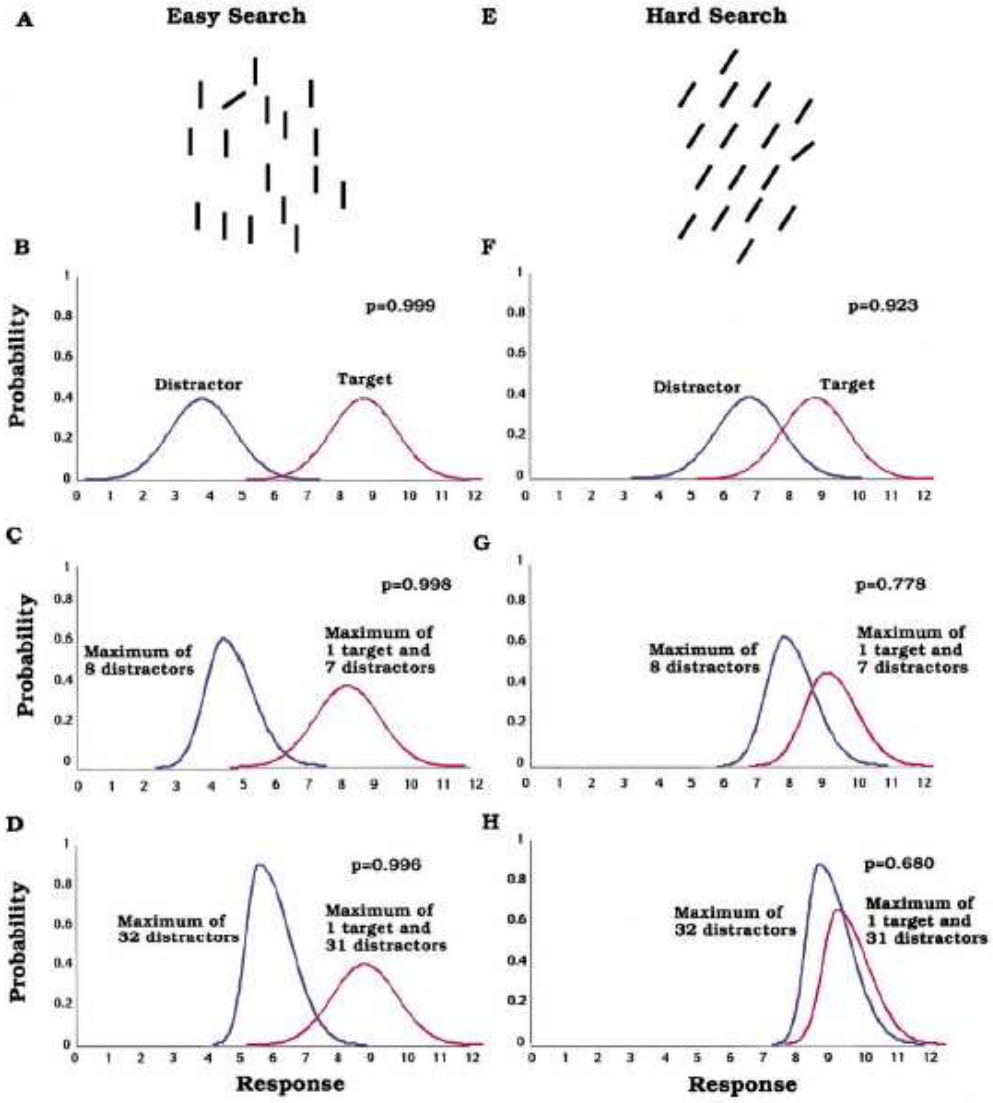

Figure 2.1: The logic of making a target-present decision when multiple potential targets are present, and perceptual processes are noisy but parallel and have unlimited capacity. The x-axis of each graph is the amount of evidence that the item perceived was the target. Note that the lowest set of figures depict a decision made based on only the single highest output from the multiple perceptual processes, since that one is most likely to be the target. This logic provides an alternate explanation of why some brief presentation searches are performed less accurately. Reproduced from Verghese (2001).

in an SDT framework map to predictions of more time. That is, I will assume a monotonic but otherwise unspecified speed/accuracy tradeoff.

Most real world search tasks contain items that differ along many different dimensions, violating the second requirement for applying pure SDTu. Although the mathematical theories become complex and require extra assumptions at this point (Palmer et al., 2000), it is again simple enough to assume that more differences make items more discriminable, and SDT can be evaluated using this rough assumption.

The last assumption, of independent processing, is also commonly violated in interesting searches. It is clear from other experimental work that item perception is not independent of surrounding items, at least away from the fovea. Donk and Meinecke (2002) have shown that closer item spacing leads to faster response times when the discrimination is an easy one. Cohen and Ivry (1991) have shown that accuracy is reduced when conjunctively defined targets are placed closer to distractors in the periphery. Additional assumptions about the interactions between perceptual processes are needed for these cases. One possibility is that the decision process calculates differences with near neighbor processes (under some particular definition of near neighbors), as in the maximum of differences model discussed by Palmer et al. (2000). It is also possible to assume that perceptual processes interact with neighbors, and physiology suggests an outline of these interactions, as discussed in sections 3.2.1 through 3.2.4.

Given the reasonable additional assumptions discussed above, SDT can explain all the experimental results listed in the introduction, with two exceptions. The first is that SDT does not in itself offer an explanation of eye movements in search. For this reason, and to avoid complex assumptions about interactions between perceptual processes in the periphery, SDT is commonly combined with so-called conspicuity area or visual lobe theories. The consipicuity area or visual lobe denotes a limited area around the center of gaze within which target detection is relatively likely to occur (Geisler & Chou, 1995; Palmer et al., 2000). This area varies with the discriminability of target from distractors.

The second piece of datum from the introduction that neccessitates additional assumptions is nonlinear search slopes (Pashler, 1987; Wolfe et al., submitted, Experiment 1 from Chapter 1). These results can be explained in terms of an eye movement, as discussed in the introduction to Experiment 1, but since the eccentricity of items does not on average vary as set size increases, a set visual lobe or conspicuity area does not explain the results. Instead the need for eye movements must be explained by interference in processing between neighboring items. Alternately we could assume that the rate of processing slowdown with extra items becomes less after a certain number of items; however this hypothesis seems unmotivated and unneccessary since eye movements do occur in the paradigms that show nonlinearities in search slopes (as shown by Experiment 5 and in section 2.4). Therefore I assume some amount of perceptual interference by crowded items, at least in the near periphery (7-15 degrees off center of gaze).

In further considerations of SDT, I will adopt all of the above reasonable additional assumptions. These assumptions are similar to those made by two prominent SDT-based theories (Geisler & Chou, 1995; Palmer et al., 2000), but add the assumption that the effective visual lobe changes with stimulus spacing (for additional evidence of a variable lobe size, see section 2.4. This type of theory can nicely explain the data discussed in the introduction. First, the continuum of search slopes arises naturally from SDT based theories, since they assume that more similar distractors produce more evidence for target presence, and therefore require a higher decision threshold, or in a continuous presentation paradigm, more time to accumulate information.

They also account fairly naturally for search slope ratios that are larger than 1:1 but not exactly 2:1, if some amount of noise is assumed in the rate of information accumulation. To avoid missing many targets, the decision process must wait long enough to allow for the possibility that the target is present, but that information happens to have accumulated relatively slowly in this particular trial. A variable rate of information accumulation is likely when realistic considerations are made about crowding effects; target information accumulates more slowly

when the target is in the visual periphery (Carrasco, Evert, Chang, & Katz, 1995) and when distractors are near the target (Cohen & Ivry, 1991). The common finding of slope ratios near 2:1 can be accounted for by including the assumptions of a visual lobe and multiple eye fixations. Existing theories posit random eye movements with tracking of visited positions, and therefore a serial self terminating search. This explanation does not run afoul of the limited memory of search processes because the conspicuity zone is usually large enough to cover the display in only a few fixations.

Efficient search for some conjunctions are also handled nicely by SDT based theories; they must only assume that perceptual processes for detecting conjunctions are possible into the periphery, when the features that compose them are distinctive enough. Put differently, the visual lobe for conjunctions of highly discriminable features becomes relatively large. The experimental results of Cohen and Ivry (1991) set boundary conditions on this assumption by showing that perceptual processes for each distractor are not independent, since even moderately close distractors interfered with target processing in their paradigm. Electrophysiological studies of representations in the visual system are in accord with these findings, since receptive fields for conjunction responsive neurons are large in the periphery; see sections 3.2.2 and 3.2.4.

SDT is less a theory of visual search than it is a set of considerations about decision making based on noisy perceptual processes. As such, it requires many additional assumptions to become a complete theory of search. The evidence presented in the next section demonstrates that these assumptions include considerations of grouping and guidance, mechanisms that are central to the other major classes of search theories.

2.2.2 Guidance Theories of Search

A second class of theory centers on the guidance of attention in search. This type of theory is exemplified by Guided Search (Wolfe, 1994), which was discussed in section 1.1.4 of the introduction. This type of theory is largely specified as modifications to the guided search

model, although more recent variants of Feature Integration Theory are also guidance theories (Treisman & Sato, 1990), and some more general theories of visual attention also assume guidance (e.g Bundesen, 1998).

The central tenet of this type of theory is that top-down strategic control can guide attention to objects based on their basic features, such as shape, color, and size. As discussed previously, this type of theory can account for all the results discussed in the introduction. To accord with current estimates of attentional movement time, such a theory must assume that multiple items are identified in parallel, as discussed in section 1.1.6. This modification is a simplified variant of the visual lobe theories discussed above. As we will see below, this type of theory must also allow for some grouping effects, the core of a third class of theories.

2.2.3 Theories of Search by Grouping

Another class of theories emphasize the role of object grouping in visual search. This type of theory offers an alternate explanation of the same results accounted for by Guided Search. One important theory of this type is proposed by Duncan and Humphreys (1989). This theory attempted to explain the large effect of distractor nonhomogeneity, (the finding that more different types of distractors make search much more difficult). The main idea is that like distractors tend to be rejected as groups. This idea was instantiated in an algorithm called SEarch by Recursive Rejection, SERR (Humphreys & Muller, 1993). It is a neural network model in which search proceeds by serially rejecting items by groups. The groups are formed by spatial proximity and form similarity. These groups are processed serially, as each one forms in turn. Distractors usually group first, so search usually proceeds by eliminating groups of distractors until the target is found or no distractors remain. SERR therefore has no top-down component.

The second theory is Spatial and Object Search (SOS) (Grossberg et al., 1994). This theory is based on detailed neural network simulations of object segmentation in lower visual areas, but it is not instantiated as a neural network, but rather as an algorithm. Like SERR, it

posits that search acts by grouping items. Unlike SERR, this grouping is influenced by top-down factors, so that for instance shared color could primarily decide object groupings if the observer were strategically attending to color. Object groupings are made in parallel. Each group is then serially compared to a target template. If it contains features that match with the target, a yes response is given; otherwise, a new grouping is formed (without any item already grouped and inspected) and search proceeds until no items are left. Object groupings are also spatially restricted, but in this case, the restriction is that a clear path between any two items is necessary for them to group together.

SOS can account for constant slopes and the common 2:1 absent-present ratio, because it is a serial self-terminating search, and the object groups over which search proceeds generally vary in number nearly directly with the total number of distractors. That is, in a standard paradigm for stimulus spacing, on average 24 distractors will produce nearly (but not quite) three times as many groups as will 8 distractors. So a constant time to search each distractor group predicts a nearly constant search slope. The same logic applies to the 2:1 slope ratio; search proceeds serially, so twice as many groups are searched (on average) on a target absent trial.

The original instantiation of SOS runs afoul of the slow time course of attention in that proposed attentional shifts take only 71 ms each (Grossberg et al., 1994), which is likely faster than attention can move voluntarily in any task (see section 1.1.5). However, relaxing the strict requirement that objects have a clear path between them in order to group together allows this type of theory to have a plausibly slow attentional movement, by allowing it to operate on larger groups.

Grouping theories of search can explain all the data discussed in the introduction, except the nonlinearities in search slopes found by Pashler (1987) and Wolfe et al. (submitted), and Experiment 1 here. However, SOS can be easily modified to explain findings in the case of the color-orientation conjunction task in which Wolfe et al. (submitted) observed nonlinearities in

search slopes. The theory already predicts that grouping will only be effective to some maximum set size; the additional crowding of randomly spaced distractors will mean that distractor-colored objects are more likely to break the straight line required for grouping. When one group cannot be formed, search becomes serial, and a nonlinearity in the target-present slope is predicted at this point. However, the theory needs extra assumptions at this point to explain the rise in search times over the range of small set sizes over which only one group is formed. It is easy to assume that the grouping process takes longer when more items are present (although this conflicts with the parallel and independent grouping mechanisms proposed by Grossberg and colleagues for SOS theory). An alternate assumption that explains the search time is that the feature identification process takes longer to work when more objects are present.

However, this explanation does not work well when searches are not conjunctions searches. In the case of the TvL searches reported in Experiment 1 here, and in Wolfe et al. (submitted), there are no distractor objects of unlike features to block grouping (or else all distractors are unlike, and grouping should not happen even at set size 2). To explain this result, an additional assumption on the limits of grouping is needed. Limiting the grouping to a maximum number of items brings the theory close to the serial-parallel theories of Pashler (1987) and Wolfe et al (submitted), although still with the important distinction of search acting on objects grouped by features. In sum, for grouping theories to explain breakpoint effects, they must adopt some of the assumptions of serial-parallel theories; namely, more time to identify more items in parallel, and limits on the number of items identified in parallel that are not likely to result from limitations in grouping as such.

Positive evidence for grouping theories is discussed in the next section. Grouping effects are clearly important for some types of visual search, and therefore a complete theory of visual search must include mechanisms for grouping. However, the mechanisms that are required to explain the data are somewhat different than the ones suggested by these theories of grouping. The data showing grouping effects and the mechanisms required to explain them are discussed

further in the next section.

2.3 Behavioral Findings

All the major types of theories, guidance, grouping, and conspicuity/detection theories, can explain the data discussed in the introduction, with the additional assumptions discussed in the sections above; however, each one fails to explain a subset of the following data. We will see that each type of theory must be supplemented with the mechanisms of the other two to provide a complete theory of search. That theory is presented in chapter 4.

However, the resulting theory is not a simple combination of the theories discussed so far. Distinct patterns of results in different types of search show that the mechanisms of each type of theory play a more prominent role in some search strategies than in others. A complete theory must specify not only the mechanisms underlying search, but must specify when and why each mechanism comes into play.

2.3.1 Search on a Subset

First I review evidence showing that grouping mechanisms are present in some searches, so that a complete theory must provide an account of such mechanisms. However, I also show that grouping is unlikely to play an important role in most searches, so that theories of grouping do not provide a satisfactory account of the entire range of search findings.

A variety of data indicates that observers are able to selectively search a subset of items at least based on color. For instance, Egeth, Virzi, and Garbart (1984) showed conjunction targets where the target was red, and varied the number of distractors that were also red. They found that search was substantially faster when few distractors were red. However, they explicitly instructed observers to search among the red items only. Zohary and Hochstein (1989) performed a slightly different type of experiment in which they also used a color conjunction search with varying ratios of distractors, but used brief displays and a stairstep procedure to determine the

display time necessary to achieve 70% accuracy in each condition. They similarly found that displays with fewer distractors matching the target color required less time, and they provided no explicit instructions

Guided Search offers an explanations of these effects. The extra salience for target colored objects is a natural outcome of the equations for calculating feature map activation (since rare items are more likely to have activations enhanced by having unlike items nearby). However, additional evidence shows that observer's beliefs about which target feature is more rare influences which subset they inspect (Bacon & Egeth, 1997). Split-brain patients seem to benefit from verbal instructions only in the language-dominant left hemisphere (Kingstone, Enss, Mangun, & Gazzaniga, 1995). The model can accommodate this finding by supposing that the amount of top-down influence on each feature map can be flexibly assigned (Wolfe, 1994).

Grouping theories can offer explanations of the above findings only by including top-down control over which group is searched first (the SOS model as stated has top-down control only over grouping criteria). The authors suggest this modification as a relatively minor one (Grossberg et al., 1994). This modification adopts one of the central tenets of guidance theories, and so moves grouping theories toward convergence with guidance theories.

2.3.2 Efficient Subset Search

There is another type of finding that guidance theories and detection theories have more trouble accommodating. Several experiments indicate that search limited to a subset of items that all have a common feature can be efficient. Three experiments show results that provide evidence for two distinct mechanisms. These results are first laid out, and their implications for the three major types of theories is laid out in the following sections.

Friedman-Hill and Wolfe (1995) presented standard conjunction searches for color and line orientation, with one modification: the two angles were randomized from trial to trial, with the requirement that they be different by at least thirty degrees. In this situation it is impossible

for the standard Guided Search algorithm to operate, since a full target template cannot be specified in advance. Figure 2.2 shows an example display from Experiment 1 of Friedman-Hill and Wolfe (1995).

This manipulation produced a distinct pattern of results: target present and target absent response times were indistinguishable. Neither the faster present response, nor the greater absent slope were observed at set sizes up to 16. Searches took substantially longer than a standard conjunction search, by about 200 ms. Searching a larger display did take significantly more time, but search did not take longer when there were more items of the target color, indicating that the longer times arise from changes in total set size, rather than from searching a subset serially. This pattern strongly suggests a parallel search; any serial search should take longer on target absent trials, and have a larger search slope on those trials.

First, Carrasco, Ponte, Rechea, and Sampedro (1998), reported that within dimension conjunction searches for color-color conjunctions (half red and half green object among red/blue and green/blue objects) can be located efficiently after practice. In this study, target absent trials become nearly as fast as target present, and have the same search slope. These results suggest a strategy in which observers first locate all the locations that have a certain color. If one display location is not included in this group, it is the target. This strategy is further indicated by the control task, in which distractors are red/blue and pink/green. Observers improved with practice on this task, but search slopes never become flat (as in previous work with a similar paradigm Wolfe, Yu, Stewart, & Shorter, 1990), and target absent trials always take longer, and have a dramatically larger slope than target present trials.

A third set of findings gives an even more vivid demonstration. DeLiang, Kristjansson, and Nakayama (2005) show that observers can efficiently tell whether there is a unique object in the display, without knowing in advance what the object is, and with little practice. Samples of their display are shown in figure 2.3. These displays provide a dramatic illustration of the principle involved. By looking at the displays with spread attention, it should be quickly

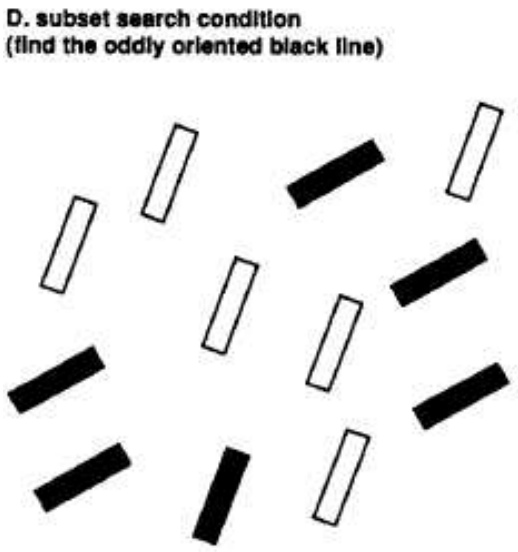

Figure 2.2: Display from Friedman-Hill & Wolfe (1995). Lines shown black were red; those in white were green. A set size of 12, target present is shown in the subset search condition of their Experiment 1. Reproduced from Friedman-Hill & Wolfe, 1995.

apparent whether there is a singleton in the display.

The pattern of reaction times obtained with these stimuli show a highly efficient search; there is no additional time needed to search displays with more items, at least up to 32. Target absent trials took no longer than target present trials at small set sizes, and little longer at larger set sizes. Although the study (DeLiang et al., 2005) does not focus on the comparison of target-present to target-absent trials, the unique nature of the experiments makes such a comparison very informative. Because in each display there are two subsets that could contain the target, this task should serve to differentiate serial from parallel accounts of subset search (at least in this particular paradigm). If each subset is searched independently, target absent trials should take substantially longer. On target present trials, the target will be found in the subset checked first on half of the trials. In target absent trials, both subsets must be searched. As argued above, the differences between target present and target absent RTs larger set sizes may well be caused by unnecessary rechecking. Whether or not that is the case, RTs are nearly the same at small set sizes, so no theory in which subsets are checked serially can explain those cases.

2.3.2.1 Guidance Accounts of Efficient Subset Search

Guidance does offer a possible account for the results discussed above. Efficient search under the conditions of DeLiang et al. (2005) et al could be accomplished by identifying the most common items, then searching for the missing ones. However, since there are two possible targets in each display, guidance would have to be directed at each in turn. We would expect that target absent trials would take significantly longer than target present trials, since both absent items must be searched for in target absent conditions, whereas only one of the two need be searched for on half of target present trials. Note that search times for present and absent are nearly equal in the smallest set size of each of the first two types of search reported in figure 2.3.

Although top-down guidance cannot explain those results, the “preattentive”, parallel

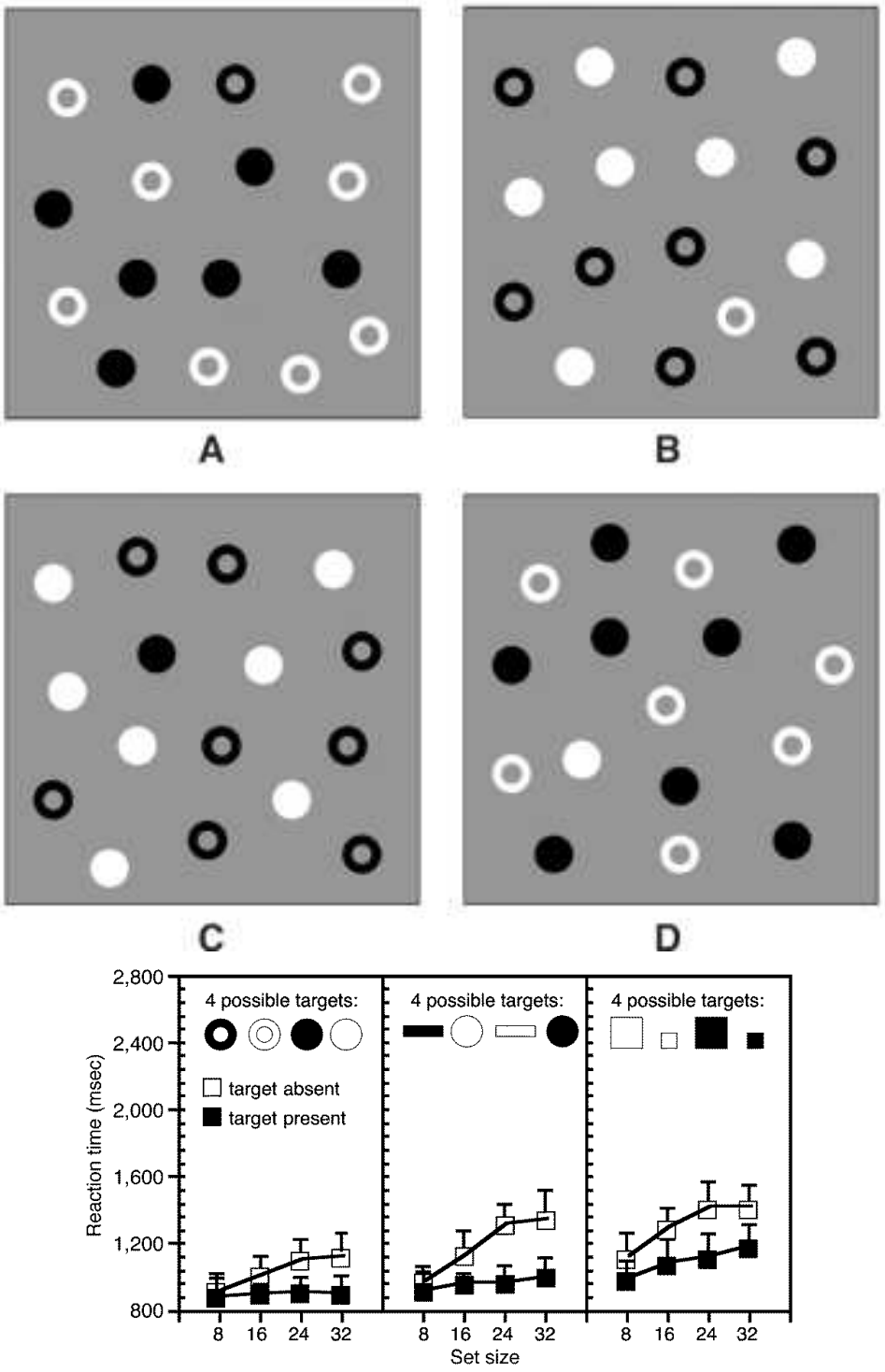

Figure 2.3: Displays that allow efficient conjunction search for without advance knowledge of target identity. Targets are defined as any item with no identical items. A black annulus is the target in A, a white annulus in B, a black circle in C and a white circle in D. Graph shows reaction time results from these and similar tasks. Note that there is no increase in target present times across set sizes in this task (first panel), but that all reaction times are around 900 ms, reflecting additional time to adopt both light and dark top-down color orientations. Reproduced from DeLiang et al (2005).

stage of guidance models could account for the above findings. Bottom-up guidance can work separately for items in different contrast polarity. It is possible that features as simple as spatial frequency are represented separately for objects lighter and darker than a background, and local difference mechanisms guide attention to the target. In terms of the Guided Search model, this amounts to adding separate feature maps for spatial frequencies defined by light and dark contrasts. This explanation is a variant of the one discussed below for signal detection theories; the difference is in whether the information is used to guide attention, or simply to make a decision on target presence; this distinction is discussed further in the next section.

Guidance also offers a possible explanation of the results of Friedman-Hill and Wolfe (1995). To locate a target of an unknown orientation among a color, it is possible to identify the distractor , then direct guidance to that and the target color. A more satisfying experimental test, and one that is needed to completely establish popout search on a selected subset, is to let the angles of the non-target-colored items vary randomly, so that their angle cannot provide a cue for guidance. Experiment 6 (section 1.7) provides this test. The fact that results are nearly identical to those from the Friedman-Hill & Wolfe study shows that guidance for both target features was not used, since it is impossible to infer the orientation of the target without identifying it in Experiment 6.

Results of the Wang et al (2005) study are not consistent with guidance to a specific target. The similarity in target present and target absent reaction times is inconsistent with this type of attention. Since two separate target templates would be needed, presumably one after the other, this is a serial account of the two subset searches. Target present trials should on average be much faster, since the target should be located with the first attentional setting on half the trials.

However, this serial explanation does fit very nicely with the results from the color-orientation efficient subset search. Friedman-Hill and Wolfe (1995) included one condition (in their experiment 5) in which the subset that contained the target was not specified known in

advance. This condition is closely analogous to the paradigm used by DeLiang et al. (2005). However, the Friedman-Hill and Wolfe study showed a signature of reaction times that indicate a serial search of the two subsets (figure 2.4) just as clearly as the results of DeLiang et al. (2005) indicate a parallel search of the two subsets (figure 2.3).

Target absent trials took considerably longer than target present trials when the target could appear in either subset. In contrast, present and absent RTs were virtually identical when the relevant subset did not change between trials. In addition, reaction times that were about one and a half times longer than searching either subset independently, while target absent times in the two-subsets condition were roughly twice those in the single subset condition. (While these are not exactly the predictions of a serial search if a constant time for response is allowed, the priming advantage of a consistent mappings in the single subset condition would stretch the predictions in this direction).

2.3.2.2 Guidance Must Accomplish Grouping to Explain Efficient Subset Searches

To explain the results of DeLiang et al. (2005) and Experiment 6 here in a guidance framework, we must assume that attention is guided to some items, but it is attention to a very large set, rather than to one or a few items proposed by guided search. To account for the parallel search within a subset shown by (Friedman-Hill & Wolfe, 1995) and in Experiment 6, Guided Search must suppose that the subset of items is first located, then fed back in to the standard feature mapping process, where it works with only a little interference from non-selected items. The one odd angle in the subset then pops out through the standard mechanisms. This explanation is a type of grouping, although it is top-down, slow, and strategic rather than bottom-up, fast, and obligatory as proposed by grouping theories. This possibility is discussed further below in section 2.3.2.5

Accounting for efficient search for a color-color conjunction when all distractors share a color (Carrasco et al., 1998) is even more problematic for guidance theories. Attention must

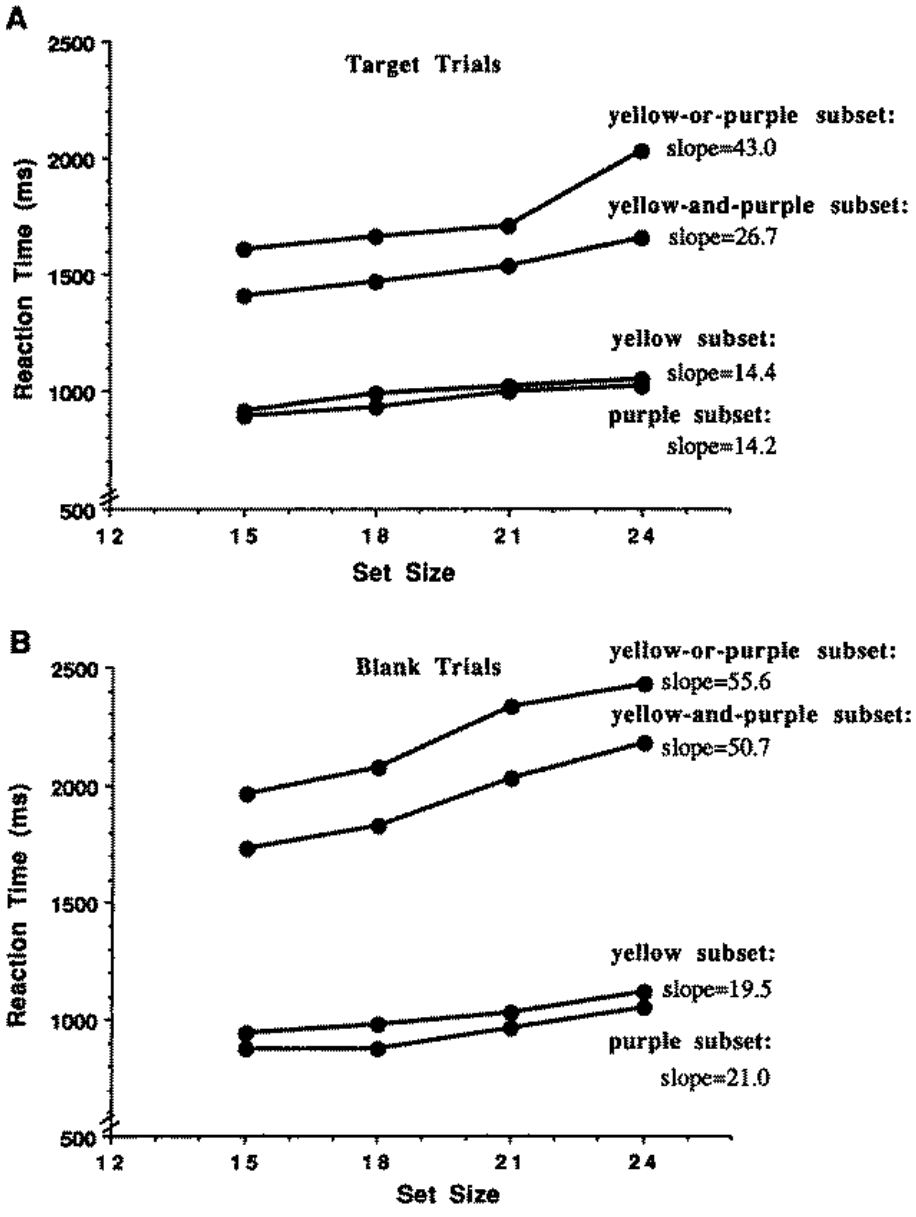

Figure 2.4: Reaction times for various subset searches. Note that searching two subsets takes a good deal longer than searching one, and this difference is greater in the target absent case. This pattern of results suggests a serial process. Reproduced from Friedman-Hill & Wolfe, (1995)

first be guided to the set of all blue items, then guided to the one item missing from that set. This operation would require a sophisticated manipulation of spatial attention. It seems likely that the human cognitive system is capable of such a manipulation, and one possible mechanism for it is laid out in the section on neural theory (section 4.2.4). However, the very short reaction times in that study make this explanation unlikely. Response times were around 600ms in target present trials on the first block, and only about 500ms for both conditions in the fourth block after about 1000 trials total of practice. Therefore it seems that perception must happen at a more global level in this situation, as suggested by grouping theories and detection theories of search.

The explanations offered here have converged with those given by grouping theories: objects are selected by groups (or group automatically in some cases), and searched in parallel for a target within that group. Note that this explanation does not explain efficient search of two subsets in parallel (DeLiang et al., 2005). To explain these results, we must assume separate feature maps for spatial frequency in positive and negative contrast polarity. This assumption is discussed further below.

2.3.2.3 Signal Detection Accounts of Efficient Subset Search

Because targets are defined solely in relation to other items, basic signal detection theories (Palmer et al., 2000; Verghese, 2003) cannot account for the findings of efficient search for a previously unknown target (Friedman-Hill & Wolfe, 1995; DeLiang et al., 2005). Instead they would have to suppose that an aggregate signal from the whole display is analyzed, and that there is a substantially different signal when an extra type of item is included.

This explanation works well in most respects for the efficient multi-conjunction searches of DeLiang et al. (2005). Because the sets are defined by contrast polarity, some lighter than the background, some darker, it is only necessary to have mechanisms that sum spatial frequencies separately for each contrast polarity. A signal that includes more different frequencies in either

the bright or dark “channels” then signals the target. As mentioned above, the fact that target absent responses are nearly as fast as target present responses suggests that the extra time beyond standard popout searches (around 300 ms longer) for similarly discriminable features is not spent in actively selecting one set, then the other (around 500 ms vs 800-900 for the unknown conjunction search).

A similar explanation is needed to account for fast location of items without a particular color. In this case as well, the hypothesized detectors are acting at the level of a group of objects, detecting in effect a large blue object that either has a gap or does not. This could be accomplished by basing decision on the output of low spatial frequency detectors that respond specifically to blue colors, again a fairly reasonable hypothesis.

However, this type of explanation does not work as well in the case of popout search on a subset using colors and orientations. The data suggest that this type of search is performed serially, with one subgroup being selected before the other, as discussed above in section 2.3.2.1. Detection theories must suppose that attention is redirected to the signature of each color of items, and that this redirection allows detection of the overall signature of orientations of that color.

2.3.2.4 Signal Detection Must Work on Groups to Accomplish Efficient Subset Searches

While the two explanations above are reasonable extensions of signal detection theories in light of known visual system architecture, they are also in effect grouping theories, since a property of the whole set of red items is detected. These additions are the central hypothesis of grouping theories, although they do not accept that grouping is automatic and mandatory, another important assumption of grouping theories.

2.3.2.5 Grouping Accounts of Efficient Subset Searches

Grouping theories account well for two of these three results; the efficient subset search of Friedman-Hill and Wolfe (1995) and efficient search for an item defined by lack of a feature (Carrasco et al., 1998) can be accounted for by either SOS or SERR with only minor modifications. These explanations are highly similar to the top-down grouping proposed by guidance theory.

However, the efficient search for unknown conjunctions shown by DeLiang et al. (2005) is incompatible with any theory in this class, since serial search among groups is a basic tenet. Grouping effects in that case happen in parallel, allowing either detection as in the modified detection theory, or guiding attention to the target efficiently, as in the modified guidance theory.

2.3.2.6 Two Types of Grouping

As shown above, explanations of search without group effects cannot account for findings of parallel subset searches. Furthermore, two types of grouping effects are necessary to account for the results. One is automatic, bottom-up, and parallel, while the other is serial and likely top-down and strategic. This analysis has shown that both mechanisms are necessary for a complete account of visual search, and has suggested some conditions in which these two mechanisms come into play.

Why would these two similar paradigms show two very distinct types of grouping effects? One possible answer is that there are dedicated detectors specific to each contrast polarity, and the difference between circles and annuli.

Because circles and annuli will cause a strong response from a simple on-center or off-center neurons, and these are known to exist at early stages of the visual system, this explanation seems likely. While there are undoubtedly some neurons that respond to conjunctions of color and orientation, these are likely to exist only at higher levels of the visual system with larger receptive fields. The fact that these neurons will usually have several objects within their receptive

field impairs their effectiveness. In addition, such dedicated neurons are rare, whereas on- and off-center cells are extremely common, making them much more effective at guiding attention, which likely cues off of neural activation at the population level. Section 3.2.2 explores these mechanisms further.

This analysis suggests that automatic, parallel, low level grouping occurs for some features (e.g. contrast polarity) and not for others (e.g. colors). Top-down grouping mechanisms seem to be important in cases where a specific strategy of grouping, as in the color-orientation subset searches. This second grouping mechanism is discussed further in section 4.2.4. Neither of these types of grouping seems to play an important role in searches where strong low level grouping does not occur, and a strategy of global grouping is not necessary.

2.3.3 Global Grouping Does Not Occur in Standard Conjunction Searches

The findings on efficient subset searches provide strong evidence that grouping does play an important role in some searches, as discussed in the previous sections. However, they do not suggest that grouping on a large scale is common in visual search. This distinction is important in that it demonstrates two different strategies used for the same type of search displays.

In subset searches for an odd orientation among many like oriented lines of the target color (Friedman-Hill & Wolfe, 1995, Experiment 6 here), observers seem to be looking for uniformity in the orientation of the relevant subset, rather than looking for a single item with different orientation. This is suggested by similar target-absent and -present trials reaction times. These experiments produced target-present and target-absent reaction times that are nearly identical at small set sizes, although they diverge at set sizes larger than 16 (Friedman-Hill & Wolfe, 1995 Experiment 2, p.539). Conjunction searches, on the other hand, show absent trials RTs that are substantially longer than present trials, with only a few exceptions.

Kaptein, Theeuwes, and Vanderheijden (1995) found faster absent responses using very small set sizes (up to 6) for a color-orientation conjunction search. They used brief (150 ms)

displays, small set sizes, and circular stimulus arrangements around fixation. Most importantly, they also instructed their subjects to search only the red subset. When these instructions were not included and displays were made less regular, they found the usual effect of longer absent trials and greater absent slopes.

Another exception was found in the same experiment that demonstrated efficient subset search. Friedman-Hill and Wolfe (1995) used a comparative condition of a standard color-orientation conjunction search, with target identity known in advance. This produced absent reaction times that are slightly longer than present times (about 50 ms different), but with slopes that were nearly the same (8.9 ms vs 7.0 ms per item for absent and present slopes). These results are not consistent with a serial search, and are much more similar to the subset search results than standard conjunction searches using similar, high-contrast stimuli (Wolfe et al., 1989).

This pattern of results suggests that observers adopted a strategy similar to that needed for subset search in those two cases. In the first case, they looked for a coherent set of red items because the short display prevented attentional movements; in the second case, half of the observers adopted this strategy because they had just completed a subset search, which forces the strategy by not allowing guided search to a specific target.

In this (hypothetical) strategy, observers look for the global uniformity of orientation among the relevant color of a target absent trial. They respond “absent” when they find this looked-for representation. They respond “present” when they fail to find such uniformity. This seems to be possible only at small set sizes; the divergence of absent and present times at larger set sizes suggests that the search either becomes serial, or that it becomes very difficult to perceive the relevant subset as a coherent item. This strategy could well be mixed with the alternate, of looking for a signature of non-homogeneous orientation within the selected subset.

Donnelly, Humphreys, and Riddoch (1991) demonstrated that visual search can proceed by looking at distractors as a coherent object. They used displays in which distractors were

shaped so as to outline the corners of a geometric shape; they found much faster target absent trials in this condition. While the evidence is not conclusive, it seems likely that observers can adopt a strategy of looking for uniform distractors in orientation-color conjunction searches with bright colors, and in displays where the items suggest a single coherent object. They do so when the strategy is especially useful given the experimental conditions, and they may adopt such a strategy on later tasks after finding it useful on a particular task. Observers could also employ this strategy in standard shape-color conjunction searches, but the data suggests that they do not, and rather follow the more basic strategy of searching for the target using top-down control.

Grouping theories suggest that automatic grouping, rather than top-down guidance, play the most important role in conjunction search. It is clear that some types of automatic, low level grouping are important to search; for instance, Treisman and Sato (1990) showed that efficient search is possible on orientation targets defined only by a texture that varied from the background, and efficient search for large scale line orientation defined only by small-scale line orientation that differed from the background (that is, all the large lines were defined by which parts of the display were composed of leftward leaning vs rightward leaning small lines).

However, existing evidence does not establish the limits of this automatic grouping, for instance, whether it is important in standard conjunction searches. It is difficult to experimentally distinguish this possibility from the mechanism proposed by the Guided Search theory, in which items are more salient when they are surrounded by different items, and are less likely to draw attention when they are grouped with like items. Both theories predict that targets are easier to find when distractors are grouped together. The existing data do not provide evidence for one or the other possibility. However, neural data show that items are more salient at the neural level when they are surrounded by unlike items, as reviewed in Appendix A. Therefore the bottom up salience proposed by GS is likely to play an important role in visual search; it remain unclear if automatic grouping has an important role, or if grouping is limited to the

type that seems to be important for subset searches; a serial, effortful grouping performed by top-down guidance.

2.4 Eye Movements

In this section I review evidence showing that search is guided. However, this review also reveals a pattern of results inconsistent with the Guided Search 2.0 model. Guidance appears to be limited by the distance from fixation, and the crowding of stimuli, as well as by the similarity between targets and distractors that is commonly discussed.

The evidence in this section is all in terms of eye movements. Saccades (eye movements) are more likely to land on distractor items that share the target's color. Records of eye movements provide invaluable evidence on the role of guidance in search, if we allow that eye movements reveal the movements of attention. I will therefore start by reviewing the evidence and arguments for a strong link between eye movements and attention.

2.4.1 The Link Between Eye Movements and Attention

It is known that there is a close link between eye movements and attention, but the exact nature of that link is still in question. Numerous studies have shown that attention can move when the eyes remain fixated, but it now seems unlikely that attention moves independently of the eyes in normal vision. Saccadic fixation times in visual search are usually between 200-300 ms. This shift rate is in line with the evidence reviewed in section 1.1.5 that calculates speed of covert attention to be between 100 and 300 ms. It is consistent to suppose that attention moves no faster than the eyes, as concluded by Ward (2001), and further that covert attention is normally followed by an eye movement to the same location unless the eye movement is suppressed by an effortful maintenance of fixation (McSorley & Findlay, 2003).

There is a great deal of evidence in favor of a premotor theory of attention, in which covert attentional shifts are the result of planning for eye movements that are suppressed (Riz-

zolatti, Riggio, & Sheliga, 1994). Additional evidence of a close link between eye movements and attention is reviewed in Liversedge and Findlay (2000). Recent evidence (Juan, Shorter-Jacobi, & Schall, 2004) has shown that attention need not detectably effect preparation of eye movements, but this study seems to be the exception that proves the rule, as they used an undemanding attentional task, and monkeys that were extensively trained in the antisaccade task. The task necessitates separating attention from eye movements, so the study shows only that the two can be separated with enough practice.

Because strictly serial theories of conjunction search posit high attentional speeds (20ms per attentional movement or faster), it has been supposed that attention may move several times during each eye fixation (Treisman & Gelade, 1980; Wolfe et al., 2000). A close inspection of eye movement data does not reveal the patterns predicted by this supposition. If attention moves serially during each fixation, it will locate the target early on some trials and later on others, when the target is found. In this case, some eye movements to the target should occur more quickly than those that do not. In two studies where this variable was examined, eye movements that located the target were not significantly faster than those that did not, despite the use of a very large number of trials in the second (Findlay, 1997; Motter & Belky, 1998b). Two different studies also found that guidance was more accurate after longer fixations, opposite of the prediction made by assuming fast attentional shifts (Hooge & Erkelens, 1999; McSorley & Findlay, 2003).

Because the assumption of fast serial attention shifts is unnecessary, and does not fit well with either behavioral or neural level evidence, I will provisionally assume that eye movements do reveal all shifts of attention. Given this assumption, eye movement provide invaluable evidence on guidance of attention in visual search. However, even without this assumption, it is uncontroversial that eye movements are always accompanied by shifts of attention; therefore eye movement studies reveal some of what attention does, even if they do not reveal everything it does. This more modest assumption is adequate to provide positive evidence of attentional

guidance from eye movement studies.

2.4.2 Guidance of Eye Movements

Analysis of eye movements during visual search provides solid evidence that attention is guided by features. However, this guidance is not entirely consistent with the Guided Search model (Wolfe 1994). Guidance by feature is very effective for one highly discriminable feature at a time. However, guidance by two features at once shows dramatic limitations with the distance of a target from the current point of fixation, and with the number of distractors close to the target.

Scialfa and Joffe (1998) found a high degree of saccade guidance when people performed a contrast-polarity and orientation conjunction search. In the experiment that examined guidance, they used the two authors as expert observers, and found that search was highly selective for the target contrast of the target; fixations fell near distractors of the target contrast three times as often as near those of the nontarget contrast. However, they also found that, with more practice, they were able to achieve high accuracy with no eye movements at all, although this accuracy still decreased with the eccentricity of the target. This finding seems to offer evidence of guidance at high eccentricities, since displays convered 37 degrees of visual field, meaning that targets were up to 17.5 degrees from fixation, yet were identified relatively quickly and accurately. It is logically possible that information about target presence was available without information about target location, but this seems unlikely. It may be that the use of contrast polarity made search easier, since several other studies have found very efficient conjunction search with contrast polarity as one feature (DeLiang et al., 2005; Nakayama & Silverman, 1986).

Findlay (1997) explored saccade guidance in a color-shape conjunction search with two concentric rings of eight items each around the starting fixation points. The inner ring was at 5.7 degrees of eccentricity, while the outer ring was at 10.2 degrees. Observers correctly fixated

targets in the inner ring on the first saccade with greater than 50 % accuracy. With targets in the second ring, they were much less accurate, but still well above chance. In addition, there was evidence that some observers were primarily using color for guidance, and others were using shape. The easily distinguishable stimuli used may have allowed more guidance by shape. Thick crosses and filled circles could be targets, but the display also contained filled squares and triangles.

The large differences between individual observers in this study suggests that strategy or motivation plays a role in the effect of guidance. The accuracy of first saccades ranged between 67% and 20% for targets in the inner ring, and between 40% and 2% for those in the outer ring. It is of some small interest that results from the study's author were substantially higher than those of any other observer; either motivation or strategies related to knowledge of the experimental hypotheses may have produced the more effective guidance.

Similar results were obtained with monkeys on a similar task (Bichot & Schall, 1999). Four to twelve stimuli surrounded the fixation point in a ring at seven degrees of eccentricity. The stimuli were red or green and either thin crosses or outline circles. The task was to locate the stimuli in a single saccade; no reward was given if the first saccade missed. Saccades were still performed quickly, about 240 ms after display onset. First saccades were accurate on 87% with four items, about 80% with six items, and about 67% with twelve items. Missed saccades were much more likely to land on items that shared the target color, about 60% or the target shape, about 35%, and only 5% landed on distractors that shared neither target. The target changed every 20 trials, so long term training with specific targets was not responsible for the effective guidance.

Another study found very limited guidance in an orientation-color conjunctive search with human subjects (Zelinsky, 1996). The relevant condition used distractors that either shared possible target features (red or green, horizontal or vertical) and those that did not (blue and yellow, 45 degree angle left or right). On target-absent trials with 17 items, only 57% of fixa-

tions fell on items that shared a feature with the possible targets, versus 43% that fell on items with no target feature. Saccades were even less selective when only 5 items were present; about 52% versus 48%.

This study had one key difference that seems to explain its very different findings. It did not use a single consistent target, even within blocks. The target could always be either a red horizontal or a green vertical bar. Thus it was impossible to direct attention to a set of unique target features. These results are consistent with the conclusion that more than one feature within a dimension cannot simultaneously guide search to any great degree. The small amount of guidance that was shown could indicate that two colors can be simultaneously used at least with practice, or could indicate merely that observers guided search to one color or orientation on some trials.

Another study looked at guidance of eye movements based on line width (Hooge & Erkelens, 1999). Targets were outline circles of thin lines; distractors were circles with gaps, made from either thin lines or thicker lines of varying width. They showed that differences in line width strongly affected guidance. Small differences in line widths (.45 degrees vs .30 degrees) produced only a small tendency for fixations to fall on distractors sharing the target's line width. When line widths differed substantially (.60 or . 75 degrees vs .30), there was more dramatic guidance, between 3 and 4 times as many saccades landing on items similar to the target.

This study also analyzed whether longer fixations produced more accurate guidance. They found that longer fixations did indeed lead to a higher probability of fixating a distractor similar to the target on the next trial. However, there was a possible alternate explanation. Fixations lasted longer on items similar to the target. Therefore the increased guidance could be the result of guidance being directed to other items that resemble the one currently in the fovea. However, the offered explanation seems more likely. It is highly plausible that guidance does function more accurately when fixations are longer. The longer fixation time should allow more

time for competition to emphasize the features supported by a top-down target template.

2.4.2.1 Eye Movements Suggest that Guidance is Limited by Stimulus Spacing

This section reviews some evidence showing that the amount of guidance is strongly dependent on item crowding. This finding provides a key link between behavioral and neuroscience evidence. The primary original contribution of this work is to show how this finding corresponds to the structure of the ventral visual stream. Many findings show that items of greater complexity are represented by neurons with larger receptive fields, and that these RFs are larger farther from the center of gaze. If we suppose that behavioral detection requires clear and ample neural representation of a target, the structure of the ventral stream predicts and explains the behavioral findings reviewed below. The data on visual system function is reviewed in chapter 3, while the link between the two is clarified in chapter 4.

A particularly relevant and thorough set of studies explored the factors affecting guidance in search (Motter & Belky, 1998a, 1998b). These studies examined eye movement data from monkeys with a great deal of visual search experience. A typical display and scan path is shown in figure 2.5. As this scan path indicates, the results were strongly in favor of guidance by feature. However, the results did not support the hypothesis of the Guided Search 2.0 model, in which conjunction search is guided by both target attributes equally and simultaneously. Search was guided primarily by color, rather than by both color and orientation.

Although the example shows fixation of nearly half of the target colored items, suggesting no guidance by orientation, this was not the typical result. On average, the monkeys made many fewer saccades than random sampling of target-colored stimuli would predict. There were only four saccades on average for displays of 48 items; displays with 96 items averaged only about 5 1/2 fixations. Clearly fixations did not proceed at random among the target-colored stimuli, but were guided to the target.

But this guidance was not at all uniform across the display. A separate analysis showed

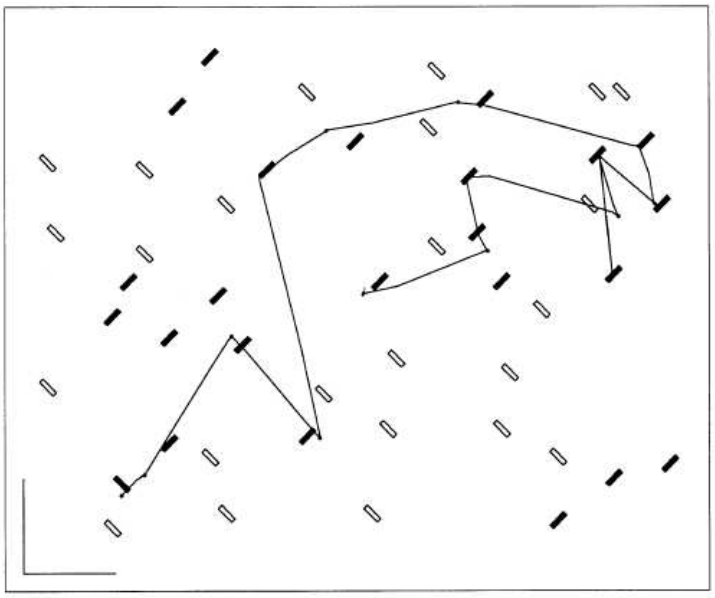

Figure 2.5: A single trial of eye movements is superimposed on the display seen. Items shown in black were red, those shown in white were green. Note that fixations (shown as dots) always land near items of the target color. Reproduced from Motter and Belky (1998a)

that targets were usually located on the next saccade when they were close to a fixation point (Motter & Belky, 1998b). The chance fell off smoothly with distance from fixation, when that distance was normalized by the average nearest neighbor distance (ANND). In the color-by orientation conjunction search of their Experiment 1, the chance of locating the target was above 50% when targets were within a radius of about twice the average distance between stimuli (see figure 2.6). Thus the spacing of stimuli was important in determining how far from the center of gaze a target could be located. These findings together suggest that search is reliably guided only within a relatively small region, and one defined by the spacing of targets (or perhaps the number of items closer to fixation than the target, see below).

While these results show that guidance is much better for the items closest to fixation, they also show that there is some guidance toward items farther away. The probability of fixating the target does not drop to chance regardless of the distance of the target from the last fixation. This is highly consistent with the neural theory presented here. The likely reason that the region of effective guidance varies with stimulus density is that receptive fields further from fixation are larger, and more likely to contain multiple stimuli (sections 3.2.2 and 3.2.4).

2.4.3 Effect of Number of Items and Stimulus Types on Number of Items Processed

The studies of Motter and Belky (1998a, b) are very relevant to the theory tested in the experiments of chapter 1. They suggest a similar but importantly different mechanism of search than the one suggested by theories based only on human behavioral data (Pashler, 1987; Wolfe et al., submitted). These differences are discussed in the next section; we first look at the consistencies. Their data are consistent with the conclusions drawn from experiment 1 here, in showing that the extent to which items are searched in parallel varies with the discriminability of targets from distractors. Additional data from the same studies suggests that the effectiveness of guidance varies not only with the spacing of items, but also with the discriminability of items.

A second experiment using one of the two monkeys varied the types of stimuli. One

Midtrial fixations First fixation after onset

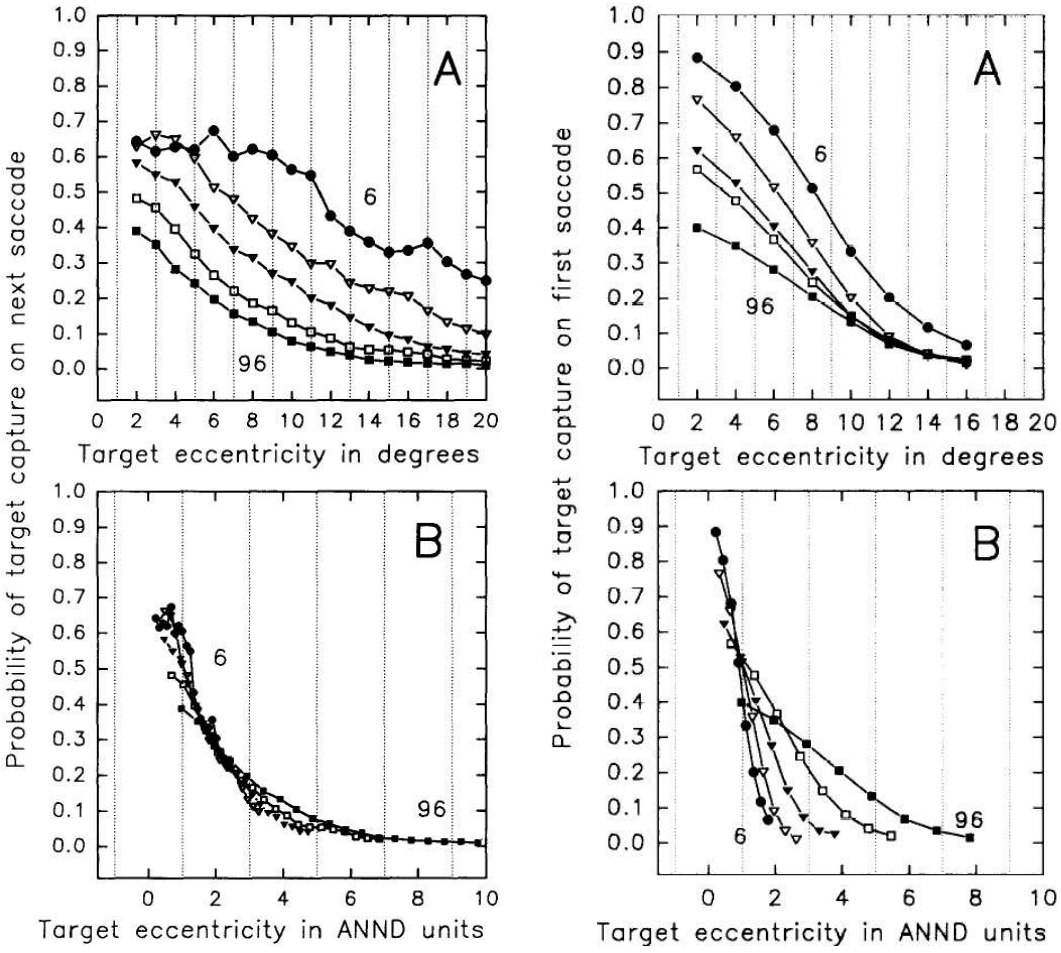

Figure 2.6: /em TOP LEFT A) Chance of fixating the target as a function of target eccentricity, for midtrial fixations. Multiple curves show different set sizes, from 6 to 92 items. Note that accuracy is worse for a given distance when there are more items present in the display and therefore on average more items close to fixation. B) The same data, normalized with the average nearest neighbor distance (ANND). The excellent fit among different stimulus conditions indicates that average nearest neighbor distance is the major determinant in the effectiveness of guidance by distance. **RIGHT** same data as left, for initial saccades. While stimulus density still effects accuracy at smaller eccentricities, it does not predict accuracy well at large eccentricities. Reproduced from Motter and Belky, 1998a

paradigm used rotated Ts and Ls very similar to those used in the experiments reported in chapter 1. In a second, displays were identical to those in the first experiment (figure 2.5), but made smaller by a factor of four. Both of these manipulations altered the curve for detection by distance. (As an aside, reducing the display to a fourth the size (1/16th area) made surprisingly little difference, and in fact improved target detection normalized by ANND somewhat, perhaps due to nonlinearities in the relation of receptive field size to eccentricity, as discussed in section. This finding helps explain the difference between the studies of Pashler (1987) and those of Wolfe et al (submitted); the prior found that more items seemed to be searched in parallel, and used a display spanning only 7 degrees of arc, as opposed to the study of Wolfe et al, that spanned 25 degrees.)

Figure 2.7 (left side) shows curves for chance of target location by target distance from the fixation point, with that distance given in terms of multiples of the average nearest neighbor distance for each set size. For the modified conjunction search (A, top left), the normalization does not work quite as well, as evidenced by the divergence of curves for different set sizes. For the Ts and Ls task (B, bottom left), the curves sit directly on top of one another, indicating that nearest neighbor distance accounts well for differences in accuracy by eccentricity across set sizes.

The average accuracy by normalized distance is markedly lower in the Ts and Ls task (figure 2.7, left bottom) than either the standard (figure 2.6, left bottom) or 1/4 scale orientation X color conjunction searches, which differ slightly from each other as well. To examine differences among stimulus types, a best fit exponential function was calculated for each of the curves shown, as well as for the modified orientation X color task. These curves are shown in figure 2.7. The different search types each produce different

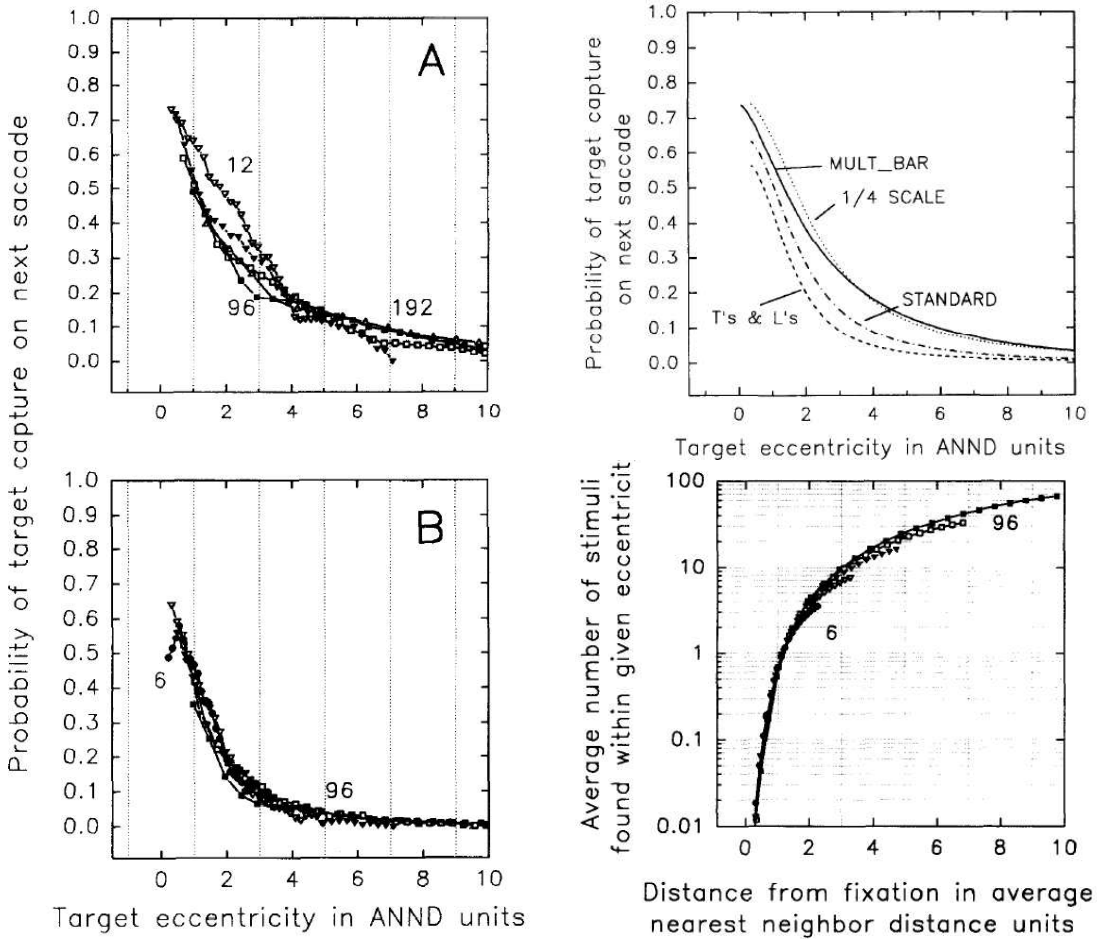

Figure 2.7:

LEFT A) Normalized data from a slightly easier variant of the color-orientation conjunction search task. B) Normalized data from a rotated L vs rotated T search task. **RIGHT TOP** Curves fitted to data from different tasks showing that the type of targets and distractors controls the relationship between target location and normalized eccentricity. **RIGHT BOTTOM** Correlation between average nearest neighbor distance and number of stimuli as close or closer to the fixation than the target. The good correlation shows that number of stimuli predicts accuracy about as well as average nearest neighbor distance. All graphs reproduced from Motter and Belky (1998a)

2.4.4 Relation of Eye Movement Data to Theories of Search

The findings of Motter and Belkey (1998a,b) are consistent with one aspect of the theory tested in the experiments of chapter one: the target can be located among a nearly fixed number of items, regardless of their distance from the fovea. The number of items closer to fixation than the target correlates well with the average nearest neighbor metric; see figure 2.7, graph on lower right. Therefore it explains the chance of locating it about as well as the average distance between items.

There is a major divergence between these data and the theory tested by the experiments here. Theories in which search processes identify a set number of items (e.g. Pashler 1987, Wolfe et al submitted) do not explain these data well. The data presented by Motter and Belky (1998a,b) require the additional consideration that more items are effectively processed when more items are present. This is also a distinction from the hypothesis tested in the experiments of Chapter 1; the number is not quite constant even within a given type of stimuli.

Figure 2.8 shows the average number of fixations for different tasks and set sizes; the number of fixations per item present decreases markedly with the total number of items in all tasks. Table 2.1 shows the average number of that are effectively processed on each fixation.[1] These calculations are made by assuming that search proceeds through half the objects on average. This is not assuming a high threshold theory of identification; the number of items processed refers only to the effective average, without assuming that processing proceeds on some discrete items and not for others. In fact processing is likely to proceed on all items at once, but with little chance of drawing attention if the item is far from the fixation point relative to the crowding of the display.

In sum, the findings reviewed above suggest that parallel processing on each attentional fixation is not limited to a certain number of stimuli as supposed by serial-parallel theories

[1] The difference between this and the estimate in Experiment 1 for number of T and L stimuli identified in parallel is explained by three factors. First the stimuli used here were about 1 degree on the long axis, as compared to about 2.5 in Experiment 1. Second six rotations were used here, as opposed to 4 in Experiment 1. Third, both Ts and Ls were targets here, interfering with practice effects

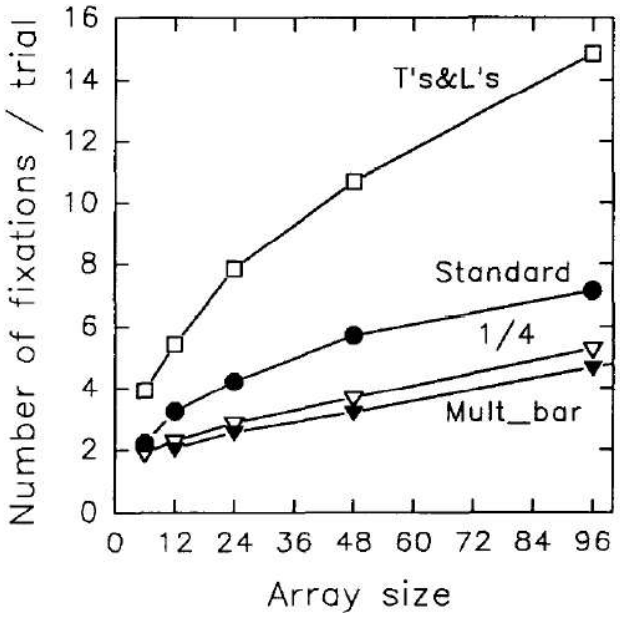

Figure 2.8: Total number of saccades to locate a target by task and set size. The number of saccades per item decreases with set size; therefore more items are on average processed when there are more items present to be processed. This finding adds another important variable to serial-parallel theories of search based on behavioral findings, notably the theory tested in the introduction. Reproduced from Motter and Belky (1998a)

Task & Set size	6	12	24	48	96
Ts & Ls	.75	1.1	1.5	2.1	3.1
Conjunction	1.5	1.8	2.9	3.2	6
1/4 size conj	1.6	2.9	4.3	6.8	9. 6
4 orientation conj	1.6	3.0	5.0	8.0	10.7

Table 2.1: Number of items effectively identified by expert monkey subjects. Top row shows set sizes; cells are estimated effective number of items identified on each fixation, for comparison to the theories tested in Chapter 1. Estimated from figure 15 of Motter & Belkey (1998b), page 1018;

.

based on behavioral results (Pashler 1987, Wolfe et al submitted, current experiments). The number of stimuli effectively processed in parallel varies with the type of stimuli, as shown by Experiment 1 here, but also varies with the number of stimuli to be processed. Every item seems to be processed, but processing is less effective when items are crowded, and when the required discrimination is more difficult. Each of these features of processing is a natural feature of the nature of representations in the ventral object recognition system, as reviewed in the next section, and the section on neural mechanisms of search with eye movements 4.2.2.

It is also possible to suppose that there are a definite and finite number of items identified on each fixation, but that these are not always the items closest to fixation. Because this number seems to change, according to the data of Motter an Belky (1998a) reported above, and the results of Experiment 1 here, that hypothesis is less plausible than supposing that all stimuli are processed with varying accuracy. This interpretation also fits nicely with functional data about the structure of the visual system, reviewed in the next section.

The findings reviewed above are consistent with a signal-detection theory, but not with any existing theory. Existing theories account for eye movements by supposing that a limited conspicuity zone is processed on each fixation (Geisler & Chou, 1995; Palmer et al., 2000). Instead it seems that the entire field of view is processed, but that items are processed with less accuracy when they are close to other items and farther from fixation. A detection theory therefore needs to explain these results by adopting considerations of stimulus crowding and variable detection with distance.

Guidance theory must also adopt these considerations of crowding and variable detection with distance to explain the results; nothing within current guidance theories accounts for the effect of stimulus spacing, although the idea of guidance falloff with distance from center of gaze is included in the sketch of a Guided Search 3.0 (Wolfe & Gancarz, 1997).

Grouping theories similarly do not account for the effective increase efficiency of search at larger set sizes. Considerations of stimulus spacing would also need to be added to this type of

theory to account for the results. Grouping theories also offer no account of the decrease of accuracy with eccentricity, and would have to add these considerations as well. The contributions of each of these theories to a more complete account of search with eye movements is discussed in section 4.2.2. The account given there is based on the neuronal basis of representation and attention, reviewed in the following section.

Chapter 3

Neuroscience Evidence

In this chapter, I review relevant neuroscience evidence, and show that it converges with the behavioral evidence reviewed above. This review supports the novel conclusion that the known structure of receptive fields explains many visual search findings. First I show that the visual features that allow efficient search are represented by neurons with small receptive fields. Second, I show that visual features that produce inefficient search are represented by neurons with larger receptive fields that include multiple items in most search tasks. Susceptibility to crowding due to size of RFs is shown to vary both with the level of the visual system (and therefore the complexity of stimuli detected) and with the distance from fixation. These findings are central to the current theory, since they offer an explanation of the results reviewed in the last sections of the previous chapter: the probability of target location is strongly dependent on stimulus crowding and distance from fixation, as well as target/distractor similarity as has long been known. Offering this explanatory link is one of the primary original contributions of this work.

Because RFs are much smaller near the center of gaze, eye movements can disambiguate representations by bringing a potential target into the fovea where RFs are relatively small. I review the evidence on attentional modulation at the neural level, and show that attention to a location can also effectively reduce the size of neural RFs, allowing detection of more difficult targets. I also show that there is ample evidence for attentional modulation by feature, and by location. These aspects of neural function are brought together in the next chapter in a

theory that is consistent with all the behavioral and neural evidence. This theory shows how a combination of attention to features and areas can provide the mechanisms of guidance and set selection that were shown to be necessary to account for behavioral data in chapter 2.

3.1 Overview of Cortical Visual Systems

The visual system is organized retinotopically so that neurons responsive to a given location are found close to other neurons with similar location response preferences (see section "Retinotopic maps"). There are a large number of retinotopically organized areas within the primate brain, so that the visual field is represented in a neural map, not once, but more than seven times.

Some of these areas are more responsive to location, motion and potential motor responses. These areas tend to be organized proceeding upward from the primary visual area of the occipital cortex, from the occipital cortex into the parietal cortex and medial temporal lobes, and toward areas representing motor movements and somatosensory maps. This progression of areas is referred to as the dorsal stream.

Another progression of visual areas seems to be more responsible for processing stimulus identity. This stream proceeds approximately around and down the head, from the primary visual cortex in the occipital lobe of the cortex, and into the lower temporal lobe. This progression of areas is referred to as the ventral stream. These pathways were hypothesized under these names by Ungerleider and Mishkin (1982) and have since been put on firm empirical footing by a large number of studies. Most areas in the ventral processing stream are interconnected to some degree by long range axonal projections. Some of these interconnections are richer than others, forming what seem to be primary information processing pathways. In the ventral stream where much experimental work has identified connections and neural representations, processing proceeds successively through areas V1, V2, and V4 in the occipital lobe, and proceeds to areas TEO, the temporal-occipital area, sometimes called posterior inferior temporal

cortex, and then to area IT, inferior temporal cortex.

Neurons in area V1, the largest visual area, are responsive to very simple stimulus features, primarily line orientation and color information, over a very small area. Neurons in the highest purely visual area, IT, are responsive to complex holistic stimulus properties, so that, for example, some particular neurons respond maximally to faces. The areas in which these neurons respond to stimuli, referred to as the receptive field (RF), become larger as one progresses up the visual system. These range from quite small in V1, responding to single line segments, to quite large in IT, encompassing most of the field of view. Receptive fields also become larger rapidly as they move toward the periphery of the visual field. Visual information is projected to primary visual cortex from the retina through the lateral geniculate nucleus of the thalamus, a subcortical region, and is transformed to some extent before reaching the cortex. For a more detailed description of this progression, and some evidence on exactly what information is represented in different areas, along with references, see section Retinotopic maps below.

3.2 Representations in the Ventral Stream Object Recognition System

The nature of the representations upon which attention and decision processes act is a key component of a mechanistic account of visual search. The overall idea is simple: items that are represented in parallel can be identified in parallel.

In order for items to effectively be represented in parallel, a substantial number of neurons must have substantially different responses to target items and distractor items. This is the case when the neurons sensitive to target/distractor differences are abundant and have relatively small receptive fields. When receptive fields contain both target and distractor items, they will respond to both, and so fail to clearly represent the target. Alternately, if only a few neurons clearly represent the target, their contribution to population responses will be too small to control either shifts of attention or responses.

There is ample evidence that features that produce fast and efficient search are represented abundently and in parallel, that is, by many neurons with small receptive fields. This evidence is reviewed in the next section. There is also evidence suggesting that features that produce inefficient search are not represented in abundently in parallel. These features seem instead to be represented by neurons with larger receptive fields that respond to multiple items simultaneously. Neural representation of these features is also likely to be much less abundant further from the center of gaze, since the receptive fields of higher level neurons usually include the center of gaze and become weaker in the periphery. The evidence on neural representation of conjunctions and of subtle feature differences is reviewed in the section following the next.

The basic idea that parallel representation allows parallel identification is obviously too simple; these representations must also be detected by response systems. These considerations are deferred until section 4.2.1.

3.2.1 "Basic Features" are Represented by Many Neurons with Small RFs

Here I review the evidence that features that give rise to parallel efficient visual search are represented by many neurons with small receptive fields. This evidence provides an explanation for the findings reviewed in section 2.4 and support for the theory of search presented in chapter 4, as discussed at the start of this section. The abundance of neurons is not directly addressed by most single-unit studies, as they do not record from neurons at random, but rather search for those that may fit a hypothesis. But the fact that neurons with particular response preferences have been identified in published reports suggests that they do exist in relatively large numbers, since it is only practical to perform even exploratory recording from a limited number of cells (in the hundreds at most), and all the studies cited found many neurons of the type reported (greater than 30, at least), In short, if more than a few neurons can be found, many more must exist.

I will provisionally use Wolfe's (1998a) definition of basic visual features as those that

give rise not only to quite efficient search, but also allow effortless texture segregation. Under this definition a simple mapping emerges. Basic features are represented by neurons early in the visual system, with small receptive fields (RF's). Here quite efficient search means less than 10 ms increase per added distractor, (also Wolfe's (1998a) definition), and a small RF means smaller than the distance between items in the paradigms used to test search efficiency and segregation.

Single cell studies have identified neurons with small RF's responding to most basic features, if we accept the list of basic features arrived at in Wolfe's (1998) comprehensive review of visual search findings. He identifies color, line orientation, motion, vernier offset (breaks in co-linearity), gloss (greater luminance to one eye), line curvature, some shapes, and depth, as defined both by stereo cues and by pictorial cues.

The first three, color, orientation, and motion, are abundantly represented in many visual areas, including V1 and most higher areas in the temporal stream. Vernier offset is also represented in V1: Cat striate neurons have been shown to respond to vernier offset with high accuracy on the order of 2-3 arcminutes (1/60 degree) of resolution) (Swindale, 1995; Swindale & Cynader, 1986). Gloss, somewhat unsurprisingly, seems completely absent from neurophysiological work. Sensitivity to curvature, however has been demonstrated in primary visual cortex; neurons responsive to specific curves have been isolated (Dobbins, Zucker, & Cynader, 1987; Versavel, Orban, & Lagae, 1990). Depth as defined by stereo cues is also represented in primary visual cortex (von der Heydt, Zhou, & Friedman, 2000). All of these results are in monkey V1 except where otherwise noted.

The case of shape is a bit more complex. Some relationships between target and distractor shapes allow efficient search, whereas others do not. A "C" among "O"s is located fairly efficiently (Treisman & Gormican, 1988), but an "E" among "S"s is much less efficient (Cheal & Lyon, 1992). One study of shape found efficient search for global shapes composed of the same elements as distractors (Nick Donnellyand Found, 2000), while another using the

same paradigm with different shapes and non homogenous (many different shapes on each trial) distractors found markedly inefficient search (Wolfe & Bennett, 1996). A particularly impressive demonstration shows quite efficient search for line drawings of a single random artifact among a heterogenous collection of animals, with a search slope of only about 6 ms/item, with search for animals among artifacts being slower (but still efficient) at 10-16 ms/item (Levin, 2001). The only reasonable explanation for this finding is that artifacts tend toward more angles, where animals tend to have more curves.

Rather than going into great detail in attempting to rank shape relationships by the efficiency of search they produce, I will note several possible explanations for the differences in search efficiency among different shape relations. Neurons representing some shape cues have been located in primary visual cortex of cats. These cues include somewhat specific angles (Versavel et al., 1990; Shevelev, 1998), and crosses and "Y" shapes (Shevelev, 1998; Shevelev, Lazareva, Novikova, Tikhomirov, Sharaev, & Cuckiridze, 2001). While these representations have not been reported in monkey V1, there is speculation that they have yet to be discovered (Das & Gilbert, 1999). In addition, neurons in area V2 show selectivity for line conjunctions of specific configurations, as well as specific radial and spiral gratings (Hegde & Van Essen, 2000). These neurons have RF sizes of 1 to 3.5 degrees each, so that some RF's would contain only one stimulus in most paradigms. Efficient search may be found when shapes are distinguished by the simple shape aspects represented by these low-level neurons with small RF's, but not for complex shape features represented by neurons with larger RF's as discussed in the next section.

Many depth cues apparently produce efficient search. Wolfe (1998a) summarizes research indicating that efficient search is possible for three-dimensional properties depicted by line drawings, apparent reflectance, shading cues, occlusion cues, slant from texture, and impossible shadows.

There are two possible exceptions to the rule that features that produce efficient search

are known to be abundantly represented by neurons with small RFS. The two such exceptions I am aware of are faces by both outline and eye-nose-mouth configuration (Hershler & Hochstein, 2005), and shape by shading (Enns & Rensink, 1990). While these representations have not been found in neurons with small RFs, they have not been searched for specifically. Both of these types of items are behaviorally important in peripheral vision, faces for social reasons and for predator detection, and shape from shading for movement purposes. Therefore it seems possible that representations for these objects will be found with further research.

Every feature that has been found abundantly represented by neurons with small RFs produces fast and efficient search. Most of these "basic features" have been found in area V1. But the intent is not to suggest that efficient search requires representation in V1, or that search operates on such an early representation.

There is ample evidence that search does not operate exclusively on representations in V1. Global shapes can allow efficient search where local elements would not (Nick Donnellyand Found, 2000; Treisman & Sato, 1990), and interfere with search that would be efficient if their constituent elements were presented without global organization (Found, 1997; Suzuki & Cavanagh, 1995). In similar cases, however, search for global shape can be less efficient than search for local shape (Boutsen, 2001), implying that search operates at multiple levels of representation.

Rather than mapping V1 representation to the idea of "basic features" employed in behavioral vision studies, I want to suggest that the neural correlate of basic feature status is representation by neurons with RF's small enough to encompass few more than one stimulus in a given paradigm, with the additional requirement that these neurons be relatively abundant. The logic behind this proposed mapping will become clearer as we consider cases in which RF's include more than one stimulus.

3.2.2 Conjunctions are Represented by Neurons with Large Receptive Fields

This section reviews evidence that target/distractor differences that produce inefficient search are first detected by neurons that have relatively large receptive fields. These large RFs include multiple

It has commonly been assumed in theories of visual function that the visual areas between V1 and IT code for features of intermediate complexity. Recently more evidence on the nature of those representations been produced. New evidence shows that neurons in area V4 respond selectively to both specific curves (Pasupathy & Connor, 1999) and to relations among these curves when they form objects (Pasupathy & Connor, 2001), as well as to different complex colored shapes (Kobatake & Tanaka, 1994). These neurons have receptive fields as small as two degrees at the center of gaze (fovea) and receptive fields up to seven degrees wide and larger at higher eccentricities.

A recent study by Hegde and Van Essen (2000) has shown that neurons in area V2 are preferentially responsive to a variety of grating types and contour stimuli. In their study, they demonstrated that cells in V2 are substantially more responsive to stimuli more complex than the cartesian gratings and oriented bars that V1 neurons seem to detect. They observed that 61 percent of surveyed cells preferred hyperbolic or polar gratings (concentric circles or spiral patterns) over the cartesian gratings preferred by V1 neurons. Moreover, they demonstrated that approximately 80 percent of surveyed neurons responded most strongly to stimuli other than the oriented bars also used as stimuli for V1 cells. Many cells in their study strongly preferred particular stimulus types, and orientations within that type. In decreasing order of number of cells with maximum response to each stimulus type, V2 neurons were found to respond to bars, 3/4 arcs, right angles, stars and circles, semicircles, Y shapes, crosses, and obtuse angles. Simple bars were the second most preferred stimulus, but they are a small minority of the sum of preferred shapes.

Each of these neurons was tested exclusively with the color of stimuli most preferred

as a single bar. Therefore no inferences can be drawn regarding the conjunctive nature of representation in V2. Receptive field sizes varied from 1 degree to 3.4 degrees, with a mean size of 1.4 degrees. The authors do not report a systematic variation of receptive field size with stimulus selectivity or eccentricity, although it is likely that they did vary with eccentricity. These RF sizes vary from about the size of 1 object to that of several objects in most visual search experiments.

Pasupathy and Connor (1999) demonstrated effects in V4 neurons similar to those found in V2 (Hegde & Van Essen, 2000), although they found these preferences to be more pronounced. They used stimuli in which a particular curve or angle was represented at high contrast within the receptive field of the neuron. The found that most neurons responded better to these shapes than to edges or to bars. They found that around 1/3 of tested neurons displayed selectivity for specific features. They found a strong bias toward contours convex to the center of the RF, implying that V4 cells might be sensitive specifically to objects within their RF.

In a followup study Pasupathy and Connor (2001) tested the responses of V4 neurons to complete shapes entirely within their receptive fields. While they found no neurons responsive to a particular global shape, many were responsive to particular regularities in those shapes, i.e. a specific curve in a specific position relative to the shape's center. Between 1/2 and 1/4 of the cells initially surveyed seemed to respond differentially to shape. In this detailed experiment they characterized each stimulus in terms of each one of its component curves; each stimulus contained from four to eight curve segments, two to four convex curves and a matching number of convex curves or angles. They analyzed the response of each cell to the component pieces of each stimulus in terms of curvature, orientation, angular position and radial position. The predicted responses calculated by considering curvature and angular position were far superior to that calculated from orientation or radial position. These results also indicated, although not conclusively, that position relative to the object center was more important than absolute position in determining V4 neurons' responses.

Most neurons were also sensitive to some degree to the curves adjoining their preferred curve. The nature of bordering curves was almost as important as the primary response feature in predicting responses, especially at small curve values rather than straight lines or sharp points. Including adjacent curves significantly improved goodness-of-fit for 94/109 cells. These neurons therefore code for a conjunction of contours, although limited inferences can be drawn about the possibility of these neurons representing conjunctions. Each stimulus was presented in the color preferred by the particular cell during initial testing. The fact that neurons displayed color selectivity as well as shape selectivity implies that they encoded these conjunctions to some degree.

In both of these studies receptive field sizes were assumed to be that reported by Gatass, Sousa, and Gross (1988), 1 degree plus .625 degrees per degree of eccentricity. Assuming that the previous estimate applies equally well to the sample showing preference for contours and contour conjunctions (a questionable assumption), these RF's would encompass between one and several stimuli in visual search experiments, depending upon the particular interstimulus spacing used, the particular set size, and the eccentricity of the items.

In contrast to these studies, Kobatake and Tanaka (1994) had previously found little preference for complex stimuli in area V2. Most cells in V2 responded maximally to simple stimuli. Only one cell was found which preferred a stimulus of rings, and another which preferred a tapered bar. However, they obtained this result using methods that would prejudice findings toward more complex stimuli. They began with complex stimuli, and simplified them to isolate the response properties of individual units. The units in V2 responded in a relatively non-preferential way to the complex initial stimuli used, and therefore were not further surveyed with simpler stimuli.

This study did, however, isolate stimulus preferences in V4 and TEO of intermediate complexity. These consisted of shapes, combinations of shape and texture, and combinations of shape and color. These neurons differed from those they surveyed in IT by responding to some

degree to simpler stimuli, having limited receptive fields, and neighboring neurons without identifiable complex stimulus preferences.

Kobatake and Tanaka (1994) found that 3/4 of anterior IT cells responded maximally to some complex feature, and responded with no more than 1/4 of that response to any simple feature (colored oriented bars or spots) About 1/2 of cells in V4 and posterior IT (probably area TEO) responded maximally to some simple features as well as responding well to complex stimuli containing them. The remainder responded maximally to complex features, like the majority of anterior IT cells.

Features in V4 and posterior IT included shapes, shape/texture combinations, and shape/color combinations. Some of the features were subjectively as complex as those found in anterior IT. Cells in V2 had a receptive field size of 1.7 with a Standard Deviation of 1.0 degree, those in V4 averaged 4.8 degrees with SD 2.5 degree; posterior IT averaged 5.4 with SD 2.8; and antererior IT averaged 16.5 with SD 6.5 degrees. While they did not report the dependence of RF size on eccentrity, the published RF diagrams make it apparent that extrafoveal RF's were considerably larger, and suggests that most RFs include the center of gaze. These findings are similar to those of the range of similar studies reviewed in Rousselet, Thorpe, and Fabre-Thorpe (2004)

These studies found neurons with clear preference for stimuli conjunctive both within and among form and color. The size of receptive fields for these cells ranged as low as a few degrees, and as high as the majority of a visual hemifield, and the complexity of conjunctions represented seems to be loosely related to the RF size of each neuron. These receptive fields would encompass a minimum of one stimulus in most visual search paradigms, with most receptive fields outside the central few degrees of space covering the space of many stimuli. The number of stimuli in each RF on average would also increase with eccentricity, as discussed in section 3.2.4 below.

A more recent and comprehensive study of RF sizes and placements in IT cortex has confirmed these results. Op De Beeck and Vogels (2000) reported a mean size of 10 degrees

with a standard deviation of 5 degrees, minimum of 2.8, and maximum of 26 degrees. They also reported that, while most RFs include the fovea, some do not; RF centers were sometimes up to 8 degrees away from the fovea.[1]

The studies reviewed have identified neurons that encode conjunctions of features that produce inefficient search, at least when distractors are inhomogeneous. These findings suggest that neurons capable of encoding either complex shape (sometimes referred to as a conjunction of simple shapes or line orientations), or conjunctions of separate features such as shape and color, have receptive fields that encompass more than one stimulus when the stimuli are outside the fovea. In the absence of attention, these neurons will respond to all of the stimuli within their receptive fields. These responses are a rough average of the rates observed in response to each stimulus individually (Chelazzi et al., 1993; Reynolds, Chelazzi, & Desimone, 1999; Chelazzi, Miller, Duncan, & Desimone, 2001 and see also section 3.3). Thus these neurons will produce an ambiguous firing rate. This situation is illustrated in Figure 3.1A. In a case where only two stimuli were present, the rate is roughly 50% similar to that produced by each stimuli. In cases where many stimuli are contained within the receptive field, response rates would bear little resemblance to that produced by any one stimulus. This situation is depicted in figure 3.1A.

3.2.3 Spatial Attention Narrows Functional RFs

Attention can act to disambiguate responses of mid-level conjunction-encoding neurons by effectively reducing the size of the receptive field. This happens in two ways. The more

[1] The generally large size of RFs in area IT raises the question of how objects are identified at all in a cluttered scene. One possibility is the existence of small RFs. A recent study has shown that RFs near the center of gaze are often small (DiCarlo & Maunsell, 2003). IT cortex neurons with complex response properties were shown to decrease their response rate to 60% of maximum with a change in item position of only 1.5 degrees, and RF size was estimated at 2.5 degrees. However, monkeys in this study were trained to identify small (less than 1 degree) items in the presence of distractors, so these RF sizes may be an exception. Another possible answer is that spatial attention is used for every identification. Rolls, Aggelopoulos, and Zheng (2003) demonstrated that IT neuron RFs shrink dramatically in natural scenes; this could be the result of either spatial attention, or of competitive mechanisms. However, other data suggest that spatial attention is not necessary in every case. However, the two studies above, among others, have found that neural responses in IT are strongest to items at the center of gaze. This property may allow IT neurons to develop relatively unambiguous representations to the fixated object through competitive processes acting without spatial attention and over a relatively short period of time.

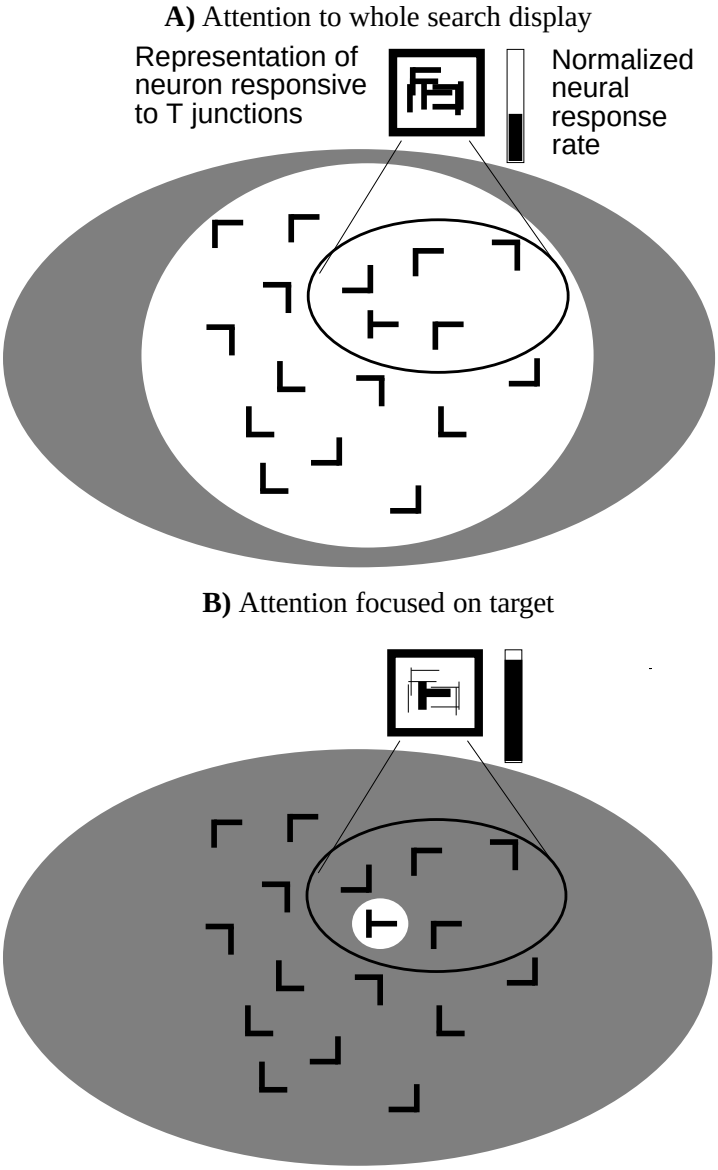

Figure 3.1: Representations in one mid-level neuron responsive to the target stimulus. All stimuli are outside of the fovea. **A)** covert attention (white) is devoted to the portion of the visual field containing the stimuli, while irrelevant areas are unattended (gray). The response of the neuron (box) is an average of that expected to any of the stimuli within the receptive field (oval). All stimuli are represented, producing an ambiguous response level. **B)** attention is now devoted only to the target, biasing the neuron's response toward that given to the target in isolation. A response can be based on the relatively unambiguous target representation. The response may be based directly on these mid-level (V4 or TEO) neural signatures, or may be passed to IT neurons, whose response is disambiguated by receiving unambiguous input

common way is simply by a shift of gaze; the smaller receptive fields in the fovea lead to an unambiguous representation of the attended item. The second way that spatial attention narrows functional RFs is slightly more subtle, and comes into play in search with covert attention.

This effect is illustrated in Figure 3.1B. Covert attention has been demonstrated to enhance the representation of attended stimuli, and reduce that of nearby (relative to RF size) items (e.g. Motter, 1993; for the likely mechanisms, see section 3.3). Covert attention has been shown to bias the response of mid-level (V2,V4) neurons so that responses are more similar to those observed when only the attended stimulus is present in the receptive field (Reynolds et al., 1999; Chelazzi et al., 2001). Thus when attention is deployed to a particular location, neurons encoding conjunctions will respond relatively unambigously, so that these neurons will drive responses further up the processing heirarchy, and from there can drive response systems. In this way a specifically attended stimulus can be clearly represented by neurons whose receptive field encompasses many surrounding stimuli.

It is likely that this is the mechanism by which visual searches are performed when displays are persistent and eye movements are not possible. This strategy need not be limited to fixating one item at a time with spatial attention. Covert attention to an area containing part of the RF should bias the response rate of the neuron toward all stimuli within the attended area. Such a situation is illustrated in figure 3.2 A). Attention to an area containing only one stimulus within a particular neuron's receptive field could be expected to bias the response of that neuron in the same way that attending to only that stimulus would (figure 3.2 B). This mechanism would allow unambiguous identification by some neurons, allowing attention to be usefully deployed to a small number of stimuli at once.

Because neither receptive fields nor attended areas are likely to have the sharply defined borders depicted here, it is unlikely that this strategy provides complete disambiguation even for single neurons. It is also important to note that the number of neurons giving a relatively unambiguous response is critical, as well as the clarity of response of each individual neuron.

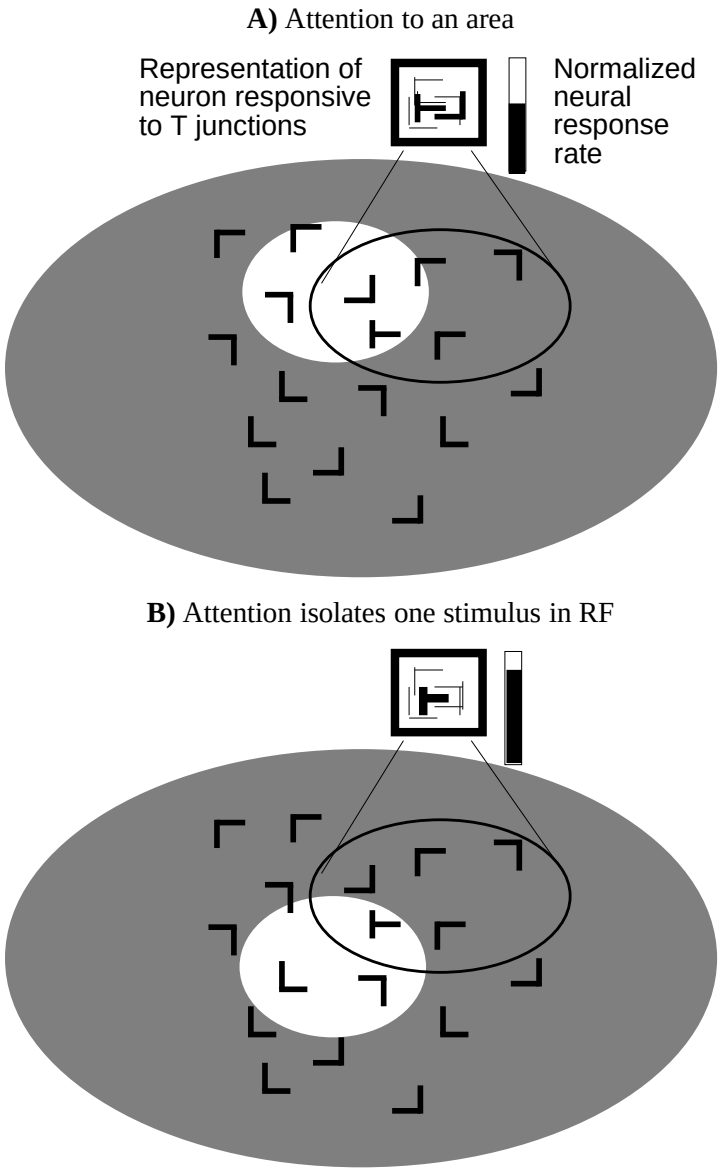

Figure 3.2: Hypothesized representations in one mid-level neuron responsive to the target stimulus. All stimuli are outside the fovea. **A)** attention (white) is devoted to an area of the visual field containing several stimuli, while the rest of the display is unattended (gray). The response of the neuron (box) is biased toward that given by the two stimuli within both the RF (oval) and attended area. All stimuli are represented, but the response is less ambiguous, allowing some identification. **B)** attention here encompasses several stimuli, but only one within the depicted neuron's RF. The response is as unambiguous as the case in which attention is devoted only to that stimulus (figure 3.1a.)

3.2.4 Receptive Field Variance with Eccentricity

The above mechanisms of attention acting on particular object representations address visual search with covert attention. However, most visual searches are performed not with covert attention, but with eye movements. The variance of receptive field sizes with complexity of preferred stimuli helps explain search with eye movements, but there is another important factor: the increase of receptive field size with eccentricity.

Taking into account the larger size of receptive fields at peripheral locations provides an explanation of eccentricity effects on search performance for conjunctions (Scialfa & Joffe, 1998), and the increase of the eccentricity effect at larger set sizes (when more than one item is more likely to be in each receptive field), (Carrasco et al., 1995). The decreased strength of large receptive fields at higher eccentricity, and the relative decrease in number of neurons representing items at large eccentricity (the cortical magnification factor) explains the negation of the eccentricity effect (at least at small set sizes) by increasing stimulus size in proportion to each item's eccentricity (Carrasco & Frieder, 1997). More central to the current theory is the link with the studies of Motter & Belky (1998a,b) that show that the chance of finding a target decreases with its eccentricity, and that this effect is dependent on average item spacing.

Receptive fields become larger at greater eccentricity. This is true at all levels of the visual system, but is a much larger factor in the higher areas. The variance in V1 is only between .5 and 1.5 degrees, and this variance is not highly dependent on eccentricity. RF sizes in V2 vary from .5 to 4 degrees; those in V4 vary from 1 to 20 degrees; in area TEO they range from 2 to 25 degrees; and in area IT they range from 2.5 degrees to 70 degrees, most of one visual hemifield (Rousselet et al., 2004). The large variance in RF sizes in higher areas is due in large part to dramatically larger RFs at greater eccentricity. For instance, in area V4 Gatass et al. (1988) report that RF sizes were approximately 1 degree plus .625 degrees per degree of eccentricity. Figure 3.3 depicts the spread of receptive field sizes laid out on a visual search array similar to the ones used in the experiments reported in chapter 1.

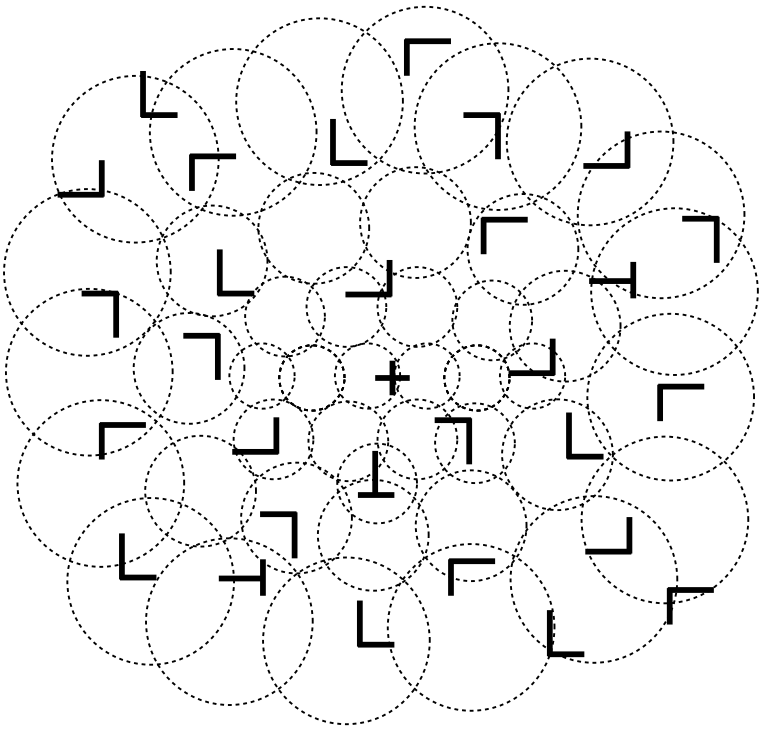

Figure 3.3: Receptive fields of neurons substantially more responsive to T junctions. The T close to fixation can be identified, since a number of neurons responsive to its particular junction will contain only that stimulus. The T on the right will likely not be identified on this fixation, since most neurons responsive to T junctions over L junctions will have many stimuli in their receptive fields, producing ambiguous representations. The T on the bottom may be identified on this fixation, since some of the appropriate neurons will contain only it, due to fewer surrounding stimuli.

The average receptive field size of neurons that distinguish a target and distractor will vary with both the distance of stimuli from the fixation point, and with the particular features that distinguish the target from distractors. A distinct neural signature at the population level is necessary to allow responses or gaze shifts to the target. Therefore, eccentricity and discriminability are important factors in determining whether a target is located and/or attended.

These are not the only factors. Another critical factor is the abundance of neurons that respond differentially to the target and distractor. Neurons that respond differentially to very subtle differences in stimuli may not exist at large eccentricities. I have not located any study that has carefully examined the precision of neural representation with eccentricity. Instead studies of IT neural response properties report that the receptive fields of IT neurons are more heavily centered on the fovea. Neurons in IT cortex have receptive fields that usually include the center of gaze (Rousselet et al., 2004). This is to be expected, since subtle discrimination for object recognition is almost always performed by fixating the object of interest.

While the population dynamics obviously play a large role, it seems that variation in average RF size with neuronal response properties and with eccentricity are a large part of the explanation for visual search with eye movements, as reviewed in section 2.4. This explanation is explored further in section 4.2.2.

3.3 Competitive Basis of Attentional Effects

The theory of visual search developed in the next chapter assumes that attention operates through top-down excitatory biasing in an environment of competition for representation. The issue of whether top-down attention operates by simple excitatory biasing (giving extra input to all relevant neurons), or whether the mechanism is more complicated, is somewhat peripheral and is therefore discussed in the appendix, section C. The assumption of attention as biased competition is somewhat more central, and the theory as well as its likely neural mechanisms is therefore discussed briefly here.

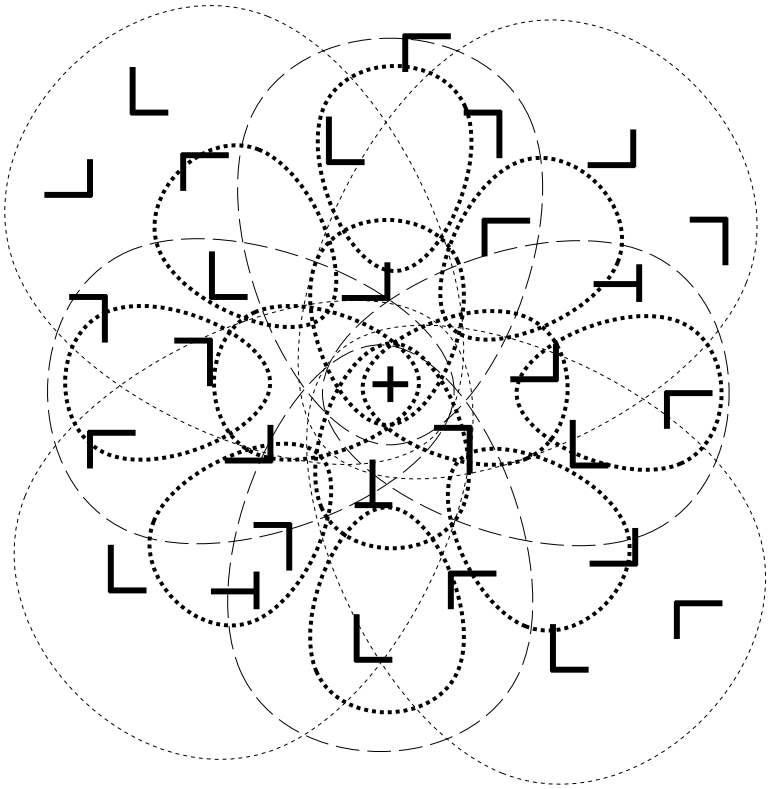

Figure 3.4: Receptive fields of neurons in IT cortex. It is possible that responses are always based on these neurons; peripheral crowding and loss of sensitivity are likely related to the large size and limited strength of RFs in the visual periphery at the level of IT cortex and related areas that perform object recognition

The theory of biased competition among stimuli for neural representation (Desimone & Duncan, 1995) is now widely supported by experimental results. Within this framework attention is viewed as the result of a competition among stimuli (and usually corresponding locations), with this competition biased appropriately by task demands. It is likely that this competition plays out among many cortical areas that have a retinotopic organization, and many others that represent space with different references. Spatial representations exist in all or most visual areas, and additionally in premotor and somatosensory cortex, in the frontal eye fields and other visually responsive areas of frontal cortex, and in subcortical areas, notably the basal ganglia and the superior colliculus. It is likely that some areas play a more central role in the competition for representation, but that a large number of areas participate in this process. This idea is discussed more fully in Appendix B in the appendix, and is particularly lucidly presented and well supported by Duncan, Humphreys, and Ward (1997).

The brain region responsible for spatial attention was hypothesized to be the parietal lobe by early neuropsychological studies of brain-damaged patients (Posner, Walker, Friedrich, & Rafal, 1984). A recent meta-analysis of patient and monkey lesion studies has identified the superior temporal lobe (Karnath, 2001). Monkey single-cell studies have identified both the lateral intraparietal area (Kusunokia, Gottlieba, & Goldberg, 2000) and the frontal eye fields (Bichot & Schall, 1999), which are important for producing both saccades (Dias & Segraves, 1999) and covert attentional shifts (Thompson, Bichot, & Schall, 2001). A meta-analysis of human neuroimaging results has implicated temporoparietial areas (Corbetta & Shulman, 2002), in accord with the results of neuropsychological studies of people with localized brain damage (Posner et al., 1984). Computational studies emphasize the possible role of areas as early as primary visual cortex in producing activation maps of saliency through lateral interactions (Li, 2002).

The large number of areas implicated in calculating the salience of stimuli is quite consistent with the proposal that a competition for representation occurs over all areas that represent

space (Duncan et al., 1997). A new integrative model has argued that the role of salience map is played collectively by the parietal eye fields, including area LIP and surrounding areas, the frontal eye fields, the superior colliculus, and the ventral pulvinar and associated areas of the thalamus (Shipp, 2004). The basal ganglia may also play an important role in this process, since the basal ganglia modulates activity in the frontal eye fields by acting on the thalamus (Alexander, DeLong, & Strick, 1986), and this modulation is likely to make use of unique learning mechanisms that give it a role in strategically manipulating attention (Hazy, Frank, & O'Reilly, in press).

Importantly, the ventral visual system is also a part of this distributed competition. Because it represents particular stimulus features, top-down support to a particular feature can be the deciding factor in the distributed competition, and determine where spatial attention is focused. Spatial attention can be guided by object representations in the ventral stream since some location information is available from the spatially specific receptive fields of neurons in lower areas and to some extent even area IT, using a coarse coded representation (O'Reilly & Munakata, 2000).

3.3.1 Attention to Features

Attention to particular features is one key mechanism of the theory proposed in chapter 4. Observers can selectively process items based on at least some features of objects; studies have focused on attentional selection by color. The behavioral evidence for the involvement of featural attention in search is discussed chapter 2, particularly in sections 2.4.2, 2.4.2.1, 2.3.1 and 2.3.2.

Attention deployed to a particular feature enhances the neural representation of stimuli with that feature. This has been demonstrated in the case of direction of motion (Treue & Trujillo, 1999) in area MT and orientation vs. color in area V4 (McAdams & Maunsell, 2000) (See the end of Appendix C for complete descriptions of this study; it suffices to say here that

adequate controls for spatial attention are used). These studies show that attention biases neural representation of a stimulus according to the presence of a specific feature in that stimulus, so that attention to that feature increases neural response associated with that stimulus. Attention to a different feature represented in the same area (e.g. opposite direction of motion in Treue and Trujillo (1999)), decreases neural response relative to an unattended condition.

It has been shown that the selective enhancement of neurons representing a target feature leads to a later (around 200 ms) dramatic enhancement of activity related to that object (Motter, 1994a, 1994b) probably as a result of spatial attention being brought to bear on it. Because attention can be directed effectively not just to locations or objects but to features of objects, attention can be deployed before objects have been identified. Featural attention biases the competition among objects so that those containing the attended feature will tend to win out and capture spatial attention.[2]

The fact that attention to features can be used to select a particular object with spatial attention allows for a stronger attentional modulation, since the effects of attention to features are more modest than those of spatial attention (McAdams & Maunsell, 2000). This mechanism plays an important role in the strategy of search by groups, as discussed in the next section.

3.3.2 Gating Mechanisms for Strategic Maintenance of Attention

The main original contribution of the theory of search proposed in the next chapter is a focus on strategy variants employed for different search tasks. Implementing these strategies requires mechanisms to selectively maintain attention and response biases. It is clear at the behavioral level that people can strategically maintain attention. The mechanisms by which this is accomplished are somewhat outside the scope of this review, but one possible mechanism is

[2] It is worth noting that the term "attention" can refer both to a cause of preferential processing, and to the result of that preferential processing. It could also be said that top-down featural biases cause items that have the biased feature to "receive attention" first. I use the term in both senses, with an attempt to specify the meaning in ambiguous cases. Saying that "featural attention causes stimuli containing the attended feature to dominate in higher level representations" (the framework of Desimone and Duncan (1995)) is the same as saying that "attention is guided toward features salient for the task" (framework of Wolfe (1994).

included in the interest of providing a complete proposal in the next section.

Working memory representations in PFC are a likely source of target templates for attentional biasing in the biased competition framework (section 3.3 above, Herd, Banich, & OReilly, IP; prefrontal activations are found in tasks requiring attentional maintenance of the task goals for performance, such as the Stroop task (Banich, Milham, Atchley, Cohen, Webb, Wszalek, Kramer, Liang, Barad, Gullett, Shah, & Brown, 2000). There is some question over whether WM is the source of attentional biasing in visual search. One study has reported no interference on search slope of a conjunction search from a concurrent visual working memory task (Woodman, 2001). However, they did in fact find a large interference effect on the intercept of the search function. They also used stimuli designed to be difficult to distinguish without focal attention, so the task may have been largely unguided in both cases. The same authors have reported that a working memory task for spatial locations does increase the search slope of a VS task performed during the memory span (Woodman & Luck, 2004). A recent theory of interactions between Basal Ganglia (BG) and Prefrontal Cortex (PFC) has proposed that the BG and related brain areas serve to decide which representations are maintained in WM in PFC (Frank, Loughry, & O'Reilly, 2001; Frank, Unpublished Dissertation; Hazy et al., in press). This hypothesis is based on an analogy to the well understood function of BG in gating motor responses. In this respect the BG aids in sequencing motor movements, by enhancing cortical representations for each movement when the context is correct (including goals, preceding movements, and sensory input), and suppressing representations of movements when the context is incorrect for their performance.

The anatomical relationship between BG and PFC areas contributing to working memory and attentional control are highly similar to those between BG and motor areas of frontal cortex (Alexander et al., 1986). The theory of working memory maintenance proposes that the BG serves a highly similar purpose for WM and attention as it does for motor movements: it enhances these representations when they are situationally appropriate according to goals, sensory

state, and current state of WM representations. This enhancement has the effect of gating appropriate representations into maintenance in WM, and suppressing inappropriate representations (Frank, Unpublished Dissertation).

Basal Ganglia updating decisions are theorized to be based on a more sophisticated reward learning system than that employed by general cortical mechanisms (O'Reilly, Frank, Hazy, & Watz, submitted). This may be the mechanism by which task instructions are converted to behaviors. However, this mechanism is well beyond the scope of this review. For the purposes here it suffices to state the hypothesis that BG mechanisms use goal information to intelligently gate into and maintain in WM representations that provide top-down bias to strategically produce both featural and spatial attention.

The PFC/BG system is also proposed to strategically maintain representations of the current strategy. These representations can in turn be accessed by the PFC/BG updating process in deciding when to strategically maintain spatial and featural representations for attention. This system can give rise to the use of distinct strategies, some of which involve a sophisticated sequence of maintaining attention and. The maintained representation of strategy can also be accessed by the PFC/BG loops that control responses, e.g. to respond that no target is present if the plan is to maintain spread attention, and no shift of spatial attention has indicated a target's presence within a period of time determined by the current strategy.

Chapter 4

A Neural Theory of Visual Search

This section ties together the behavioral and neuroscience evidence reviewed in the previous chapter, and the experimental work of chapter 1, into a framework for understanding visual search. This framework is intended to be more comprehensive than existing theories. It identifies the main strategies used for different search tasks, rather than focusing on one or two such strategies, as have all previous theories (of which I am aware). It deals with the mechanisms of search at both an algorithmic level and at a mechanistic level. It is consistent with all findings I am aware of that bear on mechanisms of visual search. Because of this breadth, and because the theory allows for different strategies of search, the theory does not make quantitative predictions for particular search tasks. The theory is not instantiated either mathematically or as a neural network model, although it is formulated in terms appropriate for guiding future neural network modeling. The theory is referred to as a neural theory of visual search, and abbreviated NTVS.

If NTVS were to be summed up in a single sentence, it would be “Visual search is complicated” or less tersely, “Visual search is the result of several distinct modes of interaction between systems for featural and spatial attention and the complex representational structure of the object recognition system”. If the original contribution of the theory were to be summed up, it would be “visual search is limited by the size of RFs of neurons representing the features that discriminate targets from distractors.” Each of the above statements needs further explication. The rest of this section describes the theory with less brevity but more clarity.

At the mechanistic level (and in brief), NTVS proposes the following. Search performance arises from an interaction of several brain systems. The process is based on competitive interactions within the ventral visual object recognition stream and the dorsal visual location and action stream, as discussed in section 3.3. These competitive interactions are strategically biased by working memory mechanisms to produce attention to locations and features. The motor system produces actions based on the high-level visual representations that result from this biased competition. These responses and the attentional biases are guided by a representation of the plan or strategy for the search. Different sequences of attention and response criteria produce several distinct strategies or modes of search.

At the broad algorithmic level, these modes are parallel search, serial search with eye movements, serial search with covert attention, and subset search with covert attention. Each overall strategy has variables that are also under strategic control, producing a spectrum of strategies. The next section briefly describes the systems underlying search performance. The following section describes the different modes of search in greater detail, at both the algorithmic and mechanistic levels.

4.1 Overview of Relevant Neural Systems

This section sums up the reviews given in the previous chapter of the systems underlying visual search behavior. It also provides sketches of the response system and the plan representation, the details of which are not central to the theory offered here. These systems and their interactions are depicted in figure 4.1.

(1) **Visual recognition system.** This is the central and most complex component. This is identified anatomically with the ventral visual stream. This system consists of a hierarchy of areas. Early areas represent simple features in specific locations, mid-level areas represent simple conjunctions of features over a number of locations, and later areas represent specific complex objects, but over much of the visual field. This

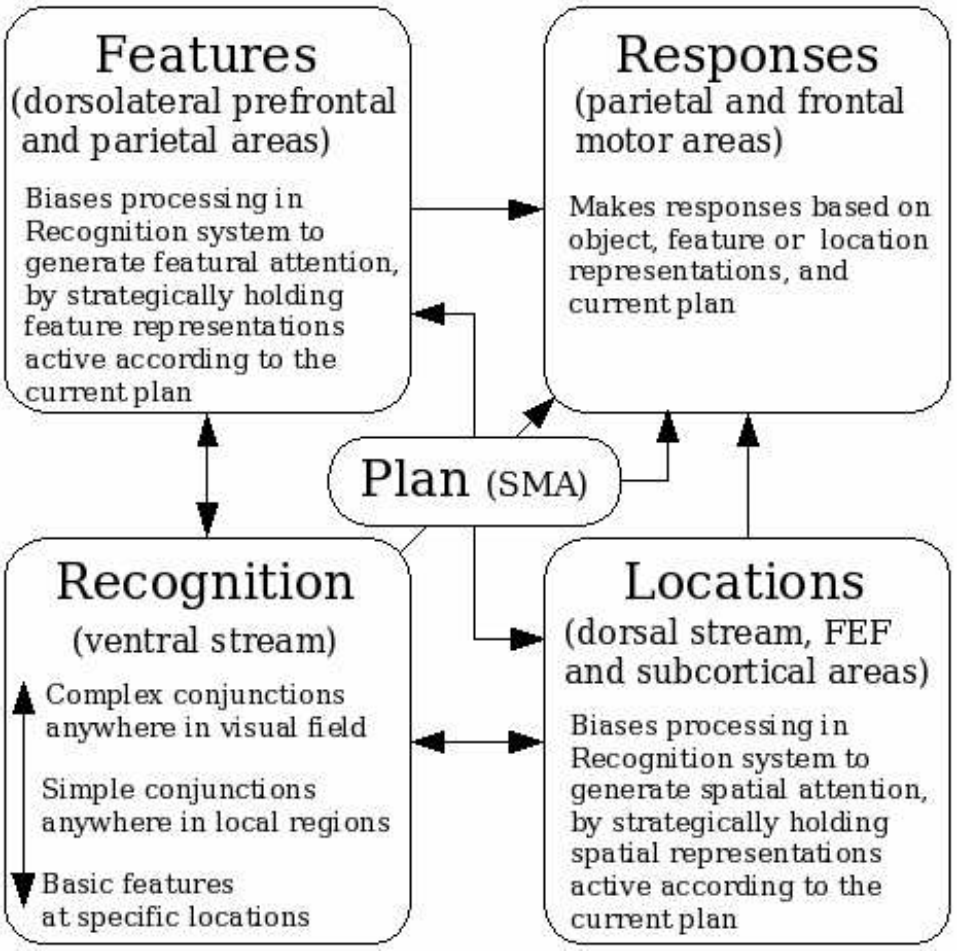

Figure 4.1: Proposed systems and interactions underlying visual search.

component is discussed extensively in sections 3.2.1, 3.2.2, and 3.2.4.

(2) **Spatial attention system.** This system instantiates spatial attention simply by representing space, although some of the areas involved serve more sophisticated roles for action-in-space for other tasks. These representations guide eye movements and direct covert attention when the eyes are fixated. Both of these types of attentional movement serve to improve representations of objects at attended locations, but at the expense of degrading representations of unattended objects.

This system always serves to focus spatial attention, but it does so based on either bottom up or top-down information. It can take bottom-up input from the recognition system, and thereby represent the locations of attended features. These representations can be sharpened by the gating functions of the Basal Ganglia (BG) portion of the system. It can also use this strategic gating function to focus attention on areas based on cuing, memory, or in a random attentional fixation strategy. The location system is identified with a complex of areas including posterior parietal cortex, frontal eye fields, and subcortical areas, as discussed in section 3.3. The mechanism by which this system focuses spatial attention is also discussed in section 3.3 and at length in Appendix B and Appendix C.

(3) **Feature biasing system.** This system represents feature information regardless of location, and is directed by top-down control. Like the location system, this system's representations bias processing in the recognition system. The effect of biasing some features over others is to make them relatively more active within the recognition system, which can in turn trigger the spatial attention system to focus attention at that location. The neural mechanisms underlying this system likely include dorsolateral prefrontal cortex and areas of parietal cortex. Importantly, BG loops similar to those strategically controlling spatial attention also exist for dorsolateral PFC, so that a target template can be maintained according to the a particular strategy. Experimental

evidence for feature based attention is reviewed in section 3.3.1 and Appendix C. The strategic application of featural and spatial attention give rise to the various strategies and strategy variants described in the next section.

(4) **Response system.** This system makes behavioral responses based on information within any of the other systems. Generally responses are made when the recognition system contains information that a given target is likely present. However, responses can also be based on the location system representing one location more strongly than the others, or on a criteria from the plan representation specifying conditions for deciding that no target is present. This system is identified with pre-motor areas in frontal cortex.

(5) **Plan representation system**. This system simply represents a plan in working memory. Sequencing and coordinating attentional shifts and responses is not performed by a single system. Instead coordinated strategic performance is the result of the other components referencing this information to perform a task effectively as a unit, and with a strategy appropriate to the task. The representation of the current strategy is maintained in working memory, identified with supplementary motor areas in prefrontal cortex. These representations are maintained by mechananisms in the PFC-Basal Ganglia loops, as discussed in section 3.3.2

4.2 Interactions of Neural Systems to Produce Strategies of Visual Search

This section describes the NTVS account of each strategy. Because the strategies, neural systems underlying them, and the evidence for both have been discussed in the previous two sections, the descriptions here serve primarily to bring these separate descriptions together. The other function of these sections is to identify some mechanisms and strategy variants that are not clearly indicated by the existing evidence, in order to suggest directions for future research.

The section on each strategy is divided into several subsections. The first briefly de-

scribes the strategy at the algorithmic level, including the type of tasks for which it is usually employed. One or more subsections provide a more detailed description of the interaction of neural systems that produce the behavioral signature of that strategy. In the last subsection on each strategy, some possible strategic variations are discussed, and unanswered questions are identified for future research.

4.2.1 Parallel search

Parallel search is the strategy employed when search displays are presented briefly, or when targets are distinct enough from distrators to allow so-called “popout”, or rapid identification of the target anywhere in the display. This strategy is often more effective in these situation because the speed of spatial attention is likely relatively slow (section 1.1.5). In these situations, attending to some items means leaving others relatively unprocessed. In this strategy, attention is spread to the whole scene, and a guess about the presence or absence of the target is made.

No featural or spatial biasing is necessary when targets are highly discriminable from distractors in certain ways that allow a relatively robust representation of the target at the higher levels of the visual recognition system on which a decision is based. However, featural attention is often used when the identity of the target is known in advance. Attention to features of the target help the target representation win in a competition for representation, and therefore become available to guide responses. Although spatial attention is slow, it can enhance the representations of some objects at the expense of others even when displays are brief, since representations persist after displays are presented (if they are not masked; see Experiment 2). This adds a small serial element to the process, as a set of parallel representations evolve over time from an initial state to a final state upon which a decision is based. However, the strategy is correctly regarded as parallel, since this evolution from in initial to final state is not repeated multiple times.

Search becomes less efficient as the discriminability of targets from distractors changes;

this effect is due to the complex nature of representations in the object recognition system, and is best discussed in the context of mechanisms, below. The considerations of Signal Detection Theory are important in determining the success of parallel search, but considerations about reduced information with stimulus crowding and distance from fixation, and the action of spatial attention in some cases, are needed to supplement existing theories.

4.2.1.1 Systems Interactions in Parallel Search

The converging structure of the ventral object recognition system (see section 3.1) allows the features of every object to have distinct representation in parallel across the field in early stages but not in later stages. Therefore each line's angle and length, each shade of color, and each texture or shading gradient likely is represented in parallel at lower levels of the visual system, where the RFs of neurons include only single stimuli. As the wave of neural activation travels up the ventral stream, each item cannot gain its own distinct representation when objects are tightly spaced. This is because the neurons at higher levels of the visual cortex have larger receptive fields, so that they respond to all objects within a given region (section 3.2.2. The size of RFs varies roughly with the complexity of the features to which a neuron responds preferentially; for instance, a neuron that responds to a vertical line better than a diagonal line is likely to have a small RF, where one that responds preferentially to an T junction than an L junction in any orientation is likely to have a much larger receptive field. The larger receptive fields of neurons responsive to subtle feature differences means that subtle differences are poorly represented when many objects are present in the display, or when objects are present in the periphery. Sections 3.2.2 and 3.2.4 review the evidence for the changes in RF sizes with representational complexity and distance from the fovea. Figure 4.2 depicts the this collapsing receptive field structure of the ventral stream.

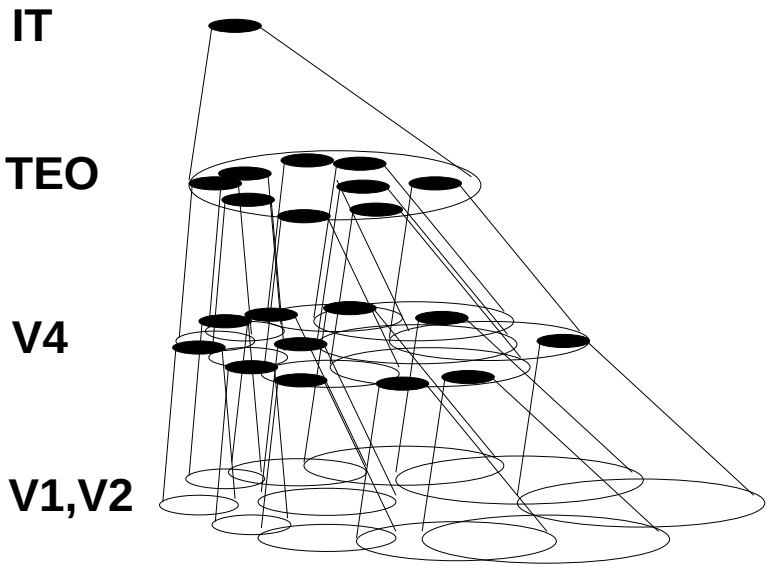

Figure 4.2: Collapsing receptive field structure of the ventral stream. A single IT neuron's inputs are depicted, to show how the RFs and representation of IT neurons are built from the RFs and representations of lower level neurons. The left side of the figure is toward the center of gaze. Filled circles represent neurons; the attached cones show which neurons they receive from. V1 and V2 have small RFs with less variance with eccentricity and are assumed to be a simple map of space. The IT neuron receives from more neurons with smaller RFs closer to the center of gaze, and so representations are stronger and more precise for objects in that region. Because the complete RF covers a good deal of the visual field, the IT neuron represents all objects within the RF at once in the absence of spatial attention. Featural attention enhances the activity of individual neurons receptive to particular features. Neurons selective for more precise features are higher in the ventral stream heirarchy, and so represent more items at once in multi-item displays. Therefore featural attention is only effective in inverse proportion to the number of items present, and the precision of the features of interest.

4.2.1.2 Effects of Attention on Parallel Representations

Featural attention acts on this structure of limited parallel representation. Attention acts to enhance the representations of attended features, likely through the simple mechanism of top-down excitatory biasing (Appendix C). If attention is directed to features that are represented clearly, the representations of items containing those features are likely to win the competition for representation and drive responses (section 3.3). This is the case in easy parallel searches. If, however, attention is directed to features that are not represented outside of the fovea, it is less likely that items containing these features will become dominant at higher levels of representation and so control behavioral responses.

The competition for representation is enhanced by the application of spatial attention. Biased competition within the ventral stream will result in neurons representing on area of space (one item or collection of items) becoming more active than others. This slight edge in competition is selectively sharpened by interactions with dorsal stream areas controlling spatial attention. One possible mechanism for this selective sharpening is the intelligent gating of working memory for space performed by the Basal Ganglia; this mechanism is discussed in section 3.3.2.[1]

The intelligently sharpened selection of a particular area in dorsal stream can beneficially bias processing in the ventral stream. Connections back to the ventral visual system further enhance the representation of that one area of space (see Shipp, 2004 and section 3.3). This mechanism ordinarily results in an eye movement to that area. Even when search is parallel and the display has disappeared before an eye movement is possible, this movement of spatial attention can help to produce clear representations of particular items.[2] Figure 4.3 depicts this

[1] Another type of possible mechanism is presented by Mozer (2002). This mechanism also sharpens representations, but does so using a relatively simple algorithm that produces one spatially contiguous area conforming roughly to the shape of the most active visual representations. This mechanism suffi ces to for the current explanation of parallel search and to explain attention in search with eye movements. However, the clear differences in subset search (section 2.3.2) seem to implicate a relatively sophisticated mechanisms of strategically controlled spatial attention. The hypothesized BG gating of WM representations of space offers such a mechanism.

[2] The enhancement of brief displays is possible even with slow spatial attention 1.1.5 since representations of individual features persist in an iconic memory of the image that may last as long as 350 ms (Bergen & Julesz, 1983).

process.

4.2.1.3 Detection of Targets Based on Representation in Parallel

If an item wins the competition for representation, it will become well represented at the top level of the visual system, on which responses are likely based in large part.[3] It is likely that more than one item can be simultaneously (relatively clearly) represented in area IT, at least for a short time (Rousselet et al., 2004). Motor systems receive input from all IT neurons, and basing a decision on the collective responses of many visual neurons is the type of decision making addressed by SDT. Therefore, the considerations of SDT (such as raising a response threshold when many items are present to contribute to false alarms signals) are important in this type of search. However, the fact that false alarms rise only modestly with set size, while missed targets rise dramatically, implies either that the threshold is raised more than the optimum amount, or more plausibly, that the addition of items to the display interferes with perceptual processing of the items, and that a target-present decision is often made only when a relatively clear representation of the target is present (at least for unpracticed observers, as in Experiment 2). This perceptual interference from extra items arises from the size of receptive fields discussed above.

The contribution of spatial attention discussed above also complicates the decision process, since it induces perceptual processes to more clearly represent one or a few items, at the expense of others. This tendency will result in some outputs from the perceptual system in which the target is clearly represented (when it was located correctly by spatial attention), and

[3] It is possible that responses are sometimes based on outputs from intermediate areas; this theory does not take a stand on that point. A review of the literature on anatomical connections between the ventral stream and motor areas would be necessary to assess the level of representation at which response occurs. Here I assume that responses are based mostly on neuronal representations in areas IT. However, if response can be based directly on representations in lower areas the overall theory does not change dramatically; a clear representation of response features is needed in either case. In the version discussed here, clear representation in lower areas must either cause a clear representation in area IT through direct competitive mechanisms, or else allow a shift of spatial attention, which biases the competition to allow clear representation of the area (and item) of interest in area IT. Neither of these steps are necessary if responses can be based directly on representations in lower retinotopic areas. However, in both cases, it is the clear representation in a lower area that allowed behavioral responses to eventually be based on that item.

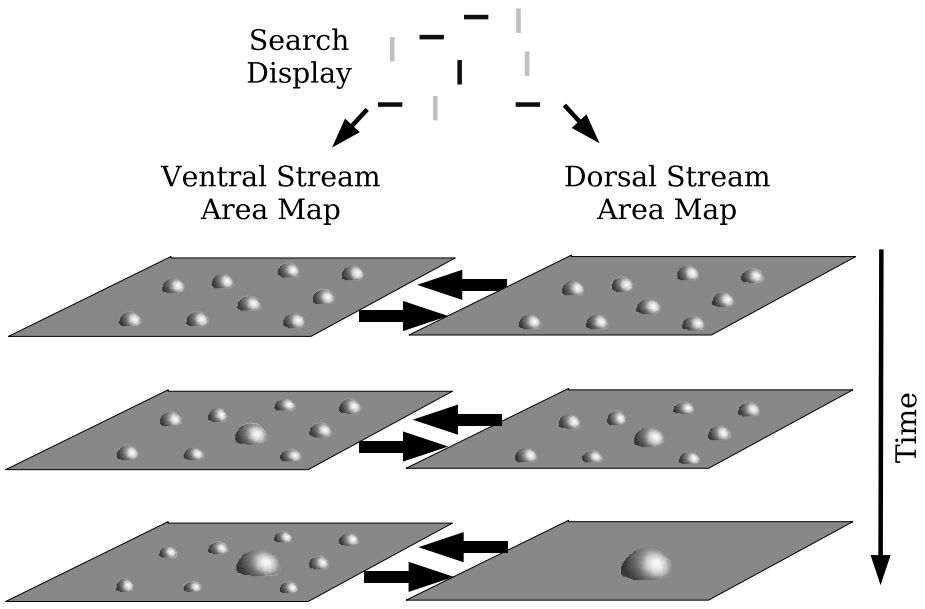

Figure 4.3: interactions between low level ventral stream and dorsal stream areas over time. Activations corresponding to particular areas in space are depicted as bumps on a smooth surface. The sequence of activation states depicts a parallel search, or a single eye fixation in a search with eye movements. The first frame shows the state of the ventral stream soon after stimulus presentation; activations exist for the features (and therefore locations) of each object presented in the display. In the second frame, attention to the target's features, combined with competition within the ventral stream, has enhanced the representation of the target item relative to nearby items. That enhancement is passed to the dorsal stream. In the third frame, gating mechanisms in the basal ganglia loop have enhanced the activation in the dominant location, and suppressed the activation elsewhere. This activity difference is passed back to the dorsal stream. The extra strength of the target representation at low levels now allows it to become clearly represented in higher level (IT) neuron that control responses. The strong activation at the target location in the dorsal stream allows an eye movement to or identification of the target's location.

other outputs that do not clearly indicate the target's presence or absence (when spatial attention settled on a different location, or did not settle at all). The fact that missed targets are more common than false alarms even in search with brief displays suggests that decisions are not based on an output with a well behaved distribution, but rather a bimodal distribution that elicits target present responses in the upper range, and target absent responses in the lower range.

The dynamics described above are modeled in the neural network model presented in Appendix E. The above description of parallel search also applies to each of the other modes of search, since parallel representations in the visual system are always present and cannot be entirely factored out of search. They are most important when attention is spread to an area, since more tightly focused attention will minimize the contribution of representations outside the attended area.

4.2.1.4 Possible Strategy Variants in Parallel Search

There are two possible variants of strategy in parallel search. One is the use of spatial attention. It seems likely that people can control the movement of spatial attention voluntarily, just as we can control eye movements; a strong attentional set for broad attention should reduce the tendency for attention to move to a likely target location. Asking observers to use one or the other strategy, combined with self reports after testing, could identify which strategy is more successful for particular classes of tasks.

A second possible strategy choice is exactly what is maintained as the target template. If the target template is strategically maintained in working memory as proposed in section 3.3.2, strategy differences in template maintenance should affect the success of search. No work I have located directly addresses this question; existing work is compatible with this hypothesis, but also has alternate explanations. These alternate explanations could be tested by giving observers varying instructions on what to look for. However, it is possible that observers refine a target template based on experience with a task; this would render instructions unable to effectively

manipulate templates.

4.2.2 Serial search by area with eye movements

This strategy is used for most searches; that is, those in which the target does not "pop out"; the display is not brief; eye movements are allowed; and the search is not specially arranged so that searching by groups is much more efficient. The eye movements in this search strategy impose a dominant serial aspect to processing, but within each serial eye fixation processing proceeds in parallel. This parallel processing is helpful in many situations, since some discriminations can be successfully made in parallel with broad attention. Even in situations where parallel processing is not helpful, some amount is inevitable, since spatial attention that is the result of a biased competition is not absolute. This hypothesized nature of attention is in contrast to that of many gating theories of attention (e.g. Cave, 1999; Bundesen, Habekost, & Kyllingsbaek, 2005).

In this strategy the eyes are moved until the target comes to attention. This is most likely to happen when the target is close to the center of gaze, and not closely surrounded by other items. A definite number of items are not identified (as hypothesized in the theories of Pashler 1987, Wolfe et al submitted, and the experiments presented in Chapter 1), but rather many items are processed superficially. This processing can allow immediate identification, or if the signal does not meet the threshold for a target-present response, can be used to guide the eyes to the target to verify its identity. Target recognition proceeds as described in the Parallel search strategy. The same parallel mechanisms described for spatial attention in parallel search provide information that can guide eye movements in serial search. The search can be guided if attention is spread, but guidance effects are small unless the target object has features that are highly discriminable and abundantly represented far from the center of gaze. This mode of search therefore spans the continuum from random eye movements to efficient parallel location of the target.

4.2.2.1 Systems Interactions in Search with Eye Movements

The main function of eye movements is to allow the sensitive portion of the visual field to process items, so as to overcome peripheral crowding effects and the general sparsity of representation away from the fovea. Search with eye movements seems to be the default mode of search. It is used whenever eye movements are possible, although it may not be the most efficient for some situations, as suggested by the results of Experiment 5 and those of Scialfa and Joffe (1998) in which practice allowed faster identification of conjunction targets without eye movements, without a loss of accuracy.

On every fixation, the mechanisms for identification and guidance of spatial attention work as described in the section on parallel search above. The shift of spatial attention described there functions to guide the next eye movement when fixation is not effortfully maintained (Liversedge & Findlay, 2000). The decrease in representation and increase in size of RFs with eccentricity (sections 3.2.4) accounts for the falloff of target location with distance from fixation and average item spacing (section 2.4.2.1). A greater number of items are effectively processed on each fixation at larger set sizes (table 2.1) because peripheral processing is reduced but not eliminated by crowding effects.

The size of area (or equivalently the number of items that are processed effectively) also varies with the difficulty of target/distractor discrimination, as hypothesized in the introduction and tested in the experiments of Chapter 1, and demonstrated in the data of Motter and Belky (1998b). This happens because the average RF size of neurons responsive to a particular target/distractor difference varies roughly proportionally to the perceptual difficulty of the discrimination, as discussed in section 3.2.2.

This model of attentional guidance in search has important differences from that of the Guided Search model. This model of guidance posits large effects of crowding and eccentricity, while the GS model does not account for crowding effects, and posits eccentricity effects only in the brief and obscure GS 3.0 (Wolfe & Gancarz, 1997). In addition, the proposed mechanisms

underlying attentional guidance are somewhat different. GS proposes that search is guided by the sum of feature maps for individual features (section 1.1.4). NTVS supposes instead that guidance happens at the level of conjunctive representations. The evidence for this is as follows.

First, there are few connections from lower ventral stream areas to dorsal stream location areas. Area V4 has few connections to the Frontal Eye Fields (FEF) that control attentional movement; most connections are from areas TEO and IT.[4] Therefore representations at the simple feature level will not be available to guide movements of spatial attention and the eyes, except by controlling representations in higher areas. Second, the effects of attention (at least for luminance and color) are found most prominently in area V4 (Motter, 1994b), despite the clear coding for these features in area V1 and V2. Third and most importantly, the multiplicative nature of crowding and eccentricity effects on conjunction search accuracy (Motter & Belky, 1998b) suggests that attentional biasing has its effects at the level of neurons with large spatial RFs.[5]

4.2.2.2 Possible Variations in Strategy for Search with Eye Movements

It is possible that the amount of guidance and the probability of locating the target on each fixation vary with the amount of fixation time. A neural network model (Herd & O'Reilly, In press, attached as Appendix 2) demonstrates that a speed/accuracy tradeoff results from the basic considerations of locating a conjunctive target through biased competition. A number of studies have linked fixation times with accuracy of saccades (McSorley & Findlay, 2003). One study (Coeffe & O'Regan, 1985) used the manipulation of asking observers to delay saccades, and showed substantially better accuracy at localizing a marked letter within a string. However,

[4] While these areas have large RFs, they can still effectively code locations of a coherent representation, through use of a coarse coding mechanism: in essence, many large RFs can identify the location of a represented item by the overlap in their RFs.

[5] The eye movement studies of Motter and Belky (1998b) reviewed in section 2.4.2.1 showed a beneficial effect of stimulus crowding for feature searches. The ordinary crowding effects due to multiple objects within the relevant RFs could be opposed in this case by the contribution of local difference mechanisms in producing popout at the neural level (see Appendix A). The second is that more neurons are responsive to feature differences than conjunctions of features, producing more reliable representations for localizing and identifying items.

none of these studies used a challenging search task; all allowed target localization with a single saccade on most trials. One exception is the study of Hooge and Erkelens (1999), but they did not calculate the relation of fixation time to target location. Instead they showed that eye movements were more likely to land on a distractor sharing a target feature after longer fixations.

This hypothesis suggests that the length of saccade fixation may be useful as a variable under strategic control: long fixations may be useful to locate targets in conditions that allow relatively efficient guidance, and fast fixations may be more useful under conditions that do not allow effective guidance. There is some evidence of these strategies in data from eye tracking (Motter & Belky, 1998b). However, the paradoxical finding that adding more distractors to a display can improve the accuracy of first saccades (probably by lengthening the initial fixation) suggests that there are more effective strategies to be discovered and employed for some searches. Although these findings are suggestive, and consistent with the speed/accuracy trade-off demonstrated in the model in Appendix D, it remains to be tested whether or not voluntarily slowing of eye movements can lead to better overall search performance for challenging search tasks.

It is likely that the use of eye movements is sometimes combined with spatial attention, so that only a limited central area around the current fixation is attended. In this way the area searched with each fixation can vary strategically. Spatial attention is probably used in this way for very difficult searches, those in which target identification is challenging even when the target is near the center of gaze. This hypothesis is consistent with the findings of Lavie and Cox (1997), who reported that a more difficult search produced a smaller effect of distractors outside the display area on a decision task performed concurrently with the search. The use of spatial attention to aid focal processing should dramatically reduce the amount of guidance in search, by biasing high-level representations toward items within the zone of attention and away from those outside.[6]

[6] The requirement for spread attention to allow guidance is distinct from the common assumption that the parallel

It is also possible that the amount of guidance used is under strategic control.[7] This is very likely to be true when spatial attention is used as described above, since it should reduce the effect of guiding processes outside of the attended region. The amount of guidance may also vary with the effortful maintenance of the target template. It is not necessary to maintain this template to identify the target, once some small amount of practice has enabled the motor system to learn responses based directly on outputs of higher visual areas.

Like the GS model, NTVS holds that template maintenance for guidance is optional, and that the template may change in intensity with effort, or that only one feature of a conjunctive target may be maintained, so that search proceeds is guided to items sharing that target feature. This guidance to one feature explains findings in which search times vary roughly with the number of target colored items; these findings are reviewed in section 2.3.1. Several authors have found results suggesting individual differences in template maintenance. Egeth et al. (1984) found that individual observers seemed to prefer to search a separate subset, so that the reaction

mechanisms that guide search are "preattentive" (Wolfe, 1994). It is consistent with several other arguments that the claim of processes truly independent of attention is unnecessary and implausible (Treisman, 1993; Di Lollo, Kawahara, Zuzic, & Visser, 2001). In this model, attention is generally not restricted to a point during object identification, and so allows peripheral processes to guide search, and focal ones to detect targets simultaneously.

[7] It is possible that, for some tasks, observers simply do not expend the effort to maintain a top-down biasing representation. This could happen if the use of a top-down template is relatively ineffective. The template is more important for items far from fixation, to overcome the tendency for high level neurons to represent items near the center of gaze. If this biasing is ineffective since colors are relatively poorly represented outside of the retina, observers may simply abandon the effortful guided component of search. This is one possible explanation of the common finding of 2:1 slopes for conjunction searches for color and form with relatively unsaturated colors (Wolfe et al., 1989). If search proceeds by an area large enough to accommodate more than about 1/5th of the display, (as is likely given estimates for number of conjunction stimuli processed in parallel (Motter & Belky, 1998a; Wolfe et al., submitted)), or if search proceeds systematically through the display, the limited memory of visual search processes (section 1.1.3.5) would be enough to allow a truly self-terminating search.

The appearance of a 2:1 ratio of search can be taken to mean that the observer has correctly calculated the number of fixations necessary to tell with relative certainty whether the target is present, and is responding "target absent" after this number (and finding the target on average with half that number). If the ratio increases at larger set sizes, as it did in Experiments 1 and 5, observers are fixating more times before signaling target absent than is necessary to determine this on average. This could be due to unreliable target detection at peripheral locations; the search is guided, and so the target is found on average with fewer than half the number of fixations necessary to view the whole display with certainty. The Guided Search model does not have a satisfying explanation of the appearance of steadily decreasing positive search slopes at high set sizes for T and L stimuli. This shape of curve is suggested by the limited number of set sizes of TvL stimuli used in Wolfe et al. (submitted) and Experiments 1 and 5. However, the data is not reliable enough to conclude this with certainty. However, the data of Motter & Belky (1998b) uses a very large number of trials from a single monkey subject in the TvL condition, and shows a clear signature of decreasing slope with set size (figure 2.8 shows number of fixations, which are very nearly proportional to the reaction time data shown in their figure 2A). The difference between T and L junctions are clearly guiding search (or equivalently, allowing detection outside the region of high conspicuity. The above analysis suggests an experimental manipulation of asking some subjects to keep a target template in mind, and testing for different detection rates.

times of some observers varied closely with the number of items sharing the target's shape, while others seemed to be searching through the set of items that shared the target color. King (2003) also showed similar significant results in opposite directions for individual observers in a conjunction search with varying number of target colored and target shaped items.

This individual variance in guidance is also suggested by eye tracking data in a conjunction search reported by Findlay (1997). They did not report statistics on these differences, but a rough post-hoc analysis suggests that there are likely to be real differences in strategy. Missed saccades fell near the objects of the target color vs the target shape on 64 vs 31, 53 vs 38, 34 vs 54, and 34 vs 46 of the trials for each observer respectively. The first two distributions are unlikely to result from chance; the odds of totals more than this different from 50-50 splits are approximately .001 for the first, and .14 for the second. The third and fourth are in the opposite direction; the odds are about .04 and .2 against distributions as far from even, respectively (odds calculated by Monte Carlo simulation). While these values are not completely convincing from post-hoc tests, they are suggestive that individual strategy influenced guidance in this task.

These results collectively suggest that observers do not maintain a representation of the complete target template, or at least do not maintain all features to the same extent. The choice of target template is another strategic choice that will produce different results from different observers and possibly different strategy instructions. Since some features will provide more effective guidance than others, the type of featural attention used is another possible way in which strategy can improve search efficiency.

4.2.3 Serial search with Covert Attention

This strategy will usually only be employed when eye movements cannot be made, since voluntary attentional movements take about as long as eye movements (section 1.1.5) and provide much less benefit. This strategy is similar to search with eye movements. At the algorithmic level, the only difference is that the movement of attention provides a much weaker

aid to discriminating targets from distractors. Therefore many searches without eye movements are likely better performed in parallel, without multiple movements of spatial attention. Covert attention provides a strong serial component when it is used; however, this component is less strong than the serial component of search with eye movements, since parallel processing in unattended regions can proceed without disruption from a change in retinotopic coordinates of each item (although it will be disrupted somewhat by the attentional shift). Processing on each attentional fixation also proceeds in parallel, for the same reasons as in serial search with eye movements: attention is an incomplete bias, and so cannot entirely stop the processing of unattended items.

4.2.3.1 Systems Interactions for Covert Search by Area

In this strategy attention is used in a specific sequence: first spatial biasing selects an area (possibly guided by an initial broad fixation, see below), then feature biasing guides further attentional movement within that area. As in search with eye movements, the same parallel mechanisms described in the section on search with eye movements are at work; both recognition and guidance operate across the field of view. However, guidance and identification are less effective outside the attended area, because spatial attention reduces processing outside the attended region. The optimum size of the attentional window varies with the discriminability between target and distractors. As discriminations become easier and allow a broader spread of attention, this strategy becomes the same as a fully parallel search.

In this strategy, random fixations of spatial attention by area are produced by the prefrontal and Basal Ganglia mechanism discussed in section 3.3.2. The spatial focus of attention acts to reduce effective RF size, as described in section 3.2.3, and improves target detection. However, the increase in accuracy is not as large as that gained with eye movements, since spatial attention does not completely suppress representations of items outside the focus of attention, and the relative sparseness of representation away from the center of gaze cannot be

counteracted by attention.

The advantage of serial fixations of covert attention is demonstrated in the neural network model presented in Appendix E. In this model, a broad attentional fixation works faster than attention to a limited area, except at the largest set size, when crowding is most severe. In that model, the advantage for random attentional fixations at large set sizes is an emergent feature. It arises from the combination of a reduced version of the collapsing RF structure of the ventral visual stream, and from an attentional excitatory bias applied to middle levels of the visual system.

4.2.3.2 Possible Strategy Variations in Covert Search by Area

.

Attentional fixations can be guided by an initial use of broad attention to allow the full effect of parallel processing, as described in that section above. This initial broad attentional fixation may be used or not as a variation in strategy; since it will require a little time, search for difficult targets may be more efficient when attention is focused at random, with no broad fixations for guidance. This variable should be under strategic control, and experiments could easily be run comparing these two strategies. The relative uselessness of covert search by area limits the practical applications of such work, but positive findings would support the hypothesized mechanisms of search.

The optimal size of attended region may vary with difficulty of discrimination; however, it is also possible that the only two sizes are used: broad attention, and attention to a point.[8] If covert spatial attention comes in only two sizes, the relative rather than absolute nature of spatial attention as a top-down bias on processing should still yield a gradient of attention. This gradient could be changed by devoting more attentional resources to focusing attention; in this case, the optimal level of focusing may also vary by task. Featural attention may be applied

[8] This assumption can still explain the continuum of search slopes. In SDT terms, evidence for the target will accumulate relatively more slowly when targets are more similar to distractors, so that more difficult targets are found more slowly even if the total number of attentional fixations is the same across stimulus set difficulties

simultaneously with area attention to improve recognition of targets within the attended area, as described in the section on parallel search, above. Eye movements may be preferred over covert attention because it is faster to redirect the eyes to potential target locations to speed verification of target presence or absence, as suggested by Experiment 5.

4.2.4 Search by Grouping

Search by grouping seems to be applied only in rare situations, as reviewed in section 2.3.2 and section 2.3.3. It seems to be employed only under specific task conditions in which a standard search with eye movements will take much longer. Experimental situations are discussed in section 2.3.2. Examples are searches in which all distractors share a common color, or when the target pops out of an attended group, as in Experiment 6. In this strategy, feature biasing is applied, and spatial biasing is applied to the result of that calculation. Once this subset of items is selected, spatial attention greatly reduces inteference from the nongrouped items. The presence of a target within a group can then be detected if the target pops out from the rest of the group. If that particular grouping does not contain the target, it can be rejected as a whole, and a new group can be selected.

Search by grouping can thus proceed as a parallel search over a single subgroup, or a serial search over multiple subgroups. However, there are multiple steps in forming each group, and these might be considered a serial process at a smaller scale. Search by grouping is thus a parallel-serial strategy, in which several steps of processing allow a single parallel search. In displays where more than one grouping is needed, this process extends to a serial-parallel-serial progression.

4.2.4.1 Systems Interactions in Search by Groups

Search by groups arises from the same object represenations and attentional mechanisms as do the other strategies of search, but they are applied somewhat differently. Search by group-

ing utilizes both spatial and featural attention, in a specific sequence. First featural attention[9] is used to enhance processing of a set of items sharing a feature that is adequately well represented into the periphery. This produces a set of activations similar to the feature map of figure 4.3, but with the activation at locations of all objects in the attended subset slightly higher than at locations of items in the unattended subset.

This activation map is passed to the dorsal stream and related areas. The relatively intelligent gating mechanisms described in section 3.3.2 then enhance all[10] locations, and suppress activation at all other locations. This process is again similar to that depicted for parallel search in figure 4.3, but the gating mechanisms emphasize all representations above a certain threshold, and suppress the rest. This gated dorsal stream activation feeds back to ventral stream, and produces much larger activations for attended than unattended items. Representations in higher areas are thus controlled by the features of attended items. These response can be used to differentiate target from distractor items, without any narrowing of attention to one target of interest. For instance, in the paradigms of Friedman-Hill and Wolfe (1995) and Experiment 6, responses could be based on outputs of high level neurons responsive to orientation textures across a large field. These neurons will respond differently when only one orientation is present in the active visual representations (the attended group) than when two orientations are present on target present trials.

[9] Object based attention could also be used in some circumstances. The two effects are similar; object based attention could be considered featural attention for high level features, such as overall shape, or even for the feature of "face-ness." The two mechanisms are enacted identically within this framework. Object based attention would be used in very specific situations, such as the paradigm of Donnelly et al. (1991) in which distractors could form coherent shapes. However, even in this paradigm it is unclear whether object-based attention was used, or whether items were simply perceived as outlines of objects through bottom up processing

[10] The data of Experiment 5 and Experiment 2 of Friedman-Hill and Wolfe (1995) indicate that subset search either becomes serial or takes substantially longer to identify target absent displays past a certain display size. This could be due to two aspects of feature representations. These are the decreases in resolution and strength of spatial representations in the periphery, due to few total neurons representing items at higher eccentricity, the cortical magnification effect (section 3.2.4). These decreases could each contribute to a difficulty in including attended items and exclude unattended items farther from the center of gaze, and require more iterations of the process, either by moving spatial attention to a completely different subset, or by "gating in" peripheral items that were missed the first time, or "gating out" unattended item locations that were erroneously included.

4.2.4.2 Possible Strategy Variations in Search by Groups

The main strategy variable in search by groups is whether the strategy is employed at all. Because this strategy is complex, it seems to be elicited only in situations where the need for it is extreme. One question for future research is whether this strategy could be more efficient in some searches that do succumb to the ordinary search by eye movement strategy. While a comparison with existing conjunction search results from Wolfe et al. (1989) suggests that the guided search strategy employed there is more efficient, it seems that the search by group strategy is more efficient for exclusion paradigms, such as the task of Carrasco et al. (1998) in which the target is always the only object without a blue segment. Therefore searches by exclusion could be performed more efficiently using a grouping by feature strategy, even when that strategy is not dictated by the task and so is not discovered naturally.

4.3 Relation to Other Theories

The relationship between NTVS and other theories of visual search has been explicated at length. However, its relation to more general theories of visual perception and visual attention has not been addressed. NTVS was developed based on neurophysiological evidence, with the goal of accounting for visual search findings. This separation from other evidence is both a weakness and a strength. A theory developed without the advantage of all available converging evidence is more likely to be incorrect. However, if theories developed independently to account for different domains of evidence converge on a similar account of vision, it is evidence for the validity of that account.

This theory has indeed converged rather closely with Bundesen's influential Theory of Visual Attention (TVA) (Bundesen, 1990). This theory is presented in terms of a formal model. The more recent computational version (Bundesen, 1998) includes considerations for spatial attention, featural attention, a limited parallel identification process, and considerations for inaccuracy with stimulus crowding. However, this theory is generally not couched in terms of

anatomy. For instance, the theory does not consider the effect of eye movements.

A recent effort to extend TVA to a Neural Theory of Visual Attention (NTVA) (Bundesen et al., 2005) focuses on explaining neurophysiological results with the theory, rather than proposing a mechanistic framework giving rise to the algorithmic behavior described by TVA. Therefore it does not include the idea of search success hinging on RF sizes of neurons sensitive to particular target/distractor differences that is central to NTVS. In addition, in NTVA the proposed mapping to neural mechanisms centers on a gating mechanism for neural transmission; no biological mechanism is proposed for this gating, and no candidate is obvious. In sum, NTVS is substantially different from the TVA theories in that it is based on neurophysiological findings, and is therefore more mechanistically detailed and plausible. However, the fact that the algorithmic level descriptions of the two theories similar between the two theories is a reassuring convergence.

The Selective Attention for Identification Model (SAIM) of Humphreys (2003) is another formal general theory of visual perception. It is instantiated as a neural network model. It primarily addresses issues of visual attention in relation to neglect and related deficits. This explanation is similar to that of the neural network model presented by Mozer (2002). The theory presented here is consistent with this explanation, although it does not predict those results. The spatial component of attention envisioned in NTVS operates similarly to that of the models above; the BG gating mechanisms proposed to drive attention in search tasks are not needed in explanations of automatic attentional effects such as cuing and neglect. The explanation of parallel search within NTVS is consistent with the extension of the SAIM model to visual search (Heinke, Humphreys, & diVirgilio, 2002), although the SAIM model describes only the parallel search portion of NTVS. The extension of SAIM to include grouping effects (Heinke, Sun, & W, 2005) describes an area that is left out of NTVS, which accounts only for grouping by top-down control 4.2.4, and does not specify the role of automatic grouping within visual search, since current experimental evidence does not discriminate between grouping and

local-differences effects proposed by the FIT and GS models.

A particularly relevant theory is the Reverse Hierarchy Theory (RHT) of Hochstein and Ahissar (2002; Ahissar & Hochstein, 2004). These authors review a great deal of evidence that indicates that visual perception seems to occur first for the information at the top level of the visual system, and only with attention and time is information represented by lower areas available to guide responses. This theory seems initially to be at odds with NTVS. RHT specifically states that popout search occurs for differences that are represented in output areas of the visual system, while NTVS is based on findings that differences that produce popout effects are represented in low level neurons with small RFs.

These two approaches are not in opposition. Ample and clear representation at low levels leads to representations at higher levels, as well. IT neurons will respond substantially differently when a single red item is added to a field of blue items, because the red item finds ample representation at low levels, and this strong representation drives some usable representation at high level.[11] In NTVS, as in RHT, item identification occurs only when items are represented in IT cortex.[12] NTVS expands on the idea offered in RHT, by laying out the mechanisms by which time and attention make possible the perception of more subtle differences represented in lower visual areas. Attention selects for representations of the item or items of specific interest, and allows these representations to control those in area IT, enabling perception of these subtle differences and responses based on them.

NTVS is consistent with the updated version of Broadbent's late-selection theory presented by Lachter, Forster, and Ruthrugg (2004). These authors argue that no identification (measured by priming) happens for truly unattended stimuli. This is similar to the load theory of attention of Lavie and colleagues stating that processing of unattended items is inversely pro-

[11] The mechanisms for popout proposed in Appendix A also help account for automatic representation of popout stimuli in higher areas. The proposed mechanism involves a shift of spatial attention triggered by greater total activity in neurons representing the area including an unlike item. This greater activity may occur during a brief period over which multiple objects are represented simultaneously, before local competition resolves in favor of some and against others (Rousselet et al., 2004)

[12] In NTVS, responses can also be based on attentional shifts, so that representation in IT is not a strict requirement for behavioral responses in visual search tasks

portional to the demands put on attention (Lavie, 1995; Rees & Lavie, 2001). NTVS proposes effect of spatial attention on representation in IT cortex that account for the failure to identify unattended items when attention is fully devoted to task-relevant items. NTVS adds the prediction that unattended items should allow priming up to the level of semantics encoded by neurons below IT, that is, some conjunctions of color and shape.

It is worth clarifying the relationship of NTVS to FIT. The central proposal of FIT is that attention is necessary to bind together features into objects. In NTVS, attention is needed only to filter out features from spatially adjacent objects in the visual periphery. In this scheme, binding happens automatically; but it happens for features of all objects with the RFs of high level neurons, so that features are incorrectly bound, producing illusory conjunctions and an inability to identify items without spatial attention or eye movements. Therefore, in NTVS attention is needed not to bind object features, but to prevent extra features from nearby objects from being inappropriately bound together

NTVS is also situated within a broader theoretical framework. The central mechanisms of NTVS are based on the biased competition theory of attention (Desimone & Duncan, 1995; Duncan, 1996). The theory proposed here extends the account by focusing on the structure of the representations among which biased competition occurs, and the varied strategies that arise from different combinations of strategically applied attentional biases. NTVS also proposes specific mechanisms for the action of attentional biases; see section 3.3 and Appendix C. The proposed account is a simple one; extra excitatory input is directed to sensory neurons that represent attended information; this input is supplied by maintained activity in neural populations representing the maintained set in the abstract. This Top-down Excitatory Biasing (TEB) mechanism is embodied in some simple neural network models of task attention (Cohen, Dunbar, & McClelland, 1990; Herd et al., IP), and more sophisticated neural network models of attentional selection and object recognition (O'Reilly & Munakata, 2000).

The TEB account of attentional biasing is supported by a recent review of attentional

effects at the neural level (Reynolds & Chelazzi, 2004). The review concludes that attention results in increased neural gain, and go on to cite two single-cell studies indicating that increased input produces an increase in the effective gain of neural responses (Chance, Abbott, & Reyes, 2002; Fellous, Rudolph, Destexhe, & Sejnowski, 2003). More simulation and experiments are necessary to test whether the TEB framework is adequate to explain the neural mechanisms of attention. However, the biased competition and TEB framework is promising in that it offers an account of attention that unifies sensory attention and attentional task control (Herd et al., IP).

4.3.1 Limitations of the Theory and Areas for Future Research

There are some interesting questions related to the topics discussed here for which I have not been able to provided even provisional answers. Foremost, the theory identifies likely strategies for visual search, but has not identified the specific situations in which each strategy is used. Identifying the particular strategies that people tend to employ for specific search tasks would be helpful for human-computer interactions and other human factors design work. A related issue is which strategies are most useful for particular types of searches. It should not be assumed that observers naturally adopt the most useful strategy, or even that they discover it with practice. Instead it would be useful to ask observers to employ a specific strategy, and measure the effectiveness of each one for each task. This information would be useful in training people for specific types of searches.

This type of study would also reveal more of the capabilities of the visual system. For instance, the fact that monkeys seldom locate a color-orientation conjunction target further away from the center of gaze than twice the average stimulus spacing does not mean that people cannot do so when they are determined to. In light of the fact that guidance is more effective with longer fixations, it may be that the strategic use of guidance is more effective in some situations than the instinctive use of rapid eye movements.

Another point of great interest to the field, for theoretical if not for practical concerns, is

the role of automatic grouping in search. The available evidence is not adequate to distinguish between guidance accounts and grouping accounts of standard conjunction searches. Clever experimental manipulations might be able to do so.

Finally, the account of detection as a result of particular receptive field qualities is only a general hypothesis at this stage. More information about neural properties is necessary to flesh out this story. In addition, this account likely leaves out important system-level effects. Attempts to simulate this account as a neural network would help to test its sufficiency.

4.4 Conclusions

The conclusions are contained in the sections on theory above, and throughout the text. At the most specific level, the conclusion is that considering the relation of receptive field size to the specificity of neural responses provides insight on the particular successes and failures of human visual search. More broadly, the current theory offers insights about strategy differences between tasks that makes sense of a complex and seemingly contradictory literature. At the broadest level, the conclusion is that integrative research can provide a more complete picture of an area of research than is gained through work that provide in-depth analysis of a limited question. The more complete picture provided here may offer guidance to future empirical and theoretical work, and to efforts to construct artificial systems for theoretical or practical purposes.

4.5 References

Ahissar, M., & Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. Trends in Cognitive Sciences, 8(10), 457–464.

Alexander, G. E., DeLong, M. R., & Strick, P. L. (1986). Parallel organization of functionally segregated circuits linking basal ganglia and cortex. Annual Review of Neuroscience, 9, 357–381.

Allman, J., Miezin, F., & McGuinness, E. (1985). Direction- and velocity- specific responses from beyond the classical receptive field in the middle temporal area (MT). Perception, 14, 105–126.

Alverez, G. A., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. Psychological Science, 15(2), 106–111.

Anderson, J., Lampl, I., Gillespie, D. C., & Ferster, D. (2001). Membrane potential and conductance changes underlying length tuning of cells in cat primary visual cortex. Journal of Neuroscience, 21, 2104–2112.

Arnott, S. R., Pratt, J., Shore, D. I., & Alain, C. (2001). Attentional set modulates visual areas: an event-related potential study of attentional capture. Cognitive Brain Research, 12 (3), 383–395.

Bacon, W., & Egeth, H. (1997). Goal-directed guidance of attention: Evidence from conjunctive visual search. Journal of Experimental Psychology: Human Perception and Performance, 23(4), 948–961.

Banich, M. T., Milham, M. P., Atchley, R., Cohen, N. J., Webb, A., Wszalek, T., Kramer, A. F., Liang, Z. P., Barad, V., Gullett, D., Shah, C., & Brown, C. (2000). Prefrontal regions play a predominant role in imposing an attentional 'set': evidence from fMRI. Cognitive Brain Research, 10, 1–9.

Bender, D., & Youakim, M. (2001). Effect of attentive fixation in macaque thalamus and cortex. Journal of Neurophysiology, 85, 219–234.

Bergen, J., & Julesz, B. (1983). Rapid discrimination of visual patterns. IEEE Transactions on Systems, Man, and Cybernetics, 13, 857–863.

Bichot, N. P., & Schall, J. D. (1999). Effects of similarity and history on neural mechanisms of visual selection. Nature Neuroscience, 2, 549–554.

Born, R. T. (2000). Center-surround interactions in the middle temporal visual area of the owl monkey. Journal of Neurophysiology, 84 (5), 2658–2669.

Boutsen, Lucand Marendaz, C. (2001). Detection of shape orientation depends on salient axes of symmetry and elongation: Evidence from visual search. Perception & Psychophysics, 63 (3), 404–422.

Brefczynski, J. A., & DeYoe, E. A. (1999). A physiological correlate of the 'spotlight' of visual attention. Nature Neuroscience, 2(4), 370–374.

Bringuier, V., Chavaner, F., Glaeser, L., & Fregnac, Y. (1999). Horizontal propagation of visual activity in the synaptic integration field of area 17 neurons. Science, 283, 695–699.

Brown, J. M. F. V., & Gilchrist, I. D. (2000). Saccade target selection in visual search: The effect of information from the previous fixation. Vision Research, 41(1), 87–95.

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

Bundesen, C. (1998). A computational theory of visual attention. Philos Trans R Soc Lond B Biol Sci, 353(1373), 1271–1281.

Bundesen, C., Habekost, T., & Kyllingsbaek, S. (2005). A neural theory of visual attention: Bridging cognition and neurophysiology. Psychological Review, 112(2), 291–328.

Buzas, P., Eysel, U. T., Adorjan, P., & Kisvarday, Z. F. (2001). Axonal topography of cortical basket cells in relation to orientation, direction, and ocular dominance maps. Journal of Comparative Neurology, 437, 259–285.

Carrasco, M., Evert, D., Chang, I., & Katz, S. M. (1995). The eccentricity effect: Target eccentricity affects performance on conjunction searches. Perception & Psychophysics, 57(8), 1241–1261.

Carrasco, M., & Frieder, K. (1997). Cortical magnification neutralizes the eccentricity effect in visual search. Vision Research, 38(1), 63–82.

Carrasco, M., Ponte, D., Rechea, C., & Sampedro, M. J. (1998). "transient structures": The effects of practice and distractor grouping on within-dimension conjunction searches. Perception & Psychophysics, 60(7), 1243–125.

Cave, K. R. (1999). The featuregate model of visual selection. Psychological Research — psychologische forschung, 62(2-3), 182–194.

Chance, F. S., Abbott, L. F., & Reyes, A. D. (2002). Gain modulation from background synaptic input. Neuron, 35(4), 773–728.

Cheal, M., & Lyon, D. (1992). Attention in visual search: Multiple search classes. Perception & Psychophysics, 52 (2), 113–138.

Chelazzi, L. (1999). Serial attention mechanisms in visual search: A critical look at the evidence. Psychological Research, 62, 195–219.

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

Chelazzi, L., Miller, E., Duncan, J., & Desimone, R. (2001). Responses of neurons in macaque area V4 during memory-guided visual search. Cerebral Cortex, 11 (8), 761–772.

Coeffe, C., & O'Regan, J. K. (1985). Reducing the influence of non-target stimuli on saccade accuracy: predictability and latency effects. Vision Research, 27(2), 227–240.

Cohen, A. (1993). Asymmetries in visual search for conjunctive targets. Journal of Experimental Psychology- Human Perception & Performance, 19(4), 775–797.

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

Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes: A parallel distributed processing model of the Stroop effect. Psychological Review, 97(3), 332–361.

Colby, C. L., Duhamel, J. R., & Goldberg, M. E. (1996). Visual, presaccadic, and cognitive activation of single neurons in monkey lateral intraparietal area. Journal of Neurophysiology, 76, 2841.

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus driven attention in the brain. Nature Reviews Neuroscience, 3 (3), 201–215.

Cowan, N. (2001). The magical number 4 in short-term memory; a reconsideration of mental storage capacity. Behavioral and Brain Sciences, 24, 87–185.

Crick, F. H. C., & Asanuma, C. (1986). Certain aspects of the anatomy and physiology of the cerebral cortex. In J. L. McClelland, & D. E. Rumelhart (Eds.), Parallel distributed processing: Explorations in the microstructure of cognition, Vol. 2 (Chap. 20, pp. 333–371). Cambridge, MA: MIT Press.

Dalva, M. B., Weliky, M., & Katz, L. C. (1997). Relationships between local synaptic connections and orientation domains in primary visual cortex. Neuron, 19 (4), 871–880.

Das, A., & Gilbert, C. D. (1999). Topography of contextual modulations mediated by short-range interactions in primary visual cortex. Nature, 399, 655–661.

Davis, G., & Holmes, A. (2005). The capacity of visual short-term memory is not a fixed number of objects. Memory & Cognition, 33(2), 185–195.

Deco, G., & Zihl, J. (2001). Top-down selective visual attention: A neurodynamical approach. Visual Cognition, 8(1), 119–140.

DeLiang, W., Kristjansson, A., & Nakayama, K. (2005). Efficient visual search without top-down or bottom-up guidance. Perception & Psychophysics, 67(2), 239–253.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. Annual Review of Neuroscience, 18, 193.

Di Lollo, V., Kawahara, J., Zuzic, S. M., & Visser, T. A. W. (2001). The preattentive emperor has no clothes: A dynamic redressing. Journal of Experimental Psychology- General, 130(3), 479–492.

Di Russo, F., Martinez, A., Sereno, M. I., Pitzalis, S., & Hillyard, S. A. (2002). Cortical sources of the early components of the visual evoked potential. Human Brain Mapping, 15 (2), 95–111.

Dias, E. C., & Segraves, M. A. (1999). Muscimol-induced inactivation of monkey frontal eye field: Effects on visually and memory-guided saccades. Journal of Neurophysiology, 81 (5), 2191–2214.

DiCarlo, J. J., & Maunsell, J. H. R. (2003). Anterior inferotemporal neurons of monkeys engaged in object recognition can be highly sensitive to object retinal position. Journal of Neurophysiology, 89(6), 3264–3278.

Dobbins, A., Zucker, S. W., & Cynader, M. S. (1987). Endstopped neurons in the visual cortex as a substrate for calculating curvature. Nature, 329, 438–441.

Donk, M., & Meinecke, C. (2002). Detection performance in pop-out tasks: Nonmonotonic changes with display size and eccentricity. Perception 31(5), 31(5), 591–602.

Donnelly, N., Humphreys, G., & Riddoch, M. (1991). Parallel computation of primitive shape descriptions. Journal of Experimental Psychology: Human Perception & Performance, 17, 561-57, 17, 561–570.

Duncan, J. (1996). Cooperating brain systems in selective perception and action. In T. Inui, & J. L. McClelland (Eds.), Attention and performance (pp. 85–105). Cambridge, MA: MIT Press.

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

Duncan, J., Humphreys, G., & Ward, R. (1997). Competitive brain activity in visual attention. Current Opinion in Neurobiology, 7 (2), 255–261.

Egeth, H. E., Virzi, R. A., & Garbart, H. (1984). Searching for conjunctively defined targets. Journal of Experimental Psychology: Human Perception and Performance, 10, 32–39.

Eifuku, S., & Wurtz, R. H. (1998). Response to motion in extrastriate area MST1: Center-surround interactions. Journal of Neurophysiology, 80 (1), 282–296.

Enns, J. T., & Rensink, R. A. (1990). Influence of scene-based properties on visual search. Science, 247(4943), 721–723.

Eriksen, C. W., & St James, J. D. (1986). Visual attention within and around the field of focal attention: a zoom lens model. Perception & Psychophysics, 40(4), 225–240.

Fellous, J. M., Rudolph, M., Destexhe, A., & Sejnowski, T. J. (2003). Synaptic background noise controls the input/output characteristics of single cells in an in vitro model of in vivo activity. Neuroscience, 122(3), 811–829.

Findlay, J. M. (1997). Saccade target selection in visual search. Vision Research, 37, 617–631.

Found, Andrewand Mueller, H. J. (1997). Local and global orientation in visual search. Perception & Psychophysic, 59 (6), 941–963.

Foxe, J. J., & Simpson, G. V. (2002). Flow of activation from V1 to frontal cortex in humans - a framework for defining ”early” visual processing. Experimental Brain Research, 142 (1), 139–150.

Frank, M. J. (Unpublished Dissertation). Dynamic dopamine modulation of striato-cortical circuits in cognition: Converging neuropsychological, psychopharmacological and computational studies.

Frank, M. J., Loughry, B., & O’Reilly, R. C. (2001). Interactions between the frontal cortex and basal ganglia in working memory: A computational model. Cognitive, Affective, and Behavioral Neuroscience, 1, 137–160.

Friedman-Hill, S., & Wolfe, J. (1995). Second-order parallel processing: Visual search for the odd item in a subset. Journal of Experimental Psychology: Human Perception and Performance, 23(3), 531–551.

Gandhi, S. P., Heeger, D. J., & Boynton, G. M. (1999). Spatial attention affects brain activity in human primary visual cortex. Proceedings of the National Academy of Sciences of the USA, 96(6), 3314–3319.

Gatass, R., Sousa, A. P., & Gross, C. G. (1988). Visuotopic organization and extent of V3 and V4 of the macaque. Journal of Neuroscience, 8, 1831–1845.

Geisler, W. S., & Chou, K. L. (1995). Seperation of low-level and high-level factors in complex tasks: visual searc. Psychological Review, 102, 256–378.

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

Grossberg, S., & Williamson, J. R. (2001). A neural model of how visual cortex develops a laminar architecture capable of adult perceptual grouping. Cerebral Cortex, 11, 37–58.

Haenny, P. E., & Schiller, P. H. (1988). State dependent activity in monkey visual cortex. i. single cell activity in V1 and V4 on visual tasks. Experimental Brain Research, 69 (2), 225–244.

Harvey, L. O. (2004). Parameter estimation of signal detection models: Rscore plus user’s manual. Unpublished user’s manual, available at http://psych.colorado.edu/ .

Harvey, L. O. (Unpublished). Detection theory: Sensory and decision processes. Unpublished class material, available at http://psych.colorado.edu/ .

Hazy, T., Frank, M., & O'Reilly, R. (in press). Banishing the homunculus: making working memory work. Neuroscience.

Heeger, D. J., & Ress, D. (2002). What does fMRI tell us about neuronal activity? Nature Reviews Neuroscience, 3(2), 142–151.

Hegde, J., & Van Essen, D. C. (2000). Selectivity for complex shapes in primate visual area V2. Journal of Neuroscience, 20, RC61 1–6.

Heinke, D., Humphreys, G. W., & diVirgilio, G. (2002). Modelling visual search experiments: the selective attention for identification model (saim). Neurocomputing, 44-46, 817–822.

Heinke, D., Sun, Y. R., & W, H. G. (2005). Modeling grouping through interactions between top-down and bottom-up processes: The grouping and selective attention for identification model (g-saim). Lecture Notes in Computer Science, 3368, 148–158.

Herd, S., Banich, M., & OReilly, R. (IP). Neural mechanisms of cognitive control: An integrative model of stroop task performance and fmri data. Journal of Cognitive Neuroscience.

Herd, S., & O'Reilly, R. (In press). Serial visual search from a parallel model. Vision Research.

Hershler, O., & Hochstein, S. (2005). At first sight: A high-level pop out effect for faces. Vision Research, 45(13), 1707.

Hochstein, S., & Ahissar, M. (2002). View from the top: Hierarchies and reverse hierarchies in the visual system. Neuron, 36(5), 791–804.

Hooge, I. T. C., & Erkelens, C. J. (1999). Perepheral vision and oculomotor control during visual search. Vision Research, 39, 1567–15.

Hopf, J. M., & Mangun, G. R. (2000). Shifting visual attention in space: an electrophysiological analysis using high spatial resolution mapping. Clinical Neurophysiology, 111, 1241–1257.

Horowitz, T. S., Holcombe, A. O., Wolfe, J. M., Arsenio, H. C., & DiMase, J. S. (2004). Attentional pursuit is faster than attentional saccade. Journal of Vision, 4(7), 585–603.

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

Horowitz, T. S., & Wolfe, J. M. (2001). Search for multiple targets: Remember the targets, forget the search. Perception & 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.

Humphreys, D. H. G. W. (2003). Attention, spatial representation, and visual neglect: Simulating emergent attention and spatial memory in the selective attention for identification model (saim). Psychological Review, 110(1), 29–87.

Humphreys, G., & Muller, H. (1993). SEarch via Recursive Rejection (SERR): A connectionist model of visual search. Cognitive Psychology, 25, 43–110.

Hupe, J. M., James, A. C., Girard, P., & Bullier, J. (2001). Response modulations by static texture surround in area V1 of the macaque monkey do not depend on feedback connections from V2. Journal of Neurophysiology, 85, 146–163.

Juan, C., Shorter-Jacobi, S. M., & Schall, J. D. (2004). Dissociation of spatial attention and saccade preparation. Proceedings of the National Academy of Sciences, 101(43), 15541–15544.

Judd, C. M., & McClelland, G. H. (1989). Data analysis, a model-comparison approach. Orlando, FL: Harcourt Brace Jovanovich.

Judd, C. M., McClelland, G. H., & Culhane, S. E. (1995). Data analysis: continuing issues in the everyday analysis of psychological data. Annual review of psychology, 46, 433–465.

Kanwisher, N., & Wojciulik, E. (2000). Visual attention: Insights from brain imaging. Nature Reviews Neuroscience, 1, 91–100.

Kapadia, M. K., Westheimer, G., & Gilbert, C. D. (1999). Dynamics of spatial summation in primary visual cortex of alert monkeys. Proceedings of the National Academy of Sciences, 96, 12073–12078.

Kaptein, N. A., Theeuwes, J., & Vanderheijden, A. H. C. (1995). Search for a conjunctively defined target can be selectively limited to a color-defined subset of elements,. Journal of Experimental Psychology: Human Perception and Performance, 21(5), 1053–1069.

Karnath, H.-O. (2001). New insights into the functions of the superior temporal cortex. Nature Reviews Neuroscience, 2 (8), 568–576.

Kastner, S., Nothdurft, H. C., & Pigarev, I. N. (1999a). Neuronal responses to orientation and motion contrast in cat striate cortex. Visual Neuroscience, 16, 587–600.

Kastner, S., Pinsk, M., DeWeerd, P., Desimone, R., & Ungerleider, L. (1999b). Increased activity in human visual cortex during directed attention in the absence of visual stimulation. Neuron, 22, 751–761.

Kastner, S., & Ungerleider, L. G. (2000). Mechanisms of visual attention in the human cortex. Annual Review of Neuroscience, 23, 315–341.

Kastner, S., Weerd, P. D., Desimone, R., & Ungerleider, L. C. (1998). Mechanisms of directed attention in the human extrastriate cortex as revealed by functional mri. Science, 282, 108–111.

Kenemans, J. L., Lijffijt, M., Camfferman, G., & Verbaten, M. N. (2002). Split-second sequential selective activation in human secondary visual cortex. Journal of Cognitive Neuroscience, 14 (1), 48–61.

King, R. A. J. (2003). Perceptual grouping selection rules in visual search: Methods of sub-group selection in multiple-target visual search tasks. PhD thesis, Georgia Institute of Technology.

Kingstone, A., Enss, J. T., Mangun, G. R., & Gazzaniga, M. S. (1995). Guided visual-search is a left-hemispher process in split-brain patients. Psychological Science, 6(2), 118–121.

Kisvarday, Z. F., Toth, E., Rausch, M., & Eysel, U. T. (1997). Orientation-specific relationship between populations of excitatory and inhibitory lateral connections in the visual cortex of the cat. Cerebral Cortex, 7 (7), 605–618.

Klein, R., & Farrell, M. (1989). Search performance without eye movements. Perception & Psychophysics, 46, 476–482.

Knierim, J. J., & Van Essen, D. C. (1992). Neuronal responses to static texture patterns in area V1 of the alert macaque monkey. Journal of Neurophysiology, 4, 961–980.

Kobatake, E., & Tanaka, K. (1994). Neuronal selectivities to complex object features in the ventral visual pathway. Journal of Neurophysiology, 71 (3), 856–867.

Kusunokia, M., Gottlieba, J., & Goldberg, M. E. (2000). The lateral intraparietal area as a salience map: the representation of abrupt onset, stimulus motion, and task relevance. Vision Research, 10-12, 1459–1468.

Lachter, J., Forster, K. I., & Ruthrugg, E. (2004). Forty-five years after broadbent (1958): Still no identification without attention. Psychological Review, 111, 880–913.

Lamme, V. A. F., & Spekreijse, H. (2000). Modulations of primary visual cortex representing attentive and conscious scene perception. Frontiers in Bioscience, 5, 232–243.

Lavie, N. (1995). Perceptual load as a necessary condition for selective attention. Journal of Experimental Psychology: Human Perception and Performance, 21, 451–468.

Lavie, N., & Cox, S. (1997). On the efficiency of visual selective attention: Efficient visual search leads to inefficient distractor rejection. Psychological Science, 8(5), 395–398.

Lee, D., & Chun, M. M. (2001). What are the units of visual short-term memory, objects or spatial locations? Perception & Psychophysics, 63(2), 253–257.

Levin, Daniel T.and Takarae, Y. M. A. G. K. F. (2001). Efficient visual search by category: Specifying the features that mark the difference between artifacts and animals in preattentive vision. Perception & Psychophysics, 63 (4), 676–697.

Li, Z. P. (2002). A saliency map in primary visual cortex. Trends in Cognitive Sciences, 6 (1), 9–16.

Liversedge, S. P., & Findlay, J. M. (2000). Saccadic eye movements and cognition. Trends in Cognitive Sciences, 4(1), 6–14.

Luck, S. J., Chelazzi, L., Hillyard, S. A., & Desimone, R. (1997). Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. Journal of Neurophysiology, 77 (1), 24–42.

Luck, S. J., & Hillyard, S. A. (1995). The role of attention in feature detection and conjunction discrimination: An electrophysiological analysis. International Journal of Neuroscience Special Issue: M. Russell Harter memorial issue: Progress in visual information processing, 80, 281–297.

Luck, S. J., & Vogel, E. K. (1997). the capacity of visual working memory for features and conjunctions. Nature, 390, 279.

Mackeben, M., & Nakayama, K. (1993). Express attentional shifts. Vision Research, 33(1), 85–90.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. Psychological Bulletin, 109, 163–203.

Mangun, G. R., Hinrichs, H., Scholz, M., Mueller-Gaertner, H. W., Herzog, H., Krause, B. J., Tellman, L., Kemna, L., & Heinze, H. J. (2001). Integrating electrophysiology and neuroimaging of spatial selective attention to simple isolated visual stimuli. Vision Research, 41, 1423–1435.

Martinez, A., Anllo-Vento, L., Sereno, M. I., Frank, L. R., Buxton, R. B., Dubowitz, D. J., Wong, E. C., Hinrichs, H., Heinze, H. J., & Hillyard, S. A. (1999). Involvement of striate and extrastriate visual cortical areas in spatial attention. Nature Neuroscience, 2 (4), 364–369.

Martinez, A., Di Russo, F., Anllo-Vento, L., Sereno, M. I., Buxton, R. B., & Hillyard, S. A. (2001). Putting spatial attention on the map: timing and localization of stimulus selection processes in striate and extrastriate visual areas. Vision Research, 41, 1437–1457.

McAdams, C. J., & Maunsell, J. H. R. (2000). Attention to both space and feature modulates neuronal responses in macaque area V4. Journal of Neurophysiology, 83, 1751–1755.

McCarley, J. S., Wang, R. F., Kramer, A. F., Irwin, D. E., & Peterson, M. S. (2003). How much memory does oculomotor search have? Psychological Science, 14(5), 422–426.

McElree, B., & Carrasco, M. (1999). The temporal dynamics of visual search: evidence for parallel processing in feature and conjunction searches. Journal of Experimental Psychology: Human Perception and Performance, 25(6), 1517–1539.

McSorley, E., & Findlay, J. M. (2003). Saccade target selection in visual search: Accuracy improves when more distractors are present. Journal of Vision, 3(11), 877–892.

Mehta, S. D., Ulbert, I., & Schroeder, C. E. (2000). Intermodal selective attention in monkeys. i: Distribution and timing of effects across visual areas. Cerebral Cortex, 10 (4), 343–358.

Mizobe, K., Polat, U., Pettet, M. W., & Kasamatsu, T. (2001). Facilitation and suppression of single striate-cell activity by spatially discrete pattern stimuli presented beyond the receptive field. Visual Neuroscience, 18, 377–391.

Moran, J., & Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. Science, 229, 782–784.

Motter, B. C. (1993). Focal attention produces spatially selective processing in areas V1, V2 and V4 in the presence of competing stimuli. Journal of Neurophysiology, 70, 909–919.

Motter, B. C. (1994a). Neural correlates of attentive selection for color or luminance in extrastriate area V4. Journal of Neuroscience, 14 (4), 2178–2189.

Motter, B. C. (1994b). Neural correlates of feature selective memory and pop-out in extrastriate area V4. Journal of Neuroscience, 14 (4), 2190–2199.

Motter, B. C., & Belky, E. J. (1998a). The guidance of eye movements during active visual search. Vision Research, 38(12), 1805–1818.

Motter, B. C., & Belky, E. J. (1998b). The zone of focal attention during active visual search. Vision Research, 38(7), 1007–1022.

Mozer, M. C. (2002). Frames of reference in unilateral neglect and visual perception: A computational perspective. Psychological Review, 109(1), 156–185.

Mueller, H. J., Heller, D., & Ziegler, J. (1995). Visual search for singleton feature targets within and acrossfeature dimensions. Perception & Psychophysics, 57 (1), 1–17.

Nakayama, K., & Silverman, G. H. (1986). Serial and parallel encoding of visual feature conjunctions. Investigative Ophthalmology and Visual Science, 27 (Suppl. 182.

Nelson, J. I., & Frost, B. J. (1978). Orientation-selective inhibition from beyond the classic visual receptive field. Brain Research, 139, 359–365.

Nick Donnellyand Found, Andrewand Mueller, H. J. (2000). Are shape differences detected in early vision? Visual Cognition, 7 (6), 719–741.

Nothdurft, H. C., Gallant, J. L., & Van Essen, D. C. (1999). Response modulation by texture surround in primate area V1: Correlates of ”popout” under anesthesia. Visual Neuroscience, 16, 15–34.

Olson, I. R., Chun, M. M., & Allison, T. (2001). Contextual guidance of attention - human intracranial event-related potential evidence for feedback modulation in anatomically early, temporally late stages of visual processing. Brain, 124, 1417–1425.

Olsson, H., & Poom, L. (2005). Visual memory needs categories. Proceedings of the National Academy of Sciences of the USA, 102(24), 8776–8780.

Op De Beeck, H., & Vogels, R. (2000). Spatial sensitivity of macaque inferior temporal neurons. Journal of Comutational Neurology, 426, 554–575.

O'Reilly, R. C. (1998). Six principles for biologically-based computational models of cortical cognition. Trends in Cognitive Sciences, 2(11), 455–462.

O'Reilly, R. C., Frank, M. J., Hazy, T. E., & Watz, B. (submitted). Rewards are timeless: The primary value and learned value (pvlv) pavlovian learning algorithm.

O'Reilly, R. C., & Munakata, Y. (2000). Computational explorations in cognitive neuroscience: Understanding the mind by simulating the brain. Cambridge, MA: MIT Press.

Palmer, J. (1995). Attention in visual search: Distinguishing four causes of a set-size effect. Current Directions in Psychological Science, 4, 118–123.

Palmer, J., Verghese, P., & Pavel, M. (2000). The psychophysics of visual search. Vision Research, 40, 1227–1268.

Pashler, H. (1987). Detecting conjunctions of color and form: Reassessing the serial search hypothesis. Vision Research, 41(3), 191–201.

Pashler, H. (1998). The psychology of attention. Cambridge, MA: MIT Press.

Pashler, H., & Harris, C. (2001). Spontaneous allocation of visual attention: Dominant role of uniqueness. Psychonomic Bulletin & Review, 8 (4), 747–752.

Pashler, H., Johnston, J. C., & Ruthruff, E. (2001). Attention and performance. Annual Review of Psychology, 52, 637–641.

Pasupathy, A., & Connor, C. (1999). Responses to contour features in macaque area V4. Journal of Neurophysiology, 82 (5), 2490–2502.

Pasupathy, A., & Connor, C. (2001). Shape representation in area V4: Position-specific tuning for boundary conformation. Journal of Neurophysiology, 86 (5), 2505–2519.

Pinsk, M. A., Kastner, S., Desimone, R., & Ungerleider, L. G. (1999). An estimate of receptive field size in human visual cortex. NeuroImage, 9, 5885.

Polat, U., Mizobe, K., Pettet, M. W., Kasamatsu, T., & Norcia, A. (1998). Collinear stimuli regulate visual responses depending on cell's contrast threshold. Nature, 391, 580–584.

Posner, M., Walker, J., Friedrich, F., & Rafal, R. (1984). Effects of parietal lobe injury on covert orienting of visual attention. Journal of Neuroscience, 4, 1863–1874.

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

Raiguel, S., Van Hulle, M. M., Xiao, D. K., Marcar, V. I., & Orban, G. A. (1995). Shape and spatial distribution of receptive fields and antagonistic motion surrounds in the middle temporal area V5 of the macaque. European Journal of Neuroscience, 7, 2064–2082.

Raizada, R. D. S., & Grossberg, S. (2001). Context-sensitive binding by the laminar circuits of V1 and V2: A unified model of perceptual grouping, attention, and orientation contrast. Visual Cognition, 8 (3/4/5), 431–466.

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. Psychological Bulletin, 114(3), 510–532.

Rees, G., & Lavie, N. (2001). What can functional imaging reveal about the role of attention in visual awareness. Neuropsychologia, 39, 1343–1353.

Ress, D., Backus, B., & Heeger, D. (2000). Activity in primary visual cortex predicts performance in a visual detection task. Nature Neuroscience, 3, 940–945.

Reynolds, J., Chelazzi, L., & Desimone, R. (1999). Competitive mechanisms subserve attention in macaque areas v2 and v4. Journal of Neuroscience, 19, 1736–1753.

Reynolds, J., Pasternak, T., & Desimone, R. (2000). Attention increases sensitivity of v4 neurons. Neuron, 26, 703–714.

Reynolds, J. H., & Chelazzi, L. (2004). Attentional modulation of visual processing. Annual Review of Neuroscience, 27, 611–647.

Rizzolatti, G., Riggio, L., & Sheliga, B. M. (1994). Space and selective attention. Attention and Performance, 15, 231–265.

Roelfsema, P. R., Lamme, V. A. F., & Spekreijse, H. (1998). Object based attention in the primary visual cortex of the macaque monkey. Nature, 395, 376–381.

Rolls, E. T., Aggelopoulos, N. C., & Zheng, F. (2003). The receptive fields of inferior temporal cortex neurons in natural scenes. Journal of Neuroscience, 23(1), 339–348.

Rousselet, G. A., Thorpe, S. J., & Fabre-Thorpe, M. (2004). How parallel is visual processing in the ventral pathway? Trends in Cognitive Sciences, 8(8), 363–370.

Sasaki, Y., Hadjikhani, N., Fischl, B., Liu, A. K., Marret, S., Dale, A. M., & Tootell, R. B. H. (2001). Local and global attention are mapped retinotopically in human occipital cortex. Proceedings of the National Academy of Sciences of the USA, 98 (4), 2077–2082.

Schein, S. J., & Desimone, R. (1990). Spectral properties of V4 neurons in the macaque. Journal of Neuroscience, 10, 3369–3389.

Scialfa, C. T., & Joffe, K. M. (1998). Response times and eye movements in feature and conjunction search as a function of target eccentricity. Perception & Psychophysics, 60(6), 1067–1082.

Shen, J., Reingold, E. M., & Pomplun, M. (2003). Guidance of eye movements during conjunctive visual search: The distractor-ratio effect. Canadian Journal of Experimental Psychology, 57(2), 76–96.

Shevelev, I. A. (1998). Second-order features extraction in the cat visual cortex: Selective and invariant sensitivity of neurons to the shape and orientation of crosses and corners. Biosystems, 48, 195–204.

Shevelev, I. A., Lazareva, N. A., Novikova, R. V., Tikhomirov, A. S., Sharaev, G. A., & Cuckiridze, D. Y. (2001). Tuning to y-like figures in the cat striate neurons. Brain Research Bulletin, 54, 543–551.

Shipp, S. (2004). The brain circuitry of attention. Trends in Cognitive Sciences, 8(6), 223–230.

Sillito, A., Grieve, K. L., Jones, H. E., & Cudeir, J. (1995). Visual cortical mechanisms detecting focal orientation discontinuities. Nature, 378, 492–496.

Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst; sustained inattentional blindness for dynamic events. Perception, 28, 1059–1074.

Smith, A. T., Singh, K. D., & Greenlee, M. W. (2000). Attentional suppression of activity in the human visual cortex. NeuroReport, 11 (2), 271–277.

Snyder, J. J., & Kingstone, A. (2000). Inhibition of return and visual search: How many separate loci are inhibited? Perception & psychophysics, 62(3), 452–458.

Somers, D. C., Dale, A. M., Seiffert, A. E., & Tootell, R. B. H. (1999). Functional mri reveals spatially specific attentional modulation in human primary visual cortex. Proceedings of the National Academy of Sciences of the USA, 96 (4), 1663–1668.

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

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. Journal of Experimental Psychology, 18, 643–662.

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.

Swindale, N. V. (1995). Responses of neurons in cat striate cortex to vernier offsets in reverse contrast stimuli. Visual Neuroscience, 12 (5), 805–817.

Swindale, N. V., & Cynader, M. S. (1986). Vernier acuity of neurones in cat visual cortex. Nature, 319, 591–593.

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

Thompson, K. G., Bichot, N. P., & Schall, J. D. (2001). From attention to action in frontal cortex. In J. Braun, & C. Koch (Eds.), Visual attention and cortical circuits. (pp. 137–157). Cambridge, Mass: MIT Press.

Tootell, R. B. H., Hadjikhani, N., Hall, E. K., Vanduffel, W., Vaughan, J. T., & Dale, A. M. (1998). The retinotopy of visual spatial attention. Neuron, 21 (6), 1409–1422.

Townsend, J. T. (1971). A note on the identifiability of parallel and serial processes. Perception & Psychophysics, 10(3), 161–163.

Townsend, J. T., & Ashby, F. G. (1983). The stochastic modeling of elementary psychological processes. Cambridge, UK: Cambridge University Press.

Treisman, A. (1993). The perception of features and objects. In A. Baddeley, & L. Weiskrantz (Eds.), Attention: Selection, awareness, and control: A tribute to donald broadbent (pp. 5–35). Oxford: Oxford University Press.

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

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

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

Treue, B. C., & Maunsell, J. H. R. (1996). Attentional modulation of visual motion processing in cortical areas MT and MST. Nature, 382, 539–541.

Treue, S., & Trujillo, J. C. M. (1999). Feature-based attention influences motion processing gain in macaque visual cortex. Nature, 399, 575–579.

Trick, L. M., & Pylyshyn, Z. W. (1994). Why are small and large numbers enumerated differently - a limited-capacity preattentive stage in vision. Journal of Experimental Psychology: Human Perception and Performance, 19(2), 331–351.

Ulrich, R., & Miller, J. (1993). Information processing models generating lognormally distributed reaction times. Journal of mathematical psychology, 37, 513–525.

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

Vanduffel, W., Tootell, R. B. H., & Orban, G. A. (2000). Attention-dependent suppression of metabolic activity in the early stages of the macaque visual system. Cerebral Cortex, 10, 109–126.

vanMarle, K., & Scholl, B. J. (2003). Attentive tracking of objects versus substances. Psychological Science, 14(5), 498–504.

Verghese, P. (2003). Visual search and attention: A signal detection theory approach. Neuron, 31(4), 523–535.

Versavel, M., Orban, G. A., & Lagae, L. (1990). Responses of visual cortical neurons to curved stimuli and chevrons. Vision Research, 30, 235–248.

Vidyasagar, T. R. (1998). Gating of neuronal responses in macaque primary visual cortex by an attentional spotlight. Neuroreport, 9 (9), 1947–1952.

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2001). Storage of features, conjunctions, and objects in visual working memory. Journal of Experimental Psychology: Human Perception and Performance, 27(1), 92–114.

von der Heydt, R., Zhou, H., & Friedman, H. S. (2000). Representation of stereoscopic edges in monkey visual cortex. Vision Research, 40(15), 1955–1967.

von Muehlenen, A., & Mueller, H. J. (2000). Perceptual integration of motion and form information: Evidence of parallel-continuous processing. Perception & Psychophysics, 62(3), 517–531.

Ward, R. (2001). Visual attention moves no faster than the eyes. In K. Shapiro (Ed.), The limits of attention: Temporal constraints in human information processing (pp. 199–216). London, England: Oxford University Press.

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

Ward, R., & McClelland, J. L. (1989). Conjunctive search for one and two identical targets. Journal of Experimental Psychology: Human Perception and Performance, 15(4), 664–672.

Williams, D. E., & Reingold, E. M. (2001). Preattentive guidance of eye movements during triple conjunction search tasks: The effects of feature discriminability and saccadic amplitude. Psychonomic Bulletin & Review, 8 (3), 476–488.

Wolfe, J., & Gancarz, G. (1997). 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). Norwell, MA: Kluwer Academic Publishers.

Wolfe, J., Yu, K., Stewart, M., & Shorter, A. D. (1990). Limitations on the parallel guidance of visual search: Color * color and orientation * orientation conjunctions. Journal of Experimental Psychology: Human Perception & Performance, 16 (4), 879–892.

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. (1998a). Visual search. In H. Pashler (Ed.), Attention (pp. 13–73). East Sussex, England: Psychology Press.

Wolfe, J. M. (1998b). What can 1 million trials tell us about visual search? Psychological Science, 9 (1), 33–39.

Wolfe, J. M., Alvarez, G. A., & Horowitz, T. S. (2000). Attention is fast but volition is slow. Nature, 406(6797), 691.

Wolfe, J. M., & Bennett, S. C. (1996). Preattentive object files: Shapeless bundles of basic features. Vision Research, 37, 25–44.

Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided search: An alternative to the feature integration model for visual search. Journal of Experimental Psychology: Human Perception & Performance, 15(3), 419–433.

Wolfe, J. M., & DiMase, J. S. (2003). Do intersections serve as basic features in visual search. Perception, 32(6), 645–655.

Wolfe, J. M., Michod, K. O., & Horowitz, T. S. (submitted). Visual search: Linear reaction time by set size functions are not really linear.

Woodman, Geoffrey F.and Vogel, E. K. L. S. J. (2001). Visual search remains efficient when visual working memory is full. Psychological Science, 12(3), 219–224.

Woodman, G., & Luck, S. J. (1999). Electrophysiological measurement of rapid shifts of attention during visual search. Nature, 400, 867–869.

Woodman, G. F., & Luck, S. J. (2003). Serial deployment of attention during visual search. Journal of Experimental Psychology: Human Perception and Performance, 29, 121–138.

Woodman, G. F., & Luck, S. J. (2004). Visual search is slowed when visuospatial working memory is occupied. Psychonomic Bulletin and Review, 11(2), 269–274.

Xiao, D. K., Raiguel, S., Marcar, V., & Orban, G. A. (1997). The spatial distribution of the antagonistic surround of MT/V5 neurons. Cerebral Cortex, 7, 662–677.

Zelinsky, G. J. (1996). Using eye saccades to assess the selectivity of search movements. Vision Research, 36(14), 2177.

Zipser, K., Lamme, V. A. F., & Schiller, P. H. (1996). Contextual modulation in primary visual cortex. Journal of Neuroscience, 16, 7376.

Zohary, E., & Hochstein, S. (1989). How serial is serial processing in vision? Perception, 18(2), 191–200.

This appendix contains information useful as a reference in reading the body of this paper, and in providing evidence for the conclusions reached there. These sections are not organized as a narrative; the only structure among them can be found in reference to the body of the work.

Appendix A

Pop-out Effects

Pop-out refers to a phenomena in which attention is directed relatively quickly and automatically to an unlike stimulus among many like stimuli, as in figure 1 a. The term is used both in texture-segmentation and search paradigms (Treisman, 1993; Wolfe, 1998a). The hypothesis of neural activity leading visual attention (section 3.3) goes some distance toward explaining pop-out effects in visual search when it is coupled with robust findings of similar-surround suppression.

Similar-surround suppression (SSS) is an effect in which neural responses are preferentially reduced when the response-evoking stimulus is surrounded by similar rather than dissimilar stimuli Many studies have demonstrated this effect. (e.g. (Hupe, James, Girard, & Bullier, 2001; Knierim & Van Essen, 1992). Therefore a vertical line surrounded by horizontal lines produces a larger neural response than does a vertical line surrounded by other vertical lines. This phenomena is referred to as cross-orientation enhancement although the enhancement in most neurons is only relative to an iso-oriented surround. This property neatly explains the pop out effect in a display such as Figure A.1. The representation of each of the horizontal lines is supressed by that of the surrounding horizontal lines, while the representation of the diagonal line is supressed less. Therefore the target location is left as the most active area in ventral stream representation.

This mechanism has been implicated in efficient search for a feature singleton (Chelazzi, 1999). It has also been shown that attention is guided to unique features in the absence of any

a) Search display b) Corresponding neural activity

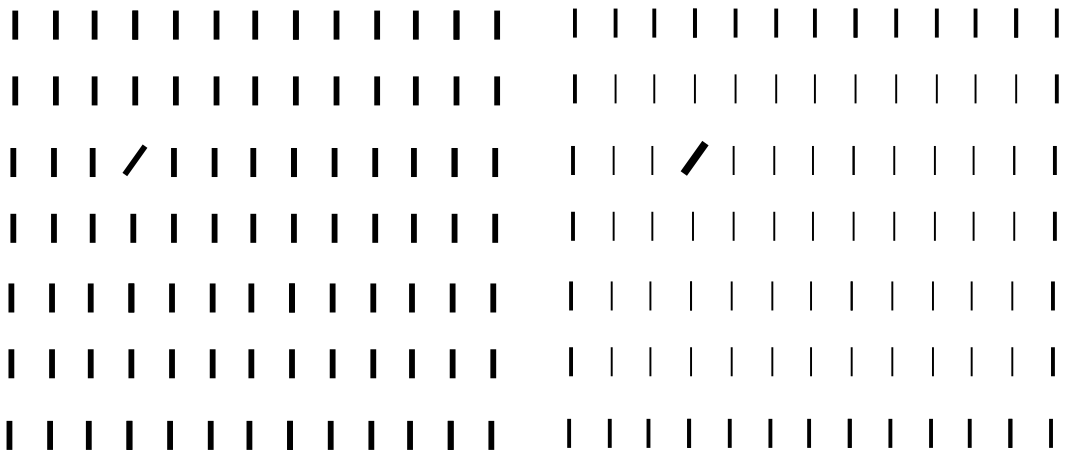

Figure A.1: Pop-out mediated by dissimilar surround enhancement effect. a) The visual display. Note how attention is drawn to the unlike stimulus. b) line weight now corresponds to activity levels of neural populations representing each line. Activity of neurons representing each horizontal line has been reduced due to the iso-surround inhibition effect; those at the edges are relatively less reduced due to few surrounding lines. The activity of neurons representing the single diagonal line is relatively less reduced, leading to that region's dominance in competition for representation.

other goal (Pashler & Harris, 2001). It also seems possible that similar-surround effects should be seen at all levels of the visual system, making this mechanism a good candidate for allocating attention to unique features in the absence of a top-down goal. The finding of similar surround suppression has been linked to so-called pop-out effects observed in very efficient search in which the target item subjectively seems to 'pop out' of the display and become the focus of attention without any effort of the observer. As such I have reviewed the basis of the effect, although it seems in the final analysis that the effect is not central to the process governing most search tasks and plays a role primarily in texture segmentation and possibly in the automatic deployment of attention in the absence of top-down goals.

A similar-surround suppression effect has been demonstrated only for orientation in area V1 (Hupe et al., 2001; Zipser, Lamme, & Schiller, 1996; Nelson & Frost, 1978; Sillito, Grieve, Jones, & Cudeir, 1995; Kastner, Nothdurft, & Pigarev, 1999a; Nothdurft, Gallant, & Van Essen, 1999; Knierim & Van Essen, 1992), motion in area MT (Xiao, Raiguel, Marcar, & Orban, 1997; Raiguel, Van Hulle, Xiao, Marcar, & Orban, 1995; Allman, Miezin, & McGuinness, 1985; Born, 2000), motion in area MST (Eifuku & Wurtz, 1998) and color in area V4 (Schein & Desimone, 1990). This mechanism supplies an explanation on one level of pop-out effects involving these features. The mechanism underlying this effect is not known, but the time course of the effect limits the possibilities.

Several analyses of orientation popout effects have determined that these effects start relatively rapidly after response onset, about 15-20 ms (Nothdurft et al., 1999; Knierim & Van Essen, 1992). This timing is long enough to allow feedback connections from higher areas to play a role, but also similar to the timing expected from horizontal connections from relatively nearby neurons. However, the effect delay is slightly longer than would be predicted by timing of closely neighboring neurons of the same layer (Bringuier, Chavaner, Glaeser, & Fregnac, 1999), although the statistical methods used would influence the calculated time when effects become apparent. Other evidence also points to feedback mechanisms. The effect seems

to depend in part on the actual orientation of the central line, an effect not easily explained through horizontal interactions. There is a tendency for the same neurons displaying a large similar-surround suppression effect to prefer their local orientation preference when the center display is orthogonal to that preference. This means they prefer a surround orthogonal to that displayed in their classical receptive field, an effect that is very difficult to explain by horizontal connections, because the overall activation level is similar in both cases so that an appeal to shifting excitation/inhibition balance does not apply. Because some neurons prefer a contrast between center and surround even when the two are reversed, it seems that the similar-surround suppression effect must be partly an effect of the attentional preference for unlike among like rather than purely a cause of attentional effects.

Similar surround supression has been likened to figure-ground segregation (Nothdurft et al., 1999), another effect of context in V1 that is clearly mediated by feedback. However, figure-ground segregation has been demonstrated to affect activity only at latencies longer than 80 ms (Zipser et al., 1996). Because the SSS effect is detectable sooner than that, it seems likely that the effect is more local, possibly involving only horizontal interactions and fast feedback. Complicating the question of whether the effect is local or feedback-based, it was recently demonstrated that the similar-surround effect is not decreased by deactivation of area V2 (Hupe et al., 2001). One word of caution regarding all of the reports of similar-surround suppression is that all results were obtained in anesthetized animals, and are therefore difficult to interpret. Feedback likely plays a larger role in the awake animal.

Although some evidence points to feedback mechanisms, there is a good explanation for the effect at a local level. A similar surround effect was obtained in a neural network model that learned connection weights from exposure to natural scenes (Raizada & Grossberg, 2001; Grossberg & Williamson, 2001). Models such as this predict that long-range horizontal projections will prefer similar orientations both for excitatory and inhibitory projections, because a basically Hebbian learning scheme that connects co-active neurons will respond to the extended

lines found in natural (and especially human-produced) scenes. Such an orientation preference has been demonstrated in striate cortex of the cat for both excitatory and inhibitory neurons (Kisvarday, Toth, Rausch, & Eysel, 1997; Dalva, Weliky, & Katz, 1997) (But see Buzas, Eysel, Adorjan, & Kisvarday, 2001 for evidence that some long-range inhibition is selective for cross-orientation).

Because both excitatory and inhibitory projections prefer iso-oriented neurons at long range (close range inhibitory projections are only slightly specific and largely nonselective Buzas et al., 2001; Das & Gilbert, 1999), the model of Grossberg and Williamson (2001) had to assume that the effects of inhibition dominated over that of excitation at higher target activity levels. There is a good deal of evidence for such a dominance in the literature. First, receptive field sizes are considerably larger at low contrasts, pointing to a different balance of excitation and inhibition at low contrast (Kapadia, Westheimer, & Gilbert, 1999; Anderson, Lampl, Gillespie, & Ferster, 2001). Also, low contrast stimuli are enhanced by colinear high contrast stimuli, while high contrast stimuli are inhibited by colinear stimuli (Mizobe, Polat, Pettet, & Kasamatsu, 2001; Polat, Mizobe, Pettet, Kasamatsu, & Norcia, 1998).

If similar-surround suppression is a learned effect based on the statistics of the natural world, we would expect to see such an effect ubiquitously at all levels of the visual cortex. In this case the mechanism could be offered as a general explanation for the tendency of attention toward distinctive items among many of the same.

However, there are other likely mechanisms that could contribute to this effect. It has been noted by Shipp (2004) that the initial wave of activation in higher visual form areas (IT and TEO) allows much more parallel representation of information than does the later state. This likely occurs because lateral competition among different representations is completed over time. In this case, the first wave of processing will briefly activate the neurons that respond to both types of items, at the one retinotopic location where both items are present. If, for instance, some neurons are activated by T junctions, and others by L junctions, both sets briefly

become active.

This activation could enable pop-out search in two ways. First, if the particular target is known as the subject of search, even a brief activation of neurons that signal its presence could trigger a response. Second, there will be extra activation in retinotopic populations that respond to both the unique object and those around it, and this extra activation could trigger an automatic shift of attention. Even if the different object is not known or even suspected in advance, there will be greater neural activity corresponding to the location surrounding that object. This would happen because populations responsive to both types of items become active in the one location (assuming that some such neurons have RFs that are large enough to encompass several stimuli). At other locations, only one population of neurons becomes active. This brief extra activity could be enough to trigger a shift of attention toward that location, and allow identification of the unique object.

Appendix B

Competition for Representation Among Visual Stimuli

A recently proposed paradigm (Desimone & Duncan, 1995) describes attention as a biasing effect in a competition among stimuli for neural representation. This metaphor seems to have much to recommend it. Desimone and Duncan (1995) present a range of findings to support their claims, and new empirical evidence for this theoretical framework continues to accrue. Much neurophysiological and experimental evidence is reviewed in Desimone and Duncan (1995), while Kastner and Ungerleider (2000) and Kanwisher and Wojciulik (2000) provide excellent reviews of human imaging evidence. I review here some of the most important studies from these reviews, as well as additional and more recent evidence.

Uncountable behavioral studies demonstrate that stimuli effectively compete for influence on behavior. The Stroop task (Stroop, 1935) and all of its many related behavioral paradigms (MacLeod, 1991) are perfect examples. Other examples include many demonstrations that increased attentional demands reduce the extent to which unattended stimuli are remembered and in some cases even the amount of neural activity associated with processing. Many studies of attention demonstrate that stimuli compete for attention so that processing a stimulus prevents processing of other stimuli to a greater or lesser degree. For an excellent overview of attentional studies see Pashler (1998). While it remains an area of active debate, no task has yet been identified which can clearly be performed completely independently of competing attentional demands (Pashler, Johnston, & Ruthruff, 2001). Lavie (1995) reviews a number of her own studies demonstrating that in many instances, varying the difficulty of a primary task controls

the amount of irrelevant information that the subjects process, so that a sufficiently demanding task produces no detectable processing of distractor items from the same input modality. This is perhaps demonstrated most strikingly by Simons and Chabris (1999). They had subjects perform a challenging task of counting basketball passes among two intermixed teams, and then noted that nearly half of participants failed to notice a woman in a gorilla suit walk through the game. These effects can also be demonstrated neurally, as reviewed in Rees and Lavie (2001)). They review a series of experiments demonstrating that subjects show less brain activation associated with distractor stimuli (moving dots) when they perform a difficult rather than easy task on visually presented letters, and, interestingly, that this effect disappears when the letters are presented auditorily rather than visually.

Neurophysiological work provides a closer look at the neural dynamics underlying the high level effect of competition among stimuli. Moran and Desimone (1985) showed that attention biases neural representation in V4 and IT towards the attended stimuli. Monkeys were precued with a target location (with a coded, not spatial, cue), and when the target and distractor were within the neuron's receptive field, the response was closer to that given only to the target then that given to the distractor. A later study from the same lab confirmed and extended those results (Chelazzi et al., 1993). In this work, they demonstrate that a neuron in V4 or IT will initially respond to both target and distractor stimuli, but switch to responding as if only the target was present within 200 ms of stimuli onset. In this paradigm, the monkey was cued with the identity of the target, then simultaneously shown both target and distractor, and required to make a saccade towards the target. The response switched to that of the target about 100 ms before the saccade was initiated.

Another key piece of evidence for attentionally biased competition for representation among stimuli is presented in a more recent and more systematic follow-up work, again from the same influential lab. Reynolds et al. (1999) performed a more detailed analysis of the response to different stimuli presented both alone and in pairs, without attention directed specif-

ically to either stimulus. What they found was that the response of a neuron in V2 and V4 to a pair of stimuli corresponded rather closely to the average of the response to each stimuli individually. This finding is key because it demonstrates that the two stimuli do indeed compete for representation; if they did not, the response would be greater than that to either stimulus individually. This finding necessitates competitive inhibition at some level of the processing of these stimuli. This inhibition is probably local and exists at all areas of visual processing, since inhibitory neurons in the cortex connect only locally, and tend to shorter axonal length than excitatory neurons (Crick & Asanuma, 1986; O'Reilly & Munakata, 2000). It is also important to note that Reynolds et al. (1999) found responses of slightly above the average to the two unattended stimuli. This is consistent with a conception of stable feedback inhibitory control (O'Reilly & Munakata, 2000).

A recent fMRI study corraborates neurophysiological findings competition within early visual areas (Kastner, Weerd, Desimone, & Ungerleider, 1998). The paradigm used was four "colorful and complex" visual stimuli all located within one visual quadrant while subjects performed a demanding task at fixation. The stimuli were presented either simultaneously or sequentially. The reported finding was a substantially higher activation when stimuli were shown sequentially rather than simultaneously. This demonstrates that the simultaneously presented stimuli were in some sense competing. This conclusion seems to depend on assumptions about the hemodynamic response that are questionable given the current state of understanding (Heeger & Ress, 2002). In a followup study (Pinsk, Kastner, Desimone, & Ungerleider, 1999) they calculate approximate receptive fields sizes in multiple human visual. This was done by varying seperation between stimuli in the paradigm described above.

They found receptive field sizes notably larger than those found in monkey neurophysiology. This finding may well indicate that there were nonlinear additive effects in the hemodynamic response. Despite the possibility of such effects, the correspondence of this study to neurophysiological findings makes it very likely that it does reveal effects of underlying neural

competion at multiple stages of processing.

Appendix C

Attention

The findings reviewed in this section and the above (Appendix B) are consistent with a view of attentional selection as a biasing effect in a competition among stimuli for neural representation. It seems likely that this bias consists of an excitatory input of the most common type, although this distinction is not terribly important for the current purposes. It is very important that the biasing effect of attention acts on stimuli in a featurewise manner as reviewed below. It is equally important that attentional effects can be observed at early retinotopic levels of the visual system (see Appendix C, section C.2), where they can influence the process of visual selection that is the main topic of this paper.

A less central but also interesting conclusion from the evidence in this section is that attentional effects are present at all areas and all temporal stages of visual processing. This result is in line with a view of attention as a desirable by-product of massive interareal connection combined with associative neural learning rules. A Hebbian learning rule would lead to higher level neurons representing a location developing preferential connections with all neurons responsive to that location at all levels of the visual system where anatomical projections exist. The fact that modulation is strongest where that modulation is most efficient in aiding task performance may point to the existence of a task learning component of the neural learning rule, and certainly points to evolutionary selection in the development of anatomical connectivity.

C.1 Neural mechanisms of attention

Current evidence points to attention as a biasing effect in a general framework of competition among stimuli for neural representation (see section 3.3. Attention has been observed to increase the firing rate of neurons representing attended information and locations in many studies, both at individual neurons and at the population level. This enhancement of firing rate is commonly considered to increase the influence of attended information on other processing, including higher areas and decisional processes.

After demonstrating that neural responses in areas V2 and V4 fall in between the responses to each stimulus individually, Reynolds et al. (1999) performed a second experiment in which one stimulus was attended, and showed that attending to a particular stimulus increased its weighting in the averaging of responses. Therefore attention could reduce the neural response as well as enhance it, depending on the response to the attended stimulus alone. This finding is a clear hallmark of competition among stimuli for representation, and therefore competition among the neurons representing different stimuli. Attention did not dictate the response of any neurons- it only biased it. This indicates that the effect of attention is additive, and argues strongly against a gating model. It is also important that this work shows that the largest effect of attention is to bias competition between responses to two stimuli when those stimuli compete for a particular neuron's response directly. In addition to replicating the results of Reynolds et al. (1999), Chelazzi et al. (2001) showed that attentional effects were reduced or eliminated when only one of the two stimuli used occupied the RF of the tested neuron. This is explained by the fact that RFs covary with the cortical location of neurons, and the finding that inhibitory neurons synapse mostly on nearby neurons, with no other detectable preference (Das & Gilbert, 1999).

Kastner and Ungerleider (2000) review results which seem to imply a diversity of mechanisms in attention. They review studies which show all of the following effects of covert attention: enhancement of neural response to an attended stimulus, the filtering of unwanted in-

formation; the abovementioned increase in baseline activity before a stimulus is presented; and the enhancement of a neuron's sensitivity to stimulus contrast. Although Kastner and Ungerleider (2000) treat these as diverse effects, all of these results can be accounted for under an account of attention as an excitatory bias signal with competitive local inhibition. The finding of increased sensitivity to contrast is the only somewhat nonlinear mapping: it is expected if a neuron with a nonlinear response curve is pushed into the sensitive part of that curve by biasing excitation (O'Reilly & Munakata, 2000).

Reynolds, Pasternak, and Desimone (2000) addressed the issue of multiplicative versus additive enhancement more directly. In single-cell work with monkeys, they found that the effects of attention in V4 were much stronger to low-contrast stimuli than to high contrast stimuli. These results strongly suggest that attention is an additive rather than multiplicative or gating effect. If the effects of attention were actually multiplicative or gating, the difference between attended and unattended conditions should be much larger for the stronger stimulus, since multiplying a larger value produces a larger change (gating would be equivalent to multiplying by 0 or 1). Instead they found that the effect of attention was to effectively increase the strength (in this case, contrast) of the stimuli. This is exactly the effect that would be expected of moderate excitatory input which excites a neuron to near threshold, into the most sensitive region of its response curve. Such an input would be expected to have little impact on a neuron which was experiencing strong excitatory input, since it would already be firing at closer to its maximum rate.

The effect of attention on visual search tasks demands that attention can enhance processing of a stimulus based on its features. Visual search studies demonstrate that observers can direct attention to a target when only one of its features is known, and that they can direct attention to a dimension such as shape or color (Mueller, Heller, & Ziegler, 1995)

Treue and Trujillo (1999) have demonstrated that features of attended stimuli can modulate responses to unattended stimuli. They find that attending direction of motion in the opposite

visual field modulates response by motion selective neurons by 13% between matching and opposite motions, and that the difference between attending opposite motion in opposite visual field and attending preferred motion within receptive field modulates activity by 25%. The difference between attending preferred and nonpreferred motion when both are within the RF simultaneously is 60% within their paradigm. They point to these results as being in disagreement with the framework of biased competition, because it is clear that features of a stimulus, rather than stimuli as a whole, compete for response competition. They propose instead a 'feature similarity gain model', as they also demonstrated that attention did not modulate tuning curves. It seems that the only reasonable way for stimuli to compete would be through selection of their features, and that one would expect objects to succeed in competition relative to their featurwise similarity to attended objects, as countless behavioural studies have demonstrated.

Similar results were reported using a task in which the monkeys decided whether a stimulus was the same or different than that presented 500 ms previously (McAdams & Maunsell, 2000). The differential modulation in this case was large despite the lack of a competing stimulus. In the spatial attention condition, activity was found to vary an average of 31% between attending to a grating patch and attending to a similar grating patch in the opposite hemifield. In the spatial plus featural attention condition activity was found to vary by an average of 54% between attending to the same grating patch and attending a colored gaussian in the opposite hemifield. These conditions demonstrate that featural and locational attention both serve to modulate activity in V4, and do so roughly additively. The observed modulations are quite large in comparison to other studies without competing stimuli within the same RF. However, two factors mitigate the apparant size of the effects. One is that neurons were deliberately selected to show attentional effects in the location plus feature condition, in contrast to previous studies. The second is that the two stimuli were relatively close together, with less separation than the diameter of each receptive field (if the schematics given are at all to scale). Thus these results seem to be roughly in line with those of other studies.

It seems that these and the results of Treue and Trujillo (1999) strengthen rather than conflict with the framework of stimulus competition laid out by Desimone and Duncan (1995). Regardless of the name of the model, all of the above results seem very much in line with a view of attention as a biasing effect in a competition for representation among stimuli, in the framework of a parallel distributed heirarchichal system. These results serve to demonstrate that this competition can be attentionally biased by features as well as by location of stimuli, an important finding for an understanding of mechanisms of visual search. If stimuli are biased by feature, and the winning stimulus becomes the focus of attention, then it is not neccessary to identify each stimulus, but only features which discriminate the desired target of attention from any other stimuli present.

C.2 Attentional Effects by Visual Area

It is important for a detailed understanding of visual search tasks to know in what areas of the brain attentional effects are observed. It is critical to the current thesis of visual search being subserved by lower areas that attention can in fact bias the competition for representation among stimuli (see sections 3.3 and Appendix B) even in low retinotopic areas.

The current evidence indicates that attentional effects are largest in area V4 and higher, smaller in V2, and difficult to detect in V1. A number of neurophysiological studies have clearly demonstrated attentional effects in area V4 (Luck, Chelazzi, Hillyard, & Desimone, 1997; Spitzer, Desimone, & Moran, 1988; Motter, 1993; Haenny & Schiller, 1988; McAdams & Maunsell, 2000; Chelazzi et al., 1993; Reynolds et al., 1999). Attentional effects in V2 are somewhat smaller (Luck et al., 1997; Motter, 1993; Reynolds et al., 1999). Attentional effects in V1 are decidedly more difficult to locate using single cell techniques, but definitely present (Motter, 1993; Vidyasagar, 1998; Haenny & Schiller, 1988). The effects of attention seem to be nearly omnipresent, extending to pre-cortical areas. A recent study by Bender and Youakim (2001) demonstrated the effects of attention within the pulvinar region of macaque thalamus,

although not within the lateral geniculate nucleus of the thalamus. These areas preceed cortical area V1 in visual processing stream. They found attentional modulation of 26% within the three pulvinar regions surveyed.

fMRI work has clearly demonstrated attentional effects in V1. Tootell, Hadjikhani, Hall, Vanduffel, Vaughan, and Dale (1998) used a high powered (3T) magnet to isolate the relatively small effects of attention in V1, and larger effects in higher visual areas. They found relatively small enhancements of activity due to their small target stimulus, but also detected decreases in MR signal in adjoining areas. A host of similar studies followed, demonstrating clearly that attentional modulation effects are present in area V1, and that these effects are in precisely the same areas that isolated stimuli activate (Martinez, Anllo-Vento, Sereno, Frank, Buxton, Dubowitz, Wong, Hinrichs, Heinze, & Hillyard, 1999; Brefczynski & DeYoe, 1999; Somers, Dale, Seiffert, & Tootell, 1999; Gandhi, Heeger, & Boynton, 1999; Sasaki, Hadjikhani, Fischl, Liu, Marret, Dale, & Tootell, 2001; Martinez, Di Russo, Anllo-Vento, Sereno, Buxton, & Hillyard, 2001). ERP's also demonstrate attentional effects in early visual areas. Luck and Hillyard (1995), for instance, have shown in human subjects a measurable enhancement of processing at attended locations in relation to unattended locations, and that this effect varies with the difficulty of the task.

Other studies indicate that attentional effects in V1 can be observed even in the absence of a stimulus. Kastner, Pinsk, DeWeerd, Desimone, and Ungerleider (1999b) showed increased activity with covert attention with no stimulus present in all retinotopic areas, strongest in V4 but present even in V1. The observed effect of attention in the absence of a stimulus seems to confirm that excitatory input is involved in attending. A similar result was demonstrated by Ress, Backus, and Heeger (2000). They demonstrated attentional effects in V1 and other areas using a paradigm of detecting a dim target, but extended these results to demonstrating increased signal in the attended condition when stimulus intensity was zero, that is, a blank screen that subjects thought might contain a stimulus. Similar effects have been reported in

single cell studies, where attention directed to a location prior to the appearance of a stimulus increases baseline firing rates (Luck et al., 1997; Colby, Duhamel, & Goldberg, 1996; Reynolds et al., 1999).

These findings are consistent with a view of attention as a biasing effect in a global competition for representation among stimuli (sections 3.3 and Appendix B). ERP studies, however, present some findings that are difficult to reconcile with this framework, and with the fMRI studies cited above. Martinez et al. (1999) and (2001) use ERP studies with the same experimental paradigm for which they report their fMRI result. The excellent temporal resolution of ERP is used to establish a time course of the attentional effects demonstrated with fMRI. The conclusion reached by Martinez and colleagues is that attentional effects are not present in the earliest portion of the response. These conclusions are somewhat troubling because they seem to conflict with the studies discussed above which indicate that attention has measurable effects in primary visual cortex (V1) even in the absence of any stimulus. There is also one single-unit study showing attentional modulation of initial responses in V1 (Motter, 1993).

Nonetheless, Martinez et al are in good company in drawing these conclusions. Many other ERP studies have reached similar conclusions. Kenemans, Lijffijt, Camfferman, and Verbaten (2002) report findings of attentional effects starting at around 120 ms, while others such as Olson, Chun, and Allison (2001; Hopf & Mangun, 2000) report attentional effects starting at 200 ms or later. The more extreme lag times in attentional onset are due to use of paradigms in which subjects cannot know before stimulus onset which stimulus is to be attended. Olson et al. (2001) used a contextual cuing paradigm in which spatial relationships between distractors had to be recognized before attention could be cued to a target location. Kenemans et al. (2002) asked subjects to attend to stimuli of a certain spatial frequency, which apparently necessitated some amount of identification before visual attention could be focused, since they recorded the first effects of attention from anterior, apparently prefrontal souces. Some other ERP studies

that report even longer onsets of attentional effects use paradigms designed to evoke shifts of attention rather than attention to item onset (Arnott, Pratt, Shore, & Alain, 2001; Hopf & Mangun, 2000). These studies show that attentional shifts start no sooner than around 200 ms, and are modulated by sources in extrastriate cortex and parietal areas, consistent with fMRI studies. These studies, while interesting, do not comment on the observed lack of attentional effects in V1 as early as stimulus-produced activity. However, even those few ERP studies that do use cues that give a location to attend before stimulus onset seem to show a lack of attentional effects in initial cortical visual processing.

Martinez and colleagues (1999,2001) report that the earliest component associated with a visual stimulus, the C1, is "not modulated by attention". They report that their statistical test of the effect revealed F(1,18)=3.6, not significant. But their F value in fact gives $P < .08$ (Judd & McClelland, 1989). These authors have improperly accepted the null hypothesis, rather than perform a more appropriate test such as a confidence interval of effect size (Judd & McClelland, 1989; Judd, McClelland, & Culhane, 1995). This practice is currently not uncommon; it allows a simplification of complex effects to a dichotomy of significant vs. non-significant. The real world, however, is more complicated, and in many cases demands a more complex analysis. This bit of legardemein concealed the fact that the near-significant difference was in the wrong direction. This curious effect is caused by the low spatial resolution of ERP recordings; the ipsilateral P1 effect obscures the negative contralateral C1. Therefore it seems that limited conclusions can be drawn from ERP data on the C1 response. Although the statistical analysis is confused, a visual inspection of the graphs provided by Martinez et al. (1999) shows that the attended and unattended ERP responses diverge as soon as a response occurs. Because the response is initially small, attentional modulation of that response is small as well. Other studies corraborate this picture.

Mangun, Hinrichs, Scholz, Mueller-Gaertner, Herzog, Krause, Tellman, Kemna, and Heinze (2001) performed a similar study, and obtained similar results. They used a task in

which subjects discriminated sizes of rectangles displayed off of fixation in the upper visual field. They used simpler stimuli to de-emphasize effects in higher areas responsible for shape discrimination, necessary in the tasks used by Martinez et al (1999,2000) and many others. Mangun et al. (2001) used PET imaging as well as ERP on the same subjects and the same task, as in the Martinez study. Here again the results for the C1 component were not significantly different according to the analysis and sample size used, but are visibly larger in the presented graphs. Thus it seems that ERP results for tasks where the location of relevent stimuli can be identified in advance of their onset, there is an existant although small effect at the earliest response latencies.

Therefore, it may be that the C1 component, with onset at 55 ms and peak at 90-92 ms and identified with the primary visual cortex (Di Russo, Martinez, Sereno, Pitzalis, & Hillyard, 2002), is modified by attention, although substantially less than the later P1 and N1 components, identified with extrastriate sources in the parietal and temporal cortices. One possibility is that initial ERP responses from primary visual cortex are largely unmodified is because attention reduces responses in surrounding regions at the same time that it enhances firing in attended regions. Many of the fMRI studies indicating retinotopic enhancement have concluded this to be the case (Tootell et al., 1998; Somers et al., 1999; Smith, Singh, & Greenlee, 2000; Di Russo et al., 2002). Vanduffel, Tootell, and Orban (2000) used a novel double labelling technique to demonstrate a relative decrease of activity in an unattended location during a spatial task relative to a feature-based task. However, the proper interpretation of this depends on an understanding of deoxyglucose labelling not currently available. In addition, single cell studies of attentional effects have consistently indicated that responses of some neurons at the attended location are suppressed by attention, although enhancement is the dominant effect. Because of the wide spatial summation by ERP recording, a simultaneous up-regulation of attended locations and downregulation of unattended locations would be impossible to detect. This interpretation of course begs the question of why extrastriate signals show a clear modulation by attention. It is possible that some areas of extrastriate cortex either do or do not produce a coherent represen-

tation based on attentional state; that the target is represented clearly in the one case, where no clear representation is formed in the other. This could explain the clear difference in extrastriate and frontal signals compared to the small effect in striate cortex. It also bears mentioning that early temporal modulation is clearly a smaller effect than feedback modulation, at least in the paradigms employed by Martinez et al (1999, 2001) and Mangun et al. (2001).

It seems odd that the visual system should not bias initial neural responses in V1 by location. Clearly neural responses at some level of the system must be biased by location, and at a fairly fine retinotopic level. It is possible that the relatively small receptive fields in V2 are the site of active biasing by location, but in a distributed system where attention is a product of communication between different functional areas, one would expect some biasing to occur at any level of the system. Martinez et al. (2001) point to the results reported by Lamme and Spekreijse (2000), showing effects of figure/ground segregation in V1, but only after around 80 ms- initial response is clearly not modulated. They hypothesize that attentional effects are similar to grouping effects in being mediated by feedback signals from extrastriate areas. It seems likely that some attentional effects are, by necessity, mediated by feedback, although this is not necessary for all attentional effects.

Most of the single cell studies showing attentional effects in V1 concur that attentional effects are primarily observed later and are the result of feedback modulation. However, these studies all employ paradigms in which the location of the behaviorally relevant stimulus cannot be predicted prior to stimulus onset. The discriminating aspects of the stimuli in these studies are not represented in V1 (dot outside of relevant figure, Vidyasagar, 1998, repeated stimulus Haenny & Schiller, 1988, and line continuity Roelfsema, Lamme, & Spekreijse, 1998.) Under these conditions, one would certainly not expect attentional modulation of V1 before some processing in higher areas occurs- the information necessary to properly modulate V1 is simply not present in the system at stimulus onset.

Two notable studies employ paradigms in which attentional effects might be expected

in area V1. Motter (1993) showed definite attentional effects in area V1. The presented plots of neural response for two V1 neurons show clear enhancement of response starting from response onset at around 50 ms post-stimulus. While timing of attentional effects is not otherwise reported, there is no reason to think that these neurons are uncharacteristic. In agreement with other studies, the effect of attention as a fraction of total response would be considerably larger after the initial stimulus-driven response. In seeming contradiction, Mehta, Ulbert, and Schroeder (2000) recently conducted a detailed study of attentional effects comparing timing and magnitude of effects across areas. The main conclusion of this study was that attentional effects are largest and earliest in higher areas (V4), with effects in earlier areas being progressively smaller and later. They do report, however, that definite attentional effects were observed as early as 35 ms post-stimulus, as early as any processing should occur in V1 (Foxe & Simpson, 2002; Di Russo et al., 2002). However, the paradigm employed was one of discriminating brightness differences in a large (10 degree) diffuse light stimulus (attended) versus discriminating changes in an auditory tone (unattended). Other results indicate that attention between modalities is neurally somewhat separate (see Rees & Lavie, 2001 for a review). In addition, evidence indicates that attentional effects are stronger when stimuli compete for the response of a particular neuron by co-existing within the same receptive field (Moran & Desimone, 1985; Motter, 1993; Treue & Maunsell, 1996; Luck et al., 1997), although still present outside of a given receptive field (Treue & Trujillo, 1999; Motter, 1993). Because there is no competition among visual stimuli whatsoever, one might predict that attentional effects based on location would be weak in this paradigm.

Mehta et al. (2000) observed no baseline firing differences as observed in many studies in which location was an important determinant of task relevant vs. task irrelevant information (Luck et al., 1997; Motter, 1993; Reynolds et al., 1999). This may indicate that strong locational attention was not necessary for successful performance of their task. While the task they used was challenging, this was primarily due to rapid presentation. Analyzing the attentional demands of the task on a level of pure utility, enhancing firing rates in V1 would be entirely

ineffective in aiding the discrimination of illumination levels in entirely suprathreshold diffuse light stimuli. The system must at some level be computing a difference between firing rates for different stimuli, and enhancing those firing rates would only be useful if the neurons involved were pushed to a more sensitive part of their response function. Such would be the case for threshold stimuli.

The above results taken as a sum seem to indicate that attentional modulation in primary visual cortex is mediated by feedback connections which largely differentially affect processing at later stages. However, in paradigms where spatial attention is necessary for task performance, and location to attend is known before stimulus onset, it seems that primary visual cortex is modulated as early as any stimulus-related activity is observed. Mehta et al. (2000) conclude that the timing and size of effects indicate that attention is a largely feedback phenomena. They do report, however, that definite attentional effects were observed as early as 35 ms post-stimulus, as early as any processing should occur in V1 (Foxe & Simpson, 2002; Di Russo et al., 2002). It seems that small early attentional effects could very well be amplified as processing proceeds to higher areas, and therefore preceeds in time. Therefore the observation of strong attentional effects in higher areas could be both an effect of strong feedback on those areas relating to task demands, and the effect of small and diffuse attentional effects refining and amplifying as they move up the chain of processing.

Appendix D

Neural Network Model 1: Speed-Accuracy Tradeoff in Parallel Location of Conjunctive Targets

This model was developed in collaboration with Randy O'Reilly; a version of this work is in press in Vision Research under the title "Serial Search from a Parallel Model".

This model explores a speed-accuracy tradeoff in the function of a parallel target location process in visual search. The time needed to perform accurate target location may be reduced dramatically if a lesser accuracy is accepted, so that several inaccurate attentional fixations are faster than a single accurate fixation.

This model does not include many of the properties of the visual system that are hypothesized to contribute to different strategies of search in the theory presented in Chapter 4. It does not include multiple layers of processing or increased RF sizes with eccentricity or complexity of representation. It does not explicitly include eye movements, although the attentional shifts discussed here can be interpreted as eye movements. It does not even explicitly model the object recognition process. Instead the model focuses on the process of target location performed by mechanisms of biased competition, and examines the speed-accuracy tradeoff and the consequences it has for strategies of search.

The FIT and GS models propose that target location in inefficient searches is the result of several serial attentional fixations. However, these results could also be the result of inefficient parallel search processes. Theories of this type are supported by a variety of evidence (Duncan & Humphreys, 1989; Chelazzi, 1999). Deco and Zihl (2001) presented a simple parallel

model that reproduced the finding of feature search times independent of number of objects in the search display, and conjunction searches times linearly dependent on number of objects. That model embodied a theory of parallel search with no serial aspects.

We constructed and further explored a computational model of this type, and discovered a relevant and probably general feature of its behavior: it worked faster if allowed to operate in a partly serial manner. We therefore offer a reinterpretation of this class of model in which it supports the Guided Search model of Wolfe and colleagues. Our interpretation supports the idea that visual search is often partly serial — a parallel process may guide attentional fixations, so that easy "pop-out" searches require only one fixation, very difficult searches may require individual inspection of each item, while intermediate difficulty searches like standard conjunction searches require only a few fixations on average. This work suggests that the degree to which search is serial varies across both task conditions, and with individual strategies.

D.1 Methods

The core of the model developed here is similar to that of Deco and Zihl (2001) in structure and basic function (figure D.1), but its performance is interpreted quite differently (see results and discussion). It includes a retinotopic feature layer, in which each unit represents a specific feature in a specific location, and location layer that represents any features at a given location. These functions match those known to exist in early ventral visual stream areas (section 3.2.1), and a complex of dorsal stream areas (section 3.3, respectively. In addition, a template layer holds on line the features of the target. This function is probably performed by prefrontal areas (section 3.3.2).

As a first step, we replicated the modeling results of Deco and Zihl (2001) using a different modeling framework. We used the Leabra modeling framework, previously used to model a wide range of psychological phenomena (O'Reilly & Munakata, 2000; O'Reilly, 1998). The Leabra framework is designed to mimic principles of cortical processing. Units are based on

the dynamics of single pyramidal neurons, and use the point neuron approximation (including ion currents and membrane potential).

The principles of the model's function can be understood in terms of spreading activation. Each trial begins with an input pattern clamped onto the input layer, and a template pattern clamped onto the PFC/template layer. Activation then spreads from these units to those they are connected to in the ventral/object layer. Those units receiving activation from both the input and template will quickly become more active.

This activity in turn spreads to the location layer, and when one location unit reaches an activity of .5, the trial is terminated. We interpret this as a commitment of spatial attention to that location. This happens only when units at one ventral/object location have become more active than those at any other competing location. The winning location is most likely to be the location containing the target, although this likelihood varies with how quickly the model is allowed to settle, as explained in the results section.

In more depth, the leabra framework functions as follows. The membrane potential V_m is updated as a function of ionic conductances g with reversal (driving) potentials E as follows:

$$\frac{dV_m(t)}{dt} = \tau \sum_c g_c(t)\overline{g_c}(E_c - V_m(t)) \tag{D.1}$$

with 3 channels (c) corresponding to: e excitatory input; l leak current; and i inhibitory input. The overall conductance is decomposed into a time-varying component $g_c(t)$ computed as a function of the dynamic state of the network, and a constant $\overline{g_c}$ that controls the relative influence of the different conductances.

The excitatory net input/conductance $g_e(t)$ or η_j is computed as the proportion of open excitatory channels as a function of sending activations times the weight values:

$$\eta_j = g_e(t) = \langle x_i w_{ij} \rangle = \frac{1}{n} \sum_i x_i w_{ij} \tag{D.2}$$

The inhibitory conductance is computed via the kWTA function described in the next section, and leak is a constant.

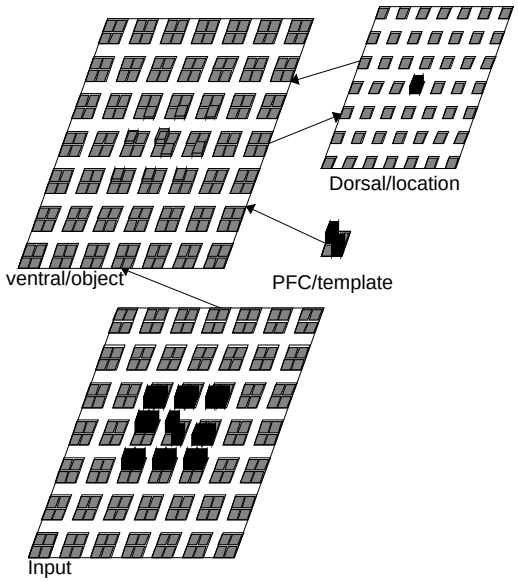

Figure D.1: Architecture of Model 1. The input layer is externally set to represent a 9-object conjunction search with the target in the center. In the input and object layers, two units of the four-unit group in each location represent different colors, while the other two represent different shapes. The four units in the PFC/template layer share this representation. The connection from input to object layer is one-to one, with uniform weights. All four units at each location in the ventral/object layer project to the one unit in the corresponding location in the dorsal/location layer, and these connections are reciprocal. Each of the four units in the PFC/Template layer connects to the one matching unit in every location in the Object/Ventral layer. The response criteria is the activation of any location unit above a threshold of .5; we interpret this response as completing the focus of spatial attention upon a certain location.

Activation communicated to other cells (y_j) is a thresholded (Θ) sigmoidal function of the membrane potential with gain parameter γ:

$$y_j(t) = \frac{1}{\left(1 + \frac{1}{\gamma[V_m(t) - \Theta]_+}\right)} \tag{D.3}$$

where $[x]_+$ is a threshold function that returns 0 if $x < 0$ and x if $X > 0$. Note that if it returns 0, we assume $y_j(t) = 0$, to avoid dividing by 0. To produce a less discontinuous deterministic function with a softer threshold, the function is convolved with a Gaussian noise kernel.

D.1.0.1 k-Winners-Take-All Inhibition

Leabra uses a kWTA function to achieve sparse distributed representations, with two different versions having different levels of flexibility around the k out of n active units constraint. Both versions compute a uniform level of inhibitory current for all units in the layer as follows:

$$g_i = g^{\Theta}_{k+1} + q(g^{\Theta}_k - g^{\Theta}_{k+1}) \tag{D.4}$$

where $0 < q < 1$ is a parameter for setting the inhibition between the upper bound of g^{Θ}_k and the lower bound of g^{Θ}_{k+1}. These boundary inhibition values are computed as a function of the level of inhibition necessary to keep a unit right at threshold:

$$g^{\Theta}_i = \frac{g^*_e \bar{g}_e (E_e - \Theta) + g_l \bar{g}_l (E_l - \Theta)}{\Theta - E_i} \tag{D.5}$$

where g^*_e is the excitatory net input.

In the **average-based** kWTA version (used for this model), g^{Θ}_k is the average g^{Θ}_i value for the top k most excited units, and g^{Θ}_{k+1} is the average of g^{Θ}_i for the remaining $n - k$ units. This version allows for more flexibility in the actual number of units active depending on the nature of the activation distribution in the layer and the value of the q parameter (which is typically between .5 and .7 depending on the level of sparseness in the layer, with a standard default value of .6). Activation dynamics similar to those produced by the kWTA function have been shown to result from simulated inhibitory interneurons that project both feedforward and feedback inhibition (O'Reilly & Munakata, 2000).

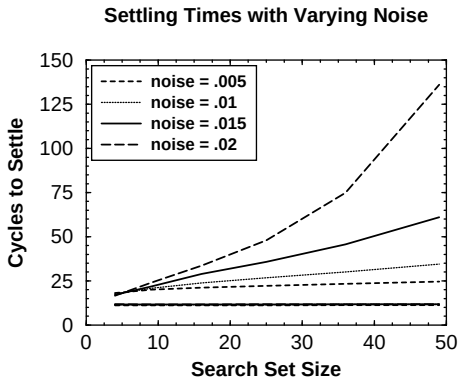

Figure D.2: Settling times for Model 1. Sloped lines are conjunction search; lower flat lines are feature searches. The amount of noise affects the settling slope for conjunction search, but does not affect the feature search settling time. We assume that human reaction times are proportional to these settling times, plus constant times for motor responses and object identification.

D.2 Results

Our model initially produced results quantitatively similar to those of the previous model, and to behavioral results. We obtained nearly flat search slopes in the feature search condition, and a linear increase in time to settle with additional distractors in conjunction search (figure D.2).

This linear increase was driven by noise: the search cost per distractor varied with the amount of gaussian noise applied to the net input current on each time step (figure D.2). According to this type of model, varying behavioral search slopes result from a larger signal/noise ratio for more easily discriminated stimuli.

The model's performance stems from the fact that only feature units that enjoy both bottom up (input) and top down (target template) inputs become active enough to influence the competition among location units. There is only one such unit in the feature search condition, the target feature in the target location. In the conjunction search condition, one target feature is present at each location, but both target features are present at the target location, allowing that location to dominate if enough evidence is accumulated to minimize the effects of noise.

Units needed two sources of input to become active because the leabra algorithm uses a

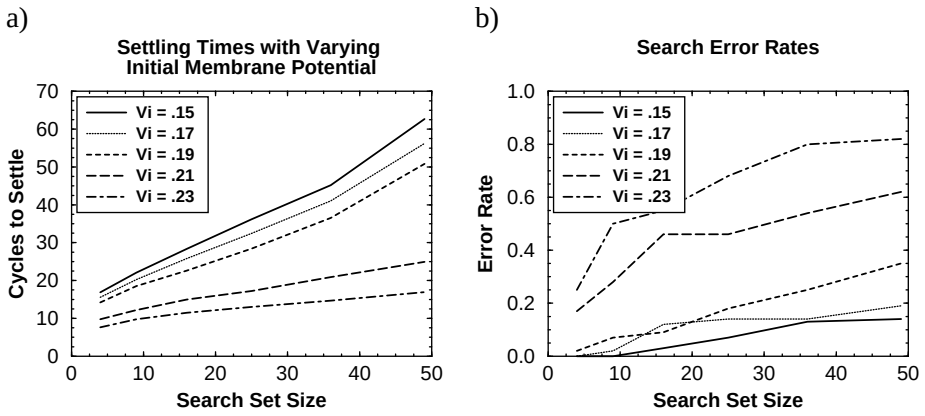

Figure D.3: Speed and accuracy of the target location process of Model 1. a) Location process times for varying starting states. Locating a potential target is dramatically speeded by larger starting membrane potentials, corresponding to a lowered threshold. b) Error rates rise rapidly as the location process becomes faster.

thresholded activation function (equation 3). Without this threshold, we would expect to see a contribution from inputs with no support from the PFC/template layer, and therefore a search cost even in the feature search condition, as is often observed experimentally. However, this cost would be very small, since only very large contributions from noise could overcome the lack of top-down support.

This model can be understood as a diffusion process model in which information is accumulated over time in a noisy environment, with more noise present for each distractor that shares a target feature. It is thus possible to speed the settling process at the cost of accuracy. Many variables could affect the system in this way. We chose to vary the starting value of the membrane potential. This has the effect of placing the system closer to settling, so that less evidence is needed to produce an attentional fixation. It also seems that this is a likely variable for online adjustment by the cognitive system; providing extra diffuse input before a trial will provide a baseline activity level, and put the system closer to its response threshold.

Raising the system's baseline activity level produced a dramatic speedup of settling, at the cost of an equally dramatic reduction in accuracy (figure D.3).

Is this reduction in accuracy disastrous for the performance of the system? It is if we

assume that every missed location is a missed trial; behavioral performance usually shows a less than 10% error rate. However, if we instead assume that the system checks the accuracy of its response with an object identification process, then chooses a new location if that object does not match the target template, then risking wrong location guesses could be a good strategy.

For simplicity we assume, rather than explicitly model, this object identification process. This identification may happen by virtue of the dorsal visual stream providing extra activation to that location in early ventral stream areas, so that higher areas respond predominantly to that information versus information from surrounding distractors. This account is in general accord with the biased competition model of Desimone and Duncan (1995), but our model does not depend on these details. We assume only that this process takes some amount of time to identify the object at the location selected by the model, gives a response if the object is the target, and triggers a new iteration of the whole process if the object is not the target.

If every missed location process results in a repeat of that process, the total search time will be given by (location time + identification time)/$(1 - P(error))$, since the series $1 + x + x^2 + x^3 + x^4$... converges to $1/(1-x)$ for $x < 1$. That series corresponds to the total number of location processes that will be completed on average when $x = P(error)$, or alternately, one plus the average number of errors per trial.

The speedup of search proved so dramatic that the system can afford one or even more missed attentional fixations, depending on assumptions about how long the identification process takes, and the signal strength and noise level. Figure D.4a gives the total search times under the assumption that an identification process takes 10 extra cycles. Even though that process is fairly costly, it can be seen that less conservative location processes are competitive with those that locate the target on the first try.

This assumption is probably still too conservative; it seems unlikely that no information is retained from the location process after the first settling process. If we assume that later location processes take 1/2 the time of the first, due to retained information, search efficiency is

biased even further toward processes that make some mistakes in the interest of a faster location process (figure D.4b). In this case, an intermediate parameter setting is the most efficient over the whole range of display sizes, while the most efficient search parameters vary with changing display size. Of course each missed location process results in a wasted object identification process, so if object identification is very slow relative to the location process, a conservative (and therefore parallel) process will be most efficient.

D.3 Discussion

We used a model in which target localization is parallel and capacity-unlimited. We replicated earlier work indicating that such a process can produce the linearly increasing reaction times with set sizes. We went on to test the speed/accuracy tradeoff within the model, and discovered that the model gained so much in speed that in some situations it was faster to obtain a correct answer by running it several times at low accuracy rather than once with high accuracy. This finding suggests that human visual search may be performed serially by default because it is faster than performing search in parallel.

This conception of search processes, based on an entirely parallel target location process, has converged with the Guided Search model (Wolfe et al., 1989; Wolfe, 1994), in which a serial search is guided by a parallel "saliency map" operation. If we assumed that the process retained all of its information instead of enough to cut settling time in half, as in figure D.4b, we would have converged closely with Wolfe's (1994) GS 2.0 model. We do not make this assumption because Wolfe's own recent work has shown that location information retention is not nearly perfect, (e.g. Horowitz & Wolfe, 2001).

This model therefore differs from Wolfe's in assuming that the time consuming parallel process must be run again for each unsuccessful attentional fixation (although some information from the previous parallel process may be retained). This follows from the following train of logic: observers generally prefer eye movements in standard conjunction tasks (Shen, Rein-

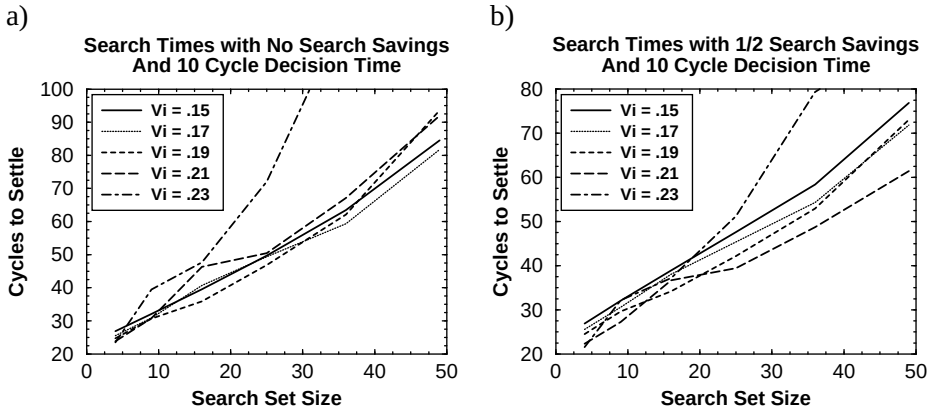

Figure D.4: a) total conjunction search times under the consideration that there is an object identification process that takes the equivalent of 10 processing cycles of the model, about the same amount of time the system takes to settle on the location of the target in a feature search. b) total conjunction search times under the same considerations for identification times, but with the more reasonable assumption that some location information is retained so that additional location processes take 1/2 the time of the first. Note that which line is lowest, and therefore the most efficient search parameters, changes between these two sets of assumptions.

gold, & Pomplun, 2003); eye movements massively disrupt representations in the early ventral stream areas; and those areas are widely identified with the feature maps that guide search (reviewed in Shipp, 2004). The implication is that the time taken by versus the accuracy of the parallel stage becomes an important tradeoff under parametric control of the observer. Thus we predict different search patterns for different strategies on the same search task, as well as among different search tasks as predicted by Guided Search.

Like the Guided Search model, this theory does not specify the conditions under which search is terminated. An effective strategy should assume that no target is present after a number of unsuccessful guesses, or after a conservative settling process does not settle in a given time. The criteria for a "no" response will vary with the internal parameters (strategy) used for the search, and the physical parameters of the search. Therefore we have dealt only with target present responses, leaving this issue to be addressed by future work.

Although we have wound up in nearly the same theoretical position as Guided Search, we have reached this position from a very different route. Guided Search assumes that a large amount of noise is inevitable in the guidance process; we have assumed that the effective amount of noise varies with the amount of time flexibly allowed to that process. Thus guidance is not inaccurate because it must be, but because it may be faster to quickly guess at and check a few locations rather than waiting for a more certain guess at the target location.

In our model, the parameters that lead to the fastest search depend on how long an identification process will take, the amount of noise in the system, and the search display size. The first two parameters can be expected to vary with the perceptual discriminability of target vs distractors, while the participants knowledge of the last can be varied experimentally. Our analysis predicts that subjects should be measurably more efficient for searches in which they know the display size before the trial.

According to this analysis, parallel neural network models lead directly to the conclusion that, under many conditions, search will have a small number of serial fixations. This conclusion

corresponds well to the finding from eye tracking experiments that participants in visual search tasks that allow eye movements show a small number of fixations in searching relatively large displays (Williams & Reingold, 2001; Brown & Gilchrist, 2000)

This type of model can potentially account for the full range of visual search efficiencies. The discriminability of targets from distractors can be modeled by changing noise amounts, by reducing the difference in input values for different stimuli, or both. Informal experimentation suggest that these changes can produce a range of search efficiencies. However, findings of slow feature search, and cases in which little or no information seems to be guiding search, require a more complex explanation. Model 2 addresses these issues in part, and future modeling work will use a more capable and realistic model of object recognition.

As a final note, this model of search has an interesting link to theories of visual search inspired by Signal Detection Theory. A major criticism of visual search theories is that they are "high threshold", that is, they do not allow for a distractor to be mis-identified as a target. High threshold theories have been convincingly rejected in the domain of simple detection (Palmer et al., 2000). The current model avoids this criticism in that the parallel stage of the model is 'low threshold'; it often misidentifies a distractor location as a target location.

However, studies of search overwhelmingly show many more misses than false alarms, implying that a low threshold model is not the only factor. We therefore theorized a second identification process, which checks the identity of the object at the selected location, and restarts search if it is not the target. In the current model this process is truly 'high threshold'; we assume it never mistakes a distractor for a target. Because false alarms certainly do occur in most search tasks, a more realistic model would include a decision process that uses a relatively high threshold for the identification process, but that does sometimes mistake a distractor for a target.

This two stage arrangement may be more efficient than simply using a high decision criteria for the parallel location process, because it directs the (likely) time consuming work of more certain identification process only to locations that are likely to contain a target. The

connection to detection theories of search is also discussed in section 4.2.1

This model has several connections to the theory laid out in the body of the dissertation. First, it serves to illustrate the principle of biased competition on which that theory is based. Second, it provides one reason that fast eye movements may be preferred over extended attentional fixations covert search: it takes less time to take fast guesses from parallel guidance processes than to wait for the system to settle into an accurate configuration. Third, it provides evidence that the speed/accuracy tradeoff hypothesized for guidance of eye movements in the theory. This hypothesis suggests that the length of saccade fixation may be useful as a variable under strategic control: long fixations may be useful to locate targets in conditions that allow relatively efficient guidance, and fast fixations may be more useful under conditions that do not allow effective guidance. This hypothesis is further discussed in section 4.2.2

Appendix E

Model 2: Tradeoff Between Broad Attention and Random Covert Spatial Attention

Model 2 explores the issue of broad attention versus randomly allocated covert spatial attention in a conjunction search. It is based on the same principles as Model 1, but has a more realistic and therefore more complex architecture.

Model 1 simulates a covert attentional search task, in which T and L stimuli surround the fixation point in a ring. This arrangement avoids the complications of varying sizes of RFs with eccentricity. To simulate this ring arrangement, the 16 locations are "wrapped around" so that the first location is considered adjacent to the last location in regard to connectivity. Like model 1, the process modeled is a parallel search guided by top-down information on the target, and by bottom up information on stimuli, as detailed in the section on parallel search (4.2.1). It is interpreted similarly to Model 1: if the model settles into an attentional fixation at a non-target location, a new fixation is calculated.

The model is thus a theory of multiple attentional fixations with no retention of information across fixations (Geisler & Chou, 1995). Total reaction time is calculated as the settling time for a single attentional fixation times the average number of fixations required to locate a target when one is present. Target absent trials are not accounted for in this model; it is assumed that the decision to terminate search and respond that no target is present is based on an estimate of how much searching is necessary to find a target when one is present under the same conditions. These stopping criteria are beyond the scope of this model.

Like Model 1, Model 2 settles on a location for attentional fixation by allowing activity generated by bottom-up input information and top down target information to meet in a competitive neural environment (section 3.3). The results of this initial competition are amplified by passing to another competitive retinotopic map, as described in the section on parallel search (4.2.1) and in shown in figure 4.3. As in Model 1, the results of this competition are interpreted as triggering an attentional shift to that location. However, in Model 2 that attentional shift is explicitly modeled, and is crucial in allowing the object identification process to correctly identify the object.

Also like Model 1, weights are pre-set rather than learned; they are set to mimic known connectivity of the visual system (some of this connectivity is detailed in sections 3.1. The basic scheme of connectivity is that there are connections between units representing corresponding locations, and between units representing the same stimuli at the same location.

The primary advantage of Model 2 over Model 1 is that it includes more realistic architecture from which crowding effects emerge, allowing for a possible advantage of attention to an area as hypothesized in section 3.2.3. This architecture is intended to capture the basic structure of the ventral visual system. Lower areas represent basic visual information about small portions of the visual scene, while higher areas represent more global information about larger areas of the visual field. This transformation occurs over many levels of visual processing; the model includes three such levels, which we labeled V1, (which is conceptualized as including V2) mid-ventral, which is identified with area V4 and area TEO, and area IT. In this model object identification happens only from area IT, and only one item can be represented by that layer. This assumption is unlikely to be accurate, since multiple objects can likely be identified in parallel, at least briefly (Rousselet et al., 2004). However, this architectural choice does not have a central role in this model; the failures in this model occur through an inability of top-down featural attention to usefully control representations in the Mid-Ventral layer when many objects are present in their receptive fields.

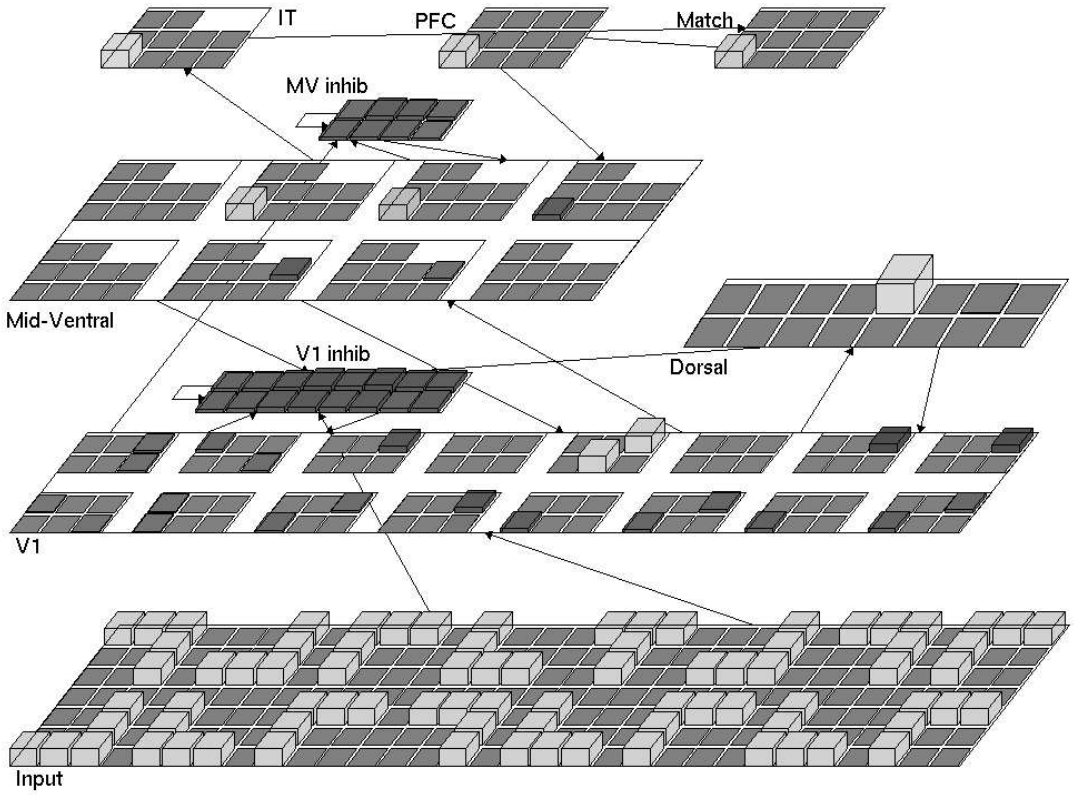

Figure E.1: Architecture of Model 2. The model is shown correctly identifying the T target in the fifth location of the top row; a single unit has become active in the Dorsal layer that represents locations. This unit activated as a result of slightly higher activation in the V1 layer, due to both top-down and bottom-up input to the units representing the target. The activation of a single unit in the Dorsal layer has further enhanced activation at the target location in the V1 layer, and this activation has suppressed activations at other locations. A clear representation thus has formed in the Mid-Ventral and IT layers, activating the Match unit, and signaling that the target is present in the display.

E.1 Details of the Model

Model 2 uses the same modeling framework detailed for Model 1; the same activation equations apply.

E.1.1 Representations and Connections

In the simplified scheme of this model, units in the V1 layer represents information about only one visual location. The six units in each location represent the three possible vertical and three possible horizontal lines. Units in the Mid-Ventral layer each represent objects in only four of sixteen locations. The ten units correspond to the four possible rotations of L objects, the four possible rotations of T objects, and two units simply represent vertical and horizontal lines. Each object unit receives projections from each of the two line units that make up that object in each of the four locations that are represented by each set of Mid-Ventral units. They receive strong projections from the central two locations, and weaker (2/3) strength projections from the two peripheral locations. This scheme is also used by the IT layer. However, this layer represents objects at all locations; it therefore receives in a 1:1 pattern from units in each of the eight Mid-Ventral groups.

E.1.2 Competitive Dynamics

Activation levels in the early and mid-ventral layers are governed by an inhibitory layer. This is in contrast to the usual use of the kWTA mechanism in Leabra simulations. The kWTA function provides a “set point” inhibitory dynamic, so that a set number (k) of units are usually activated regardless of the total level of activation from inputs. This practice is very efficient for most models, since using inhibitory units is time consuming both computationally and for ease of use. In this case it was necessary to use unit inhibition to provide an “elastic” inhibitory dynamic, so that more input results in more total activity.

The “elastic” dynamic is critical to the function of the model. In order for the early

retinotopic areas (modeled here as early and mid ventral layers) to function as the “feature maps” posited by many theories, there must be more activity in the locations which contain target features than those that do not, so that a sum of total activity in each location can cue attention correctly. These maps are incomplete feature maps in that there is no lateral mechanism to produce the similar suppression effect observed in early ventral stream areas (Appendix A above).

The IT and Dorsal layers use the standard kWTA algorithm, since it is useful for them to “sharpen” the results of competition in their input layers, producing only one active unit when the input layers have several units partially active. In this way they identify the dominant activations in the previous layers. This function is a surrogate for identification of several objects in parallel in the real function of IT cortex. In the Dorsal layer this sharpening function is a surrogate for and for the gating effects of strategic attentional control in the case of the Dorsal layer.

The competitive dynamics in this model arise from top-down featural attention projected from a maintained representation in the PFC layer, to the Mid-Ventral layer. Physiological and behavioral evidence for the effects of featural attention at this level, but not directly to lower areas, is given in section 4.2.2. The progression of a successful settling process is described in the caption of figure E.1, and are not repeated here.

In this model, attention to an area was implemented by providing a small excitatory bias to either the whole V1 layer in the broad attention condition, or to a subset of only four adjacent locations, chosen at random. This excitatory bias was enough to give a significant advantage to representations within this area, so that when a subset of locations was attended, attention would almost always settle within the biased region, resulting in the identification of one of the objects there.

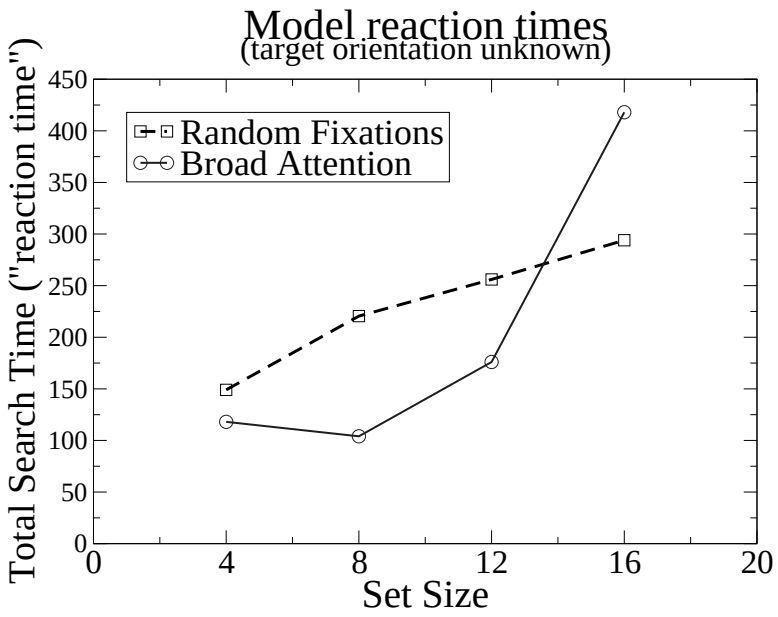

Figure E.2: Results of Model 2

E.2 Results

Results of the simulations of search for randomly rotated T among randomly rotated Ls are shown in figure E.2 The model shows the expected results of larger search sets taking longer to search. The more important result is that in the more difficult searches, the model actually performs more efficiently with an attentional bias to a quarter of the display rather than the same bias to the whole input. This is a somewhat counterintuitive result, since the attentional bias is assumed to be random, and therefore only has a chance of locating the target one time in four. It is worth noting that this result was obtained without adjusting parameters specifically for the purpose; parameters were adjusted only to obtain an intermediate level of search performance, so that the model obtained better than chance but worse than perfect search performance with a wide attentional field. The model thus shows that there is a computational advantage for such a mode of search in a system sharing some of the computational constraints of the human visual system.

E.3 Discussion

The results of Model 2 demonstrate that crowding effects can make covert search by area a more efficient strategy than using broad attention. This effect occurs when crowding passes a critical threshold. Beyond that point, top-down guidance by feature becomes insufficient to overcome the mixed representations in middle areas of the ventral system. An attentional fixation to an area overcomes this difficulty by supporting some but not all of the stimuli within each of two Mid-Ventral layer locations. This partially disambiguates each of these representations, and allows top-down featural attentional[1] to gain control of the Mid-Ventral layer representations. These representations then support the actual target location in the V1 layer. This small extra activation causes the target location to often (but not always, since noise is present) to be-

[1] This effect could equally be considered object based attention. It is specified at a level of the whole object, and affects individual features only through the reciprocal connections between object representations and the features that make up those objects.

come the focus of a finer-grained spatial attention through the dorsal layer, and with that extra support, to control representations in the upper layers, and so to control output responses in the Match layer.

As with Model 1, the importance of this model lies in the emergent nature of the results. The model was designed only to perform the target location task with less than perfect accuracy, with broad attention. The additional manipulation of a limited attentional fixation produced the observed results without any further parameter changes. The results are therefore an emergent product of the model's structure (which mimics in basic form that of the visual system), and the hypothesis of covert attention to an area. As such, this model provides some support to the hypothesis proposed in the section detailing search with covert attention (4.2.3). Specifically, it suggests that the optimum strategy for search with covert attention may differ for different set sizes of the same stimuli. Although this model is strictly of covert attention, the same principal applies to search with eye movements, and is in line with the conclusions drawn from Experiment 1.