The Dissertation Committee for Thomas Lafayette Thornton Jr.
Certifies that this is the approved version of the following dissertation:

Attentional limitation and multiple-target visual search

Committee:

David L. Gilden, Supervisor

Wilson S. Geisler

Lawrence K. Cormack

W. Todd Maddox

Randolph Blake

Attentional limitation
and multiple-target visual search

by

Thomas Lafayette Thornton Jr., B.A.

Dissertation

Presented to the Faculty of the Graduate School of

the University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

December, 2002

Table of Contents

Introduction

Questions concerning the mind's ability to simultaneously apprehend more than one object date back to Plato and Aristotle and continue to fuel large branches of cognitive inquiry. In the domain of visual psychophysics these types of questions have motivated research spanning nearly four decades. The majority of work on this issue has sought to determine how performance for simple visual discriminations deteriorates with increases in the number of items that must be attended. In the argot of the information processing age, the issue is recast in terms of measuring capacity limitation under conditions of divided attention. If increases in the number of items that must be attended to has little to no effect on discrimination performance, processing is said to be *capacity unlimited*, reflecting the finding that increases in stimulus load seem to place minimal demands on the processing of individual stimuli. On the other hand, if increases in the number of attended elements leads to deterioration of discrimination performance (e.g. discrimination takes longer or is realized with less fidelity), processing is instead said to be *capacity limited*.

In addition to measuring capacity limitation psychophysical research has sought to discriminate the *style* of processing employed in the analysis of multi-element displays. The organizing question here centers on how processing resources, independent of capacity limitation, are allocated over displays containing more than one relevant element. Specifically, the goal has been to determine if a particular discrimination is accomplished in parallel, such that multiple elements are processed simultaneously over space, or instead in a serial manner, such that discrimination proceeds on a one-at-a-time basis (Neisser, 1967). Though many experimental methods have been used in service of answering these types of questions, the *visual search* paradigm is far and away the most popular of extant techniques.

The standard visual search method consists of measuring an observer's ability to find a specified thing (the "target") when it is present among some number of distracting things ("distractors"). Typically performance is quantified by measuring either changes in time to respond, or changes in the accuracy of responding, as distractor number increases. This general class of experiment dominates much of existing attention research and has remained a favorite technique for two reasons. First, the method itself is easy to implement, and represents an improvement over early attempts to measure attentional bandwidth in which performance limitations were contaminated by memory decay (e.g. full report methods; Sperling, 1960). Second, the data structures provided by visual search are simple, and have for the most part afforded straightforward connections to existing theory.

Despite its relative simplicity and popularity, standard visual search is known to have a number of methodological flaws. These flaws include a general inability to distinguish serial processes from parallel, limited capacity processes, and the confounding of attentional effects with lower level sensory limitations (Carrasco & Frieder, 1997; Duncan & Humphreys, 1989; Geisler & Chou, 1995; Palmer, 1995; Palmer, Ames, & Lindsey, 1993; Palmer & McLean, 1995; Townsend, 1972). Moreover, the standard methodology on the whole provides no clear data-driven means for discriminating processing style at all (Wolfe, 1998a). In this dissertation I seek to circumvent these shortcomings by investigating visual search in the context of a multiple-target search method (see van der Heijden, La Heij, & Boer, 1983).

Multiple target search is designed to augment standard visual search methods by including conditions in which more than one target appears in a stimulus display. This method is superior to the standard methodology for three reasons: 1.) it has the power to distinguish serial and capacity limited, parallel processes, 2.) it attenuates low-level confounds by using a small number of stimuli presented near the fovea, and 3) it provides a rich set of response time and accuracy data that can be used to obtain graded measures of capacity limitation. Distinctions between processing style during search become clear in the context of this method because serial and parallel models make unique predictions for displays containing more than one target.

Measures of capacity limitation are obtained using a computational model of search to reproduce empirical data patterns. The key motivation for incorporating a computational model in this context is to provide a tool wherein complex constellations of response time and error data are reduced to a few simple ideas about attention and decision.

One of the primary goals of the work laid out in this paper is to investigate attentional limitations during visual search using a novel multiple target search method. A significant fraction of this investigation will consist of simulating response time and accuracy data using a psychologically motivated model of information processing. Because I am generally interested in how attention limits search performance, I have conducted a large ensemble of redundant-target search experiments in which targets and distractors are defined along a number of fundamental and representative stimulus dimensions. The model parameters that effectively simulate the average data from each of these experiments will then allow me to generate a space of varied capacity limitation in which search tasks can for the first time be quantitatively ordered along a continuum. Only by extending our conceptions of search quality beyond dichotomous notions such as *efficient/inefficient* or *serial* and *parallel* can we begin the more difficult work of understanding how it is that stimulus structure determines attentional limitation. I begin with a brief review of the standard methods used in most visual search experiments before introducing the MTS methodology that forms the core of this work.

Singleton search

In the standard version of visual search, observers are asked to make speeded decisions about the presence or absence of a single target element that may or may not be present among a variable number of distractor elements. Because only a single target is presented, we will refer to this general method as *singleton* search. In *singleton* search, the target is present on some proportion of the trials, usually 50%, and on the remaining trials only distractors are displayed. Observers make one of two responses to signal the target being either present or absent in the display. Target and distractor stimuli are often differentiated along a single visual dimension such as color, size, or orientation, or instead, in terms of conjunctions of specific values along several dimensions. The dependent variable that is most often used is response time (RT), which is measured as a function of the number of elements presented in a given display (i.e. set size). Dependent measures of accuracy have also been used in the context of briefly presented displays (Bergen & Julesz, 1983; Palmer, 1994; Shiffrin & Gardner, 1972; Verghese & Nakayama, 1994), and much of the logic and theory are similar to methods using RT.

In order to determine the style of attentional processing inherent in a specific target-distractor discrimination, RTs to respond target present and target absent are compared across variation in set size. If RT varies little with increases in set size, the search process is thought to be parallel in nature, in so far as the addition of distracting information seems to have little effect on performance. On the other hand, if there are proportional increases in RT with set size, the search is thought to be conducted in a serial fashion, reflecting the idea that search performance deteriorates systematically with set size because on average more elements must be searched through in order to find the target. Graded measurements of capacity limitation based on the data coming from singleton-search methods have for the most part been of secondary interest, with researchers instead attempting to categorize a host of target/distractor discriminations within the dichotomous structure defined by parallel or serial processing.

The singleton search method has been used primarily to provide evidence for or against the existence of specialized "feature" detectors in the visual system. The logic of this approach has been quite simple: IF a given image feature supports parallel visual search, THEN that feature is assumed to have privileged representation in the visual system (for a comprehensive review see Wolfe, 1998a). In addition to divining those image features that correspond to "features" of perception, the method has played a prominent role in the testing of various theories of search, most notably Treisman's *feature integration theory* (Treisman & Gelade, 1980), and Wolfe's theory of *Guided Search* (Wolfe, Cave, & Franzel, 1989). More recent utilizations of the method have included work on attentional control and strategy (Pashler, 1988; Theuuwes, 1992), as well as various attempts to understand how grouping interacts with the search process itself (Humphreys, Quinlan, & Riddoch, 1989; Rensink & Enns, 1995; Snowden, 1998).

Despite its lasting ubiquity, the singleton search method has not been beyond reproach. Recent work over the past decade has made it clear that a number of confounds exist which dramatically attenuate the power of the method to reveal underlying processing architecture (Carrasco & Yeshurun, 1998; Duncan & Humphreys, 1989; Geisler & Chou, 1995; Palmer, 1995). Specifically, discrimination of targets and distractors will necessarily decline with increases in set size if large displays are used. This is because larger set sizes increase the average eccentricity of search elements, and for many stimulus differences, increases in eccentricity are associated with losses of information at the sensory level. In addition, set size is often also confounded with stimulus spacing. In general, the more items in a display area, the closer they will be to one another, thereby allowing nearby stimuli to mask one another in a way that might degrade performance. Furthermore, these masking effects increase dramatically in the periphery, and so have more of an effect as larger set sizes increase the average element eccentricity. To the extent that low-level effects such as these exist in a search task, it will be impossible to clearly argue for capacity limitation or an underlying architecture. In fact, a large percentage of the variation in search efficiency across an ensemble of basic stimulus domains can be explained by differential loss of resolution off the fovea, with no cause to invoke higher level limitations on processing (see Geisler & Chou, 1995).

Though many of the confounds just mentioned can be eliminated by carefully generating stimulus displays in which eccentricity and spacing effects are controlled or counteracted (Palmer, Ames, & Lindsey, 1993; Carrasco & Yeshurun, 1998), there remains one serious methodological flaw in the singleton search method that can only be circumvented by a change in the design. Since the time of its inception, singleton visual search has been known to be incapable of making a basic distinction in processing architecture – namely, the standard method can not distinguish a serial process from a parallel, limited capacity process (Townsend, 1972). This is because at the level of mean RT, these two processing models can effectively mimic one another, both in target–present and absent RT functions, given appropriate choices of model parameters (see Townsend, 1974). The intuition here is that even when all elements are processed simultaneously and in parallel, it is still possible to produce RTs equivalent to those produced by a classic, serial search if the rate of parallel processing decreases with set size. The inverse relation between processing speed and set size is but one way to implement a limit in the capacity of a parallel process, but the point is clear – if the fidelity of parallel processing decreases with stimulus load, response times will increase, all else being equal, and the magnitude of this increase will be some function of processing fidelity and set size. In addition, it has been shown that even in the absence of processing limitations, a parallel model can still yield serial-like response time functions due to the increased likelihood of "false alarms" as distractor number increases (Palmer & McLean, 1995). That the standard, singleton search method can not discriminate these logically distinct allocation schemes renders the approach virtually useless, save its ability to make coarse distinctions in search efficiency (Wolfe, 1998a).

An alternative search method

In recent years, a number of alternative methodologies have been proposed that bypass some of the shortcomings of singleton visual search (Townsend, 1990). The multiple target search method (MTS) as pioneered by van der Heijden, La Heij, & Boer (1983) is one such example. This method is a related extension of standard singleton search that has the power to reveal spatial parallelism in the presence of inefficient processing. MTS augments the singleton search design by including trials in which multiple targets are presented. Specifically, trials are included in which a variable number of identical targets is factorially combined with a variable number of distractors. Most importantly though, MTS includes key conditions known as "pure" target trials, in which every one of *n* elements in a display is a target. These pure target trials turn out to be critical in terms of distinguishing serial and parallel, limited capacity models. The logic by which this distinction is made is relatively simple. The inclusion of pure target conditions allows for an analysis of how RT changes with increases in target redundancy. Specifically, we look for the presence of redundancy gain*s* in RT; namely, a benefit in target "present" response times the more targets there are in a pure target display. The presence of a redundancy gain in RT is diagnostic regarding search architecture because no serial process can produce such a gain. The intuition behind this assertion is as follows: 1) a serial process will terminate search with a target "present" response as soon as a target is found, 2) in pure target displays all elements are targets, and 3) the first element visited in the search will always be a target. Because of these conditions, the prediction for an idealized serial process is that RT to respond target "present" should be invariant over increases in target number. Thus, if we find evidence for a redundancy gain in the pure target RTs we can rule out serial models of search. More importantly, we have evidence in favor of spatially parallel processing, even when that evidence coexists with sharp RT costs to respond target "present" for single-target displays (i.e. the standard stimulus conditions examined in the singleton search method).

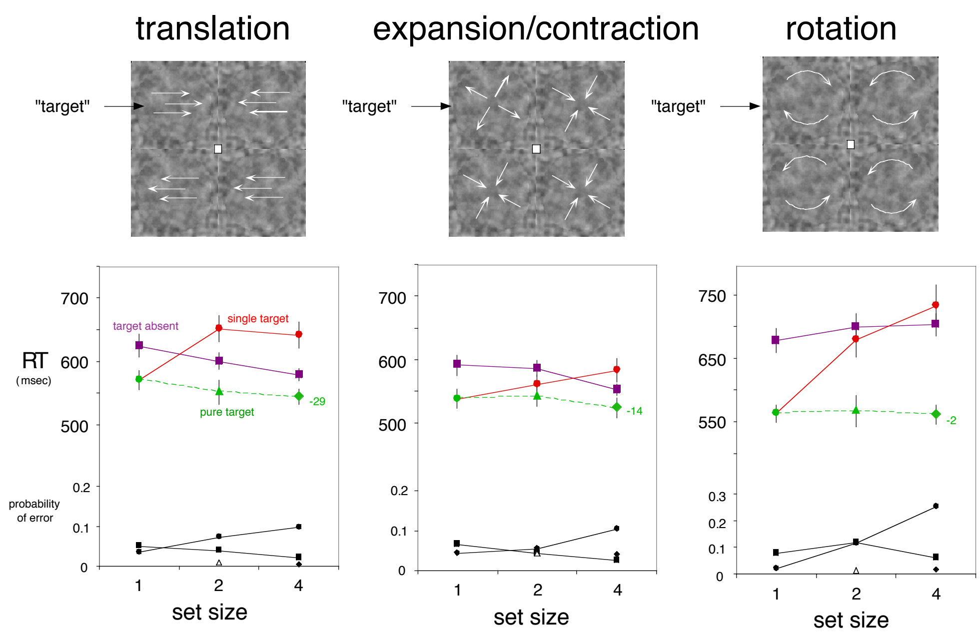
Previous work using multiple targets has been primarily interested in the mechanism(s) by which target redundancy improves performance (Diederich & Colonius, 1991; Egeth, Folk, & Mullin, 1988; Fournier & Eriksen, 1990; Miller, 1982; Mordkoff & Yantis, 1991; Schwarz, 1994; Townsend & Nozawa, 1995). Specifically, the focus has been on using estimates of single and double target RT cumulatives to distinguish whether redundancy gains arise from *statistical facilitation* (Rabb, 1962), or from *coactivation* of sensory channels (Miller, 1982, 1986, 1991). Because the basic focus has been on the etiology of the redundancy gain itself, much of the previous research has used simple target detection tasks, has fixed set size to be no more than two, and has designated targets arbitrarily (for example, a tone and a simple visual marker may both be "targets" in a single experiment). The redundant target method used here employs a set of different techniques. First, because I am primarily interested in measuring capacity limitation, I investigate discrimination performance as a function of set size. Second, the MTS methodology defines a target as a single unique thing – when displays contain more than one target they always contain multiple repetitions of this one thing (Egeth & Mordkoff, 1991; Mordkoff, Yantis, & Egeth, 1990; van der Heijden et al., 1983).

The MTS method has a number of advantages over standard search methods in addition to the inclusion of multiple target conditions. The specific method I propose is based on a relatively sparse range of set sizes: on any given trial only 1, 2, or 4 elements are presented. Though this restricted range inevitably limits generalization to larger, more dense stimulus arrays, it provides a number of advantages that give the method leverage in making claims about attentional processing. Namely, the use of a small number of elements allows for all stimuli to be presented near the fovea. In my experiments, search displays consist of elements configured along a virtual circle centered on fixation. This allows individual element eccentricity to remain constant across variation in set size, thereby reducing low-level confounds (Geisler & Chou, 1995). In addition, small set sizes attenuate masking effects by maintaining ample, and constant space between individual elements (Palmer et. al, 1993). More importantly though, MTS is an improvement over single-target methods because it provides richer data sets. By including conditions in which targets and distractors are displayed in a variety of combinations, this method simultaneously provides estimates of costs due to stimulus load, as well as benefits from target redundancy. The particular pattern of RT costs and benefits that arise in the use of this method will for the first time provide a principled means with which to distinguish *serial* processes from those that are *parallel*, but limited in capacity.

Early work

The initial use of the MTS method arose from efforts to understand how the sensing of motion sign deteriorates with increases in the number of attended motion regions (Thornton & Gilden, 2001). Gilden and I examined left-right translation, expansion/contraction, and rotary motion in separate

Figure 1



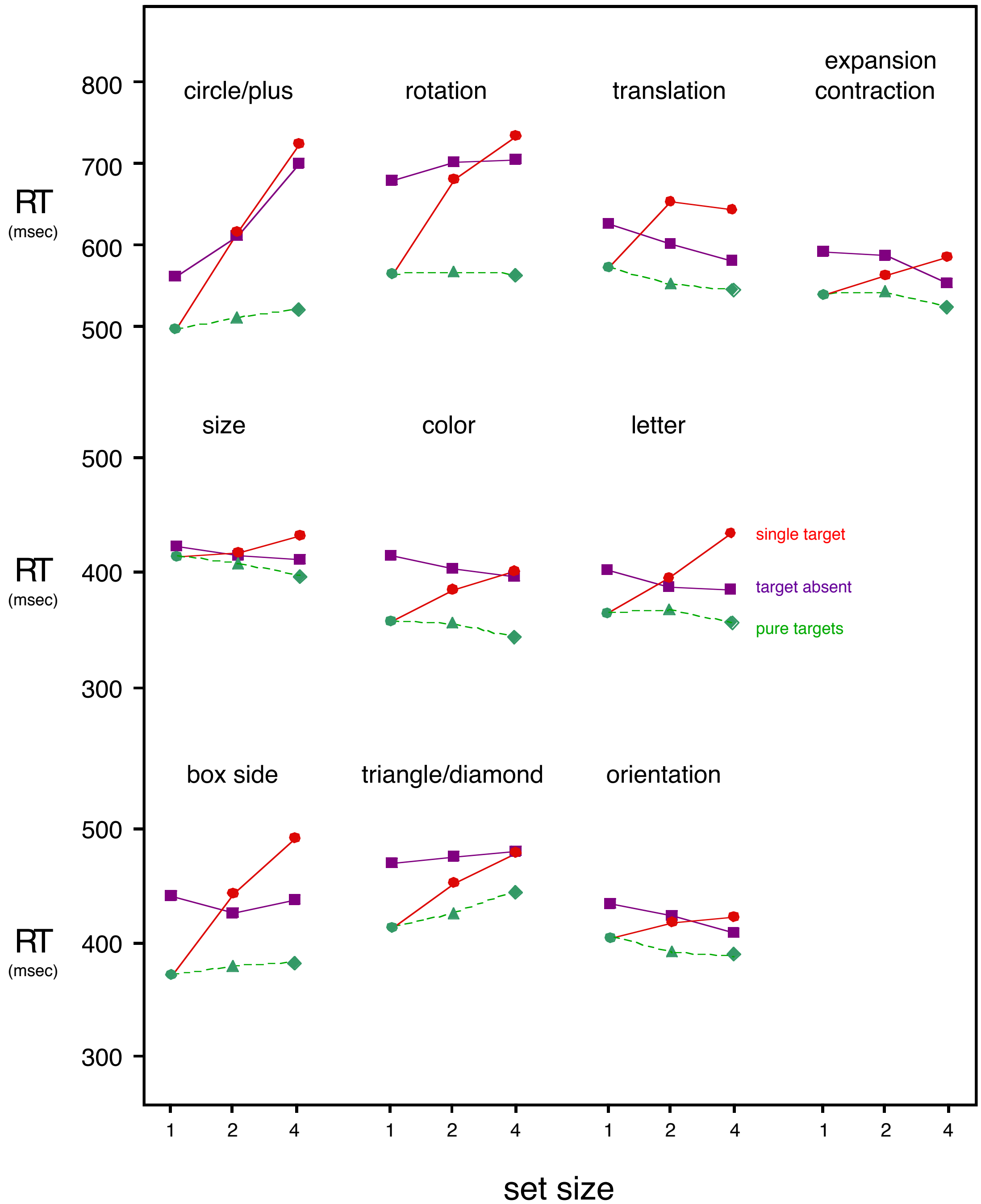
experiments by having observers search for a specific motion sign (i.e. rightward translation, expansion, clockwise rotation). One, two, or four target motion directions were presented among a variable number of distracting directions, such that the ensemble of displays contained all possible combinations of target and distractor directions. The results from these experiments are shown in figure 1.

The average median response times for correct trials only are plotted along with associated error rates. Points connected by a red line represent trials in which a single target direction was present in a display. Increases in set size for these single target trials thus represent increases in the number of presented distractors, and as such are the conditions typical to standard singleton search. The dashed, green lines represent conditions in which every element in a display moves with target sign. These conditions represent the so called *pure* target trials that give MTS the power to distinguish search architecture. Increases in set size here represent increases in the number of targets. Points connected by the purple lines are target absent trials in which displays contain only distractor motion directions – increases in set size across these lines represent increases in the number of distractor elements present in a display.

There are several important points to take away from these data. First, all three motion experiments show some degree of capacity limitation, in that response times increase for single target trials as distractor number increases. Secondly, there is notable variation in the magnitude of redundancy gain across motion type (benefits in response time for four targets relative to a single target are inset in green). For translation, there is a significant redundancy gain of nearly 30 milliseconds; for expansion/contraction, a significant gain of 15 milliseconds, and for rotation an insignificant gain of 2 milliseconds. Based on these gains we would conclude at first glance that translation and expansion/contraction direction are processed in parallel, while the processing of rotation direction is accomplished via a serial process (i.e. there are sharp costs in single target RTs with set size, and no redundancy gains). However, closer inspection of the target absent response functions makes a clear interpretation of processing style based solely on the target–present data questionable. For all three motion types, the purple target absent functions appear to "mirror" the green, pure target functions (the only real difference we see between target present and absent response times is in terms of a constant absolute offset in RT that is typical of two-choice decision). When there are decreases in RT for pure target trials with set size (translation, expansion/contraction), so too are there similar decreases for the associated target absent conditions. When pure target RTs are invariant with set size (e.g. rotation), then likewise, the slower target absent RTs are also invariant with set size. This "mirroring" between conditions is especially troubling in the case of our target absent conditions because flat or decreasing RT functions are not predicted by either serial or parallel models of search. Both standard serial and parallel models of processing predict that RT should increase with set size for target absent trials. In the case of a serial process, this prediction is rather straightforward. Target absent responses can only be made after all elements have been identified as non-targets, and thus increases in set size necessarily lead to increases in target absent RTs. In the case of a parallel process (capacity unlimited or not), the prediction is also that target absent RTs should increase with set size due purely to statistical considerations. Namely, the response to signal target "absent" can only be made once all elements have been identified, and thus the composite RT for *n* distractor elements will be limited by the slowest of the *n* processes.

Figure 2 plots the single-target, pure target, and target absent RTs for additional pilot experiments using targets and distractors drawn from a wide variety of basic stimulus domains (error rates are not shown in this figure, but are qualitatively similar to the data shown in figure 1 – the motion sign experiments in figure 1 have been re-plotted). The exact details of the various stimuli utilized in these experiments are not important; this figure is introduced here simply to highlight the differences and similarities that arise in the use of the MTS method. The pattern of mirroring between pure target and target absent response times made clear in this figure appears to be universal in the context of multiple target search, and consistently appears in all my experiments. This commonality is all the more striking given the substantial variation in absolute RT and redundancy gain magnitude seen across experiments (Thornton, & Gilden, 2000). Though there have been previous reports of target absent trials manifesting RT-by-set size slopes comparable to those found for single target trials (e.g. Enns & Rensink, 1990;

Figure 2



Pashler, 1987), there has been little to no evidence from the literature of flat or decreasing target absent RTs. Typically, shallow or flat target absent RT functions have been taken to indicate configural effects and the strong grouping of similar distractors (see Humphreys et al., 1989). Such an account runs aground in the case of composite data patterns consisting of flat target absent RTs, sharp increases in single target RTs with set size, and no redundancy gains for pure target trials (see rotation, triangle/diamond, and box-side data in Fig. 2). This specific pattern of data is in fact more consistent with *serial* than *parallel* processing. Additional investigations have revealed that the mirroring we observe remains intact despite manipulations that attenuate grouping, and is present nonetheless for stimuli in which texture segmentation experiments provide little evidence of inter-element grouping.

One possible explanation for the ubiquitous mirroring seen in these data is that observers are systematically scaling response criteria with set size in such a way so as to reduce RT cost (Zenger & Fahle, 1997). Any adjustment of criteria that reduces response time will necessarily increase error rate. Examination of the error rate patterns in figure 1 in this context provides additional evidence for the idea that observers may be able to reduce target absent RTs by trading error. For all motion experiments (see Fig. 1), as well as for the majority of experiments shown in figure 2, there is a general trend for single target misses to increase with set size, while false-alarms hold constant or decrease with set size. This pattern of error is especially evident in the case of search for rotation direction, where misses for a single target among three distractors approaches 25%. Note that a bias to say target “absent” that increases with set size will simultaneously reduce the effect of set size on target absent RTs, and increase the rate of misses. This particular pattern of errors, in which miss-rates rise far more sharply with set size than false-alarms is not peculiar to multiple target search, but appears generally throughout the use of supra-threshold search methodologies (e.g. Rensink & Enns, 1995), and may reflect the adoption of a near-optimal strategy (see Zenger & Fahle, 1997 for a treatment of error patterns in singleton search).

Clearly, there will be trading of speed for accuracy in any decision task in which both RT and error are free to vary. In the typical search experiment though, because the average error is constrained to be no more than 10%, considerations of speed/accuracy tradeoffs arise only in so far as they are necessary to insure that target present RT patterns are not simply accounted for by associated changes in target miss-rates. The problem that emerges in considering a speed/accuracy tradeoff account of these target absent data is that there is absolutely no reason to limit such trading to specific conditions – once this Pandora's box has been opened, any clear interpretation of RT data is going to have to also consider the error data. The problem of understanding how trading is expressed across conditions is made all the worse in the case of multiple target search simply because there are so many additional stimulus conditions to consider (9 RT conditions and the 9 associated error rates).

One general approach that has been taken in this context has been to form speed/accuracy trading functions by measuring both aspects of performance under varying degrees of incomplete processing (Meyer, Irwin, Osman, & Kounios, 1988; Pachella, 1974; Wickelgren, 1977). While this approach remains a possibility, it involves estimation of separate trading functions for each condition of interest thereby limiting its feasibility to simple experimental designs. Other approaches applicable to visual search have included methods in which accuracy levels are fixed and RT is not measured (e.g. Palmer & McLean, 1995), or methods in which stimulus presentations are kept brief, and accuracy is instead the dependent measure (e.g. *threshold search methods*; Palmer et. al, 1993). What these approaches have in common is that they either constrain or ignore variation in one performance variable in order to gain clarity in the interpretation of the other. The approach I have decided to take does not constrain RT or error, but rather makes their interaction a point of inquiry.

In order to understand how the specific patterns of RT and error that arise during multiple target search are linked to limitations in capacity, I create a process-based model of visual search. This model jointly simulates response times and error rates, and has a number of "tunable" parameters of key psychological import. The model is a variant of a random walk model that has been modified for the search domain (Link, 1975; Ratcliff, 1978). Specifically, I parametrically manipulate sensory, attentional, and decisional variables in the context of a multiple random walk model (for related models see Palmer & McLean, 1995; Ward & McClelland, 1989). This approach is relatively novel in terms of visual search models in that the model can simultaneously account for both RT and error patterns. By utilizing the entire pattern of RTs and error rates to constrain the model, performance variables are integrated and thus provide the necessary power to reveal graded levels of capacity limitation not realizable by techniques that focus solely on RT or error. It is important to emphasize here that the use of a model in this context is not an *end* that provides an exhaustive and veridical account of the visual search process itself, but is rather a *means* through which the complex constellation of RT and error patterns seen in the data are ultimately reduced to a few simple ideas about the stimulus, attentional limitation, and decision.

Overview of the dissertation

There are several primary goals that frame this dissertation. First, I seek to implement a multiple target search method that is a relatively straightforward extension of traditional search methods. The method proposed here represents an improvement over existing methods in that it bypasses the major shortcomings inherent in singleton search, and provides measures of costs under divided attention as well as benefits from target redundancy using a relatively small number of stimuli. Second, the method will be used to characterize the processing limitations that exist across a large ensemble of fundamental stimulus discriminations. To do so, I introduce a model of MTS that effectively integrates RT and error data. The model is simply a tool that simulates multiple-target search data, and the values of its parameters provide insight into how much decisional flexibility and capacity limitation are necessary to account for specific patterns of RT and error. With this goal in mind, I will examine a number of alternative ways of instantiating capacity limitations during search in an attempt to uncover the etiology of the pattern of *decrements* and *gains* in RT and accuracy that occur during the unconstrained MTS task. Specifically, I will attempt to account for the search data using two class of models: the *D-type* model, in which attention multiplicatively attenuates perceptual evidence prior to decision [this model can be shown (see II.A.1.f) to be formally equivalent to an account of search in which there are no capacity limitations and the only source of variation in response variables across set size arises from statistical and decisional limitation]; and the *A-type* model, which has an additional source of internal noise that creates limitations above and beyond those that characterize the *D-type* model.

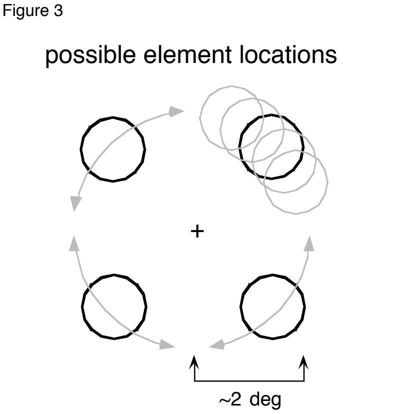
This dissertation is organized around two major phases of work. In the initial phase I present the MTS methodology in detail, along with results from 26 search experiments. This experimental ensemble was designed to capture the range of stimulus contrasts typically used in the study of visual search. The ensemble is defined by four distinct classes of search experiments in which targets and distractors differ: 1) along simple *featural* dimensions, 2) in terms of *emergent form* or *shape*, 3) in terms of *relative position*, or 4) in terms of *motion* direction.

In phase II, I introduce and implement two general models of MTS so as to extract measurements of capacity limitation from response time and accuracy data. The various parameters that give the model flexibility are motivated and discussed in terms of their psychological relevance, and the general protocol for finding the parameter values that best account for particular patterns of RT and error-rate is explained in detail. I then present the parameters of the model that best simulate the data from each of the MTS experiments, and compare how individual search tasks are distinguished in terms of discriminability, decisional flexibility, capacity limitation, and internal noise.

I. Multiple target search

The 26 experiments detailed in this section represent the bulk of my work using the multiple target search method. As mentioned in the introduction, some of the pilot experiments were part of a related line of inquiry that sought to determine attentional limitations in the sensing of motion direction (Thornton & Gilden, 2001). The remaining set of experiments were chosen on the basis of two major criteria. First, I sought a set of experiments that would allow me to calibrate the multiple target method by introducing target/distractor (*t/d*) discriminations presumed to vary in the demands they place on attentional resources. Second, I wanted a stimulus set that approximated much of the variety of *t/d* differences seen in the literature. The goal here was simply to form a data base that would sufficiently capture the range and variety of RT and error patterns manifest under multiple target search. In addition, various experiments were chosen to examine how variations in discriminability, *distractor* heterogeneity, and element assignment manifest themselves in the MTS data, and subsequently how they get reflected in the model parameters.

Figure 3



For all multiple target search experiments, displays contained either one, two, or four elements. Individual elements were configured about a central fixation point along a virtual circle whose radius varied from 1.5 to 2.5 degrees of visual angle. Elements were drawn at canonical locations along the virtual circle (+45, -45, +135, -135 deg), and for most experiments the entire display was randomly rotated about fixation to remove configural effects by choosing a uniform deviate from the interval +/– 25 deg. For some of the experiments additional radial jitter (~.5 deg) was added individually to each element to remove effects due to element collinearity. A schematic of the general protocol for display generation is shown in figure 3.

Table 1

set size	# of targets: 0	1	2	4
1	$\frac{1}{12}$	$\frac{1}{12}$	ø	ø
2	$\frac{2}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	ø
4	$\frac{3}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$

ø : not applicable

All search elements were suprathreshold, in general of high contrast, and varied in size from .5 to 3 degrees of visual angle at a viewing distance of 57.3 cm. Individual trials consisted of displays containing one, two, or four elements, each of which could be a target or distractor. In all, there were nine basic types of stimulus displays; displays containing three targets were excluded from this design. Displays consisted of either all distractors (target absent trials), all targets (*pure*, target present trials), or some combination of a variable number of targets and distractors (mixed, target present trials). Table I shows the relative probabilities of encountering each type of stimulus display. This particular design matrix is necessary to insure that across set size the probability of encountering a target present display is equal to the probability of encountering a target absent display.

Nine different observers participated in each motion sign experiment (nine additional observers served in experiments 3, 16, and 17 for a total of 18 observers), and eight to nine observers participated in each of the remaining experiments. For all experiments, stimulus displays were preceded by a brief fixation interval (~500 msec), and were present until response.

I.A. Stimuli

In the following section I present the methodological details of the individual experiments that form the MTS ensemble. In all, the ensemble contained 26 experiments, each of which could be broadly categorized into one of five basic stimulus regimes: *feature* search, *form* search, search for *motion-sign*, search with *heterogeneous distractors*, or *relative position* search. For reference, each experiment description is accompanied by a typical example of a set size 4 search display (the element assigned as "target" consistently appears in the upper left quadrant of each figure).

I.A.1. *Search for simple "featural" differences*

The following 5 experiments consist of searches based on targets and distractors that differ along a single *feature* dimension. This class of experiment should make near minimal demands on attention as these stimulus dimensions 1) are known to have privileged representation in the early visual system, 2) generally lead to highly efficient patterns of spatially parallel search, and 3) support vivid texture segmentation. In the context of MTS, these types of search provide a strong benchmark for calibrating the general method. Accordingly, the expected pattern of data for this class of experiment is characterized by large redundancy gains in pure target RT, shallow set size effects for displays containing mixtures of targets and distractors, and relatively low error rates. Experiments 1 and 2 examine search for targets that differ from distractors in terms of *size* and *spatial frequency*. Experiments 3 and 4 examine search for targets that differ from distractors solely in terms of *orientation*. Experiment 5 examines search based on *color* differences.

Experiment 1: size (target=big)

Target and distractor stimuli consisted of Gabor-type elements that were fixed at an orientation of -45 deg, and had effective bounding envelopes that were inversely proportional to spatial frequency. Target elements had a spatial frequency of 2 cpd and an envelope subtending approximately 1.5

deg; distractor elements had a spatial frequency of 4 cpd and an envelope of approximately .75 deg. Thus, size and spatial frequency were purposely confounded. For both the size (exps. 1 & 2), and the orientation (exps. 3 & 4) experiments below, Gabor stimuli were presented against a background of low contrast, static Gaussian noise (the purpose here was primarily aesthetic–the noise background effectively masked edge artifacts due to the thresholding of individual Gabor elements).

Experiments 1 & 2
Spatial frequency / size

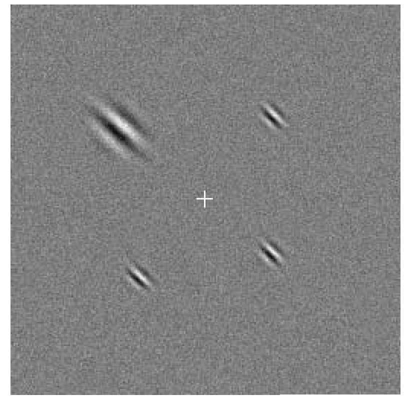


Figure 4

Experiment 2: size (target=small)

The stimuli for this experiment were identical to those in experiment 1 except that here the roles of target and distractor were interchanged. Together experiments 1 and 2 probe performance asymmetries during multiple target search. A number of previous studies support the notion that search for *big* among *little* is accomplished in a qualitatively distinct, and more efficient manner than search for *little* among *big* (e.g. see Treisman & Gormican, 1988). Here we test whether this advantage extends to a MTS design that minimizes the eccentricity, density, and target uncertainty effects known to plague singleton search.

Experiment 3: orientation (Δθ=45˚)

The third "feature" search experiment consists of Gabor-type stimuli of fixed size and spatial frequency (i.e. 2 cpd, 1.5 deg envelope). These were used to create targets and distractors that differed solely in terms of element orientation. Targets were oriented at -45 deg and distractors at 0 deg (i.e. vertical). The *Δ*θ of 45 degrees was chosen to produce a highly discriminable orientation gradient between targets and distractors.

Experiment 3
Orientation *(Δθ=45˚)*

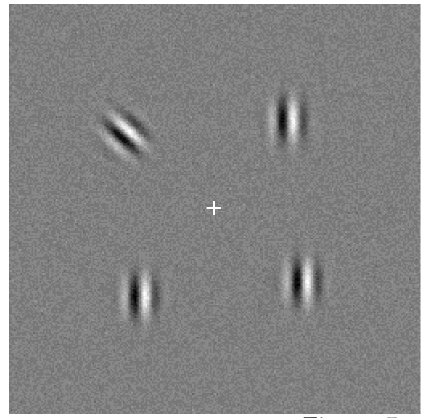


Figure 5

Experiment 4: orientation(Δθ=15 ˚)

The stimuli for this experiment were similar to those in experiment 3, except that here the orientation gradient (*Δ*θ) was reduced by 1/3 by giving the distractor Gabors a base orientation of 30 degrees; the targets remained oriented at 45 degrees. The orientation gradient was reduced to examine how decreases in target/distractor discriminability are manifest in MTS. More specifically, experiments 3 and 4 will provide a partial test of the psychological reality and consistency of the model parameters associated

Experiment 4
Orientation *(Δθ=15˚)*

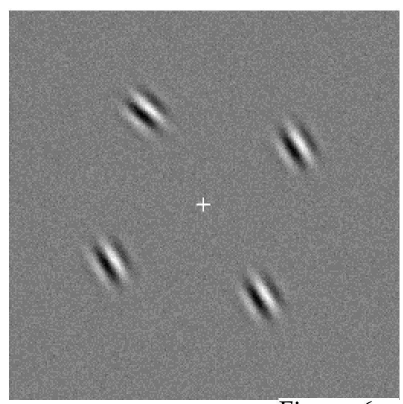


Figure 6

with element discriminability (section II.A.1.b). The question here is whether these parameters will pick up an explicit manipulation of similarity. The 15 degree gradient seemed a natural choice given previous evidence that it borders a transition point between *efficient* and *inefficient* search slopes (Wolfe, 1998b).

Experiment 5: color

Experiment 5
Color

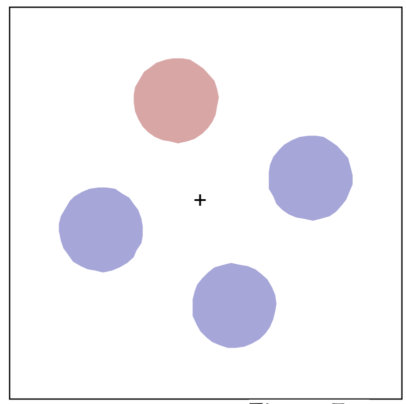


Figure 7

The final feature search experiment consists of search for *reddish* target disks (radius of ~1.33 deg), among *blueish* distractor disks. The actual target and distractor disks were created to be more similar than as depicted in the adjacent figure (weights assigned to each monitor gun were as follows – red: [.75 .65 .65]; blue: [.65 .65 .75]). The elements were made more similar here because in earlier pilot work, highly saturated colors were found to produce extremely fast RTs in which most of the data were compressed near the floor. By reducing the color contrast of the elements it should be possible to move the RTs off the floor, therein revealing the magnitude of both the redundancy gains, and the effect of distractor number.

I.A.2. *Shape and search for "emergent" form*

The 7 experiments in this section consist of search displays in which targets and distractors are distinguished either in terms of *shape* (exps. 8, 9, 10, and 11), or in terms of an "emergent" *scene-based* property (exps. 6, 7, & 12). By *scene-based*, I mean a stimulus property that is not present per se in the stimulus, but is *constructed* using assumptions about either grouping, lighting, or 3D projection and layout. The subset of experiments employed here are particularly interesting because they typically lead to efficient patterns of search in the singleton method, even though in all cases targets and distractors are physically distinguished only in terms of the relative arrangement of parts. Because such findings run counter to results for typical relative position searches, they provide support for the notion that these sorts of *scene-based* representations are available early. Unfortunately, all the previous work has been tied to singleton search, a method whose only real diagnostic has been the single-target slope. As I have mentioned repeatedly throughout the introduction, interpreting processing style and/or the magnitude of capacity limitation using this measure is problematic at best. Here I reexamine the nature of search for emergent form using the MTS method. If search for these types of stimulus differences is truly parallel and efficient, we should expect to see joint patterns of redundancy gains in the pure targets and shallow set size effects for mixed element displays.

Similarly, the model-based estimates of attentional limitation should be on par with those for paradigmatic "popout" tasks (e.g. experiments 1 – 5).

The remaining subset of experiments examine the role of *shape* in search. In addition to rounding out the MTS ensemble, these experiments are of special interest given the current debate as to the nature of *shape* processing. It remains an open question as to whether or not *shape* information is available early and in parallel during search (see Wolfe, 1998b). Experiments 8 and 9 test this question using novel target shapes with random contours that can be parameterically manipulated to be more or less similar to circular distractor shapes. Experiments 10 and 11 examine search for familiar letters that are distinguished by shape and other uncontrolled for stimulus differences (e.g. orientation components, curvature, etc.). These experiments are also interesting because it is not entirely clear how a confluence of stimulus attributes of the sort that distinguish one letter from another will affect the search process.

Experiment 6
3D cubes

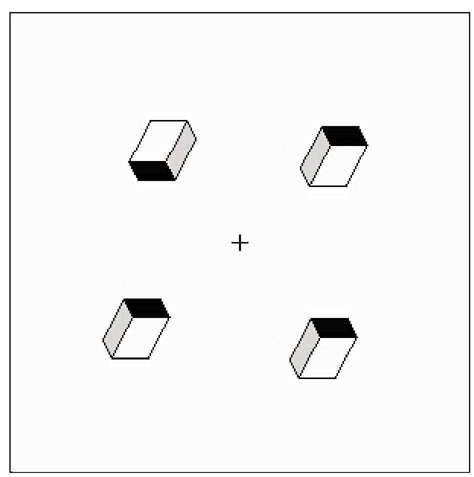


Figure 8

Experiment 6: 3D cubes

In this experiment I examine search for simple line drawings that support the perception of "cubes" oriented in depth. Though target stimuli consist of the same component pieces as distractor stimuli, they differ from distractors in terms of their implied orientation in depth. These stimuli are similar to the targets and distractors used by Enns and Rensink (1990). In a number of studies these stimuli were found to yield relatively shallow RT-by-set size slopes. Appropriate coloring has been added to the faces of the target and distractor "cubes" so as to further enhance the perception of 3-dimensionality.

Experiment 7: closure

This experiment uses search stimuli similar to those originally investigated by Pomerantz and Pristach (1989). These sorts of stimulus difference are appropriate in the present context because the two arrangements of curved lines lead to distinctions in emergent form (the arrangements produce either an "oval"-like shape or an "hourglass"-like shape). Target and distractor stimuli were constructed so as to minimize any overall difference in visual extent. Again, though the relation between targets and distractors make this search formally equivalent to a relative position search, we may find moderate to low levels of

Experiment 7
Closure

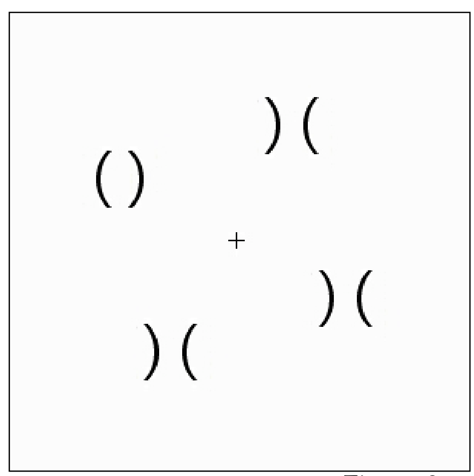


Figure 9

capacity limitation if search processes can capitalize on the emergent distinctions in form. These types of stimulus differences have previously produced shallow RT slopes in the context of singleton search.

This type of search also has import for my ongoing investigations into search for motion sign. Specifically, these stimuli share a sort of formal structure with search for the sign of *expansion/contraction*. To see this, note that the sign of a complex motion (e.g. rotation, expansion/contraction) can be simplified as a particular arrangement of at least 2 motion vectors. Thus, expanding motion can be represented by a vector of *leftward* flow organized to the left of a vector having *rightward* flow. In this scheme, contracting motion can be similarly represented by simultaneously flipping both flow vectors about an axis orthogonal to their flow directions (i.e. simultaneously changing the sign of each component, but preserving their organization). This type of symmetry relation between expanding and contracting flow is equivalent to the representations of target and distractor stimuli in the static search presented here. It is generally true, at least for static searches, that when targets and distractors are 1) composed of parts that have directionality (e.g. arrows), and 2) share a symmetry relation of the sort described above, they will also differ in emergent form. One question then, is whether the previous advantage that was found for searches based on the sign of expansion/contraction over search for rotation sign (Thornton & Gilden,

2001) was due to some emergent property associated with the inherent symmetry relation between stimuli. If this relation is what makes expansion/contraction look more like translation, than we should expect experiment 7 to also produce data with strong redundancy gains and shallow set size effects.

Experiment 8: random shape (high Δ)

Experiment 8
Random shape *(high Δ)*

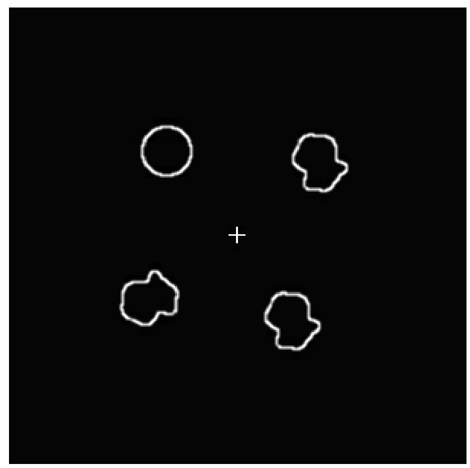


Figure 10

The following two experiments examine search when targets differ from distractors solely in terms of shape. For these experiments I have created a base "distractor" shape by forming a closed random contour. The random contour is a sum of a set of 8 sinusoids whose frequencies are harmonics of the contour length (the sinusoids have randomly chosen phases and power that falls with frequency as $1/f^2$). The amplitude of the contour is varied to create shapes that increasingly deviate from circularity. In the adjacent figure I show a set size 4 display containing the random distractor shape (~1 deg) and the circular target used in experiment 8. For experiment 8 the distractors were created to be highly discriminable from the circular target, and were randomly oriented within and across trials to remove fixed orientation cues.

Experiment 9: random shape (low Δ)

The primary advantage of using random-contour distractors is that the amplitude of shape non-circularity can be easily manipulated. In this experiment, I use the same base distractor contour as in experiment 8, but here I have reduced the amplitude of deviation from circularity by a factor of 2. This was done to provide a comparison shape experiment in which only target/distractor discriminability differed. In this way, experiments 8 and 9 are related to the previous experiments (3 & 4) examining discriminability in the context of orientation gradients. Here I explore how explicit manipulation of the shape gradient gets reflected in the patterns of MTS data, and subsequently, in the relevant model parameters.

Experiment 9
Random shape *(low Δ)*

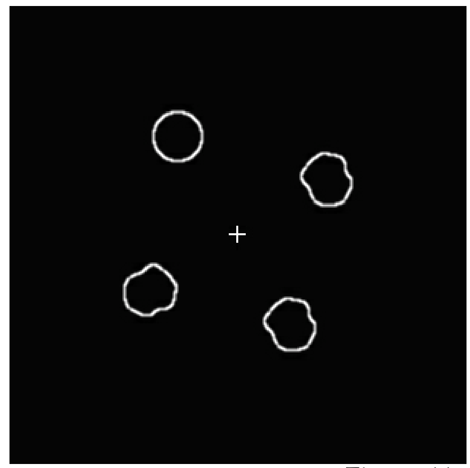


Figure 11

Experiments 10 & 11: letter search (static, dynamic)

The two following form experiments consist of search for the letter "A" among "B" distractors. I chose these letter stimuli because they differed in

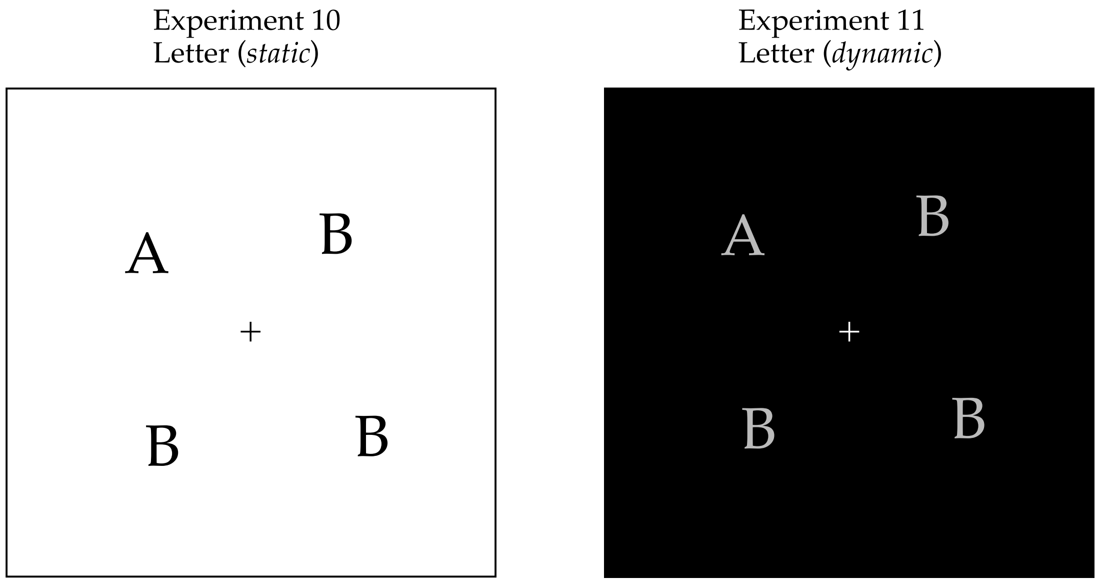


Figure 12

terms of line orientation and curvature, and were thus easily discriminable. In the static version of the experiment, high contrast, black letters (~1.15 deg) were presented against a white background. In the dynamic version of letter search, letter stimuli (~1.3 deg) were ramped up in luminance from black to white against a black background (the cycle from minimum to maximum luminance took 1 sec). Dynamic letter search was included in the experiment ensemble because I sought to investigate how set size effects and redundancy gains are affected when information about stimulus identity is protracted in time.

Experiment 12: implied lighting (up/down)

In the adjacent figure I show the target and distractor stimuli that form the final investigation into how emergent properties affect MTS. These stimuli are shaded so as to appear phenomenally as “bumps” and “pockets” within a gray, background surface. The perception of surface curvature that attends these stimuli is thought to arise from the interaction between stimulus shading and intrinsic assumptions about an overhead lighting source (Ramachandran, 1988; Sun & Perona, 1996). Thus, by changing the sign of shading from light-on-top to dark-on-top I can easily effect changes in perceived surface curvature.

Experiment 12
Implied lighting (*up/down*)

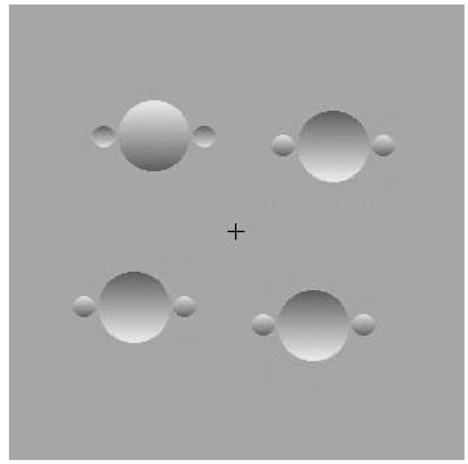


Figure 13

For this experiment targets were elements with large central “bumps”, and distractors were elements with large central “pockets”. I have added small, reverse-shaded inducers to the sides of the larger “bumps” and “pockets” so as to enhance the perceptions of curvature when displays contain only a single element. Note that these search stimuli, despite a striking distinction in 3D curvature, nonetheless are related in the image by a simple symmetry flip. At present it is an open question as to whether this sort of stimulus difference can be used efficiently in search. Interestingly, the best evidence in support of the notion that shape and reflectance are computed early in visual processing has come from texture segmentation demonstrations which reveal strong region formation on the basis of these sorts of stimulus differences. However, it is true that segmentation and search methods do not always agree (Wolfe, 1992), and it remains to be seen how redundant sources of curvature sign are used in search.

I.A.3. *Search for motion-sign*

The following 7 experiments investigate the acquisition of motion-sign during multiple target search. Experiments 13, 16, 17, & 18 were run as part of previously published study that formed the beginnings of my investigations into MTS (Thornton & Gilden, 2001). These experiments are included in the ensemble because they extend the method to the motion domain, and in addition provide a heterogeneous set of data characterized by capacity limitation and redundancy gain magnitudes that differ substantially across motion type. The remaining subset of experiments introduced here represents a replication of a study (exp. 14), an extension using more naturalistic objects (exp. 19), and part of an explicit manipulation of motion-sign discriminability (exp. 15).

Experiment 13
Translation I

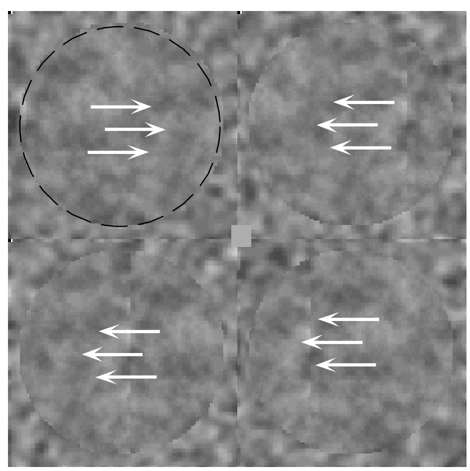


Figure 14

The majority of the motion experiments (exps. 13-17) consisted of animated textures moving behind apertures (~3 deg). Individual animations consisted of sequences of random texture that were constructed to have the same power spectra as natural landscapes (power $\propto 1/f^2$). These animations supported the percept of continuous and coherent motion, and were presented for one second against a matched background of uncorrelated dynamic noise. In experiments 13, 16, & 17, empty apertures (e.g. set sizes of 1 or 2) were replaced with uncorrelated dynamic noise. In the remaining experiments, the moving elements were presented either against a gray background (exps. 14, 15), or against a static noise field (exps. 18, 19). Stimuli were generally equated across motion type for spatial extent, contrast, spatial frequency content, and perceived speed.

Experiments 14 & 15: translationII (16 frames, 3 frames)

The following two experiments served as replications of experiment 13, which formed part of an earlier investigation. The stimuli used here were again moving textures matched in speed and size to those in the previous experiment. However, these stimuli

Experiments 14 & 15
Translation II

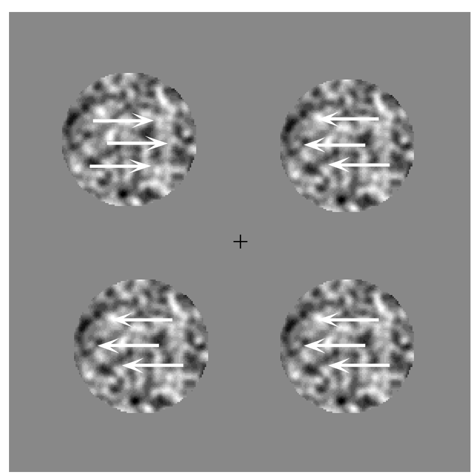


Figure 15

differed from those in experiment 13 in that the moving textures were scaled to have higher contrast, and the motion elements were presented against a gray background without dynamic noise.

Experiment 14 consisted of a 16-frame motion sequence (total duration was ~800 msec), thus providing a replication of experiment 13. Experiment 15 was identical to experiment 14 in every respect, except that the motion displays were limited to 3 frames of motion (total duration was ~150 msec). Duration was shortened so as to provide a simple analogue in the motion-sign domain of the previous orientation and shape experiments exploring discriminability effects in MTS. Presumably, decreasing the overall motion duration by a factor of 16/3 should lead to targets that are more similar to distractors. Again, the question here concerns how such changes in discriminability will manifest themselves in MTS, and more importantly, whether the MTS models are correctly specified so as to reveal these differences in the right parameters.

Experiment 16: expansion/contraction

In the case of expansion/contraction, I ran counterbalanced versions of the experiment in which 9 separate observers searched for either expansions among contracting distractors, or contractions among expanding distractors. I decided to use both directions as target because previous psychophysical

Experiment 16
expansion/contraction

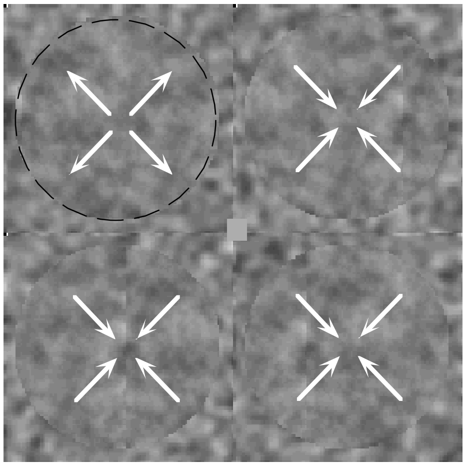


Figure 16

work suggested some form of anisotropy, though the sign of the asymmetry differed across studies (Ball & Sekuler, 1980; Edwards & Badcock, 1993; Graziano, Anderson, & Snowden, 1994; Harris, Morgan, & Still, 1981; Reinhardt-Rutland, 1994; Takeuchi, 1997). Previous results with MTS revealed little differences in search performance as a function of target choice – and so the data I will analyze here will generally reflect an average over choice of target. However, later in section II.C.7 I will use the MTS models to explicitly look for any performance asymmetry. The models are extremely useful in this regard because they permit an integration of response variables that is not possible by other means. Thus, it is possible that a true asymmetry in parameter estimates could exist for the different target assignments, even in the absence of clear differences in either the pattern of RT or error rates.

Experiment 17: rotation 2D (textures)

Like many of the previously described experiments, this experiment was part of a previous investigation into the nature of attentional limitations during search for motion sign (Thornton & Gilden, 2001). Experiment 17 employs rotating textures matched to those described in experiment 13 and

14 (visual extent was 3°; 16 frame sequences lasting ~800 msec). Eighteen observers searched for clockwise rotating targets among counterclockwise rotating distractors. Nine observers participated in the initial study, that was then replicated with a set of nine additional observers. The data reported here is based on the average of these two data sets.

Experiment 17
Rotation 2D (*textures*)

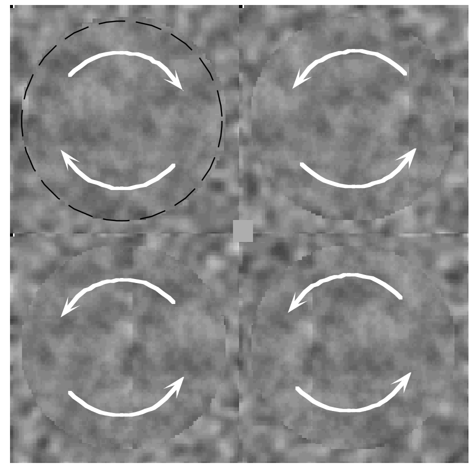


Figure 17

Experiment 18: rotation 2D (pinwheels)

This experiment served as a control experiment that ruled out the possibility that translation and expansion/contraction sign were superior to rotation in MTS because only these motions had accretion and deletion of texture at the aperture border. In this experiment stimuli were black "pinwheels" that rotated against a static, Gaussian noise background. This type of rotating stimulus leads to constant accretion and deletion of texture as the solid wings of the pinwheels move

Experiment 18
Rotation 2D (*pinwheels*)

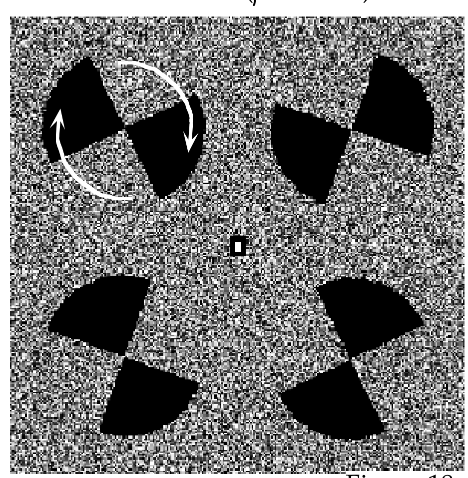


Figure 18

across the fixed background. Results using these sorts of rotary stimuli replicated an earlier study using aperture-bounded, rotating textures (see experiment 17 above), such that there were no redundancy gains, and large set size effects in the mixed display RT and error rates. If anything, rotating pinwheels appear to produce a more capacity-limited search in that pure target RTs are seen to reliably increase with set size.

Experiment 19: rotation 3D (coins)

Experiment 19 represents an extension of experiments 17 and 18 – these experiments examined search for rotation sign when the rotations were confined to the fronto-parallel picture plane. Here I ask whether the high levels of capacity limitation found previously for planar rotation also exists for more realistic, 3D objects that rotate in and out of the picture plane.

Experiment 19
Rotation 3D (*coins*)

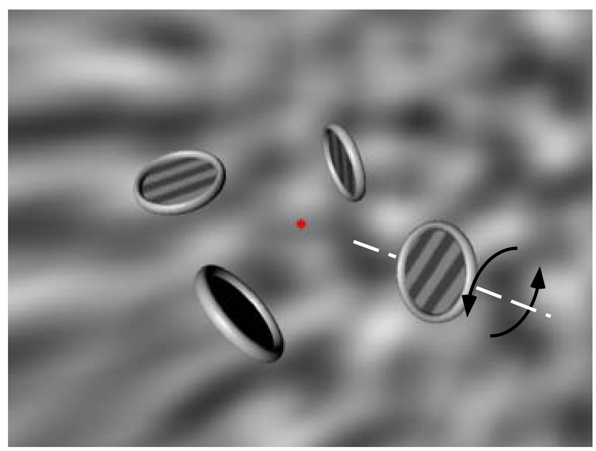


Figure 19

Target and distractor stimuli were based on realistic "coin"-like objects created using a sophisticated rendering program (Cinema4D, Maxon). The individual objects rotated about oblique axes oriented in depth, and were presented against a smoothed, noise background. Unlike the previous motion experiments, these stimuli differed slightly in size within display owing to perspective projection, contained realistic lighting cues and specularities that strengthened the 3D percept, and were presented in continuous motion until response.

I.A.4. *Heterogeneous distractors*

Here I introduce 3 search experiments that seek to investigate factors that might affect criterial scaling. One of the key components that allows the MTS model of search to account for flat or decreasing target absent RT functions is the parameter C (see section II.A.1.d). This parameter controls the extent to which sub-criterial but consistent sources of evidence guide decision. My pilot investigations using the model indicated that it successfully accounted for RT and error-rate data using a value of $C > 2$. This commonality suggests that observers adopt a general strategy of criterial relaxation during MTS. It is of interest to determine whether this tendency is simply a function of the homogeneity of the target and distractor sets that form the bulk of the MTS ensemble. Specifically, the experiments proposed in this section attempt to answer the following question: will observers continue to use multiple lines of weak evidence to guide decision when those lines of evidence arise from a set of heterogeneous elements? It may be that the search strategy embodied in the parameter C is only viable when a) distractors differ from targets in a single unique way, and b) distractors and targets form homogeneous sets.

A large body of search experiments consistently implicate stimulus heterogeneity as a major determinant of increased search difficulty and set size effects (Duncan & Humphreys, 1989; Humphreys et al., 1989). If stimulus heterogeneity has a similar effect in multiple-target search we should expect redundancy gains to diminish and target absent RTs to increase with heterogeneity. If such effects exist in the data, we should likewise expect to see correspondent change in the parameter C – that is, as target absent RT by set size slopes increase, the value of C extracted in fitting the model should be lower, indicating the model's reluctance to move secondary criteria far from the primary criteria set by T.

Experiment 20: triangle in rotated diamonds

This experiment is based on target/distractor stimuli consisting of simple polygons. Targets were designated to be filled triangles, while distractors were filled diamonds. In all cases, stimulus displays consisted of randomly oriented examples of these two base stimuli. Both elements were created to be of roughly the same size, and were chosen because they appeared highly similar. In pilot work I created multi-element textures with abutting regions composed of all triangles or all diamonds. These textural regions segmented only weakly, suggesting that in the context of multiple target search these types of stimulus differences would produce little to no

Experiment 20
Triangles in rotated diamonds

Experiment 21
Triangles in polygons

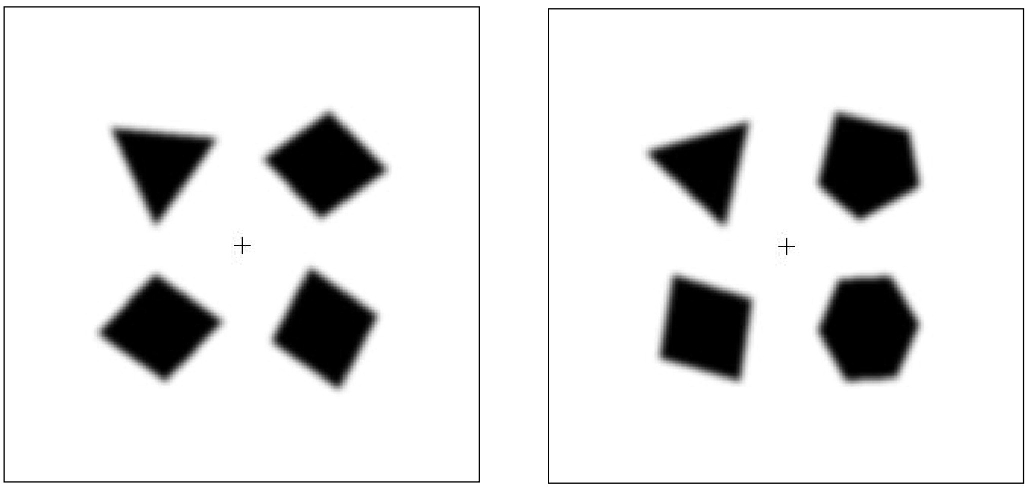


Figure 20

race gains in the pure target trials, and large increases in single target and target absent RTs with set size.

I have included this particular experiment in the section examining search with heterogeneous distractors because pilot work has indicated that the random orientations imposed on the polygon stimuli cause a sense of heterogeneity – the *mirroring* between pure and target absent RTs was reduced in this type of search. This breakdown of the symmetry suggests that observers are less willing to combine multiple sources of sub-criterial information, and that the MTS model will accordingly use a reduced value of C in accounting for the data.

Experiment 21: triangle in polygons

This experiment builds on experiment 20 by increasing the distractor heterogeneity so that it now includes random variation in shape. Specifically, a variable number of targets (filled black triangles) were presented among a variable number of distractors chosen randomly from a heterogeneous set of filled polygons (either a pentagon, diamond, or hexagon matched in phenomenal size to the target). The members of the distractor set were chosen to be highly similar to the target triangle. Again, all search elements were randomly oriented within and across displays to remove simple cues as to target presence.

Experiment 22: conjunction (color by orientation)

This experiment consists of a basic conjunction search, so named because target elements share a single feature with each member of a set of heterogeneous distractors. As such, the target can only be distinguished from the distractor set by virtue of a unique <u>conjunction</u> of features. This class of search is interesting for two primary reasons: first, it represents a heterogeneous distractor search and so in keeping with the previous arguments allows us to look for correspondent change in the parameter *C*; and second, this type of search experiment typically produces sharply increasing RT by set size functions in singleton search. Surprisingly, there have been few studies looking at conjunction search tasks with multiple targets. Ward & McClelland (1989) did investigate conjunction search in the context of a random walk model similar to the model proposed here. Their search data were fit by a model in which target-distractor discriminability decreased dramatically with set size.

The particular conjunction experiment I have used utilizes a color x orientation stimulus set that is known to consistently yield large set size effects (Nakayama, Wang, & Kristjansson, 2000; Wolfe, 1998b). Stimuli consisted of a variable number of target elements (vertical, white) that were presented among distractors based on random draws from a 2-element set containing black, vertically oriented rectangles and white, horizontally oriented rectangles. In this way, each distractor shared at most 1 *feature* with the target, and thus search could only proceed via feature combination. Practically, the target absent displays were created so as to maintain distractor heterogeneity on every trial. This is an important control because without it would be possible for observers to reject displays solely on the basis of all elements being black, or oriented horizontally.

Experiment 22
Conjunction (color X orientation)

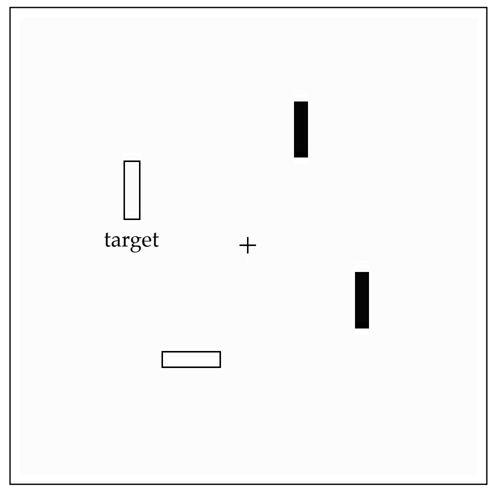


Figure 21

I.A.5. *Search for relative position*

The following set of experiments examine search when targets and distractors can only be distinguished in terms of the *relative position* or configuration of features or parts. The experiments detailed here are of particular interest for the following reasons. First, there is a large body of search work attesting to the difficulty that discrimination of relative position imposes on search (Enns & Rensink, 1990; Logan, 1994; Moore, Egeth, Berglan, & Luck, 1996; Palmer, 1994; Poder, 1999; Saarinen, 1996; Wolfe, 1998b; Wolfe & Bennett, 1996). In addition, these sorts of stimulus differences support little to no texture segmentation when targets and distractors define unique regions (Beck, 1966; Geisler, Stern, Thornton, Kuyel, & Ghosh, 1998; Malik & Perona, 1990; Renstschler, Hubner, & Caelli, 1988; Sagi, 1995). Taken together, the search and texture segmentation findings suggest that discrimination of relative position will manifest sharp costs in RT as distractor number increases, and little, if any redundancy gains across pure target conditions. My pilot investigations using MTS are in accord with this work as there is consistently a high degree of capacity limitation when targets and distractors are distinguished only by a symmetry flip. In addition, understanding why certain types of relative position judgements are searched more efficiently than others will be revealing as to the relation between stimulus structure and attention.

Experiment 23: missing side

Experiment 23
Missing side

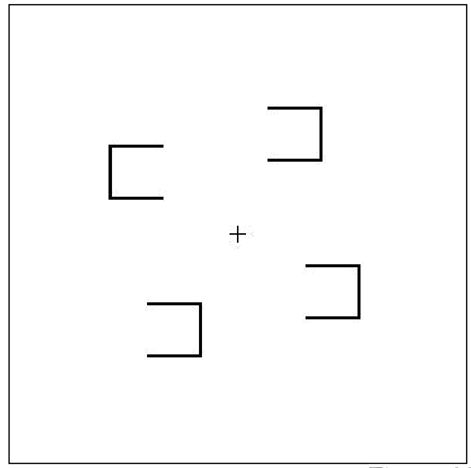


Figure 22

The stimuli for experiment 23 consisted of a three-sided box (~1.2 deg)–the target had its right side missing, the distractor its left side. These stimuli are a nice choice in the study of relative position because ,while they are extremely discriminable in the psychophysical sense, they engender very weak segmentation as abutting regions in a visual texture (Geisler et. al, 1998). The expectation for MTS is that search will be highly capacity limited with patterns of data characterized by flat or increasing *pure* target RTs, large set size effects , and high miss rates for displays containing single targets and 3 distractors.

Experiment 24: implied lighting (left/right)

Experiment 24
Implied lighting (*left/right*)

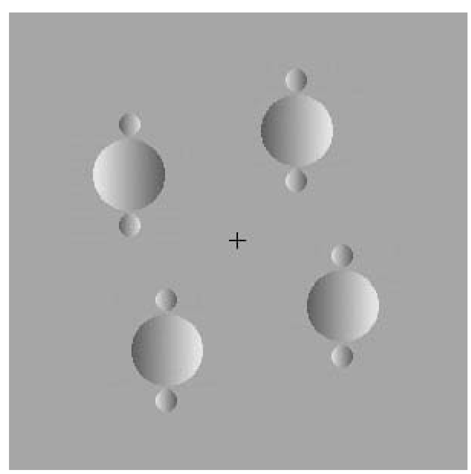


Figure 23

The stimuli for experiment 24 are identical to those used in experiment 12 except that the entire display has undergone a 90° rotation. This rotation removes the percept of curvature that attended the stimuli of experiment 12, presumably because there is no internalized assumption about side lighting (Sun & Perona, 1996). With this rotation, the target and distractor stimuli are no longer distinguished by an emergent property, and so they now truly differ only in terms of the relative polarity of shading. As the adjacent figure makes clear, search for left/right polarity is extremely difficult, an intuition that has support from both singleton search (Heathcote & Mewhort, 1993), and texture segmentation studies (Julesz, 1981; Rentschler, Hubner, & Caelli, 1988). In terms of model-based estimates of attentional limitation, we should expect then to find a large dissociation among the data from experiment 12 and its rotated counterpart.

Experiment 25: broken cube

In this experiment I use the same stimuli created for experiment 6 which examined search for cubes oriented in depth, but here have separated the cube faces in such a way so as to attenuate the percept of extension in depth. If this emergent percept has a role in search, then we should find a dramatic increase in

Experiment 25
Broken cube

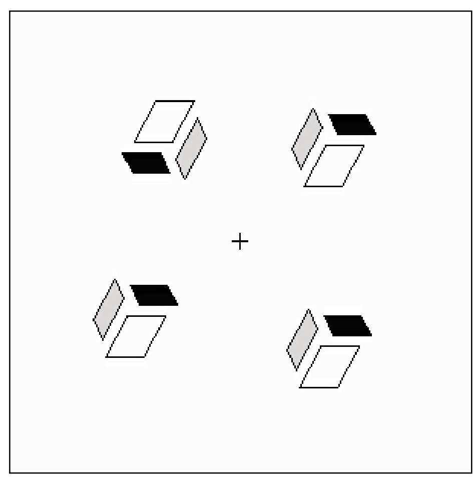


Figure 24

capacity limitation. In the context of singleton search this exact manipulation engendered large increases in the RT-by-set size slopes relative to the slopes for intact cubes (Enns & Rensink, 1990).

Experiment 26: circle-plus

The final experiment in the MTS ensemble used target/distractor stimuli composed of different arrangements of a plus sign and a circle (each pair subtended ~1.2 x 2.2 deg). The target was arbitrarily chosen to be plus to the left of circle; the distractor was created by flipping the target element about its minor axis (i.e. plus to the right of circle). Again, these stimuli are highly discriminable, yet in the context of multi-element displays lead to inefficient patterns of search characterized by large set size effects (Logan, 1994; Moore, Elsinger, & Lleras, 2001).

Experiment 26
Circle-plus

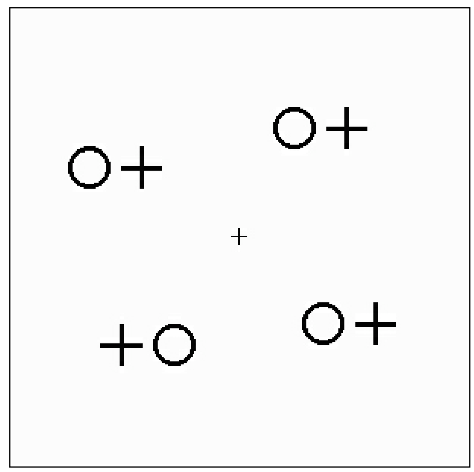


Figure 25

I.B. Data analysis

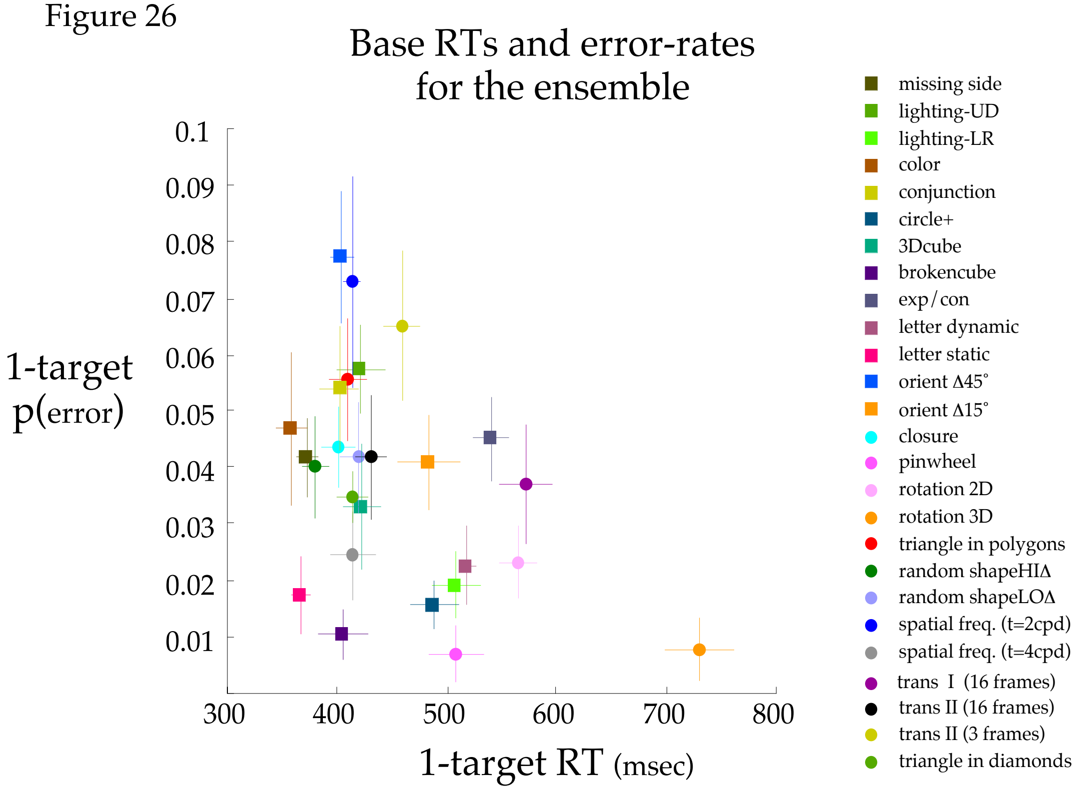
For all 26 experiments, observers completed 288 trials of practice, before providing either 576 trials (motion sign experiments), or 864 trials of data (the remaining 22 experiments). In the preparation and analysis of these data I excluded all trials on which errors occurred, and trials with RTs greater than 1500 or less than 150 milliseconds. I then computed both within-observer medians and standard errors for all MTS conditions (9 RT and 9 error-rate conditions).

I.C. Results

A summary of some of the more theoretically important RT patterns and error rates across the entire ensemble of 26 experiments is shown in the following three figures (the data from the pilot experiments shown in figure 2 is included these figures). I have opted for summary figures in order to highlight variation across the ensemble and to make inter-experiment comparisons easier. Complete RT and error-rate data for all 26 experiments appears in appendix IV.

Figure 26 plots the average error-rate ("misses" and "false-alarms") in

Figure 26



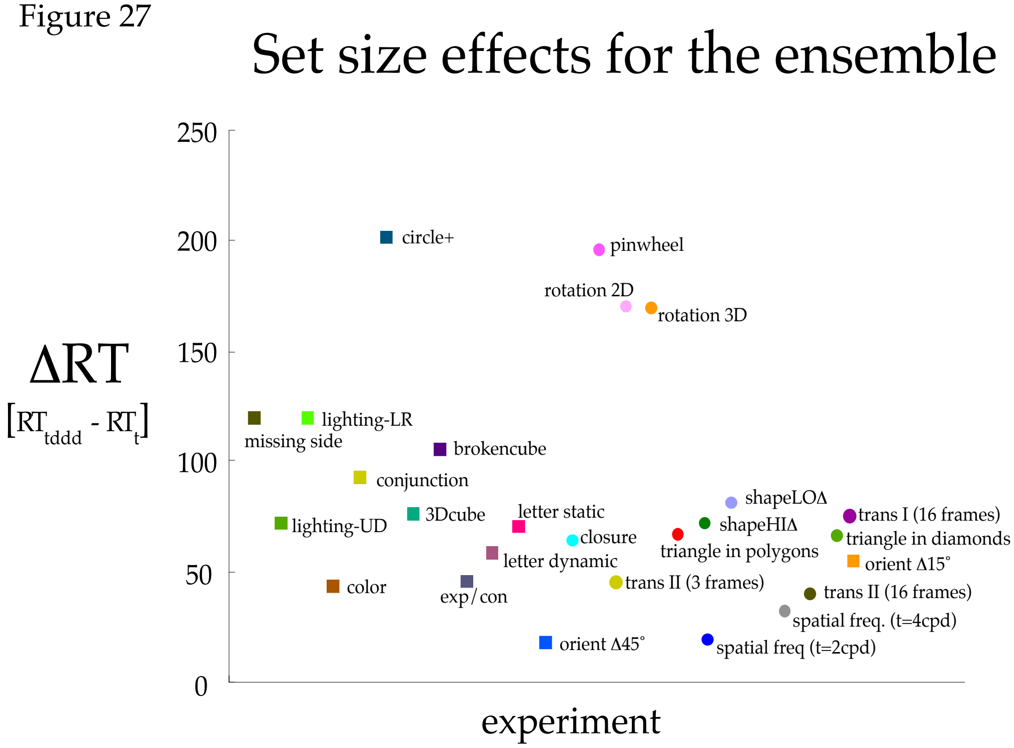
deciding that a single target is present in a display, against the average RT for single target displays (± standard errors are included). There is considerable variation in both response measures, with average error rates ranging from 1 to 8%, and average RTs ranging from 350 to 750 msec. In general, the longest base RTs are due to experiments defined by dynamic displays (i.e. the various motion experiments and the dynamic letter experiment). For these experiments, target/distractor discrimination depended on an underlying frame-rate that placed a limit on the minimal amount of time necessary for decision. Excluding these dynamically-defined experiments, the ensemble is characterized by very similar base RTs of approximately 400 msec. The base target absent data are not shown, but in general consisted of slightly higher error rates and RTs about 50-100 msec. slower than those associated with target present decisions.

The next figure plots the difference in RT for a 1 target/3 distractor display (*tddd*) relative to the single target base RT (*t*) for all 26 experiments. This RT difference (denoted the 1-target ΔRT) is a measure of the rate that RT to find a single target increases with set size, and is directly related to the RT-by-set size functions of singleton search. Recall, that in the context of the singleton search method, these so called "set size" effects play a central role in attempts to distinguish *serial* from *parallel* processes, and are indicative of the underlying search *efficiency* (Wolfe, 1998b).

There are three key points regarding the *set size effects* that emerge in the use of multiple target search. First, like the data from singleton search experiments, the 1-target ΔRTs are continuously distributed and are thus inconsistent with a strict division of search into *serial* and *parallel* categories (Wolfe, 1998a). Second, there is an intuitive ordering of search tasks, such that experiments based on simple featural differences of the sort known to have early cortical representations (i.e. spatial frequency, orientation, translational motion) yield smaller set size effects than searches based on conjunctions of features or the sensing of relative position. Finally, the set size effects are generally larger than those found in corresponding singleton search experiments. There are a number of possible explanations concerning why MTS might yield larger RT costs with set size than standard approaches (e.g. inclusion of a set size 1 condition, use of a restricted range of set sizes, etc.). Later in section II.C.5, I use an idealized Bayesian model of search containing no attentional limitation to show that adding multiple targets to a singleton search design naturally leads to larger ΔRTs with set size.

The centerpiece of the multiple target method lies in the use of *pure* target trials to signal “redundancy gains” in RT. These trials give the method

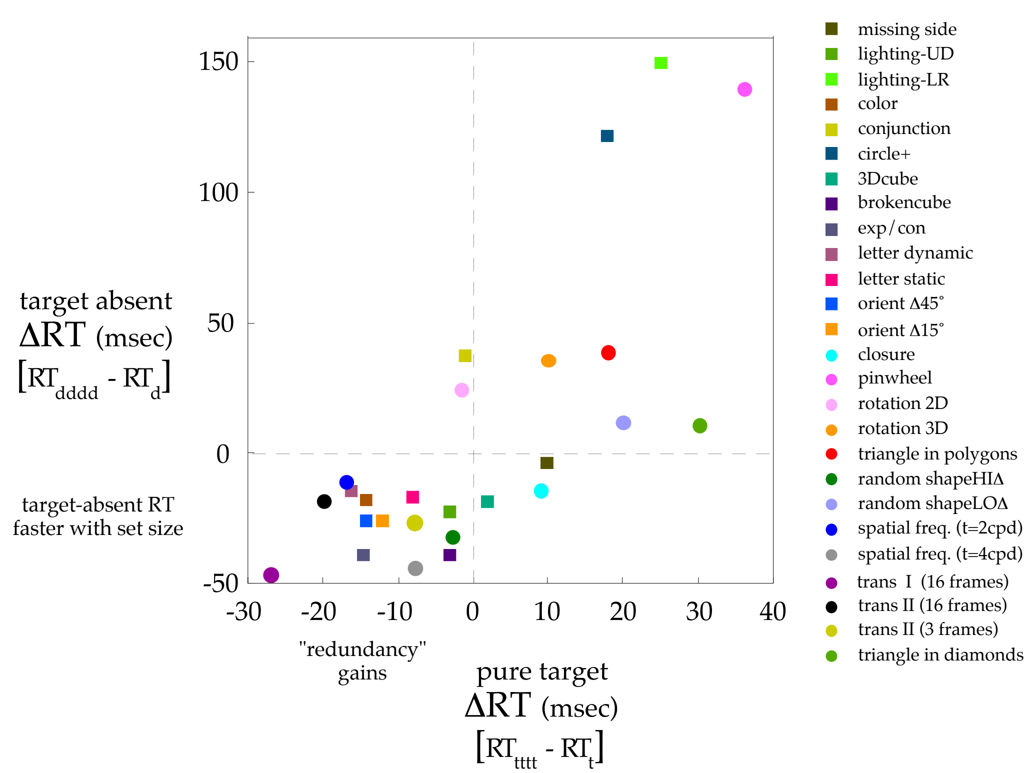
Figure 27



its power in deciding whether search is consistent with a *serial* process, or whether instead it indicates a *parallel*, but capacity-limited process. In figure 28 I have plotted the magnitude of redundancy gain obtained for each experiment in the ensemble against the corresponding measure for target absent displays. Specifically, the abscissa plots the average benefit in RT that occurs when displays contain exactly 4 targets (*tttt*), relative to displays which contain only a single target; the ordinate plots the average RT difference for displays containing exactly 4 non-targets (*dddd*), relative to displays containing only a single distractor. For both measures, positive values indicate a slowing of response time as more elements are presented, while negative values indicate a speeding up of response time with element number.

There are several important points regarding these two measures. First, as was true for the case of the 1-target ΔRTs, there is continuous variation in both the *pure* target and the target absent ΔRTs across experiments. For example, some experiments yielded negative *pure* target ΔRTs (faster) as more nontargets are added to displays (orientationΔ45°,

Figure 28 comparison of pure and target-absent slopes



color, translation), while others had constant (missing side, polygons, random shape), or even large positive ΔRTs (lightingLR, pinwheels, circle+). Similarly, for some experiments the *pure* target ΔRTs were negative in value, indicating the presence of "redundancy gains" as more targets are added to displays, while others were characterized by no such gains, or the even stranger pattern of slower RTs with increases in target number (heterogeneous shape, conjunction, rotation). Increasing RT with set size for the *pure* target trials is inconsistent with a *serial* process, and suggests that these types of search are highly demanding of attention, such that any benefits conferred by the addition of multiple targets is outweighed by the costs associated with an increase in set size.

The second point is that there is general linear covariation between the *pure* and the target absent ΔRTs (r=.728). This relationship provides an ensemble wide replication of the "mirroring" that was found in the preliminary uses of the MTS method (see figure 2). As mentioned in the introduction, this covariation represents a pattern of RT that cannot be accounted for using simple *serial* or *parallel* models of search, and thus provides a major constraint on the modelling of these search data. Decreasing target absent RTs with set size implicate a trading of speed for accuracy that is contingent on set size, and any successful model will therefore have to incorporate considerable decisional flexibility if it is to account for MTS data.

The final point to note regarding figures 27 and 28 is that there is a consistent relationship between the magnitude of an experiment's set size effect, and the magnitude and sign of the associated redundancy gain. Those experiments with large set size effects tend to have no observable redundancy gains, while those experiments with smaller set size effects tend to produce reliable redundancy gains. In general, the former pattern is consistent with a *serial-like* search, while the latter pattern (shallow set size effects + redundancy gains) suggests a *parallel*, somewhat capacity-limited search. However, because of the trading of speed and accuracy indicated by the target absen*t* data, it is not possible to categorize experiments using any single response measure.

In the remainder of this dissertation, I describe a modified random walk model of search that incorporates sensory, attention, and decision-based parameters. I then use the model to simulate the corpus of MTS data. The settings of the model that best account for each experiment's data provide measures of information quality, attentional limitation, and criterial flexibility. These measures are possible only through the use of a model that takes stock of the joint pattern of RT and error-rate.

Phase II: Modelling multiple-target search

In order to successfully account for multiple target search data we need a model in which 1) information about element identity accumulates stochastically over time, 2) the rate at which information accumulates is adjustable and dependent on attentional limitation, and 3) decision criteria can be manipulated to examine speed/accuracy tradeoffs. The sequential sampling class of models fits these requirements admirably in that these types of models simultaneously produce both response times and errors, and are parameterized by stimulus variables that control the rate of information accumulation, and decisional variables that determine how much information is required for response.

The core component of the model I use to simulate MTS derives from the basic random walk model (Laming, 1968; Link, 1975; Stone, 1960; Townsend & Ashby, 1983). This specific type of sequential sampling model has been enormously successful in the modeling of 2-choice decision (Ratcliff & Rouder, 1998), and versions of it have been previously applied to the visual search domain (Palmer & McLean, 1995; Ward & McClelland, 1989). In addition, this type of model permits decisions to be made on the basis of partially accumulated information (see Ratcliff, 1988), and represents an extension of signal detection theory across time. Moreover, the geometry of the random walk process itself naturally leads to positively skewed distributions of performance variables, a feature common to empirical RT distributions. However, I in no way wish to imply a commitment to this particular model, and recognize that other types of sequential sampling models that possess similar properties represent plausible choices for the modeling of multiple target visual search (e.g. the Poisson race model, Pike, 1973; Townsend & Ashby, 1983; Van Zandt, Colonius, & Proctor, 2000).

I extend the random walk model of decision to the visual search domain by introducing multiple independent "walkers", one for each element present in a stimulus display (for related models see Palmer & McLean, 1995; Ward & McClelland, 1989). In the standard conception of the random walk model of search, each walk evolves stochastically over time through the formation of a running sum based on repeatedly sampling random deviates (usually Gaussian, though not required). The walk drifts in Brownian fashion because at each discrete time step another random deviate is added to the existing sum (statistically, the sum grows in direct proportion to the number of deviates, and varies in proportion to the square root of the number of deviates). By definition this type of random walk model is a member of the class of discrete-time, continuous-state, sequential sampling models.

In the model, each random walk represents a record of the time-varying evidence about an individual element's identity. Target walks are formed by summing deviates from a distribution of positive mean, and will on average move toward a positive, target present criterion; likewise distractor walks are formed by summing deviates from a similar distribution whose mean is of opposite sign, and will on average drift toward a negative, target absent criterion (typically the means of the target and distractor increment distributions are symmetrically placed about zero, and both distributions have identical variance, though this is by no means required). In the context of discrimination and decision the deviates drawn from the "target" and "distractor" increment distributions are construed as varying amounts of information about stimulus identity obtained per unit time. The overlap of these distributions can be manipulated to account for variation in underlying target/distractor confusability. The *n* walks, one to represent each element in a display, drift between two primary decision criteria. A target "present" decision is made when the first walk crosses a positive criterion, or a target "absent" decision is made when <u>all</u> walk(s) have crossed a negative criterion. Response time arises naturally within this architecture, and is simply proportional to the number of deviates required for a given sum to exceed a response criteria. In general the average number of deviates required for a walk to reach a criterion is given by the following simple equation (Smith & Mewhort, 1998):

(1) $$\frac{T(1-2e)}{S}$$

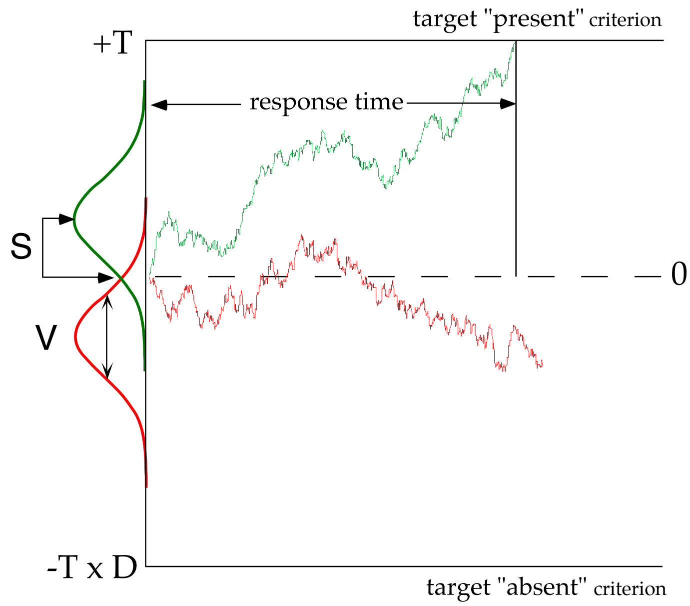
where T is the criterion, e is the error-rate, and S is the mean of deviates composing the walk. Errors occur when a walk is absorbed at the wrong response boundary; for example, one of n "distractor" walks reaches the target–present criterion prior to all the walks crossing the target absent criterion – in this instance a "false alarm" would be recorded. The following expression from Karlin and Taylor (1974) can be used to obtain the error-rate and is based on the probability that a single walk is absorbed at some criterion B prior to absorption at a criterion A, given an initial starting point x:

(2) $$p(abs) = \frac{e^{\frac{-2Sx}{V^2}} - e^{\frac{-2SA}{V^2}}}{e^{\frac{-2SB}{V^2}} - e^{\frac{-2SA}{V^2}}}$$

where S and V^2 are the mean and variance of the deviates. Schematic details of the standard random walk model of search are summarized in figure 29. In the figure a single "target" and a "distractor" walk are shown (plotted in green and red respectively), along with representations of the underlying increment distributions that give rise to each random walk (note that the moments of these distributions are exaggerated in the figure).

Figure 29

Basic random walk discriminator



There are several key assumptions that define the particular random walk model of search proposed here. First, I assume that all search processing is parallel, and possibly limited in capacity. These particular assumptions are made primarily because multiple-target search data appear to be wholly inconsistent with serial classes of search. In addition, there is growing evidence for the notion that many simple target/distractor discriminations make some demand on attentional resources (see Joseph, Chun, & Nakayama, 1997). Second, I assume that variation in search performance across trials and stimuli arises through the joint effect of costs due to stimulus load (i.e. set size effects, either based on decision or decision + perception), and benefits due to redundant targets (i.e. statistical facilitation or pooling accounts). Finally, I assume that the information about each element in a search display can be represented as an independent random walk. This assumption, though no doubt questionable, is reasonable given that the search elements in all my stimulus displays are spaced 1-2 degrees apart.

Before continuing, I would like to point out that the random walk model I am using to explain my visual search data is formally equivalent to a general class of models that sum likelihood ratios. In the literature there is a long history of random walk models that accumulate stimulus likelihoods over time using the *Sequential Probability Ratio Test*, so called *SPRT* models. These models are well understood and have been shown to be optimal in terms of minimizing RT for a given error-rate (Stone, 1960; Laming, 1968; Ashby & Townsend, 1983). In the SPRT model, the increments to the walk are themselves log likelihood ratios based on a single sample of evidence taken at time τ. Specifically, the SPRT model is based on increments given by

(3) $$L_i = \ln \frac{p(x_i \mid N(+S,V)}{p(x_i \mid N(-S,V)}$$

In the context of deciding any given element's identity during search, x_i denotes the current sample of evidence drawn from either the target [$N(S,V)$], or distractor [$N(-S,V)$] distribution. The L_is are summed to form a random walk in which decisions are made on the basis of the combined likelihood. Interestingly, the SPRT walk model differs from the class of walk models I am using only in terms of a single scale factor. This is because the distribution of log likelihoods is related to the distribution of Gaussian-distributed x_is by a constant factor $2S/V^2$ (for a proof see Appendix I). Because of this simple relationship, it is possible to conceptualize the generic Gaussian-increments model I am using as a *SPRT* model that categorizes search evidence in an optimal way.

I begin a detailed exposition of the structure of the quantitative models used to account for MTS data with the framework that formed the basis of my early investigations. This framework instantiates attention as a multiplicative scalar that attenuates evidence about element identity in a strictly Weberian fashion (i.e. attention scales both the average size and variability of evidence samples in like fashion). This manner of conceptualizing attention, though extremely successful in accounting for MTS data, presents a number of problems that I will discuss in detail later – most notably, multiplicative scaling of perceptual samples can be recast as a decision–limited model without attentional limitation in the sense put forth by Marilyn Shaw and John Palmer (Shaw, 1984; Palmer, Ames, & Lindsey, 1993). Accordingly, I denote the random walk model of search that instantiates attention as multiplicative scaling as a *D-type* model. This is not to say that the *D-type* model does not embody components of attentional limitation, only that in the random walk framework it is indistinguishable from a certain class of model that has no such limitation. In a later section, I will introduce and simulate a general class of *A-type* models that instantiate attentional limitation in a way that is not reducible to decisional manipulation. To anticipate, the majority of the search experiments reported here are best explained using a *D-type* model, while only a small handful of highly capacity-limited searches required an additional *A-type* model.

II.A. Model structure

The basic model of MTS is unique among existing models of search in so far as it incorporates two novel parameters identified with attentional limitation and decisional flexibility. First, the model has an attention parameter to represent varying degrees of capacity limitation. This parameter which I denote ε, allows the model to control the rate at which the random walks drift toward response criteria by multiplicatively scaling the increment distributions with set size. The specific assumption here is that capacity limitation in visual search can be modeled as a decrease in the amount of information acquired about stimulus identity per unit time. This construel of capacity limitation is not new and has often been invoked to explain how parallel, limited capacity models can produce increasing RT by set size functions of the sort once thought to uniquely signal a serial process. Basically, the intuition is that a parallel model that simultaneously processes all elements in a display can nonetheless yield increasing RTs with set size if the rate of processing varies inversely with set size (Townsend, 1974; 1990). This class of model differs from previous approaches in that I do not fix the extent to which processing is slowed with stimulus load, but rather allow it to enter as a free parameter that varies from no effect of set size on processing(i.e. unlimited capacity), to a perfect inverse relation in which rate of processing is divided by set size (i.e. highly capacity limited).

The second novel component of the model is the use of a free parameter to control the extent to which decision criteria are relaxed with set size. This additional parameter is motivated by the need to account for the invariant or decreasing RTs I find for target absent decisions as a function of set size (i.e. the target absent "mirroring" of pure target RTs seen in figures 1, 2, and 8). Recall, that both serial and parallel models of search predict that RTs to respond target absent should necessarily increase with set size. In multiple target search this does not happen because observers appear to routinely adopt a strategy that allows them to respond using incomplete information. Thus, flat or decreasing target absent RTs indicate that observers can take advantage of consistent, but sub-criterial amounts of evidence. The logic is as follows: even though no single walk may have reached the target absent criterion; if all n walks are simultaneously below some relaxed criterion it becomes increasingly unlikely that a target is present in the display – therefore respond target "absent". The strategy is implemented in the model via the parameter C which inversely scales the two primary response criteria as a function of set size. As set size increases there will be two secondary response criteria, one for target present decisions and one for target absent decisions. These criteria are by definition scaled to be closer to the origin than the primary criteria, and are therefore less stringent, allowing decisions to be made prior to any one walk reaching the primary response criteria. Later in section II.C.5 I explore the optimality of such a decisional strategy in the context of a Bayesian model of MTS.

II.A.1 The *D-type* model

There are five primary parameters that define the Gaussian-increments random walk models of MTS. Three of these parameters are generic to a large class of random walk and diffusion models (Link, 1975; Ratcliff, 1978; Townsend & Ashby, 1983), while the remaining two are the novel parameters previously mentioned.

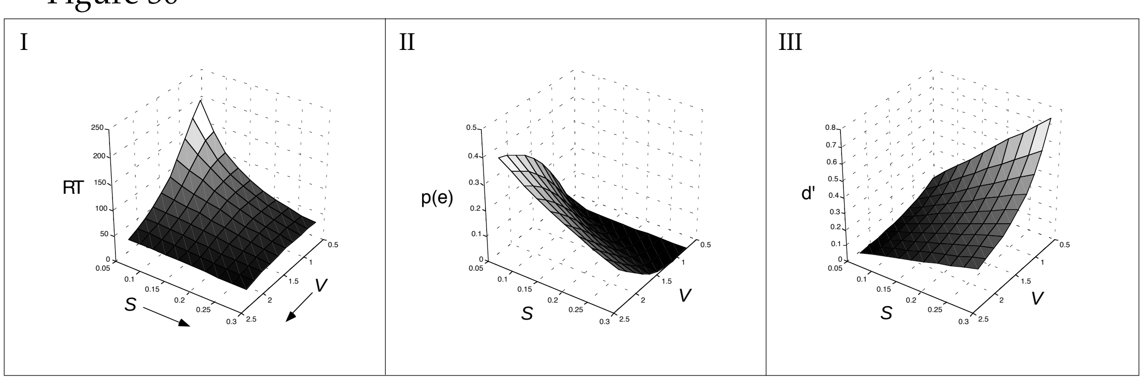
II.A.1.a *T* (*Primary response criterion)*

T represents the amount of evidence necessary to decide an element's identity and thus sets the distance of the target present and target absent response criteria from the walk origin (the target absent criterion is by default equal to the target present criterion, but of opposite sign; the walk origin is set to zero). A target present response is initiated as soon as one of the *n* independent walks reaches the boundary set by *T*; a target absent response is initiated as soon as <u>all</u> of *n* walks reach the boundary set by -*T*. In an earlier incarnation of the model, *T* was free to vary and *V*, the parameter representing increment variability was fixed. Because of simple scaling relationships among *T* and the mean (*S*) and variability (*V*) of the increment distributions, one of these parameters can be fixed without loss of generality. To see this, note that *T* is really just a distance that can be expressed as a linear combination of *S* and *V* [i.e., T= ZV + *S*, where *Z* is a normalized distance]. For all the modeling work reported here I have chosen to fix *T* at 15, and to let *S* and *V* enter the model as free parameters. The exact value of *T* has no real psychological meaning in and of itself, but produces appropriate base RTs and error rates in the context of the other parameter values.

II.A.1.b *S & V(increment mean and variability parameters)*

The model includes a sensitivity parameter *S* that corresponds to the magnitude of the mean of the target (+*S*) and distractor (-*S*) increments, and a variability parameter *V* that represents the shared standard deviation of those increments. For a fixed level of *V*, as *S* increases both the time-to-absorption

Figure 30



(~*RT*) and the probability of an error will decrease. For a fixed level of *S*, as *V* increases *RT* will decrease and the probability of an error will increase. These relationships are shown in panels 1 and 2 of figure 30 in which I plot the RT and error-rate surfaces over the ranges of *S* and *V* used in the simulations (the height of each surface was determined using the simple analytic equations for absorption time and error-rate given earlier).

Taking a ratio of these two parameters provides a measure of *signal–to–noise* that is proportional to the base discriminability of a target increment from a distractor increment (panel 3). Due to the statistics of the random walk, the *signal-to-noise* ratio of the accumulated increments grows with time such that the discriminability goes as the square-root of the number of increments currently in the running sum. By allowing both *S* and *V* to vary freely across simulations it is possible to compare the *signal-to-noise* ratios of different experiments. It may be that the experiments in my ensemble differ in terms of the inherent discriminability of targets and distractors. We know that underlying element discriminability plays a significant role in determining RT-by-set size search functions (Duncan & Humphreys, 1989; Palmer et al., 1993; Geisler & Chou, 1995). Interestingly, it is the case that in fitting the full model simultaneously to all the RT and error-rate conditions provided by multiple target search, the quantity S/V acts more like a measure of the *confusability* of the *n*-element display, rather than a measure of single-element discriminability in the usual psychophysical sense. This is an important distinction – distractors and targets may be highly discriminable in a 2AFC paradigm, and yet still be able to effectively mask each other in a multi-element display. In section II.C.3 I discuss this distinction in detail.

II.A.1.c *D (present/absent asymmetry parameter)*

The parameter *D* is included so as to introduce an asymmetry into the target present and target absent response criteria. *D* effectively scales the target absent criterion to be some multiple of the target present criterion (i.e. target absent criterion = -*T x D*). *D* is constrained to be greater than or equal to one – this is done primarily to account for the ubiquitous finding that most target present decisions are generally made faster than associated target absent decisions. The finding that positive decisions are made faster than negative decisions appears to be a universal in cognitive judgement (Baddeley & Hitch, 1974; Clark & Chase, 1972; Kosslyn, 1975; Lewis & Anderson, 1976; Sternberg, 1969; Treisman & Gormican, 1988), and may reflect in part an inertia to respond in the negative. Here I have modeled this RT inertia by moving the target absent criterion further from the origin, thereby requiring that additional evidence accumulates before a "no" response can be made.

I have investigated two alternative ways to implement a bias to respond target “present”, neither of which has proved successful in accounting for MTS data. First, I conducted an informal exploration of a model in which a fixed amount of increment-bias, *b* is added to each increment in the walk (Ashby, 1983) – for $b > 0$ the model favors target “present” responding over target “absent” responding. This type of model is equivalent to 1) shifting both increment distribution means by *b* so that they are no longer symmetric about the origin, or 2) tilting the decision criteria in state space such that $T_\tau = -b \times \tau$ (where τ represents some discrete point in time). While this type of model can produce asymmetries in RT, it makes two false predictions: that miss-rates should decrease with set size, and that false-alarms should increase with set size. These types of error patterns occur rarely, and never together in MTS.

I have also modelled the target “present” response bias as a simple *inertia* to respond in the negative, possibly arising from the mental set of “looking for a target”. The idea here is that it takes slightly longer to respond target “absent” simply because the observer is primed to hit the target “present” key, and thus there is a concomitant overhead associated with re-orienting oneself to hit the “absent” key. This *inertia* is modelled as a late RT-cost (denoted *Z*) that independently augments only the target absent RTs by some small, fixed amount. Again, this type of bias successfully accounts for asymmetries in response time because for the most part the actual target absent RTs differ from the target present RTs only in terms of a constant. However, this type of bias can not lead to different error rates across conditions and so can not account for the majority of MTS experiments in which there is such an asymmetry.

The general approach I have chosen to take regarding modelling the target "present" response bias is to first let the parameter *D* attempt to account for any asymmetries in the RTs and error rates. Then, I investigate whether an additional additive RT-cost at the level of response preparation (i.e. the parameter *Z*) is necessary to capture the residual asymmetry that remains in any given experiment. In principle, the additional parameter *Z* will be nonzero whenever the RT asymmetry is larger than the corresponding asymmetry in the error rates.

II.A.1.d *C (criterion scaling parameter)*

The parameter *C* controls secondary response criteria as a function of display set size. Specifically, the model has an additional set of criteria that are available when set size is greater than one – a secondary criterion for target present responses, and a secondary criterion for target absent responses. These criteria are by definition less conservative than the two primary criteria set by *T* and *D* because they are closer to the walk-origin (because secondary criteria are placed closer to the origin, they permit responses to be initiated sooner; however, because they are more lax they can also lead to increased error). I include secondary criteria in order to give the model the ability to make decisions on the basis of multiple random walks – though each walk may be sub-criterial in relation to the primary criteria, the secondary criteria allow a response to be initiated when walks are consistent in the magnitude and sign of their displacement from the origin (i.e. the model can use a preponderance of weak but consistent *evidence* in favor of a particular response). It is important to make it clear that even though these criteria are less conservative in terms of the absolute amount of accumulated information required for any one walk to reach them, a response is initiated only if all of the n walks simultaneously exceed one of the secondary criteria.

Secondary criteria are implemented in the model in the following manner. For C on the interval [1,4], and $n = 2$ or 4, we define the secondary target present and target absent criteria to be $T \times C^{-n/2}$ and $-T \times D \times C^{-n/2}$ respectively, where n represents display set size. This relation effectively places the secondary criteria geometrically closer to the origin as set size increases. When $C = 1$, there are no secondary criteria and all responses are based on the primary criteria – this represents the standard formulation of a random walk model with invariant decision boundaries (Townsend & Ashby, 1983). When C is greater than 1, response criteria are appropriately scaled with set size depending on the magnitude of *C*. A specific example may prove clarifying here. If we fix the base target present criterion *T* to equal 15, and the criterion scaling parameter *C* to equal 2, then for a set size of two the secondary target present criteria will equal 7.5, and for a set size of four the secondary target present criteria will equal 3.75.

II.A.1.e *Decision rules*

When $C > 1$ and set size equals 2 or 4, the model effectively has four decision criteria, two primary and two secondary. These criteria map walk positions into responses using one of two general decision rules. Responses can be initiated by either the first walk to reach a primary criterion, or instead when all *n* walks exceed a more relaxed secondary criterion.

> <u>Primary criteria</u>. In the case of model responses initiated by absorption at one of the primary criteria, there are different decision rules for target present and target absent responses. The model responds target "present" as soon as the <u>first</u> of *n* walks is absorbed at the primary positive criterion. Alternatively, the model responds target "absent" as soon as <u>all</u> *n* walks are absorbed at the primary negative criterion. These decision rules reflect the underlying task that defines multiple target search – namely, target "present" responses are to be

made any time a display contains at least one target, and target "absent" responses are to be made only when there is no target element in a display.

Secondary criteria. The decision structure is slightly modified in the case of decisions made at secondary criteria. The positive and negative secondary criteria allow a preponderance of evidence to guide decision. Here, both target present and target absent responses are initiated using an identical decision rule: the model responds target "present" as soon as all *n* walks exceed the secondary positive criterion and likewise, responds target "absent" if all *n* walks are simultaneously below the secondary negative criterion. It is this symmetry in the decision rule that manifests the symmetry between *pure* target and target absent RTs seen throughout MTS.

In sum, the addition of secondary criteria allows the model to make target present responses in one of two ways: a target "present" response is initiated as soon as one walk reaches the primary positive criteria, or when all walks reach the secondary positive criteria. A target "absent" response is always initiated in a manner consistent with exhaustive processing; that is, only when all walks exceed the secondary negative criteria.

II.A.1.f ε *(attentional limitation parameter)*

The parameter ε determines the extent to which increases in set size attenuate the processing of multi-element displays. Specifically, ε determines how fast any given walk drifts toward the response criteria. Drift rate is attenuated simply by multiplicatively scaling each sample of evidence prior to its accumulation. Multiplicative scaling of the magnitude of each sample implies that both the mean and standard deviation of the underlying increment distribution are scaled identically by the factor $n^{-\varepsilon}$, where n represents display set size, and ε is defined on the interval [0,1]. This range for ε was chosen because it parameterizes a large variety of search models. For example, when $\varepsilon = 0$ there is no attenuation of processing rate with set size. At the other extreme, where $\varepsilon = 1$, processing rate is dramatically attenuated with set size (the walk drift rate is cut by a factor equal to $1/n$). By allowing ε to take on intermediate values that fall between these two extremes, I can manipulate the extent to which set size affects processing, the goal being to discover the values of ε that best account for performance variation in an ensemble of search experiments.

Scaling perceptual samples

Symmetric scaling of the mean and standard deviation of the walk increments by $n^{-\varepsilon}$ was done for several reasons. First, it seemed intuitive to conceive of attention as having a multiplicative influence on stimulus information, much as increasing resistivity has a multiplicative influence on current flow. Along these lines, it is also true that "turning down" or limiting the range of a process (i.e. decreasing the average step size) can yield a matched decrease in the instantaneous variability of that process depending on the underlying probability distributions. The second reason I chose to use this type of scaling is because it preserves the intrinsic discriminability of target and distractor increments, the aim being to incorporate attentional effects independently of the underlying stimulus quality. Finally, and most importantly, of the various ways of implementing attention that I have investigated, only multiplicative scaling of the increments can simultaneously account for both the RT and error-rate patterns typical of MTS.

Scaling criteria

In the context of a random walk between absorbing barriers it is possible to conceptualize the effect of ε in an altogether different way that has no relation to attention whatsoever. It is true, given the linear relationship between T, S, and V (see section II.A.1.a), that a model that symmetrically scales S and V with set size is formally equivalent to a model that scales T with set size, but leaves S and V unaltered. Thus, in the context of a random walk account of decision, comparing a cumulative sum of samples (which have been attenuated by the scalar quantity $n^{-\varepsilon}$) to a fixed response criterion

T, is equivalent to comparing an unattenuated accumulation of samples to a more conservative criterion $T_C = n^{\varepsilon}T$. What this means practically, is that there is a symmetry in the random walk model such that ε can be conceived of as either an attentional scalar ($n^{-\varepsilon}$), or as a criterion scalar ($1/n^{-\varepsilon}$), such that there is no difference in the predicted time-to-absorption or the error rates. Scaling criteria to be more conservative as a function of set size is not a new idea, and such a strategy will yield set size effects without attentional limitation as observers attempt to control error in the face of increased statistical decision noise (see Palmer, 1994, Pashler, 1998). In fact, recasting ε as a criterion scalar rather than an attentional effect makes some sense in that its effects are synonomous with a speed-accuracy tradeoff – larger values of ε lead to longer RTs and lower error rates, while smaller values lead to shorter RTs and increased error.

One major problem remains though with conceiving of ε as a decisional scalar: it dramatically increases the number of criteria assumed to be under the observer's control. Recall that the current model incorporates a novel decisional parameter C in order to account for the flat or decreasing target absent RTs that are ubiquitous in MTS. This parameter creates a set of additional criteria that are more relaxed relative to the base criterion T (section II.A.1.d). If we also conceptualize ε as a decisional parameter, the observer now has explicit control over 10 separate criteria! An obvious alternative to this growing plethora of criteria is to split up the free parameters such that ε remains under attentional control, and the observer is left to manage the 2 primary criteria and a set of secondary criteria. This seems more psychologically plausible and is the framework I will adopt from here on out. That said, I acknowledge that parsimony is better served by seeing how much can be explained by a single mechanism of limitation (i.e. *decision*), prior to invoking a role for additional mechanisms (e.g. *attention*). With the caveat in mind, I will denote the random walk model with multiplicative scaling as a *D-type* model in recognition of its equivalence class, while continuing to construe the parameter ε as an estimator of attentional limitation. Whether it makes sense to think of ε as a true attentional parameter remains at present an open question – we will revisit this issue later in section II.C.2 where I show that the majority of MTS search tasks can be reliably and meaningfully ordered along the ε-continuum.

Alternative models of attention

Multiplicative scaling of increments is not the only way that attention might operate in the context of a sequential sampling model. In the following section I introduce and compare the qualitative predictions of models that implement attentional limitation in a variety of different ways. The reader who is less interested in the details relating and distinguishing these various conceptions of attention can skip ahead to the final paragraph of this section (*Summary of models*).

II.A.2 *The A-type model*

In addition to the *D-type* model which conceives attention to have a multiplicative influence on perceptual samples, here I explore a class of models that implement attentional limitation via the addition of internal noise. This class of models is denoted *A-type* because the attentional effects they employ cannot be reduced to decision or a speed/accuracy tradeoff, and so they serve to break the symmetry that undermines the clear interpretation of ε as an attentional parameter in the *D-type* model. This class of models is based on the same underlying multiple random walk architecture and shares all the stimulus and decisional structures that define the *D-type* model (i.e. *S*, *V*, *C*, *D*), differing only in terms of the way that attentional limitation is implemented.

II.A.2.a *Dual noise source model*

One potential shortcoming of the *D-type* model is that it predicts no difference in the quality of single-element discrimination as a function of set size. In fact, the *D-type* model predicts that overall error-rate should decrease as ε increases, and thus is not relevant to search methods that center around brief presentations of search displays with accuracy as the DV (e.g. the *threshold–search* paradigm, Verghese & Nakayama, 1994; Palmer et. al, 1993).

To circumvent this shortcoming, and to break the attentional/decisional symmetry of the *D-type* model, I introduce a stimulus-invariant source of additive noise into the standard model. The use of multiple noise source models is common in the psychophysics literature, and additive internal noise is often used to model general inefficiencies in processing due to unexplained sources of neural/perceptual variability (Burgess, 1985; Pelli, 1990; Dosher & Lu, 1998; Gold, Bennett, & Sekuler, 1999; Weiss, Simoncelli, & Adelson, 2002).

The additional noise source I introduce into the model consists of a sequence of zero-mean, Gaussian deviates [i.e. $N(0,\sigma_{INT})$] that are added to each of the n perceptual samples at each moment in time. If we let $W_{\tau,k}$ represent the current sum of evidence for the k^{th} element in a display at time τ then we have:

(4) $$W_{\tau,k} = \sum_{i=1}^{\tau} [x_i + N(0,\sigma_{INT})]$$

where x_i represents the current perceptual sample drawn from the appropriate target or distractor increment distribution, and $N(0,\sigma_{INT})$ is a deviate from a zero-mean Gaussian.

Effectively, the additive noise increases the base variability of the increment distributions without affecting the average step size. More importantly though, there are now two sources of variability in the random walks, one that can be scaled by attention and one that is invariant. The parameter V corresponds to the standard deviation of the walk increments themselves, and in keeping with the *D-type* model, both V and S are attenuated by the factor $n^{-\varepsilon}$. The quantity σ_{INT} refers to an internal additive noise source that does not depend on n or ε. The key idea behind this so called dual-noise source model is that attention still operates multiplicatively, but now only on the stimulus noise. By allowing V to scale with set size and attention, while keeping σ_{INT} constant, the model now predicts a decrement in single-element discriminability. The magnitude of this decrement depends on both n and ε, and implies that the *A-type* model can no longer be reduced to a criterion shift. The prediction for the discriminability of a single random walk as a function of S, V, σ_{INT}, n, and ε appears below:

(5) $$d'_n \approx \frac{2Sn^{-\varepsilon}}{\sqrt{(Vn^{-\varepsilon})^2 + \sigma_{INT}^2}}$$

Note that the current number of accumulated increments, τ does not appear in the expression and simply serves to augment discriminability by a factor of the square root of τ.

There are several additional advantages to the use of a dual-noise source model that make it an attractive alternative to the *D-type* model in certain contexts. First, because this model has an invariant additive noise component, attention now attenuates the mean of the increments (S), more than the total variability of the increments [let $V_{TOTAL}=(V^2+\sigma_{INT}^2)^{1/2}$]. To see why, note that S gets attenuated by $n^{-\varepsilon}$, whereas only the component V in V_{TOTAL} is subject to n and ε (see eq. 5). In the context of a random walk, this leads to increased errors with set size. Thus, a model with internal noise, all else being equal, predicts higher miss rates than a similar *D-type* model. Recall, that several of the MTS experiments were characterized by large increases in RT with set size, flat or increasing *pure* target trials, and very high miss rates that approached 25% for $n = 4$ (see *rotation*, *circle-plus*, and the *implied lighting* experiments). The ability of the dual-noise model to produce long RTs along with high miss rates should allow it to better simulate the data from this class of experiments.

A second advantage in the use of a dual-noise model is that it is more general: it contains the basic *D-type* model as a special case of an *A-type* model with zero internal noise. This relationship is convenient because it means I can compare the performance of the *D-type* model to a matched *A-type* model simply by changing a single parameter (σ_{INT}). From here on, I will denote specific versions of the *A-type* model as $A(\sigma_{INT})$ to signify the magnitude of internal noise injected into each model. In this notation the *D-type* model of section II.A.1 is synonomous with an *A*(0) model.

Relation to the fixed sample size model

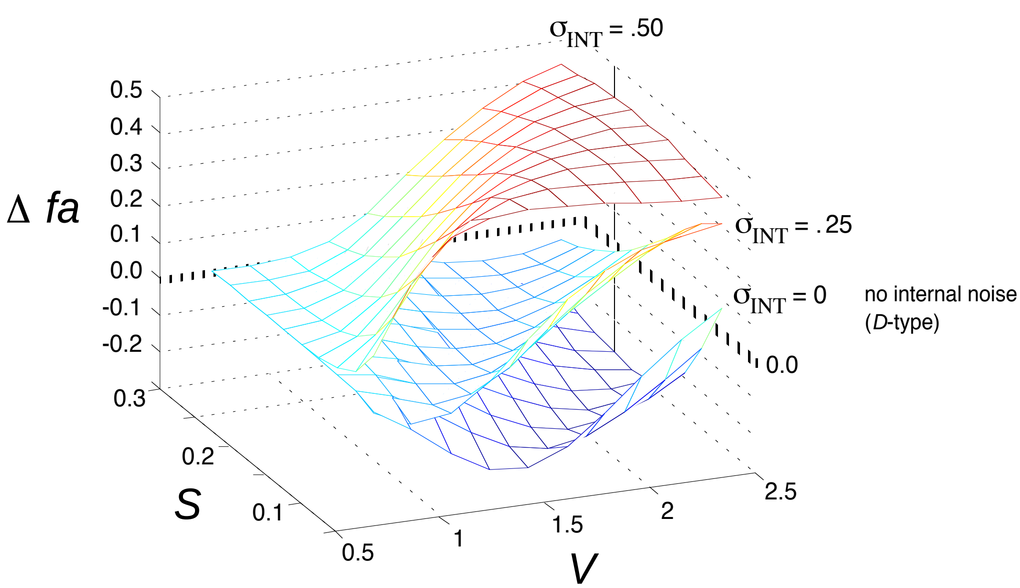
Another advantage of incorporating $A(\sigma_{INT}>0)$ models into my investigations is that for high levels of σ_{INT}, these models approach the performance of the *fixed sample size* model of Shaw and colleagues, especially when ε is high (Shaw, 1980; Palmer, 1994). The *fixed sample size* model figures prominently in attempts to determine whether increases in difference-thresholds with set size are the product of statistical limitations, or instead are large enough to indicate perceptual limitations. Previous work indicates that for the case of simple stimulus dimensions (orientation, size, luminance) the threshold-by-set size functions could be entirely accounted for by decision noise; that is, the set size related decrements in performance were too shallow to indicate capacity limitations on perceptual processing. In stark contrast, only a handful of experiments, generally characterized by targets and distractors differing only in terms of configural arrangement, produced increases in thresholds as a function of set size large enough to match the predictions of a *fixed sample size* account of attention. I can use the *A*(.5) model to approximate this sort of limitation, and thus can determine whether any of the experiments in the MTS ensemble can be accounted for using a *fixed sample size* model of attentional capacity.

One shortcoming of the *A*(.5) model though is that it cannot produce false-alarms that decrease with set size. Recall, that this is the modal pattern of error across the MTS ensemble, and is easily captured by an *A*(0) model. What this means in general is that if we demand a good qualitative account of the patterns of misses and false alarms, then an *A-type* model with high internal noise will fail to account for these data. In the adjacent figure, I highlight this fact by plotting the relationship between the slope of the predicted false-alarms over set size for three *A*-type models parameterized by σ_{INT} = 0, .25, and .5. Delta false-alarm surfaces (~slope of the probability of a false alarm across set size) are plotted as a function of the increment parameters *S* and *V* at an intermediate fixed level of ε = .5 (variation in the parameter ε had little qualitative effect on the relation among these surfaces – as ε approached 0 all three surfaces converged on a strictly positive surface similar to that shown for σ_{INT}=.5). The ordinate represents the change in the proportion of false alarms in going from *n*=1 to *n*=4. All three predicted Δ false-alarm surfaces were computed analytically using the expression in equation 2 (sec. II), and the following expression for the probability of at least one walk being absorbed at the incorrect boundary prior to any single absorption at the correct boundary:

(6) $$\Pr(fa,n) = 1 - [1 - \Pr(fa,1)_{\varepsilon,n}]^n$$

[where $\Pr(fa,1)_{\varepsilon,n}$ represents the probability of a single false-alarm given model-appropriate scaling of the increment moments by $n^{-\varepsilon}$]. The validity of this expression was verified via partial simulation of the actual models.

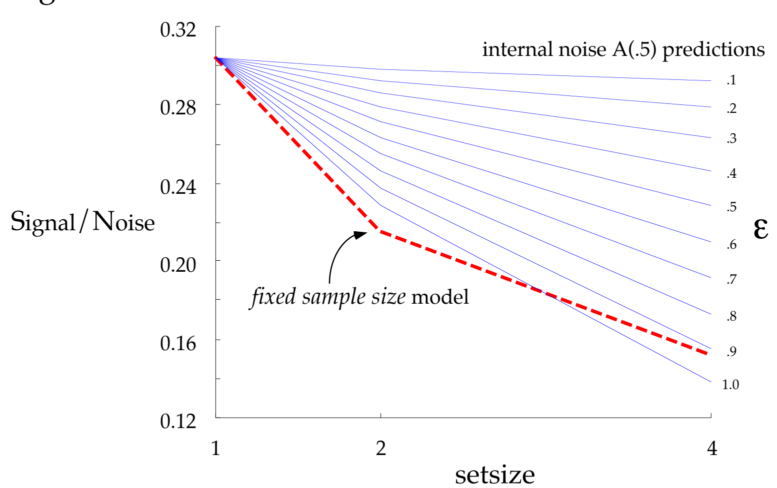
Figure 31



There are a number of important insights into the models revealed in this figure. First, the *A*(0) model is seen to produce large negative slopes (i.e. false-alarms that decrease with set size) over a large portion of the *S*-*V* space. This property is due to the way that attention is implemented in the *D-type* model (multiplicative scaling). The figure also reiterates the point that there are no combinations of *S*, *V*, and ε for which the *A*(.5) model is able to produce decreasing false-alarms – in every case this model manifests invariant or increasing false-alarms with set size (this holds for all ε for realistic ranges of error). The *A*(.25) model is intermediate and, depending on the choice of parameters, can manifest mildly negative, invariant, or moderately large and positive false-alarm slopes.

The fact that the *A*(.5) model yields only nondecreasing false-alarms with set size is diagnostic and offers a preliminary indication of the class of experiments that can be fit using this type of capacity limitation: if an experiment yields false-alarms that reliably decrease with set size it cannot be explained using the *A*(.5) model, and similarly cannot be accounted for using a *fixed sample size* model of attention. However, it is important to realize that just because a given search task is not fit by a sample size model does not mean that it is not capacity limited. In fact, the *A*(.5) model

Figure 32



parameterized by ε clearly reveals that the *fixed sample size* model of Palmer and Shaw may represent an extreme level of capacity limitation. In figure 32, I plot predictions of the $A(.5)$ model (thin lines, parameterized by ε) for the single element signal-to-noise ratio (SNR) as a function of set size. These predicted decrements in SNR were computed using average values of S and V and equation 5. The inset dashed line represents the prediction of the *fixed sample size* model in which the SNR goes as the square root of n (for a more detailed development see Appendix III). As the figure makes clear, it is only at the highest levels of ε that the $A(.5)$ model approaches the performance decrements associated with a *fixed sample size* account of limitation. It is interesting that implementing capacity limitation as a slowing of the rate at which information accumulates plus internal noise can yield discriminability functions considerably shallower than those predicted by a *fixed sample size* conception of capacity. In the context of threshold search there were very stimulus domains that produced threshold-by-set size functions large enough to meet the predictions of the *fixed sample size* model (e.g. see Palmer, 1994; Poder, 1999). This may simply reflect the fact that the majority of stimulus domains are only limited by decision (Palmer et al., 1993; Palmer, 1994). However, it also possible that the lower limit imposed by a *fixed sample* account of capacity limitation is too severe. The relationship between the discriminability curves produced by the dual-noise model and the *fixed sample size* model show that it is possible to create a range of capacity limited predictions less severe than the bound set by a fixed sample size. A recent extension of Palmer's work has in fact attempted to quantify capacity limitation using a conceptualization that is continuous and considerably less extreme than the *fixed sample size* framework (see McLean, 1999). In the following section I introduce a more general probability-sampling model that includes the *fixed sample size* model as a special case. I then go on to show how the dual-noise model (i.e. $A(.5)$) and probability-sampling model are qualitatively equivalent in their predictions.

II.A.2.b *Probability-sampling (time-scaling) models*

Another way to conceptualize attention in the context of sequential sampling models is as a limitation on the sampling rate: specifically, as attentional load increases, the probability of including an additional sample of evidence into any random walk decreases, such that the accumulation of evidence consists of fewer samples per fixed period of time. This notion differs in principle from that of the *A-type* models for which the probability of sampling evidence at time τ is always 1, and limitation enters only to scale the quality/size of the current piece(s) of evidence. Conceptually, a *probability-sampling* model of attention leaves the quality and size of all samples unaltered, and only attenuates the likelihood of the random walk receiving a sample at each moment in time.

For ease of comparison with the general class of *A-type* models, I let the probability p of taking a sample be expressed as $n^{-\lambda}$, where n is set size and λ is an attentional parameter in the range [0, 1]. [In this context, λ and ε both represent attention and are defined similarly as parameterizing power functions of set size. For clarity though, I will always use λ to denote attentional limitation in the *probability–sampling* model, and ε to denote attentional limitation in the *A-type* model]. Defining p in this way is intuitive because as limitation (i.e., λ) rises, the value of p falls gradually as a power function eventually reaching a value of $1/n$ for the case of $\lambda = 1$. When the probability of receiving a sample is $1/n$, the *probability–sampling* model reduces to a *fixed sample size* model of attention. To see this equivalence, note that for an n-element display, the *fixed sample size* model allows only S_{tot}/n samples per element because by definition there are only S_{tot} total samples available to be divided among the display elements. Similarly for the case of a *probability-sampling* model, if we assume that at any there are at most S_{tot} samples available to analysis (up to time τ), then the expected number of samples accumulated in any single random walk will be pS_{tot}, or S_{tot}/n when λ=1. Thus, both the *probability-sampling* model (λ=1) and the *fixed sample size* model predict identical effects on the expected number of accumulated samples at time τ. However, the two models differ on a trial-by-trial basis because the *probability-sampling* model has a source of variability that the *fixed sample size* model does not: the *probability-sampling* model predicts S_{tot}/n

samples per element on average, while the *fixed sample size* model predicts S_{tot}/n samples on every trial. Despite this difference the two models yield identical distributions of walk states as a function of time, and times-to-absorption and error rates that are basically indistinguishable (see Appendix III for details).

There are a number of features that favor the *probability-sampling* model over a *fixed sample size* account of capacity limitation. First, the *probability–sampling* model is parameterized by ε, and so like the class of dual-noise models it can yield predictions that span a range of capacity limitation. Second, the *probability-sampling* model reduces to the *fixed sample size* model as ε approaches 1. Finally, the class of *probability-sampling* models is conceptually related to a framework that implements attention via time scaling. The idea here is that capacity limitation alters only the time scale of processing for the *n*-element display, such that each individual random walk has to wait longer to accumulate a sample of evidence. Thus, increasing $\Delta\tau$, the average time between samples in the context of perfect sampling (i.e. p(sample)=1), is identical to the *probability-sampling* approach in which $\Delta\tau$ is fixed and the probability that a sample is received gets reduced.

II.A.3. *Summary of model comparisons*

Beyond the core *D-type* model, there are two distinct approaches to implementing capacity limitation in a random walk framework, both of which are not reducible to decisional scaling:

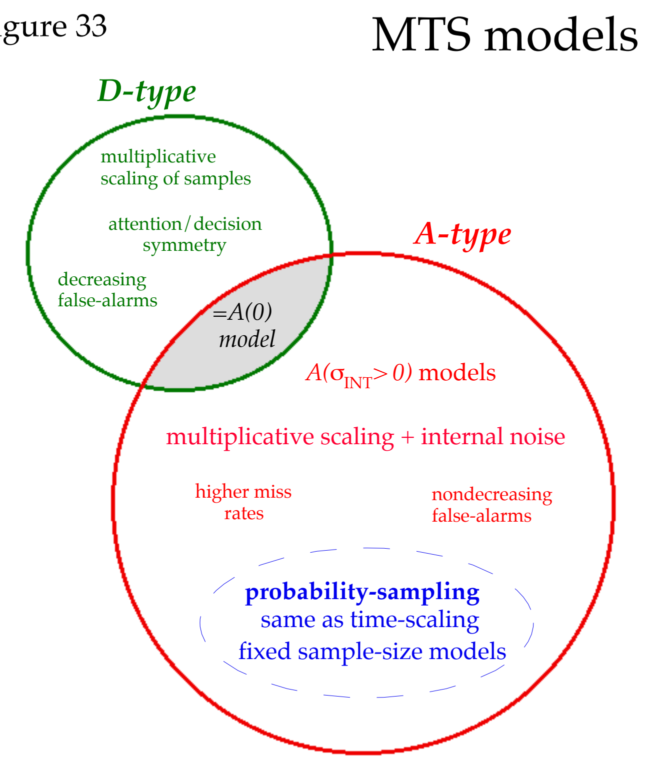
i) the dual-noise or $A(\sigma_{INT} > 0)$-type models, and

ii) the *probability-sampling* (time-scaling) models of the previous section.

Though these two approaches are conceptually quite different in terms of the manner in which set size effects are presumed to arise (one assumes additive noise degrades performance, the other that fewer samples are available per unit time), at present, they cannot be distinguished based solely on fits to MTS data – both approaches can account equally well for experiments characterized by sharp increases in RT with set size, high miss-rates, and flat or increasing false-alarm rates. The important point to realize here is that both models can be aligned to predict nearly identical effects on the scaling of the moments of the increment distributions. Equivalent scaling implies that both models will yield roughly identical absorption times and errors given that all remaining parameters are shared in common, and I have verified this via simulation.

A summary of the relations between the various models outlined in this section is depicted in figure 33. For the purposes of this dissertation I will not be concerned with deciding which model of attention is correct. Moreover, because of the many equivalences highlighted above, it is somewhat redundant to investigate fully both the dual-noise model and the *probability-sampling* model. In recognition of this, I have chosen to investigate attentional limitations during multiple target search using mainly the dual-noise framework. Specifically, I will compare MTS data to predictions generated by a model with no internal noise, that is the *D-type* or *A*(0) model, and a model containing mild to moderate amounts of internal noise [i.e. an *A*(.25,.5) model]. Later, where appropriate, I also explicitly simulate a matched *probability-sampling* model to verify that it produces fits similar in quality to those produced by the dual-noise *A*(.5) model.

Figure 33



II.B *Details of numerical simulations*

The approach I take in modeling multiple target search is two-fold. First, I explicitly simulate the search process for a large number of unique combinations of parameter values. Second, I compare the simulations to empirical search data in order to find the parameter values that best represent data from a given search experiment. In practice the model is simulated over an entire 5-dimensional parameter space defined by the basic model parameters (S, V, C, D, ε). This consists of parameterizing the model using a single 5-tuple from the space, and then conducting a Monte Carlo simulation based on 2000 simulated search trials. This number of simulations was chosen because it was computationally feasible, and produced relatively reliable predictions (for typical parameter values, the average standard deviation of simulated RTs and error rates across repeated simulation of 2000 trials was 1-4 walk steps and 2-5 tenths of a percent respectively). Each simulation yields a set of 9 average RTs and 9 error rates. The simulated RTs and error rates are saved to disk, and another 5-tuple is chosen and simulated. In this way I can simulate the model over the entire range of parameter combinations to form a "prediction" space. I then search for that region of the prediction space that best matches a specific set of empirical data.

The computational models used here to characterize multiple target search cannot be approached analytically, primarily because they include a complex nonlinear decision structure (based on *T*, *C,* and *D*) that is not amenable to simple analytic decomposition (Laming, 1968). Running the model in an explicit simulation is the only way to determine exactly what the model will do in any given situation, though in simple limiting cases I do check that the simulations give mathematically correct answers using analytic formulations for distributions of the "first passage" in a random walk based on Wald's identity (e.g. eqs. 1 and 2). What numerical simulation lacks in elegance, it far makes up for in terms of flexibility – only by foregoing the world of analytic model fitting is it possible to explore the variety of complex architectures embodied in the class of *A-type* models. For ease of exposition I will continue to use the term model *fitting* throughout this dissertation to represent the joint process of simulation, followed by search for "best" *fit* within a prediction space.

In practice, I simulate the model over a discrete 10 x 10 x10 x 10 x 10 x 3 grid of possible parameter values (10 linearly spaced values each of the parameters *S*, *V*, *C*, *D*, ε, & 3 levels of σ_{INT}), thus giving me 100,000 distinct parameter 5-tuples at 3 different levels of internal noise. For each 5-tuple, I explicitly simulate the random walks 2000 times x 9 stimulus conditions for a total of 18,000 trials in order to get a stable estimate of the model's expected

RT and error-rate for each condition. I then repeat this process to yield 3 complete five-dimensional prediction spaces for each level of σ_{INT}. To ascertain the reliability of the model fits to data, I then refit the model to the data from each experiment using a re-sampling procedure (see sec. II.A.6).

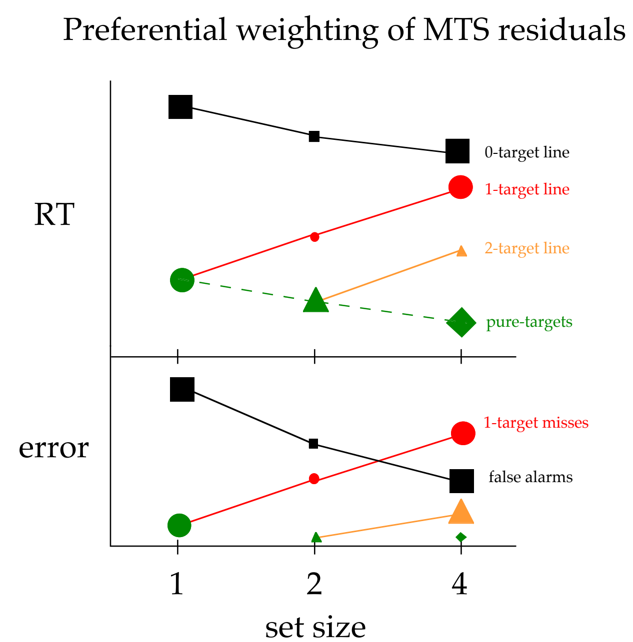
II.B.1 *The objective function*

I define the best simulation of a data set to be that point in the discrete prediction space that is simultaneously closest to the RT and error-rate data. The specific objective function I use to determine the distance between simulation and data is defined to be the combination of the sum of squared-error between data and model-generated RTs (SSQ_{RT}), and the sum of squared-error between data and model-generated error rates (SSQ_{ERR}). In this way, both the RT and error-rate jointly determine a minimal distance in the prediction space: *distance* = $SSQ_{RT} + SSQ_{ERR}$. I have explored a number of alternative ways of combining the SSQs (i.e. *euclidian*, *city-block*), and in general the various choices made little difference – all methods converged on the same local region in the parameter space.

Because the SSQ_{RT} and the SSQ_{ERR} are based on different scales of measurement, a decision must be made regarding how to weight each term prior to forming the total sum of squared-error (SSQ_{TOT}). In general there is remarkably little guidance provided in the literature on this issue, no doubt a reflection of the relative paucity of work using objective functions that take

RT and error-rate residuals as arguments. One tact that has been adopted in this regard is to scale the RT and error-rate residuals so as to approximately

Figure 34



give them equal weight (Maddox & Ashby, 1996; Van Zandt et al., 2000). I adopt a somewhat different approach and choose instead an automatic weighting procedure in which each RT and error-rate residual is weighted by its associated variability. Specifically, I scale each of the 18 residuals that comprise the SSQ_{TOT} (9 RT residuals and 9 error-rate residuals) by the inverse of the standard error associated with each condition (standard errors are based on between-observer variability). This type of scaling converts the deviations of *model* from *data* into RT and error χ^2s. The use of chi-square residuals is not uncommon in the least-squares fitting of descriptive functions and has the nice property that it approximates maximum-likelihood techniques.

One difficulty in using a single-valued objective function comprised of 18 components (9 RT residuals + 9 error residuals) to quantify overall fit is that often the region of the prediction space that minimizes SSQ_{TOT} yields simulations that fail to capture important aspects of a data set. For example, in certain cases the *fitting* procedure may choose as "best", a simulation that has significant redundancy gains when there are in fact no such gains present in the data. This arises because the fit to the *pure* target RTs is being sacrificed for a better fit to some other aspect of the data. The problem here is that we want the model to account for the data in general, but more importantly to make accurate predictions about those conditions which are the most theoretically interesting. I deal with this problem by preferentially weighting certain components in the SSQ_{TOT}. Specifically, the objective function preferentially weights the individual χ^2 residuals to emphasize the fitting of the single-target RTs and miss-rates, the *pure* target RTs (i.e. redundancy conditions), and the false alarms. There were thus 11 *key* residuals which were weighted a factor of 10 greater than the remaining 7 components. In figure 34 I depict those conditions which received greater weight in the determination of the SSQ_{TOT} using larger symbols. The weighting procedure proved successful in that it yielded fits that captured the *key* qualitative patterns in the data, while maintaining a good account of the remaining RT and error conditions (smaller symbols).

II.B.2 *Transforming output to RTs*

There is an additional complexity that must be dealt with in forming the SSQ_{RT} component of the objective function: empirical RTs have units of seconds, while model RTs have units of number of steps. In order to quantitatively compare model to data it is necessary to bring both onto a common scale. One common means of placing model RTs on a scale commensurate with RT data is to assume that each step corresponds to a millisecond, and then to use an additive constant to bring model and data into accord. The key idea behind including an additional RT component is not new, and reflects a theoretical decomposition of the observable RT into one component due to perception and decision, and another component due to response preparation and execution (Luce, 1986).

Bringing the model and RT data onto a common scale is accomplished via the parameter M_R. This parameter effectively shifts all model values by a fixed constant amount. In this way the model and data can be quantitatively compared and there is no alteration of the simulated relationships between conditions (i.e. the slopes and relative intercepts). I attach no psychological relevance to the absolute value of M_R in that it is entirely determined by the values of T, S, and V. That is, M_R can be pushed around by varying the core model parameters, and this can be done in such a way so as to preserve the qualitative pattern of RT and error-rate. The value of M_R is thus not unique and serves only as an aid in determining the SSQ_{RT}.

In addition to M_R, I also include a target absent response cost parameter *Z*. This parameter is used to soak up any additional asymmetry in responding not accounted for by the decision parameter *D*. This constant shifts all the model target absent output equally and independently of the target present output. Thus, the value of *Z* that emerges in fitting the model to data represents the additional amount of time necessary to initiate a target absent response, above and beyond any RT asymmetry produced at the level of decision criteria. Moreover, allowing a non-zero value of *Z* insures that the SSQ_{RT} is not dominated by residual error due simply to the model not adequately accounting for slower target absent responses. At present, it is unclear why some experiments produce decrements in target absent responses too large to be captured by criterial asymmetry. Incorporating a nonzero value of *Z* into the RT-transformation is of practical use in the fitting of these experiments because it emphasizes good qualitative fits by effectively keeping the residual target absent response costs from corrupting the SSQs.

It is important to note that the RT transformation is always applied after the prediction space has been simulated. Specifically, at each point in the prediction space I find the values of M_R and *Z* that minimize (in the least squares sense) the deviation of model from the RT data. In this way, the RT transformation is applied independently to each qualitative model prediction in the space, and the final residual error after this transformation forms the SSQ_{RT}.

II.B.3 *Alternative sources of error*

The random walk models that are simulated here differ from *high-threshold* models because they assume that errors arise as a result of the accumulation of noisy perceptual samples. A *high-threshold* model (e.g. Wolfe's *Guided Search* model) asserts that errors arise only via guessing after the search process has "timed out", and moreover, that distractors are never confused for targets in the classic signal detection sense. In addition to noise and guessing, errors during search could also arise from any number of alternative sources (e.g. various sensory distractions, fatigue, motor programming errors, …, etc.). These sorts of errors are no doubt partly responsible for the overall levels of error in an MTS experiment. The simplest way to incorporate these alternative sources of error into the model(s) is to add on a constant probability of error to all of the error rates predicted by the model. Rather than utilize an additional parameter, I assume that a good estimate for this composite source of unexplained error is each experiment's actual four target error-rate. Thus, prior to forming the SSQ_{ERR} between model and data I augment all the error rates in the prediction space by the four target error-rate of the experiment that is being fit. This simply shifts all the simulated error rates vertically, much as M_R shifts the simulated RTs, leaving all the qualitative patterns and interactions intact. Though using the four target error-rate in this way effectively removes a degree-of-freedom from the data, it has the nice property that it also removes an unexplained source of variability from the SSQ_{ERR}.

II.B.4 *Re-sampling data*

In my earlier work, the model simulations were *fit* to the observer-average data from each experiment. This provided a single estimate of each of the 5 parameters that characterized performance in any given search experiment. It would be nice to know both how reliable these parameter estimates are so as to make meaningful comparisons across experiments, and how sensitive particular choices of parameters are to variation in the data. I have attempted to address both these problems by using a re-sampling procedure that allows me to form confidence limits on the parameter estimates.

For each experiment in the ensemble I form 20 re-sampled sets of RT and error data. Specifically, I create data for 9 *pseudo*-observers by randomly sampling from a set of normal distributions with moments defined by the actual observer-averaged data (9 RT distributions and 9 error-rate distributions). I then average over the RTs and error rates of the *pseudo*-observers to yield a single set of re-sampled data. This process is then repeated 19 times to yield a complete set of 20 re-sampled versions of each experiment.

The model *fitting* procedure described previously is then applied to each of the re-sampled data sets. Because some of the model *fits* will be worse than others, due to the random nature of re-sampling, I calculate a SSQ_{TOT} for each of the 20 *fits*. I then take a weighted-average of each parameter across *fits* using the normalized inverses of the SSQ_{TOT} of each fit as the weights. This re-sampling procedure results in a weighted-average and standard error for each parameter value for each experiment. Importantly, I have verified that the re-sampling procedure produces *fits* that differ very little from those obtained by fitting just the average data. In this way, I get the same parameter values as I would have if I had only fit the average data, but by virtue of re-sampling I also get confidence limits around each parameter value.

There are two remaining points to address regarding the re-sampling procedure. First, I chose to *fit* re-sampled data over *fitting* individual observer data because single observer data sets were generally too noisy and often had conditions with zero errors. I have investigated *fits* of the model to single-observer data for a subset of experiments and find that in general the observer-average 5-tuple was similar to that obtained under the cleaner re-sampling procedure. Second, the manner of re-sampling I have used is conservative regarding issues of *fit* reliability because I re-sampled the RTs independently of the error rates. The architecture of the random walk model of decision mandates that RT and error must trade. Because of this fact, faster RTs generally occur in the context of higher error rates, and vice-versa. Unfortunately, there is no quantitative metric to specify the trading of seconds for percent-correct and so I can only re-sample RT and error independently.

II.B.5 *Determining the magnitude of* σ_{INT}

At this point, the fitting procedure specifies the "best" set of 5 basic parameters (*S*,*V*,*C*,*D*, & ε) given appropriate RT-transformation parameters M_R and *Z*. What it does not do is decide whether an experiment is fit better by the *A*(0) model with no internal noise, or an $A(\sigma_{INT} > 0)$ model that includes internal noise. This is accomplished by finding that level of σ_{INT} that minimizes each experiment's SSQ_{TOT}.

In order to decide what level of internal noise best accounts for a given data set, I simulate a full 5-parameter model for each level of σ_{INT} under consideration. This consists of repeating the complete fitting procedure outlined above for each of 3 levels of σ_{INT} (0, .25, and .5). In this way, I obtain the best-*fit* average 5-tuple for each level of internal noise giving me three 5-tuples for each experiment in the ensemble. In order to claim that any given experiment was fit better by a model with a level of $\sigma_{INT} > 0$, I use the following simple statistical procedure. This procedure is possible because there is variability associated with evaluating how well the various MTS models fit each of the 20 re-sampled versions of an experiment. Because each model is *fit* to 20 re-sampled data sets per experiment, there will be a separate distribution of SSQs associated with each model. These various distributions can be compared to look for reliable differences in the average SSQ. Specifically, I use repeated measures *t*-tests to look for significant differences among each model's SSQ_{ERR} (the SSQ_{RT} differed very little across choice of σ_{INT} and so was uninformative in this context). Only if a given level of σ_{INT} produced significantly better fits to an experiment's error data, relative to a model with no internal noise (i.e. the *A*(0) model), was that experiment said to require the addition of internal noise. Thus, I computed two *t*-tests: one between the SSQ_{ERR} of the *A*(0) and the *A*(.25) models; the second between the SSQ_{ERR} of the *A*(0) and the *A*(.5) models. For cases in which both tests revealed significantly better fits (i.e. a lower SSQ_{ERR} for either the *A*(.25) or *A*(.5) model), I computed a third test between the SSQ_{ERR} of the *A*(.25) and the *A*(.5) models.

A summary of the general *fitting* algorithm is given below:

1.) Form a SSQ_{TOT} value for each point in the prediction space. This consists of combining the SSQ_{TOT} value obtained after applying the RT-transformation to the model RTs, with the SSQ_{ERR} value obtained after augmenting the simulating error rates to reflect unexplained sources of error.

2.) I then use an automatic procedure to explicitly search through the entire prediction space to find that 5-tuple associated with the smallest SSQ_{TOT}. The process of searching through the prediction space for the parameter 5-tuple with the lowest SSQ_{TOT} is carried out separately for each of the 20 re-sampled versions of an experiment. These 5-tuples are then averaged across sample *fits* using the 20 SSQ_{TOT} as weights. In this way I obtain a weighted average "best" fitting 5-tuple to characterize each search experiment in the MTS ensemble.

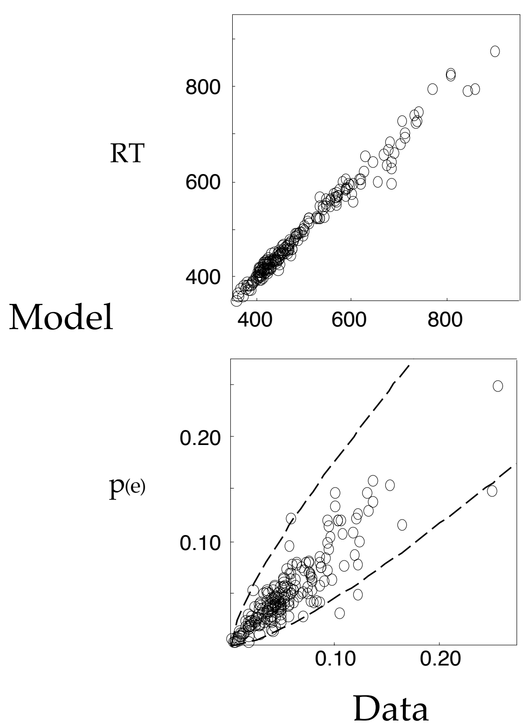
3.) To decide the level of internal noise that best describes a particular experiment, I repeat steps 1 and 2 for each choice of σ_{INT} and conduct paired *t*-tests using each model's distribution of SSQ_{ERR}.

II.C. *Modeling results*

The basic model of MTS proved quite successful in accounting for the RT and error-rate data across the experiment ensemble. In the upper panel of the following figure I show the relationship between the entire set of observed average RTs (i.e. 9 conditions x 26 experiments), and model predictions. In the lower panel I show a similar plot relating the ensemble error-rate data to the associated model predictions. The lower panel also includes upper and lower bounds (thick black lines) derived using the standard error of the proportion (±1.96 x sep). In general the model simultaneously accounts for a large percentage of the RT data (median ensemble r = .952), and the error-rate data (median r = .933), though the fits to the error rates are substantially more variable than those for the RTs. This general pattern, in which simultaneous fits to RT and error-rate favor the RT fits, is not uncommon in the use of sequential-sampling models

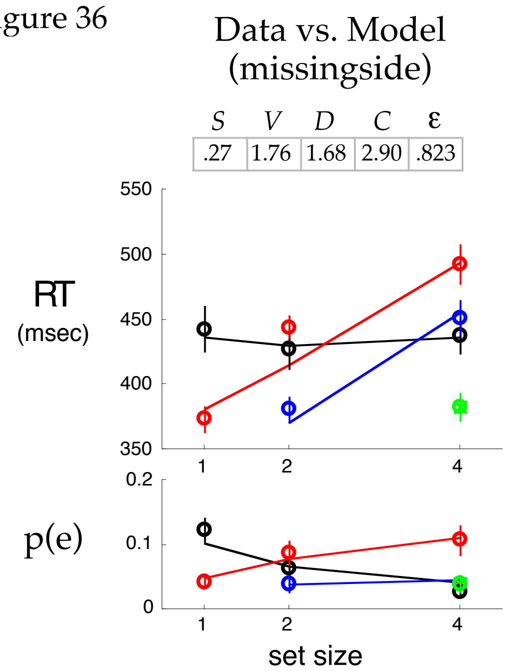
Figure 35

Comparison of MTS data and model fits (n=468)



(see Van Zandt, et al., 2000 for a similar fitting asymmetry), and may simply reflect the fact that the model's decision structure is too rigid, or that empirical error rates are multiply determined. In addition, a poorer account of the error rates is to be expected given that the model-generated RTs are modified by an additional 2-parameter transformation (i.e. M_R and Z). More importantly though, all the predicted error rates fall within the bounds set by the intrinsic variability of the data.

Figure 36



In figure 36 I show a representative example of the model's ability to simultaneously capture the complex interaction between RT and error typical in the use of MTS. This figure plots both the RT and error data (points), and the "best" fit model predictions (lines) for a single experiment and is typical of the fits produced by the model. The complete set of MTS results and model comparisons is included for reference in appendix IV.

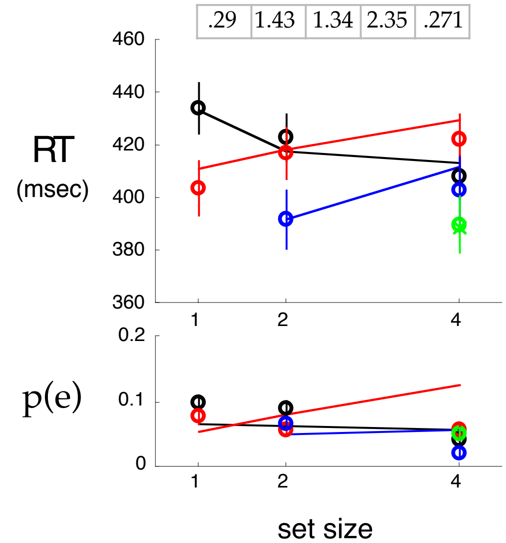
Although the model was generally quite successful in accounting for the empirical data, there was a single notable exception. In the following figure I show the data (solid symbols) and model predictions (lines) for the *orientation* ($\Delta\theta$=45˚) experiment. While the model does a fine job of accounting for the pattern of RTs (r = .944), the fit to the errors is abysmal (r ~0). The primary reason for this failure is that there are no regions of the prediction space for which both misses and false-alarms are seen to simultaneously decrease with set size. In fact, there currently appears to be no parameter combinations that support this pattern of errors, though my search has by no means been exhaustive.

Further evidence for the unique status of the *orientation* ($\Delta\theta$=45˚) experiment is that <u>no</u> other experiment in the MTS ensemble was found to produce this pattern of errors. In fact, search based on a similar, but less discriminable target/distractor set (i.e. the *orientation* ($\Delta\theta$=15˚) experiment) failed to produce such a pattern, and was accordingly fit better by

Figure 37

Data vs. Model
(orientation $\Delta\theta$=45°)

S	*V*	*D*	*C*	ε
.29	1.43	1.34	2.35	.271



the model. More to the point, these data are not simply a sampling anomaly, in so far as I have completely replicated the $\Delta\theta=45^{\circ}$ experiment using an additional group of 8 observers. That this pattern emerges when feature differences are great, may indicate the adoption of a general strategy based on local *feature gradients* (Beck, 1966). By that account, observers may be less concerned with looking for a target, instead choosing to make decisions on the basis of display *homogeneity* versus display *heterogeneity*. This idea is partially supported by the actual pattern of error rates: the miss-rate for the set size 4/1 target condition is equal to that of the set size 4/4 target condition. If one is in fact "searching" for targets, it is hard to understand how a display containing 4 targets can be mistakenly rejected as often as a display with 3 distractors and only 1 target. In general you expect both speed and accuracy to increase as more targets are added to displays. On the other hand, if one has adopted a strategy based on "searching" for displays that contain target/distractor mixtures, then it is intuitive that *homogeneous* displays (*pure* and target *absent*) should be more readily confused. Aside from the use of secondary criteria, the current MTS model does not make use of inter-walk comparisons, and so it may not be too surprising that it fails to account for these kinds of data. The question of why search based on large differences in orientation produces a novel pattern of misses that decrease with set size is at present open, and may be related to increasing sensitivity to orientation discontinuities as element number increases (Nothdurft, 1991; 1993).

The real power of the use of the MTS model is not just to *fit* data, but rather to reduce the complex patterns of RT and error into psychologically tractable entities. In the following sections, I systematically compare how model parameters corresponding to information quality (*S* and *V*), decisional integration (*C*), and attentional limitation (ε and σ_{INT}) vary across the experimental ensemble. In this way, I can reveal meaningful differences among search tasks that are not readily apparent via the independent analysis of RT or error. A complete listing of all the average parameter values used to *best* account for each search experiment is included in Appendix V.

II.C.1 *Variation in* σ_{INT}

The first major division of search tasks revealed by the modelling is in terms of σ_{INT}. Recall, that each experiment was simulated using three *A-type* models: one containing no additive noise (i.e. an *A*(0) or *D-type* model), one containing a moderate amount of noise (an *A*(.25) model), and one containing a higher level of noise (the *A*(.5) model). The models were tested using repeated measures *t*-tests to compare the SSQ_{err} at each level of internal noise (all the models were basically indistinguishable in terms of the SSQ_{RT}). For all comparisons, the "best" fitting model was assumed to be the basic model with no internal noise. In this way, a model with internal noise was deemed "best" only if it significantly reduced the residual variation relative to the *A*(0) null (a subsequent test was performed on the residuals of the *A*(.25) and *A*(.5) models if both provided significantly better fits than the *A*(0) model). Results of these tests revealed that only a small subset of the ensemble required σ_{INT} to be greater than 0. Specifically, the three rotation experiments [*rotation 2D (textures)*, *rotation 2D (pinwheels)*, and *rotation 3D (coins)*], and the *implied lighting (left/right)* experiment were fit significantly better by a model with $\sigma_{INT} = .5$, relative to fits produced by models with lower values of internal noise. The *conjunction* and *circle-plus* experiments were fit best by an *A*(.25) model with intermediate levels of internal noise. The remaining group of 20 experiments were fit best using a model that incorporated no additional source of internal noise (i.e. the *D-type* model).

One of the primary advantages of using a model with additive noise is that it gives the standard *D-type* model the ability to account for high miss-rates in the context of sharply increasing RTs. Accordingly, the subset of experiments singled out as requiring the addition of internal noise were exactly those experiments having some of the largest set size effects in terms of miss-rates and RT. For example, this subset produced miss-rates close to 25% for displays containing 3 distractors and 1 target, as well as associated increases in RT of 150 to 200 msec. In addition, this subset of experiments differed from the remainder of the ensemble in that only these experiments manifested a relatively constant false-alarm rate across set size. This stands in contrast to the general ensemble-wide pattern of false-alarms that decrease with set size. As pointed out earlier during the comparison of the various MTS models (see section II.A.2), the class of *A-type* models having a $\sigma_{INT} > 0$ is incapable of producing decreasing false-alarms, though this class can easily yield invariant or increasing patterns of false-alarms with set size. With this limitation in mind, it is really no surprise that the *A*($\sigma_{INT} > 0$) models outperformed the *D-type* model on only a small handful of search tasks – the vast majority of the experiments in the MTS ensemble have decreasing false-alarms and thus cannot be simulated by a model with high levels of internal noise.

It is worth reiterating, that though the parameter σ_{INT} provides a qualitative division of search tasks, this division derives entirely from differences in the SSQ_{err} (most notably from the residuals associated with the set size 4/1 target miss-rates, and the overall slope of the false-alarm rates); the residuals associated with the patterns of RT had no effect in deciding the appropriate magnitude of σ_{INT}, in so far as all the models were basically indistinguishable in terms of how well they accounted for the RT data. Thus, I am somewhat cautious in making strong claims relating variations in the parameter σ_{INT} with qualitative differences in attentional processing.

Nonetheless, the model comparisons appear to provisionally categorize the ensemble of search tasks into two distinct classes: those that require the injection of internal noise, and those that do not. From here on, I will denote these two classes of search as the *A-class* and the *D-class*, respectively.

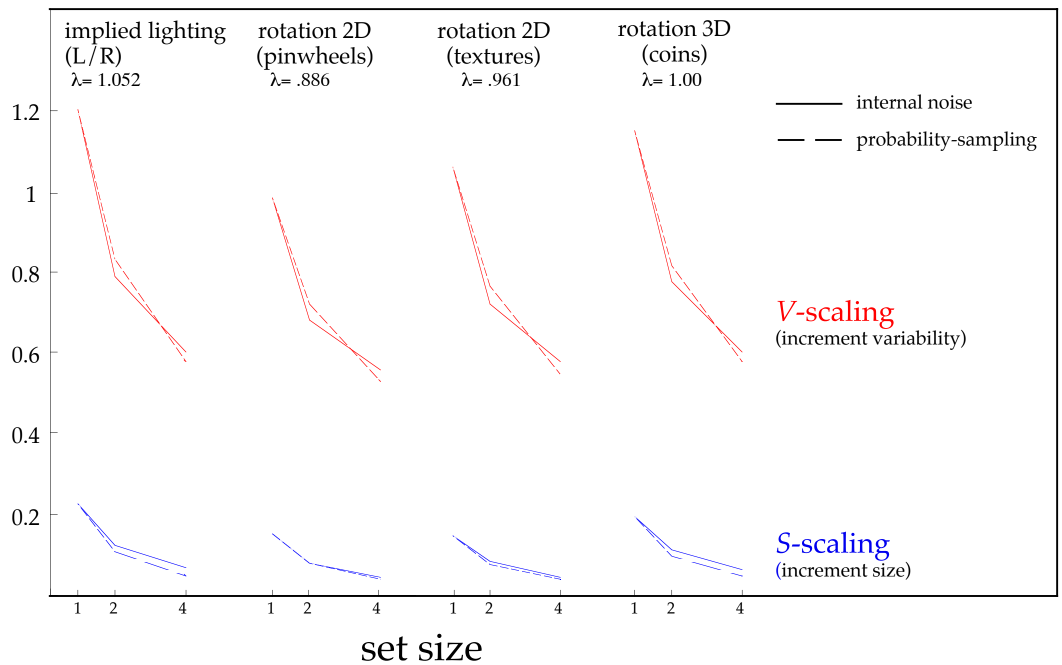
Probability sampling models

In section II.A.2.b we saw that there is a near equivalence between the dual-noise source model parameterized by σ_{INT}, and the probability-sampling model parameterized by λ. Here, I explicitly show that the data from the majority of *A-class* experiments can also be simulated using a matched probability-sampling model. I reiterate here, that the probability-sampling model is not suitable to account for the remaining set of 20 *D-class* experiments, for the same reasons that prevented an account based on internal noise: neither the dual-noise nor the probability-sampling models can produce a pattern of false-alarms that decrease with set size (recall, that this is the general pattern associated with the *D-class* of search tasks).

The general approach for simulating a probability-sampling model of MTS is quite simple. First, I use curve-fitting routines (Matlab) to find the value of λ that comes closest to reproducing the scaling effects of the "best" fit *A-type* model. I then simulate a matched probability-sampling model based on that λ, and a parameter set matched to that of the *A-type* model (i.e. the same *S*, *V*, *C*, *D*, M_R, and *Z*). Specifically, I find that λ which minimizes the deviations between: 1) the function governing the attenuation of the increment means and standard deviations with set size based on the "best" fit $A(\sigma_{INT})$ model, and 2) the same function based on the probability-sampling

Figure 38

Comparison of increment scaling for internal noise and probability-sampling models



model parameterized by a particular choice of λ. For the range of set sizes used in MTS, it is often possible to find a value of λ that closely mimics the scaling effects of the appropriate $A(\sigma_{INT})$ model. In the adjacent figure I show the results of matching λ to mimic the increment scaling functions of the best fitting *A-type* model for a subset of four experiments (the scaling functions for the step size *S*, and variability *V* are plotted in blue and red respectively). As the figure makes clear, the matched probability sampling model comes remarkably close to replicating the scaling functions of the internal noise model at these set sizes and levels of σ_{INT}. All else being equal, any model that identically scales the increment distributions with set size, will produce identical predictions in RT and error-rate. This was in fact the case, as the simulations of the full probability-sampling model based on the values of λ shown in figure 38 generated RT and error fits that were indistinguishable from fits of the dual-noise model with a σ_{INT} of .5. That is, the probability-sampling model provided an excellent account of the joint patterns of RT and error data for all three *rotation* experiments, as well as the *implied lighting (LR)* experiment [λ = .961 for *rotation2D (textures)*, λ = .886 for *rotation2D (pinwheel)*, λ = 1.00 for *rotation3D (coins)*, and λ = 1.05 for *implied lighting (LR)*]. That all four data sets required a matched probability-sampling model with similarly high values of λ (avg. λ = .974) , bears further witness to the extreme levels of capacity limitation inherent in these types of search – at these levels of λ the probability-sampling framework approaches a *fixed sample size* model of attentional limitation in which the probability of sampling an element goes as the inverse of the set size (Shaw, 1980; Palmer et. al, 1994).

The probability-sampling model was not as successful in accounting for the patterns of data associated with the two remaining experiments in the *A-class* which were fit by a model with σ_{INT} = .25 (i.e. the *conjunction* and *circle-plus* experiments). This failure arises primarily because the equivalence between the $A(\sigma_{INT})$ and probability-sampling models breaks down at lower levels of internal noise. Specifically, the $A(.25)$ model can still capture mild decreases in false-alarms with set size that the probability-model cannot.

There are several important conclusions to be drawn from the results of these simulations. First, the *A-class* of search experiments can be further divided into those that permit a probability-sampling account of attention (the various *rotation* and *implied lighting (LR)* experiments), and those that do not (the *conjunction* and *circle-plus*). In general, it appears that any data which can be fit by an $A(.5)$ model can also be explained in terms of probability-sampling. However, the same equivalence does not hold for experiments fit by an $A(.25)$ model, though the results are based on only two experiments. Second, there is nothing at the level of these MTS data to decide among an account of attentional limitation based on increment scaling (the $A(\sigma_{INT})$ model), and an account based on probabilistic sampling of unscaled increments (probability-sampling). At present, all one can say is that a subset of the MTS ensemble is best simulated by a model with internal noise, and that within that subset, only those experiments fit with relatively high levels of internal noise have a matched model based on probability-sampling.

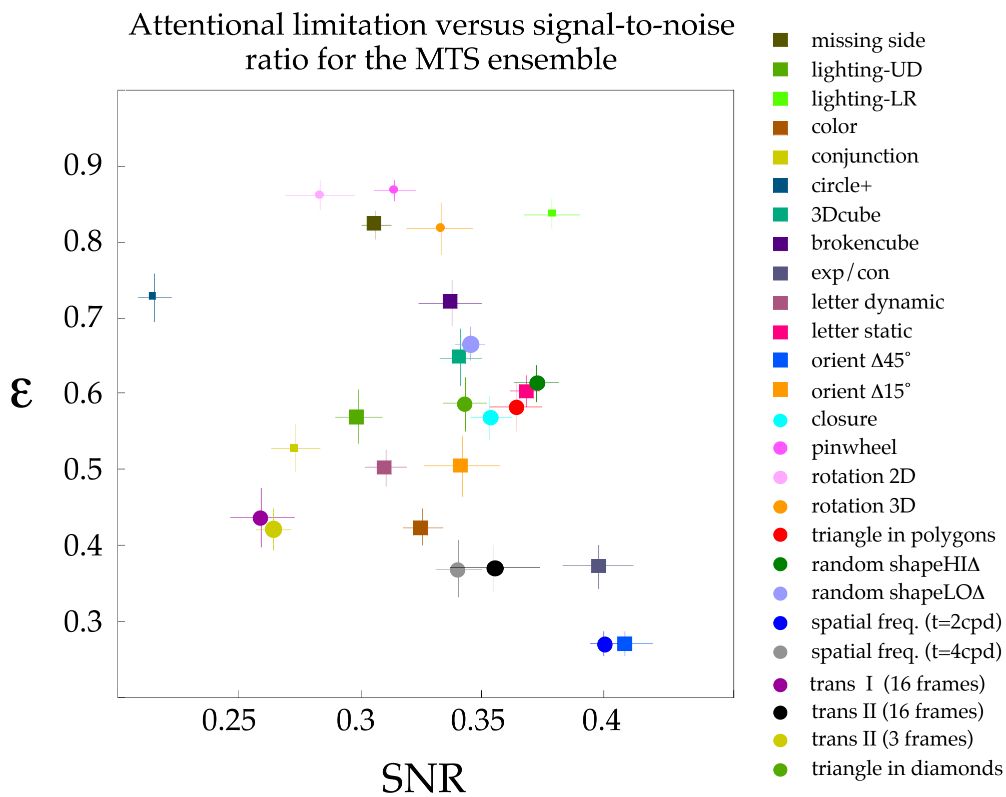
Interestingly, there is a common thread that relates all 6 experiments categorized within the *A-class*: in every case, targets and distractors can be construed as differing in the relative arrangement of component *features*. This relationship is especially clear for the *circle-plus*, and *implied lighting (LR)* experiments. In these experiments distractor stimuli differed from target stimuli solely in terms of a mirror-reversal. For the remaining variants of the *rotation* and *conjunction* experiments, targets and distractors can be construed as differing in terms of feature configuration. For example, searching for a *clockwise* rotation among a field of *counter-clockwise* distractors can be adequately conceptualized in terms of different arrangements of motion energy *features* of the sort represented in V1 (see Sekuler, 1992; Takeuchi, 1997 for an analogous way of conceptualizing complex motion direction). Similarly, elements in the *conjunction* experiment can be defined as configurations of orientation and color.

However, it is important to realize that not every member of the class of *relative position* experiments required an internal noise model. For example, the *missing side* and *broken cube* experiments are paradigmatic examples of search for relative position, yet both were fit best by a *D-type* model without additive noise. Nonetheless, it is striking that every member categorized as *A-class* requires a configural analysis to distinguish targets from distractors. This class homogeneity is all the more interesting because there is a growing body of visual search results to highlight the especially demanding nature of these sorts of discriminations (Logan, 1994; Pashler, 1998; Wolfe, 1998b). For example, in the context of measurements of search accuracy during briefly presented displays, only stimuli differing in the relative arrangement of features produced increases in threshold large enough to invoke perceptual limitations (Palmer, 1994; Poder, 1999). If we take seriously the distinctions between the *D-type* and *A-type* models as being *decisionally* and *attentionally* limited respectively (see section II.A.1.f), then we have here an account of supra-threshold search that is in agreement with the basic findings for search defined at threshold – namely, that the majority of visual search effects can be accounted for in terms of decision (i.e. the *D-class*), with true perceptual limitations arising only for discriminations of relative position (the *A-class*). Despite the apparent simplicity of this summary of the MTS results, it requires a *D-type* model of search burdened with an implausible number of criteria. Moreover, as will become evident in the following section, an ensemble analysis based solely on σ_{INT} ignores the undeniable fact that the *D-class* of search experiments can be meaningfully ordered in terms of ε.

II.C.2 *Variation in* ε

The ultimate goal in the modelling of MTS is to extract attentional limitations using a model that simultaneously accounts for both set size effects in RT and accuracy. In the adjacent figure I show the values of ε

Figure 39



required to simulate the data from all 26 experiments in the ensemble. Specifically, figure 39 plots the average best-fit value of ε against a measure of target/distractor discriminability [basically, a signal-to-noise ratio (SNR) = $2S/V$]. Note that the 6 smaller points within the scatterplot denote the subset of experiments discussed previously that were fit best by an $A(\sigma_{INT}>0)$ model (i.e. a model with additive internal noise). The fits clearly reveal that the entire MTS ensemble can be reliably distinguished using a parallel model and a continuum of capacity limitation. The reality of this continuum argues against the categorization of search tasks using simple-minded dichotomies such as *serial/parallel*, or *efficient/inefficient* (Wolfe, 1998a,b).

The figure also reveals that a substantial amount of variation in ε was necessary in order to capture the entire ensemble. Specifically, some data sets were accounted for by relatively low levels of ε of around .3, indicating that for this class of experiments there was very little effect of set size on the acquisition of stimulus information. At the other extreme, were experiments that required near maximal levels of capacity limitation (i.e. ε ~ .9). For this class of search, the rate at which stimulus information was accumulated in each random walk was effectively reduced by a factor approaching the inverse set size. Intermediate to these two extremes were those experiments that required moderate values of attentional limitation. It is noteworthy that in the context of MTS, every task in the ensemble was simulated using some degree of capacity limitation. This is consistent with recent evidence that there are no discriminations that can truly be termed "preattentive" in that even the simplest stimulus discriminations (e.g. orientation) appear to make some demand on attentional resources (Joseph, Chun, & Nakayama, 1997).

In addition, figure 39 shows that the observed variation in ε is theoretically well-ordered, in that it segregates and groups experiments in a psychologically intuitive way. For example, searches based on differences along primary featural dimensions tend to inhabit regions of the space characterized by mild limitations, such that ε is typically less than .4 (e.g. *spatial frequency*, *orientation*, *color*, *translation*, and *expansion/contraction*). At more moderate levels of limitation ($.4 < \varepsilon < .7$), are those search tasks based on either shape discrimination, or the discrimination of emergent form (e.g. *letter dynamic/static*, *closure*, *implied lighting (up/down)*, *triangles in polygons*, etc.). Finally, the highest levels of ε are reserved for the most difficult class of search tasks based on discriminations of relative position or rotation direction. In general, these tasks appear to be doubly limited in that they are simulated using high levels of both ε and σ_{INT}.

There is also evidence that the local ordering of experiments by ε is theoretically consistent. For example, consider the relative difference in attentional limitation found for the two *implied lighting* experiments. Recall, that the *up/down* version of this experiment consists of apparent differences in surface curvature (*bumps* or *indentations*) that emerge by virtue of shading and the visual system's presumed bias to assume an overhead lighting source (Ramachandran, 1988). This experiment is contrasted with the *right/left* version (identical target/distractor stimuli rotated by 90°) for which there is no emergent distinction in curvature, and hence the search task here reduces to a strict discrimination of relative position. The stimuli from these two experiments are known to qualitatively dissociate when they are embedded in textures; the *up/down* stimuli produce vivid intra-region groupings and inter-region segmentations, whereas the *left/right* stimuli lead to camouflage. Results based on the extraction of ε via the model provides a similar dissociation among these stimulus sets in that the *up/down* task is found to be moderately limited in capacity (avg. ε= .568), while the *left/right* task was found to suffer extreme limitations on par with the most difficult types of search (avg. ε= .85, σ_{INT} = .5).

Other evidence that ε is meaningful in terms of the small-scale ordering of search tasks it provides comes from a comparison of the *3Dcube* and *broken cube* experiments. Recall, that these experiments are based on experiments by Enns and Rensink (1990) that examined the role of perceived orientation in depth during search. The two experiments differ as follows: the *3Dcube* experiment consists of simple, 2D rendered "cubes" that appear to have an orientation in depth, whereas the *broken cube* experiment takes these same cube-like stimuli, but separates the faces of each *cube* so as to effectively eliminate perceptions of depth. Results from a series of studies based on singleton search revealed that the *broken* cubes were much less efficient relative to search for intact cubes (Enns & Rensink, 1990). Again, the relative values of ε extracted using MTS are consistent with these results. Search for intact "cubes" was significantly less capacity-limited (avg. ε= .647), than search based on a set of identical, but *broken* stimuli (ε= .721). Later, in section II.C.6 I will show that ε can also reliably pick up explicit manipulations in target/distractor similarity that are expected to make search more difficult (Duncan & Humphreys, 1989).

On the whole, both the *global* and *local* ordering of experiments via ε provides strong evidence that the multiple target search method and model are well-calibrated: stimulus differences that lead to targets that "pop-out" in a field of distractors and produce vivid textural segmentations have near minimal levels of ε, while those differences that support camouflage have higher levels of ε. Furthermore, the subset of search tasks requiring an analysis of relative position were found to require the highest levels of ε. This result agrees with the qualitative categorization of experiments via the noise parameter σ_{INT}, and provides additional evidence that these sorts of discrimination are singularly demanding of attention (Palmer et. al, 1993; Pashler, 1998; Poder, 1999).

As alluded to earlier (section II.A.1.f), there are two possible explanations concerning why ε should vary across experiments. These two explanations arise because there exists an inherent symmetry in the random walk framework between the scaling of increments and the scaling of criteria. The first explanation is *attentional* in nature, and holds that ε controls the rate at which information is accumulated via a multiplicative scaling of the perceptual samples as a function of set size and capacity limitation. By this account, the variation seen in figure 39 arises because search stimuli differ in terms of the demands they make on attentional resources.

The second explanation is *decisional* in nature and holds that ε controls the magnitude of the primary and secondary response criteria as a function of set size. By this account, the variation in ε arises because search stimuli have intrinsic differences in single-element discriminability (i.e. the discriminability of a single target from a single distractor). Experiments with lower single-element discriminabilities will necessarily have higher error rates for a fixed criterion (T) – to control for differences in the intrinsic error, ε multiplicatively adjusts the magnitudes of the response criteria as a function of set size. Though this account places ε at the level of decision, it is still true that the magnitude of criterial scaling has to be a function of set size. In a sense then, even this account is attentional in that the value of ε is ultimately determined by interactions in the multi-element data. The real question then that remains, is not whether the parameter ε operates via scaling the criteria or the perceptual samples, but rather why this scaling varies in a psychologically sensible way across the ensemble. In the next section I argue that the observed variation in ε across the ensemble is <u>not</u> determined by intrinsic differences in stimulus discriminability, but rather by attentional limitations.

II.C.3 *n-element confusability*

A central issue in the comparison of visual search results across disparate stimulus domains concerns whether observed differences in patterns of RT and/or accuracy are due solely to limitations at the level of attention, or rather arise from lower level differences in the intrinsic discriminability of target elements from distractor elements. It is well documented, all else being equal, that basic differences in the discriminability of single elements is sufficient to explain much of the variability in set size effects on RT and accuracy during singleton search (Duncan & Humphreys, 1989; Palmer, 1994; Geisler & Chou, 1995). The MTS model(s) employed here contain parameters presumed to relate to single-element discriminability and attentional limitation, and so it is possible to explore how the values of these parameters covary within experimental fits.

Returning to figure 39, we see that there is in fact a slight relationship ($r^2 = .139$) between the magnitude of each experiment's estimated SNR (the ratio of S to V) and its associated value of ε – low SNRs tend to be associated with high levels of attentional limitation, while higher SNRs are seen to cooccur with decreased levels of attentional limitation. This relationship is in fact quite weak, being driven primarily by a cluster of experiments having near minimal levels of ε (e.g. *spatial frequency*, *orientation*). Assuming that the model-generated estimates of these SNRs are veridical, the relationship in the figure suggests that the parameter ε may be partly a measure of intrinsic information quality, rather than a pure measure of attentional limitation. To the contrary, I will present evidence that the covariation between these measures results not because ε is a function of target/distractor d′, but instead because the model-extracted SNRs, like ε, are reflections of the *confusability* of the *n*-element display.

There are a number of important points that argue for the psychological reality of ε as a measure of attentional limitation, independent of any intrinsic variation in discriminability. First, the apparent relationship shown in figure 39 is for the most part illusory. If we exclude the *spatial frequency* and *orientation* experiments from consideration, the relationship between SNR and ε virtually disappears (r^2 = .016), indicating the high leverage of these extreme data points. Furthermore, the observed covariation between SNR and ε arises as a direct consequence of the architecture of the model. To see this one must realize that the upper right quadrant of the SNR-ε space produces data simulations that have no errors. This is a result of the value of the primary criterion T, and the ranges of S and V employed in these simulations. Because the experiments in the MTS ensemble have a 5% error-rate on average, they cannot be fit using the combinations of *S*, *V*, and ε that define that quadrant. Moreover, because of the way attention is conceived in the MTS model, increases in ε generally lead to slightly lower errors and longer RTs (though this relation can be somewhat reduced by the addition of large amounts of internal noise, see section II.A.2.a). In attempting to account for both the error-rate and the RT data, the model emphasizes those parameter combinations that do in fact lead to errors. In practice, the model negotiates the tradeoff by increasing *V* (increment variability) to offset increases in ε (*V* and ε are correlated across experiments .6, while *S* and ε are uncorrelated). A similar relationship holds for the extreme lower left quadrant of the SNR-ε space. There is a paucity of fits in this region precisely because the parameter values there lead to patterns of data that are nonexistent in the use of MTS (either very low error rates, extreme false-alarm rates, or the absence of *pure* target/target absent mirroring).

The second point that argues against ε as a correlate of sensory limitation arises from an explicit assessment of target/distractor discriminability. I have conducted a series of informal psychophysical calibrations on a subset of the target/distractor pairs used in MTS, and the results contradict any clear relationship between empirically-measured d’ and ε. For example, using a procedure drawn from Geisler and Chou (1995), the *rot2D* and *translation*I stimuli were found to be equally discriminable across a large range of stimulus eccentricities (eye movements were controlled, contrast was minimal, and presentations were kept brief, ~120 msec.). That these two target/distractor sets proved equally discriminable, while at the same time yielding polar opposite values of attentional limitation (*rot2D*, avg. ε=.86; *translation*I, avg. ε=.38), provides strong evidence that intrinsic stimulus quality is dissociated from attentional limitation in the context of MTS.

In a similar series of psychophysical investigations based on single-element discrimination, I found that the stimuli in the *missing side* and *circle+* experiments were much more discriminable than those in the *spatial frequency*, *color*, or *orientation* experiments (all stimulus presentations were brief and were preceded and followed by a high contrast pattern mask; complete psychometric functions were measured as a function of presentation time). These results are all the more striking given that the highly discriminable *missing side* and *circle+* stimuli manifested extreme levels of *confusability* in the context of multi-element search (avg. ε was .88, compared to an avg. ε of .275 for *spatial frequency, color,* and *orientation*).

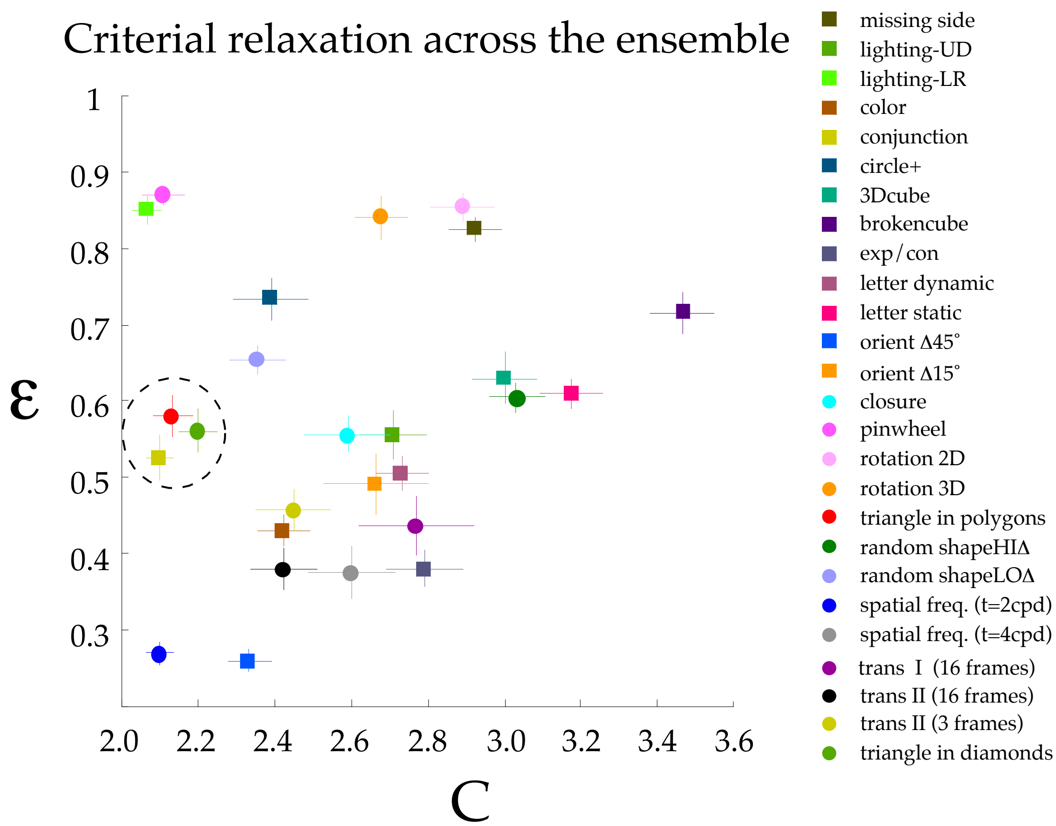
Finally, I fit a modified random walk model to the single-element RTs and error rates for all 26 experiments using simple analytic expressions for the time-to-absorption and the probability of an error absorption (eqs. 1 & 2). These fits produced SNRs that were entirely uncorrelated from the SNRs produced by the *full* model fits to the 18 RT and error-rate conditions. Furthermore, the single-element fits were highly dependent on the magnitude of M_R (the additive constant reflecting motoric and unexplained factors in the RT), yielding SNRs that were roughly invariant across experiments. This similarity is really not too surprising, given that with few exceptions, there was little variation in the single-element RTs and error rates across the experiment ensemble.

To summarize, the convergence of the previous arguments makes a strong case that ε is a measure of attentional limitation. When data are accounted for by a simulation based on higher levels of *V* and ε , it is not because a single target element is less discriminable than a single distractor element, but rather because the *n*-element target present displays are less discriminable than the corresponding *n*-element absent displays. For example, attentionally demanding search tasks generally yield high miss-rates and sharply increasing RTs as a function of set size. The model accounts for the RTs by increasing ε and fits the associated high level of misses by increasing *V*, which in turn decreases the apparent discriminability. By this account, the apparent relationship between ε and the SNR arises simply because processing costs and *n*-element *confusability* are dual concomitants of attentional limitation.

II.C.4 *Variation in C*

The MTS models I use here are novel within the visual search literature in that they utilize a set of secondary criteria by which responses can be made on the basis of multiple lines of incomplete, but consistent evidence. Recall, that these criteria are implemented via the decisional parameter *C* which in principle, "relaxes" the amount of information necessary to initiate a response

Figure 40



when displays contain more than one element. For values of *C* greater than 1, set size effects on target absent RTs will be reduced in such a way that these conditions will "mirror" the *pure* target conditions. In addition, increasing *C* naturally leads to higher miss rates as it becomes more likely for the single target-generated random walk to exceed the more liberal secondary criterion associated with an "absent" response. In the following figure, I show the average value of *C* along with the associated value of e necessary for the model to simulate the 26 experiments in the ensemble.

First, note that though there is a substantial amount of variation in *C* across the ensemble fits, there appears to be no discernable correlation between this parameter and the model's estimates of attentional limitation: in general, the best fit value of *C* is independent of the value of ε. Second, there appears to be a clear logic governing some of the inter-experiment variation in *C*. For example, some of the variability seen in figure 40 may be explained by the twin notions of distractor heterogeneity and search difficulty. Consider the three experiments enclosed in the dashed circular region (*conjunction*, *triangles among diamonds*, *triangles among polygons*). This subset is distinguished from the majority of remaining experiments in the ensemble in that for these experiments non-targets were drawn from a heterogeneous set (due to differences in orientation, identity, or feature conjunction). Heterogeneity in non-targets appears to attenuate the integration of information, in that the average value of *C* for these three experiments is nearly minimal (M=2.156±.074sem), and reliably less than the remainder of the experiments (M=2.686±..022sem). That experiments with heterogeneous distractor sets are fit by lower levels of *C* argues for the theoretical consistency of this parameter, and suggests that observers may have a harder time integrating information when elements are not easily alignable. Specifically, it appears that the random orientation of a single distractor shape (*triangles among diamonds*) engenders a sense of heterogeneity on par with that produced by distractor sets consisting of multiple unique elements (the *triangles among polygons* and *conjunction* experiments). In this regard, one could also speculate about the findings for the *implied lighting (LR)* and *rotation2D (pinwheels)* experiments: both data sets required minimal levels of *C* near 2. This result could reflect a perceived sense of heterogeneity in these distractor sets, such that observers become less willing to combine evidence from multiple sources of rotation or luminance phase. Though this is an intriguing conjecture, and one that has support from studies indicating very little region segmentation for these sorts of differences (Beck, 1966; Gilden & Kaiser, 1992; Julesz & Hesse, 1970; Malik & Perona, 1990; Renstschler, Hubner, & Caelli, 1988; Sagi, 1995), two other related experiments yielded average values of *C* that were reliably higher and more consistent with the ensemble average (see *rotation2D (textures)* and *rotation3D*).

Variation in *C* also appears to be partially related to the attentional demands of the search task. Specifically, the magnitude of *C* is reliably lower for searches based on simple psychological differences of the sort that lead to strong texture segmentation and spatial grouping (e.g. *spatial frequency*, *orientation*, *translation*, and *color*), relative to more attentionally demanding searches based on form, or relative position of features. One possible reason that *C* appears to be somewhat modulated by search difficulty may simply be that featural differences lead to low-level groupings that diminish the need for additional integration at the level of decision.

Regardless of the exact cause of these local variations in criterial relaxation, the most important psychological finding that emerges from these analyses is that every single experiment in the ensemble was fit using a value of *C* in the range 2 and 3. This is a testament to the psychological reality of the idea embodied in this parameter – that observers routinely reduce their criteria in a systematic way with set size so as to integrate information across independent walks. In the next section I argue that one reason this integration is ubiquitous in MTS, is because it reflects an optimal-like decision strategy.

II.C.5 *The optimality of using partial evidence to guide decision*

On the whole, there is one hallmark pattern in the data that distinguishes MTS from other search paradigms: target absent RTs are consistently either invariant, or decreasing with set size. Nowhere in the use of MTS do I find target absent RTs that increase with set size in the 2:1 ratio predicted by a *serial, self-terminating* process. In fact, as noted throughout this paper, the MTS target absent data are entirely inconsistent with both *serial* and *parallel* classes of search. Moreover, there is no model that posits complete identification before a response is initiated that can predict target absent RTs that decrease with set size.

To successfully account for the error and RT data, the MTS model(s) necessarily incorporate secondary response criteria so that decisions can be made on the basis of multiple lines of incomplete evidence about element identity. This addition by its very definition embodies a speed/accuracy tradeoff – one can use multiple sources of partial evidence to decrease the cost associated with target absent responding, but only at the cost of increasing the miss rate. Given this state of affairs, it is natural to ask whether this sort of decisional opportunism represents an optimal strategy?

To answer this question I have formulated a random walk model of MTS that makes decisions using a Bayesian observer. The Bayesian observer is optimal because its decisions are made by choosing whichever response alternative maximizes the likelihood of the current stimulus information (for a detailed derivation see Appendix II). In the context of MTS, the Bayesian observer achieves optimal performance because it embodies perfect knowledge regarding the distribution of target and distractor increments, the statistical properties of a general random walk, and the exact prior probabilities of the various trial types. With this information, the Bayesian decision-maker can be used to a formulate the optimal accuracy for distinguishing target present from target absent displays at a fixed moment in time. Specifically, I implement a Bayesian framework for multiple target search in the following manner:

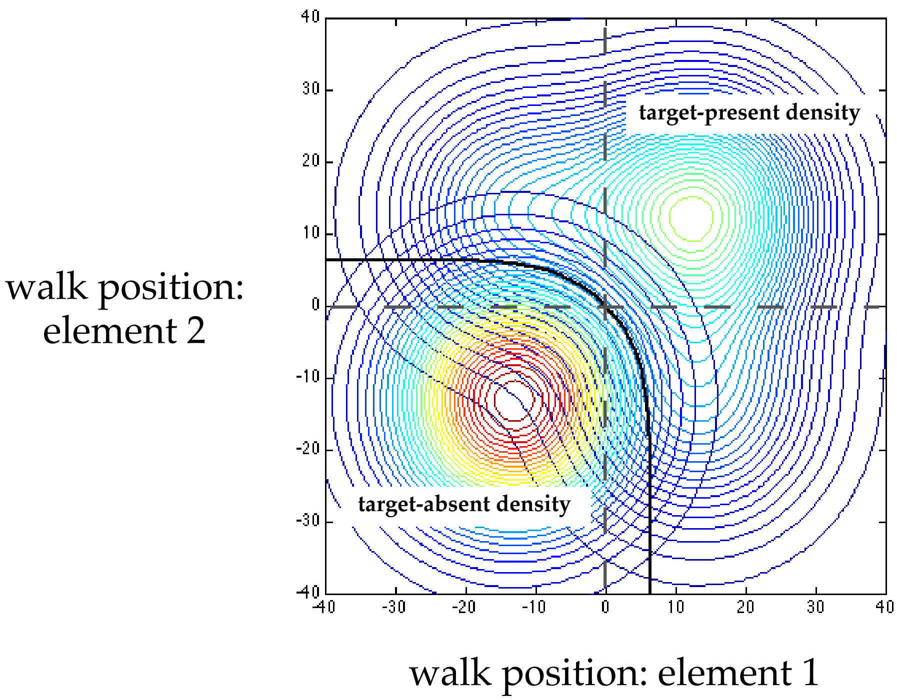
i. Each element in a search display is represented as an independent random walk with increments drawn from a Gaussian distribution (i.e. $N(\pm S,V)$). For all cases, the parameters S and V were fixed to average values based on the model fits (altering these values had no qualitative effect on the results).

ii. The distribution of walk states, in the absence of absorbing barriers can be modelled as asymptotically Gaussian, such that at time $\boldsymbol{\tau}$ the walk states are distributed as $N(\pm S\tau,\ V\boldsymbol{\tau}^{1/2})$.

iii. There is no attentional limitation, and hence all processing is capacity <u>un</u>limited and independent of set size (note that attentional limitation can easily be incorporated within the front-end of the model).

In the following two figures I plot the results of using the Bayesian model of MTS. Figure 41 shows a representation of the 2D probability density functions (i.e. $n = 2$) of possible walk states after 50 time steps have

Figure 41



elapsed for both the target present (upper-right), and the target absent case (lower-left). In this figure the densities are depicted as a series of two-dimensional level curves each representing the iso-probability of the joint position of two independent random walks in the state space. Note that the target present density is formed by taking a weighted average of the densities for the three types of target present trials given a set size of two (i.e. $walk_1$=*target* & $walk_2$=*distractor*; $walk_1$=*distractor* & $walk_2$= *target*; $walk_1$=*target* & $walk_2$=*target*). Similarly, the target absent density is the probability of the joint position of two distractor-generated walks. The thick black line represents the optimal decision bound assuming no bias for distinguishing target present from target absent, and is simply that region of the space for which the difference between the target present and target absent densities is zero. The optimal classifier will respond target "present" whenever the joint state of the two random walks is above or to the right of the decision bound, else it will respond target "absent".

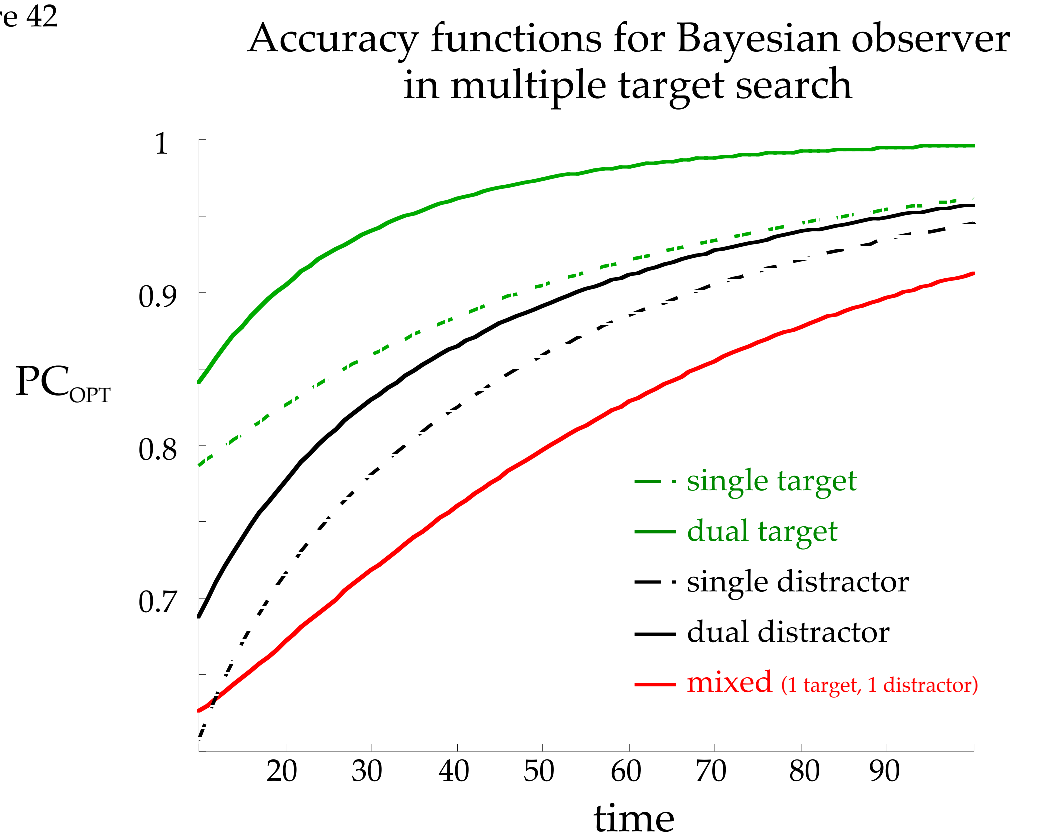
There are two important points concerning the shape of this decision bound. First, it is non-linear in the state space, and is in fact remarkably similar to the structure of the secondary criteria implemented in the standard model via the parameter *C*. Second, the $n = 2$ bound is shifted closer to the target present density relative to the decision bound for a set size of 1 [the n=1 bounds are the dashed gray lines in the figure and are included for comparison – they are centered at zero and represent the standard un-biased criteria of signal detection for the case of overlapping 1D Gaussians]. All else being equal, this shift in the 2D bound means that small positive accumulations in one of the walks will be outweighed by large negative accumulations in the other. This is the first indication that an optimal classifier will make target "absent" decisions without exhaustive processing.

In the next figure I use the optimal 2D bounds to show how accuracy grows with time for all the n=1 and n=2 conditions of MTS. Specifically, the optimal percent correct (PC_{opt}) is plotted as a function of the number of time steps (τ), for single-target, single-distractor, mixed (i.e. 1 target + 1 distractor), dual-target (i.e. $pure_{n=2}$), and dual-distractor trials (i.e. 2 distractors). The accuracy curves were formed by calculating PC_{opt} across a range of linearly-spaced values of τ using the expressions in Appendix II.

There are a number of important points contained in these curves. First, all the accuracy-by-time functions grow with τ in an approximately square-root fashion. This is because the signal/noise ratio of any random walk grows as $\tau^{1/2}$. Second, the curves are well ordered with time such that the same ordinal relationships between the accuracy of various conditions are preserved with time. For example, the accuracy for single element trials is always greater than for mixed trials. Because the relation between τ and PC_{opt} is monotonic, it is also the case that the single element trials will reach a criterion level of accuracy sooner than the mixed trials. Thus, it is clear that even for a capacity-unlimited model with an optimal decision structure, there will be set size effects in visual search. This cost in RT has been termed *decisional* limitation in the literature, and arises from the fact that larger set sizes lead to an increase in statistical uncertainty that exists prior to any influence of attention (Palmer & McLean, 1995, Palmer, 1994, Eckstein, 1998).

Now, consider the single and dual target curves. For all τ, the dual-target curve (n=2) lies above the single-target curve (n=1) implying a redundancy gain in accuracy (and therefore, in time as well) for displays with

Figure 42



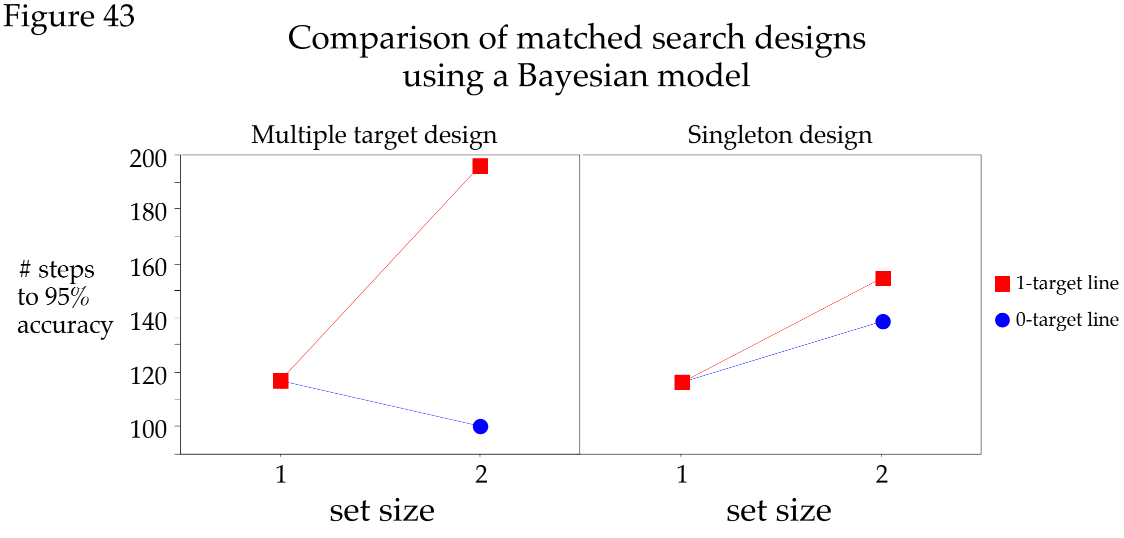
two targets, relative to displays with one. The magnitude of this redundancy gain is seen to grow slightly with τ, and reflects a benefit in response due entirely to the statistics of the minima of independent processes (for a discussion of these so called "race" gains see Raab, 1962; Miller, 1982). More importantly, a similar relationship is seen to hold for the case of dual distractors relative to a single distractor. As for the case of redundant targets, the two distractor curve lies above the single-distractor curve for nearly all τ, thus implying higher accuracy and faster responding relative to the single distractor condition. This relationship clearly indicates that an optimal classifier will generate target "absent" responses that decrease with set size in the context of a multiple target search design.

For comparison, I have also used the Bayesian classifier in a singleton search design and here the results are strikingly different. In singleton search there are no multiple target trials and increases in set size simply reflect the addition of distractors. For this type of design, the optimal classifier predicts that target absent RT do <u>increase</u> with set size. This is the intuitive result assuming exhaustive processing in which target absent responses are limited by the slowest of n processes. Thus, it is clear that the ubiquitous "mirroring" between target absent RTs and *pure* target RTs in MTS is due in part to the addition of *pure* target trials in the experimental design – this addition alters the overall target present density and thus the structure of the optimal decision bound.

In addition to revealing a difference in the sign of the target absent slope, the Bayesian framework also predicts that the slope of the single-target function should differ across experimental designs. Specifically, the Bayesian model predicts a larger set size effect on RT in a multiple target design than in a matched singleton design. To see why, note that the optimal decision bound for MTS is shifted towards the target present density (Fig. 41). As mentioned earlier, this shift leads to faster target absent responding relative to the set size = 1 condition. The displacement of the decision bound also leads to an increase in single-target misses that in turn amplifies the single-target RTs relative to the predictions for singleton search. This difference across designs is informative in so far as I always find larger set size effects in my MTS experiments than expected on the basis of the corpus of singleton search results. Most singleton search experiments that have used highly discriminable target and distractor elements indicate that *featurally*-based searches (i.e. size, color, orientation, translation) produce very shallow slopes on the order of milliseconds. For comparison, my *feature* experiments with equivalent stimuli yield substantially larger set size effects that are on average about two to three times greater than expectation. Though multiple target search differs in a number of important ways from the standard singleton search experiment, the Bayesian predictions indicate that the addition of multiple-target trials is at least partly responsible for the general augmentation of set size effects that arise in the use of this method.

In figure 43 I summarize these relationships by using the accuracy-by-time functions in figure 42 to interpolate expected RTs across conditions and search designs assuming a fixed level of error. In the left panel, I plot the pattern of predicted RTs for a fixed 5% error-rate for a multiple target search design. In the right panel I show the predicted RTs for a matched singleton search design. Summarizing the previous discussion, the target absent slopes are seen to differ in sign and magnitude across experimental designs, and the single-target set size effects are clearly larger for a MTS design.

Figure 43



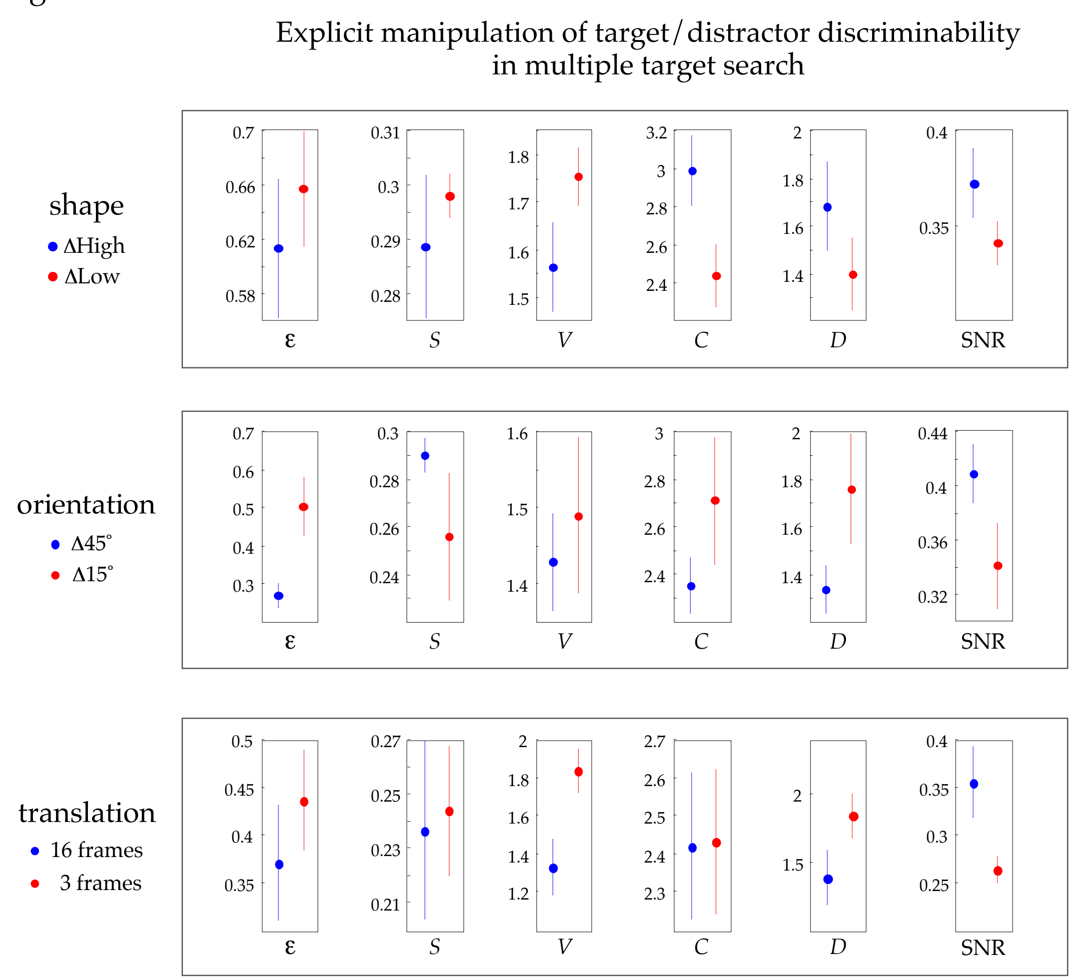
II.C.6. *Explicit manipulation of the signal-to-noise ratio*

In preliminary *fits* to MTS data I found very little variation in the SNR (i.e. $2S/V$) across a pilot ensemble of ten or so experiments. Moreover, the ensemble-wide homogeneity in apparent discriminability was found to coexist with large and significant differences in the attentional limitation parameter ε. There are two possible explanations concerning why the SNR was found to vary so little across that ensemble: 1) the various search stimuli in those experiments were in fact equally discriminable, or 2) the model was failing to pick up actual differences among experiments, possibly due to the structure and implementation of that particular MTS model (the value of the primary criterion T was free to vary, the increment variability, V was fixed at 1, and the parameter S had only two potential values in the simulation space). It is quite possible that real differences in discriminability were obscured by the coarse structure of the space and/or by variation in the decisional parameter T. In my current work I have attempted to deal with this problem as follows. First , I have substantially altered the structure of the model: now T is fixed and the parameters S and V are free to take on 10 potential values each. Thus, I let the increment parameters vary freely and then calculate the SNR from their best-fit values. Recasting the MTS models in this manner proved successful in that now the model fits reveal substantial, and reliable variation in the SNR across the 26 experiments in the ensemble (e.g. see figure 39). Second, and more importantly, I have included experiments in the current ensemble that are designed to explicitly test whether the MTS model can pick up manipulations of target/distractor (*t/d*) discriminability. The key idea here is to look for measurement consistency in the model – if two experiments differ in terms of *t/d* discriminability, then the model should accordingly reveal that difference in the estimates of each SNR.

There are three specific search domains included in the ensemble that are relevant here (for each domain 2 matched experiments were run in which *t/d* similarity was explicitly manipulated for a total of 6 experiments). The first 2 experiments are based on *shape* discrimination in which observers search for a target circle(s) among non-circular, random contour distractors. The extent to which the distractors deviate from circularity is manipulated across experiments, such that in the first experiment the distractors are highly, non-circular (experiment 8), and in the second experiment the amplitude of this non-circularity is reduced by half (experiment 9). The second set of experiments are based on *orientation* discrimination in which observers search for highly tilted Gabor targets (45°) among either vertical distractors in experiment 3 (i.e. Δ45°), or moderately tilted distractors (30°) in experiment 4 (i.e. Δ15°). I chose to contrast the 45 degree difference in orientation with search based on a 15 degree gradient because several studies have implicated a qualitative change in search efficiency around that boundary (see Wolfe, 1998b). The final set of experiments is based on the discrimination of *translation* direction. In the first version of the experiment (exp. 14) observers searched for rightward drifting textures among leftward drifting distractors; the motion stimuli consisted of 16-frame sequences of perceptually continuous motion (~800 msec of motion). In the comparison experiment (exp. 15), observers conducted the same task, but now with 3-frame sequences of motion (~150 msec of motion). For both experiments the various motion sequences were 100% coherent, and only the total number of frames in a motion sequence was manipulated. The idea was to hopefully introduce *t/d* discriminability differences by limiting the duration of the stimulus. Overall sequence length was chosen as a potential means of controlling discriminability here because it was easy to implement, did not alter overall stimulus visibility, and in the limit of a single frame sequence (i.e. *static*) insured that the *t/d* discriminability went to 0.

Both comparison experiments in each stimulus domain were analyzed separately using the *D-type* model (there were no cases in which an *A-type* model produced significantly better fits). In the adjacent figure, I show the average values of the 5 primary model parameters that best simulated the data from each experiment, along with the SNRs computed as twice the ratio of *S* to *V* (the 95% confidence limits for each parameter are included for reference). As the figure makes clear, all 3 stimulus domains revealed reliable differences in the model-based estimates of the SNR that were in direct agreement with expectation given the experimental manipulations. Specifically, for each member of the experiment pair in which the *t*/*d*

Figure 44



discriminability had been explicitly reduced, the model-extracted estimate of the SNR was significantly lower. This is an important finding because it indicates that the model can pull out small differences in *t*/*d* discriminability when they are known to exist.

Closer inspection of figure 44 reveals that the causes of the observed changes in the SNR differ across the stimulus domains depending on whether *S* or *V* are altered. For the *shape* and *translation* experiments the effective differences in the SNR are the result of associated differences in the increment variability parameter *V* (there were no significant differences in *S* for these experiments). In contrast, the discriminability manipulation engendered a significant change in the average value of S, but not V for the orientation domain. Thus, for the translation and shape experiments, decreases in *t/d* discriminability via reductions in presentation time or contour non-circularity appear to be associated with higher levels of *noise*, while attenuations of the *t/d* orientation difference are revealed in the model by reductions in the increment *signal*.

Interestingly, there are also differences in the estimates of ε as a function of the various *t/d* discriminability manipulations. For all 3 domains, decrements in discriminability were associated with higher values of ε, though only for the orientation experiments was this difference significant. Though it would be nice if manipulations of *t/d* discriminability were entirely independent of variation in the other model parameters, in the case of ε it is somewhat expected – there are both strong intuitions and empirical evidence to suggest that visual search becomes less efficient as target and distractors are made more similar (Duncan & Humphreys, 1989).

The overall picture that emerges then is that the MTS model is able to reliably distinguish explicit variation in *t/d* discriminability primarily via changes in the random walk increment parameters *S* and *V*, and less so in the attentional limitation parameter ε. That these explicit manipulations of single element similarity have a consistent effect on the model estimates of attentional limitation are somewhat troublesome, and suggest that at least some of the variability in ε across the ensemble may be due to inter-experiment differences in discriminability. While these sorts of uncontrolled differences across the ensemble may explain some of the local variation in ε, they are surely insufficient to explain the large-scale differences that separate featural types of search (avg. $\varepsilon = .35$), from searches based on the analysis of relative position or feature configuration (avg. $\varepsilon = .8$). In fact, many of the most limited types of search have *t/d* discriminabilities equal to or exceeding those in the feature class. For example, consider search based on the sign of translating and rotating textures (exps. 13-19) – we know from the psychophysical calibrations (sec. II.C.3, p. 118) that these experiments are approximately matched in discriminability, yet the two occupy qualitatively distinct categories of attentional process. Moreover, virtually all the relative position experiments contained maximally discriminable *t/d* pairs that nonetheless engendered highly *confusable* search displays.

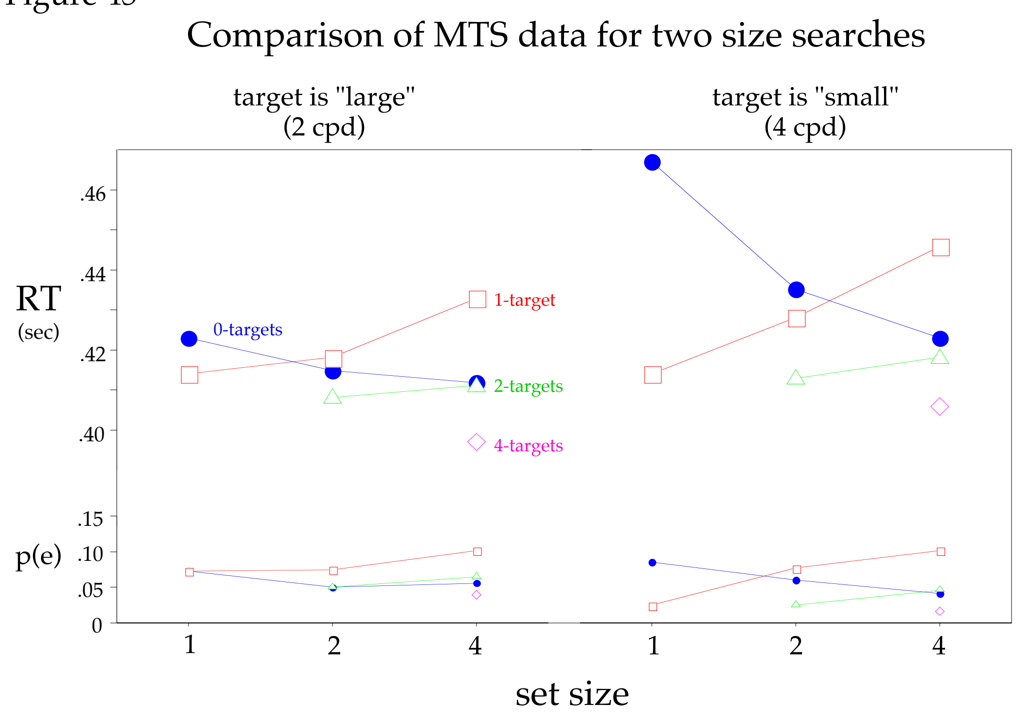
II.C.7 *Search asymmetries?*

There is a long history of search asymmetries using singleton designs (Treisman & Souther, 1985; Treisman & Gormican, 1988), and there remains continued interest in whether reversing the roles of targets and distractors alters search performance in an interesting way. These types of experiments typically look for dissociations in search efficiency depending on which of two base stimuli is designated as "target". More generally, a search asymmetry exists whenever there is a performance difference (e.g. speed, accuracy, efficiency, etc.) between search for an *A* among *B*s, relative to search for a *B* among *A*s (where *A* and *B* denote arbitrary search elements). For example, in the context of single-target search with extensive set size manipulations, search for a large item among small items is efficient and produces flat RT x set size slopes, while search for a small item among large items is relatively inefficient, producing much steeper slopes. Similar findings of asymmetrical search performance depending on choice of target have been reported for stimulus pairs distinguished by spatial frequency, orientation, and closure (Treisman & Gormican, 1988; Carrasco, McLean, Katz, & Frieder, 1998), and more recently for letter inversion (Wang, Cavanaugh, & Green, 1994), and discontinuities in contour curvature (Kristjansson & Tse, 2001). While early theorizing posited an explanation of the basic phenomenon in terms of either signal-to-noise ratios, or search for the presence/absence of canonical features (Treisman & Gormican, 1988), recent work has begun to question the ubiquity of such asymmetries once low-level confounds have been adequately accounted for (Carrasco & Frieder, 1997; Geisler & Chou, 1995).

Interestingly, there appears to be some weak evidence for corresponding asymmetries in the context of these MTS experiments. I discuss two experiments in the ensemble that are relevant here: *expansion/contraction* and *spatial frequency/size*. These two stimulus domains are pertinent in this context precisely because there is previous evidence for performance asymmetries in the context of singleton search (Treisman & Gormican, 1988; Takeuchi, 1997). In the case of *expansion/contraction*, Takeuchi (1997) has argued that search for an expanding target among contracting distractors is *parallel*, while the converse assignment appears to be accomplished *serially*. For search based on size there are a number of studies to indicate that *large* items are more readily sensed among *small* than the converse assignment (Beck, 1982; Carrasco, McLean, Katz, & Frieder, 1998; Gurnsey & Browse, 1987; Treisman & Gormican, 1988).

To explore the role of target assignment in MTS two independent versions of *expansion/contraction* and *size* search were run so that each member in a set of 2 basic search elements had the chance to play the role of "target". Figure 45 shows the RT and error-rate data for the two search experiments

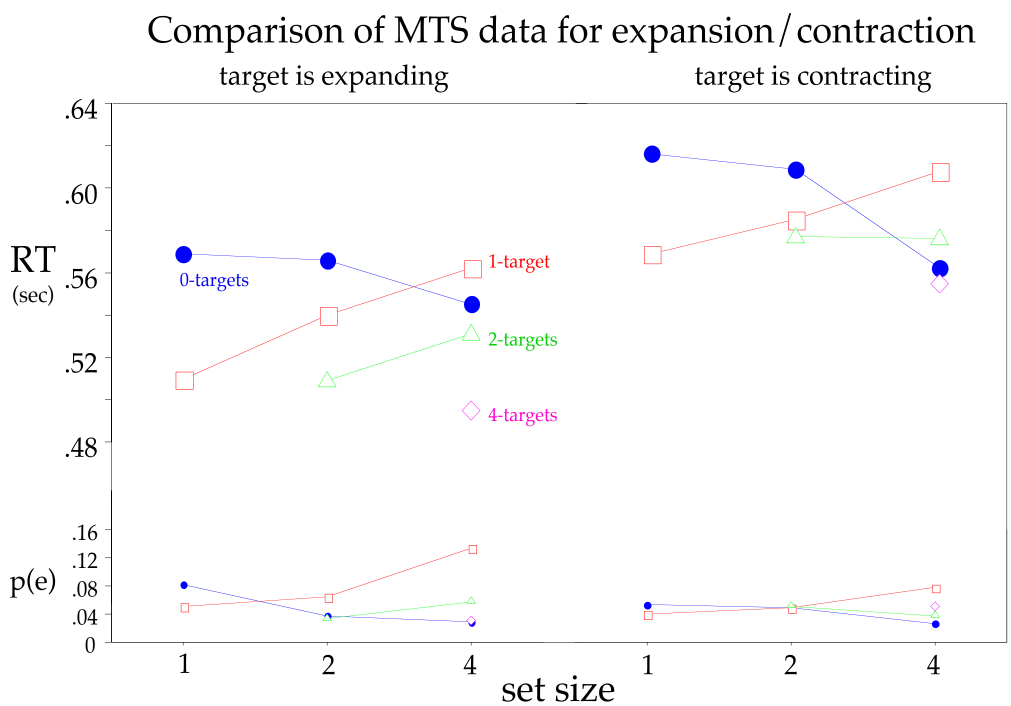
Figure 45



based on *size*. The data on the left represent search performance when targets are “large”, 2 cpd Gabors among “small”, 4 cpd distractors; the data on the right represent performance for the converse assignment. Clearly, there is a noticeable difference in the overall pattern of responding depending on target assignment. Search with “large” targets yields larger redundancy gains, and smaller set size effects for both RT and error-rate relative to search with “small” targets. Though this qualitative inspection suggests an asymmetry favoring “large” targets, the complexity of the interactions between the various RT and error-rate conditions undermines drawing any firm conclusions. For example, how do the larger benefits in target absent RT with set size, in addition to the lower *pure* target error-rates for "small" target search figure into comparisons of search efficiency? These sorts of questions reiterate the importance of using a model-centered analysis by which the dimensionality of the data can be reduced.

In figure 46 I show a comparison plot for the case of two search

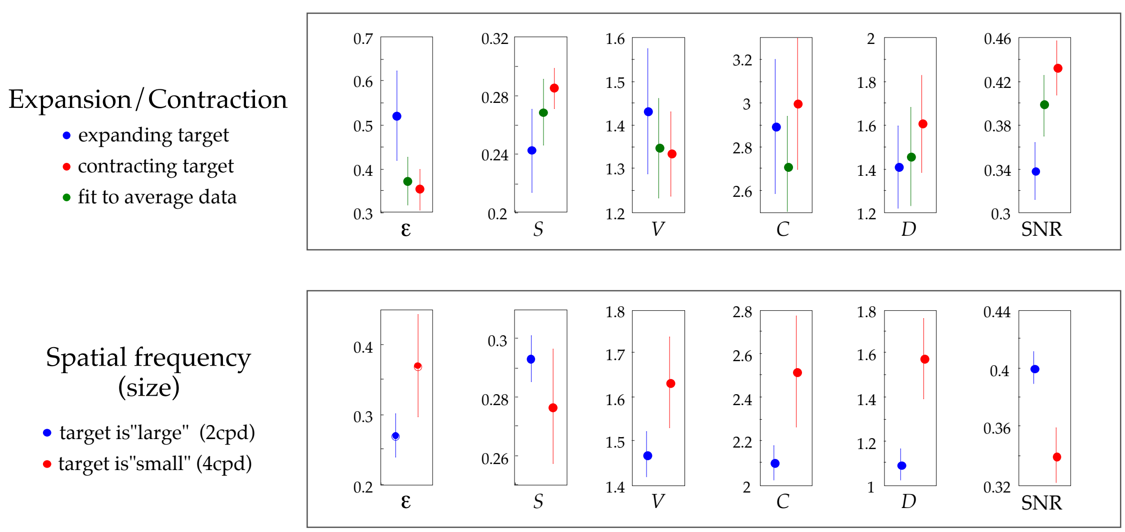
Figure 46



experiments based on the *expansion/contraction* of visual texture. Relative to the previous figure, there appears to be little qualitative difference between search with *expanding* targets and *contracting* distractors, and the converse assignment of *contracting* targets and *expanding* distractors. In a previously published paper, this apparent similarity motivated a subsequent pooling across target assignment, and the model fits to the *expansion/contraction* data reported here derive from this pooled data. Despite such qualitative equality, closer inspection reveals that there are in fact a number of small differences

Figure 47

Asymmetry effects in multiple target search



across the two experiments, most notably in the target absent RTs and miss-rates. Specifically, search with *expanding* targets produced larger misses with set size, while search for *contracting* targets manifested steeper target absent slopes. Again, we can use the MTS model to ask whether there are reliable differences in attention limitation (i.e. ε) or information quality (the ratio of *S* to *V*) to suggest a search asymmetry.

In the upper panel of figure 47 I show the results of using the *D-type* model to re-fit the two versions of the *expansion/contraction* experiment separately. The blue points represent the average values of 5 key parameters required to account for search with *expanding* targets, the red points represent the values required to account for search with a *contracting* targets, and the green points represent the parameter values that fit the data averaged over target assignment (the errorbars in the figure are the 95% confidence limits based on data re-sampling). A comparison of the various parameter values reveals two significant differences in the model fits across choice of target: search for *contracting* targets required a reliably lower value of ε, and a higher value of the SNR (due to a significant increase in *S*) relative to search for *expansion*. Taken together, these differences seem to suggest that search for *contraction* is superior attentionally to search for *expansion*. Though this finding runs counter to an earlier related investigation based on singleton search which found an apparent advantage for *expansion* at the level of the RT-by-set size slope (Takeuchi, 1997), there are other psychophysical studies to indicate either an advantage for either *expanding* motion (Ball & Sekuler, 1980), or *contracting* motion (Edwards & Badcock, 1993). Thus, without more work it seems suspect to draw any firm conclusions regarding the direction, if any, of a search asymmetry in the analysis of expanding/contracting motion.

In the second explicit investigation into potential search asymmetries , two versions of a spatial frequency/size based search were run. The 5 basic model parameters necessary to best simulate each data set appear separately in the lower panel of figure 47. Again there are asymmetries apparent in ε and the SNR, such that search for *large* targets requires lower values of ε and higher values of the SNR relative to the comparison search with *small* targets (here the change in the SNR arises via significant differences in *V*). This joint effect in the parameters implies that search for *large* among *small* may be attentionally privileged, a finding that has consistent support from a number of other studies (Beck, 1982; Carrasco, McLean, Katz, & Frieder, 1998; Gurnsey & Browse, 1987; Treisman & Gormican, 1988).

In addition to speaking to issues of asymmetry, these results also have bearing on the nature of the psychological referent of the SNRs that come out of the model fits. The question here is whether the SNR extracted by the model reflects the single-element *t/d* discriminability in the usual psychophysical sense (i.e. d'), or instead is a reflection of the discriminability of target present displays from target absent displays. If the latter situation holds, than we should expect the SNR to covary with ε in that both are sensitive to how set size affects performance. Recall, that in section II.C.3 I argued that the SNR primarily reflects *n*-element confusability. The matched comparisons documented here offer perhaps the strongest piece of evidence in support of that argument. The MTS model reveals apparent differences both in attentional limitation and in the SNR as a function of target assignment, despite the fact that the search elements in these experiments do not change, and hence neither does the single-element discriminability. This constraint implies that the observed variation in parameters is not the result of intrinsic differences in element discriminability. A more plausible interpretation is that together, *S* and *V* get pushed around in accounting for MTS data so that they come to reflect the quality of information available for categorizing the average multi-element display.

Though the explicit investigation of the role that target assignment plays during MTS reported here has revealed statistically reliable anisotropies, two important caveats are in order. First, it is important to keep in mind that these differences in attentional limitation are relatively small, especially when placed in context with the larger variation in ε across the experiment ensemble. For both experiments examined here, changing target assignment failed to produce the large qualitative change in processing of the sort previously reported using singleton search designs (Takeuchi, 1997; Treisman & Gormican, 1988). Another problem with these results is that for the expansion/contraction comparison, the MTS analysis provides evidence of an asymmetry that is entirely juxtaposed to earlier claims (Takeuchi, 1997). Finally, it is difficult to make a strong case for search asymmetries in the absence of a within-observer design. For both sets of experiments reported here different observers were used in each experiment. Though the MTS method with an *n* of 8 or more typically leads to reliably indistinguishable replications within experiment, it remains uncertain how much of the difference across these manipulations of target identity are due to different sets of observers. Clearly, an important area of future work will consist of the application of a repeated measures design in MTS to determine whether the asymmetries at the level of the best-fit model parameters documented here are reliably signaling veridical differences in search performance.

II.C.8 *Summary: attentional limitation and stimulus structure*

Section C has primarily focused on comparing the 26 experiments in the MTS ensemble in terms of their variation along the more theoretically interesting parameter dimensions. These comparisons have provided important insights into the way that information quality, attentional limitation, and decisional flexibility conspire to produce the complex pattern of RTs and errors that emerges during MTS. The ultimate goal guiding this research has always been to first *measure* how attentional limitation varies across a large number of important stimulus dimensions, and second, to *explain* this variation at the level of the stimulus structure. While this dissertation represents a significant stride towards a careful measurement of attentional limitations during search, I can only offer a tentative, first attempt at explaining the structural etiology of these limitations.

In the adjacent figure I have provided a summary of the ensemble variation in ε and σ_{INT}. This summary relates the model-based measurements of attentional limitation to intrinsic stimulus structure by grouping the ensemble into broad categories of search. This organization of the experimental results provides evidence for three global regimes of attentional limitation in search:

1. A *featural* regime, characterized by relatively minor attentional limitations ($\varepsilon < .5$). This regime is populated by the class of so called *preattentive* stimulus differences of the sort known to have early cortical representations; for example, searches based on differences in spatial frequency, orientation, color, etc. Search in this regime generally leads to a pattern of data consisting of mild set size effects on RT, low miss-rates, and redundancy benefits in both *pure* and target *absent* responding.

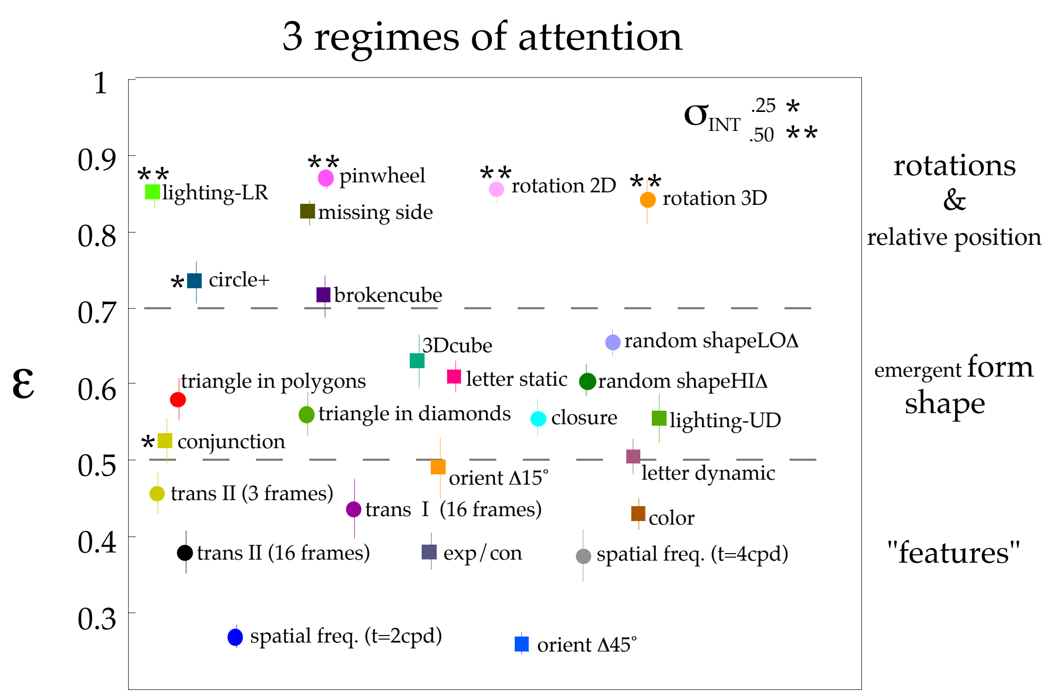
2. A *form* regime, characterized by intermediate amounts of attentional limitation ($.5 < \varepsilon < .7$). This regime contains searches based on simple shape differences and discriminations of emergent form. The pattern of data for search in this regime consists of moderate set size effects on RT, low miss-rates, and no redundancy benefits in the *pure* target trials (occasionally there are small benefits associated with set size for the target *absent* RTs).

3. A *relative position* regime, characterized by the highest levels of attentional limitation ($\varepsilon > .7$, $\sigma_{INT} > 0$). This regime consists of those experiments that require a configural analysis based on the relative arrangement of features (e.g. mirror reversals, part-whole organization, and direction of rotary flow). As such, these experiments are the most demanding of attentional resources, often require nontrivial amounts of

additive noise (i.e. $\sigma_{INT} > 0$), and generally manifest a pattern of data in which high miss-rates coexist with large RT costs for all *mixed*, *pure*, and target *absent* conditions.

Beyond these global distinctions, ε also appears to have sufficient structure as a metric to support the local ordering of experiments within regimes. For example, in several cases I have found that increasing target/distractor *confusability* leads to a consistent and reliable increase in ε (the *shape*, *translation*, and *orientation* experiments). Similarly, I find that small changes to search stimuli that remove or attenuate differences in emergent

Figure 48



structure (e.g. the sign of surface curvature, or implied orientation in depth) lead to significant increases in attentional limitation as estimated by ε.

In contrast to the notion that attention is ordered along a continuum, there is another analysis of the MTS results that arises by considering variation in the internal noise parameter, σ_{INT}. Recall, that the data from only a small handful of experiments was simulated using an *A-type* model having additive internal noise. This general model of attentional limitation is distinct from the *D-type* model that successfully simulated the large majority of MTS search data. In the random walk framework, the *D-type* model is equivalent to a *decision-limited* model that has no perceptual limitations, while the *A-type* model cannot be so reduced, and hence represents a framework that incorporates true perceptual limitations arising from either 1) the injection of additive noise, or 2) via a reduction in the probability of stimulus sampling (see section II.A.2 for details). This distinction in terms of *decisional* (*D-type*) versus *perceptual* (*A-type*) limitation offers another way to categorize the MTS ensemble: experiments fit best by the *D-type* model can be explained without invoking any limitations in perception, whereas experiments requiring an *A-type* model cannot.

The picture then that emerges from this level of analysis, and one that is consistent with previous results using *threshold search* (Palmer et. al, 1993;

Verghese & Nakayama, 1994; Poder, 1999), is that most search tasks are not perceptually limited; additional *perceptual* limitations are necessary only for those tasks that require some form of configural analysis (see Pashler, 1998 for a similar account). As compelling as this view is, there are good empirical reasons to question a parsing of visual search data into two qualitative categories (see sections II.C.2 and II.C.3). Most notably, partitioning the ensemble into *decision-limited* and *perception-limited* classes of search ignores the reliable ordering of *D-class* experiments in terms of ε. To assert that this ordering simply reflects differences in criterial scaling begs the question of why some experiments require much larger increases in criteria with set size than other. The interpretation I favor is that in the context of supra-threshold stimulus differences, visual search tasks are in fact ordered along a continuum of attentional limitation, with a special population of the most highly limited searches having additional source(s) of internal noise.

Conclusions

In closing I would like to summarize some of the more general implications that follow from my work with the multiple target search method and model.

I. Simple search is always conducted in *parallel*.

The MTS methodology as laid out in this work concerns *simple* visual search, that is search for supra-threshold targets in sparse, controlled displays that minimize the need for multiple fixations. For the 26 experiments contained in this dissertation, a model of search based on parallel, capacity-limited processing was able to account for all the joint patterns of RT and error. Moreover, in no case could the data be adequately explained using a *serial* model of search. Though serial models can explain isolated aspects of the search data, they are incapable of simultaneously accounting for joint patterns of decreasing false alarms, increasing misses, and shallow target absent RTs that mirror the pure target RTs. The obvious implication of this failure over such a broad and various ensemble is that there are effectively no serial processes in visual search. As such, this work can be added to a growing body of research that has called into question the reality of seriality in simple visual processing (Eckstein, 1998; Palmer, 1994; Pashler, 1987; Wolfe, 1998a). The consistent point that emerges in modelling these MTS data is that variation in capacity limitation, in conjunction with a complex decisional structure is sufficient to reproduce patterns of *inefficient* search once touted as the best evidence for fast, sequential deployments of attention. Of course, this implication applies only to *simple* visual searches of the sort examined here, and surely does not hold for more *complex* types of search in which sequences of multiple fixations impose seriality via a scan path.

II. Parallel search can be distinguished at 2 levels.

The previous claim, that parallellism is universal in simple search, is not new and has in fact been made previously by Pashler and others who provided evidence that the processing of small "clumps" of elements (e.g. $n <$ 4 –8) was accomplished in parallel, with serial processing emerging only when search displays contained multiple clumps (Grossberg, Mingolla, & Ross, 1994; Humphreys, Quinlan, & Riddoch, 1989; Pashler, 1987). This is exactly the distinction laid out previously between *simple* and *complex* visual search. However, my work extends Pashler's by making it clear that though all search is parallel at small set sizes, there are nonetheless, reliable and important distinctions that remain. The results reported here indicate that even within a relatively limited range of set sizes, the MTS models can distinguish search tasks in terms of two levels of limitation.

First, search tasks can be reliably ordered along a continuum of attentional limitation in terms of the parameter ε. Tasks low in this measure tend to have large redundancy gains in pure target RTs, decreasing target absent RTs, and generally shallow set size effects in RT and miss rate. Tasks high in this measure manifest little redundancy gains, and produce patterns of RT historically associated with *serial-like* processing. As I mentioned in the previous sections, the ordinal structure contained in the ensemble estimates of ε provides a global categorization of search tasks that is psychologically meaningful.

Second, the MTS models also provide a level of description that singles out those tasks which are limited by additional sources of additive noise. The results derived from the current ensemble isolate a small handful of experiments as being additionally limited in this way. This special class appears to be reserved for two specific types of search experiment: 1) those that involve targets and distractors that differ in the relative position of parts, and 2) those that involve the sensing of rotation sign. In the use of MTS there appears to be a general association between configurally-based search and internal noise, such that when target elements can only be distinguished from distractors via a configural analysis, the resultant data requires an internal noise model (or equivalently some form of sampling limitation).

III. Attention operates multiplicatively on the evidence available to decision.

Twenty of the 26 experiments in the MTS ensemble were best simulated by a *D-type* model with no internal noise. Recall, that the *D-type* model instantiates attentional limitation by multiplicatively scaling the perceptual evidence (i.e. the walk increments) based on a power-law relation between set size and ε. Scaling the increments in this way amounts to a Weberian account of attentional limitation in that attention does not alter the intrinsic discriminability of perceptual samples from targets and distractors (this is because multiplying a random deviate by some factor scales both the mean and standard deviation identically). Other manner of implementing capacity limits in search were investigated extensively (see section II.A.2), and in general these implementations could not reproduce the patterns of RT and error typical of multiple target search, though they did prove successful in a limited regime. The general success of Weberian scaling in the context of random walk models of search has two interpretations. The first is that attention has its effects on RT and error as a form of *resistance*, that constrains the flow of perceptual evidence to higher decisional processes. In this conception, attention acts *late* in the sense that it does not change the quality of the perceptual information, but only scales the evidence available to the decision maker. In appendix I., I show that the type of random walk model used here is formally equivalent to a SPRT model that sums stimulus likelihoods (Laming, 1968). Thus, the *D-type* way of implementing attention is consistent with a multiplicative attenuation at the level of the incoming sequence of likelihoods.

The second possible interpretation is that there is no attention in the *D-type* model, and that the Weberian scaling is really the result of set size dependent adjustments to decision criteria (see section II.A.1.f for a detailed exposition of this point). Multiplicative scaling of the random walk increments is entirely consistent with a tightening of response criteria

(criteria = Tn^{ε}). If this interpretation is correct, it implies that the 20 experiments fit best by the *D-type* model are not limited by attention – all response variation derives from decisional sources. Though this way of conceptualizing the *D-type* model brings the experimental results into line with the previous work of Palmer and colleagues (Palmer et. al, 1993; Palmer, 1994; Poder, 1998; Eckstein, Thomas, Palmer, & Shimozaki, 2000), it does not explain the range and consistency of the ordering provided by ε.

There were 5 experiments in the MTS ensemble whose data were best simulated by an *A-type* model. This class of model augments the Weberian scaling of the *D-type* model by including an additional source of invariant noise. Because only the samples of perceptual evidence, and not the noise deviates get scaled by $n^{-\varepsilon}$, this model does predict changes in single-element discriminability with set size. As such, the *A-type* model breaks the symmetry that plagues the *D-type* model and accordingly its effects cannot be mimicked by criterial variation. It is intriguing that the *A-type* model, with its irreducible source of attentional limitation, proved to be generally incapable of accounting for the majority of the MTS data.

Now, if we focus just on the *A-type* model, it becomes increasingly difficult to decide <u>how</u> attention limits search performance. This is because, in the context of a sequential-sampling model of the sort used here, the *A-type* model with its dual noise sources is nearly equivalent to a number of other ways of instantiating attention (section II.A.2.b). For example, the type of limitation imposed by Weberian scaling plus noise can also arise from a *sampling* scheme in which attention serves to modulate the probability of integrating a perceptual sample (or likelihood) into the running accumulation of evidence. Moreover, when the sampling probability goes as the inverse set size, the probability sampling model is nearly identical to the *fixed sample size* model of Shaw and colleagues (Shaw, 1984). Similarly, it is easy to show that a *sampling* approach can be recast in terms of *time-scaling*, in which the probability of integrating a sample is fixed, and now attention is conceived to scale the time between samples (see section II.A.2.b for details).

Clearly, the data and models as described are inadequate to decide among these various alternatives. One potentially fruitful area for future work centers around investigating model predictions at larger set sizes where many of the near-equivalences between the various conceptualizations of attention break down (i.e. for $n >> 4$). Another possibility is to use the Bayesian model of search described in section II.C.5 and appendix II. The different ways of implementing attention could each be incorporated within the Bayesian framework to see which model generates predictions that square with intuition. For example, as attentional limitation grows, either via an increase in ε, σ_{INT}, or λ, even an optimal decision maker will manifest set size effects that are meaningfully related to the parameters. It remains to be seen whether the equivalence between the *A-type* model and the *sampling* models will hold in the context of the Bayesian decision framework.

IV. Observers are "optimal"-like in their integration of information.

The final major implication that emerges from this work is that in the context of MTS, observers appear to be near-optimal in their decision making. The most striking feature of search with multiple targets is that target absent RTs are often found to decrease with set size, a pattern that is not predicted by any basic model of search. In the Bayesian context, this pattern arises naturally as a direct consequence of an optimal decision maker integrating all the available information in order to rationally arrive at target "present" and target "absent" decisions (see II.C.5). Observers in the MTS experiments also seem to use partial evidence to guide their decisions. The best evidence for this is that the complex patterns of RT and error characteristic of MTS can only be simulated by a model that allows response criteria to relax with set size. This optimal-like strategy is embodied in the MTS models via the parameter *C*, and evidence for its psychological reality exists in so far as every experiment in the ensemble required *C* to take on a value somewhere between 2 and 3 (see Fig. 40, section II.C.4). The specific interpretation of this commonality is that observers effectively reduce their criteria by a factor of about 2 in order to exploit the information available via the MTS design.

Presumably, the fact that the MTS design includes both multiple targets and multiple distractors leads observers to adopt a decision structure that emphasizes the potential information available in the joint state of *n* walks. In contrast, the strategy embodied in *C* is surely nonoptimal for a singleton search design that inherently emphasizes an "oddball" search strategy (see Zenger & Fahle, 1997 for a treatment of optimality in singleton search). In MTS, the relaxation of criteria with set size has two important data signatures. First, there is the ubiquitous *mirroring* evident between the target absent and pure target RTs – this symmetry is a direct result of the symmetry in *present* and *absent* decisions afforded by the introduction of secondary criteria (sec. II.A.1.e). Second, there is an apparent magnification of set size effects in MTS, such that single target RTs increase more sharply with *n*, relative to the functions reported for comparable search tasks within a singleton design (see Fig 43).

In general, it is the subtle tradeoffs between speed and accuracy that have revealed the parameter *C* to be a necessary component in the successful modelling of decision in multiple target search. These tradeoffs are an unavoidable part of any unconstrained search experiment, and if nothing else, they highlight the importance of always conditionalizing patterns of RT on the underlying error. In the context of the work reported here, this conditionalization was possible only via the structure provided by the computational models.

Appendix I.

Relation between a SPRT walk model and the standard Gaussian-increments model

For a Gaussian-increments random walk, let each x_i be a random deviate drawn from a Gaussian distribution N(S, V); define W_τ as the cumulative sum of τ of these deviates. For the class of SPRT models, the random walk consists of a sequential sum of probability ratios defined on these increments. Specifically, we define Z_τ to be the sum of the log of the ratio of likelihoods that an increment x_i was drawn from either the target or distractor increment distribution. Thus, the likelihood ratio for the *i*th step is

(I.1)

$$r_i = \left(\frac{\Pr\,(x_i \mid \text{target})}{\Pr\,(x_i \mid \text{distractor})} \right)$$

and

(I.2)

$$Z_\tau = \sum_{i=1}^{\tau} \ln\,(r_i)$$

Let S_t, V_t, S_d, and V_d represent the first two moments of the target and distractor evidence distributions respectively. We can then re-express the likelihoods as follows:

(I.3a)

$$\Pr(x_i \mid \text{target}) = \frac{1}{\sqrt{2\pi V_t^2}} \exp\left(\frac{-\left(x_i - S_t\right)^2}{2V_t^2} \right)$$

(I.3b)

$$\Pr(x_i \mid \text{distractor}) = \frac{1}{\sqrt{2\pi V_d^2}} \exp\left(\frac{-\left(x_i - S_d\right)^2}{2V_d^2} \right)$$

Substituting equations I.3a and I.3b into I.1 and taking the natural logarithm yields

(I.4) $$\ln(r_i) = \frac{(x_i - S_d)^2}{2V_d^2} - \frac{(x_i - S_t)^2}{2V_t^2}$$

For the standard case in which the target and distractor distributions are assumed to have means symmetric about the origin and equal variances, the expression above simplifies further such that $\ln(r_i) = \frac{x_i 2S}{V^2}$. This simple relationship reveals that a random walk based on the sum of increment log-likelihood ratios differs from a random walk based on the raw increments themselves only by the scale factor $\frac{2S}{V^2}$.

It is also easy to show that summing the log-likelihood ratios at each time step in a random walk is equivalent to computing the log-likelihood ratio of the current random walk state W_τ:

(I.5) $$Z_\tau = \sum_{i=1}^{\tau} \ln(r_i) = \ln\left(\frac{\Pr(W_\tau \mid \text{target})}{\Pr(W_\tau \mid \text{distractor})} \right)$$

Because of this equivalence, the Gaussian-increments random walk can be construed as a model in which decisions about element identity are made in an optimal way using the likelihoods of the current accumulated evidence.

Appendix II. *MTS using a Bayesian observer*

Here I describe in detail a random walk model of multiple-target search that is based on Bayesian estimation. This search model is optimal in the following sense: it knows the prior probabilities of the search stimuli, and incorporates knowledge of the respective distributions of target and distractor evidence at time zero, and the statistics relating how a general random walk evolves in time. With this knowledge the model makes target *present* or *absent* decisions based on the likelihood of the current state vector of the *n* random walks.

To begin, consider performance in a general classification task in which there are two response alternatives, c_1 and c_2 and an observer must correctly choose among them given a particular amount of perceptual evidence $\boldsymbol{S}$. In the case of stimulus noise or substantially overlapping categories errors are unavoidable and the optimal classifier will choose among the two alternatives using the following decision rule

(II.1) $\quad$ *if* $\Pr(c_1 \mid \boldsymbol{S}) > \Pr(c_2 \mid \boldsymbol{S})$ then choose c_1, else choose c_2

In the case of MTS we formulate the available stimulus information in terms of independent random walks that diffuse in time. Thus, we can replace $\boldsymbol{S}$ with the n–dimensional random variable $\boldsymbol{Z}$ (for simplicity, the following derivations will all be for n=2, but are easily extended to larger set sizes). The random variable ***Z*** represents the vector of current accumulated evidence about the identity of each element in a search display, and thus is comprised of the joint state of two random-walks at time τ (I drop tau in the remainder of this derivation for simplicity).

(II.2)

$$\text{Let } \mathbf{Z} = \langle z_1, z_2 \rangle, \quad z_k = \sum_{i=1}^{\tau} x_{i,k}, \text{ such that } x_{i,k} \in N(\pm S, V)$$

where $N(S,V)$ represents the appropriate Gaussian distribution of walk increments, and the sign of S depends on the actual identity of the k^{th} element in the display (recall that the mean of the distribution of target and distractor increments is $+S$ and $-S$ respectively). Given the two possible outcomes in MTS, target-present (*Tp*) or target–absent (*Ta*), then the decision rule in equation 1 becomes

(II.3) $$\textit{if} \ \Pr\left(Tp \mid \mathbf{Z}\right) > \Pr\left(Ta \mid \mathbf{Z}\right) \ \text{respond "}\textit{target present}\text{"}$$

This equation can be equivalently re-expressed such that decisions are based on whether the ratio of the two posterior probabilities exceeds some criterion (when there is no bias and the priors are equal this criterion will be 1).

(II.4) $$\frac{\Pr\left(Tp \mid \mathbf{Z}\right)}{\Pr\left(Ta \mid \mathbf{Z}\right)} > 1$$

Using Bayes formula we can formulate the posterior probabilities in terms of the likelihoods and the priors,

(II.5) $$\Pr\left(Tp \mid \mathbf{Z}\right) = \frac{\Pr\left(\mathbf{Z} \mid Tp\right)\Pr\left(Tp\right)}{\Pr\left(\mathbf{Z}\right)}$$

Because the prior probability of the evidence vector Z is common to both posteriors, the optimal decision rule given by equation II.4 can be re-expressed as

(II.6) respond target "present" *if* $$\frac{\Pr\left(\mathbf{Z} \mid Tp\right)}{\Pr\left(\mathbf{Z} \mid Ta\right)} > 1$$

In words, equation II.6 says that the optimal decision strategy is to respond target "present" whenever the likelihood of the vector of walk-positions given *Tp* is greater than the likelihood of the vector given *Ta*.

For set size $n = 2$, the probability of the vector given a target-present display must include the combined probability of the vector for all stimulus configurations in which one or two targets is present, i.e. the set {*td*, *dt*, and *tt*}, where *td* represents a configuration in which display elements 1 and 2 are a target and distractor respectively, etc. Thus, the combined probability of Z, given a target-present stimulus, is the sum of the individual probabilities of the vector given each stimulus configuration, weighted by the probabilities of each stimulus configuration:

(II.7)
$$\Pr\left(\mathbf{Z} \mid Tp_{n=2}\right) =$$
$$\Pr\left(\mathbf{Z} \mid t_1 d_2\right)\Pr\left(t_1 d_2\right) + \Pr\left(\mathbf{Z} \mid d_1 t_2\right)\Pr\left(\mathbf{Z} \mid d_1 t_2\right) + \Pr\left(\mathbf{Z} \mid t_1 t_2\right)\Pr\left(t_1 t_2\right)$$

Incorporating the actual priors, and the assumption that the current information about each element is represented by an independent random walk, equation II.7 is equal to

(II.8)
$$\Pr\left(z_1 \mid t\right)\Pr\left(z_2 \mid d\right)\frac{1}{8} + \Pr\left(z_1 \mid d\right)\Pr\left(z_2 \mid t\right)\frac{1}{8} + \Pr\left(z_1 \mid t\right)\Pr\left(z_2 \mid t\right)\frac{1}{4}$$

where z_1 and z_2 represent the accumulated evidence about each element's identity as per equation II.2.

Given a fixed moment in time τ, the distribution of possible walk-positions for a single element without absorbing barriers follows a Gaussian density and is distributed as $N(S\tau, V\tau^{1/2})$ for a target element, and $N(-S\tau, V\tau^{1/2})$ for a distractor element. Thus, we can explicitly write down the probabilities in equation II.8

(II.9.a)
$$\Pr\left(z_1 \mid t\right)\Pr\left(z_2 \mid d\right) = \frac{1}{2\pi\tau V^2} e^{\frac{-(z_1 - S\tau)^2 - (z_2 + S\tau)^2}{2\tau V^2}}$$

(II.9.b) $$\Pr(z_1 \mid d)\Pr(z_2 \mid t) = \frac{1}{2\pi\tau V^2} e^{\frac{-(z_1+S\tau)^2-(z_2-S\tau)^2}{2\tau V^2}}$$

(II.9.c) $$\Pr(z_1 \mid t)\Pr(z_2 \mid t) = \frac{1}{2\pi\tau V^2} e^{\frac{-(z_1-S\tau)^2-(z_2-S\tau)^2}{2\tau V^2}}$$

In similar fashion we can also explicitly write down the probability of Z given *Ta* for *n*=2

(II.9d) $$\Pr\left(\mathbf{Z} \mid Ta_{n=2}\right) = \Pr(z_1 \mid d)\Pr(z_2 \mid d) = \frac{1}{2\pi\tau V^2} e^{\frac{-(z_1+S\tau)^2-(z_2+S\tau)^2}{2\tau V^2}}$$

To calculate optimal performance for distinguishing target-present displays from target-absent displays we define *L* to be the natural log of equation II.6 and integrate equation II.8 over the region of two-dimensional walk-positions for which *L* is greater than 0. In practice this space is discretized and the optimal percent correct at a given time τ is based on summing over that region. Once we have the optimal criterion (i.e. *L*=0), we can then calculate the optimal percent corrects for each specific stimulus configuration. For example, below we express the optimal percent correct for both a 2-target display (eq. II.10.a), and a 1 target/1distractor display (eq. II.10.b):

(II.10.a) $$PC_{opt}(\tau, tt) = \frac{1}{2\pi\tau v^2} \iint_{L>0} e^{\frac{-(z_1 - s\tau)^2 - (z_2 - s\tau)^2}{2\tau v^2}} \, dz_1 \, dz_2$$

(II.10.b)
$$PC_{opt}(\tau, 1target/1distractor) = \frac{1}{4\pi\tau v^2} \iint_{L>0} \left[e^{\frac{-(z_1 - s\tau)^2 - (z_2 + s\tau)^2}{2\tau v^2}} + e^{\frac{-(z_1 + s\tau)^2 - (z_2 - s\tau)^2}{2\tau v^2}} \right] dz_1 \, dz_2$$

Equations of this sort express the optimal percent correct at a fixed moment in time. Alternatively, we can ask at what time do different stimulus configurations reach a common level of accuracy. To do this I first calculate PC_{opt} for a range of τ's, and then invert the relationship to find that τ that yields the desired level of accuracy.

Appendix III. *Sampling models*

In this appendix I show that a probability-sampling model of attention is equivalent to a model that scales the increment parameters *S* and *V* independently. Despite a formal equivalence, these two models are quite different conceptually. The increment-scaling model is an extension of the basic *D-type* model of attention in which attention scales both the size (*S*) and variability (*V*) of the evidence samples that feed the random walks. The scaling of both increment parameters in the *D-type* model is implemented identically as a function of set size and attention (both parameters are multiplicatively scaled by the factor $n^{-\varepsilon}$). The increment-scaling model detailed here extends the *D-type* model by letting *S* scale as $n^{-\varepsilon}$, and *V* as $n^{-\varepsilon/2}$. Thus, *V* gets scaled with set size by cutting ε in half (given the power law relation, this is identical to scaling *V* by the square-root of the scale factor that is used to scale *S*). In contrast, consider the probability-sampling model that is introduced in section II.A.2.b. This model conceives of attention as modulating the probability that a sample is received and integrated into decision, but does not alter *S* or *V* with set size. Instead, it is now the probability of receiving a sample of evidence that is jointly determined by attention and set size ($p = n^{-\lambda}$, where λ is an attention parameter on the interval [0,1]).

The equivalence of these two distinct conceptualizations of attention can be seen in the context of each model's predictions regarding the distribution of random walk positions at an arbitrary time *k*. The comparison predictions are derived below for $n > 1$, and for simplicity both the attentional parameters ε and λ will be set to 1, though the equivalences hold throughout. I point out here that for $\lambda = 1$, the probability-sampling model reduces to the *fixed sample size* model which has figured predominantly in attempts to decide the nature of limitation for visual search at threshold (e.g. Shaw, 1984; Palmer et. al, 1993). The *fixed sample size* model assumes that a finite number of perceptual samples gets divided equally among the *n* stimuli that comprise a search display, such that the number of samples available per element goes as $1/n$. This also describes the proportion of samples available to each random walk under probability sampling with a λ of 1. Thus, any equivalence between the increment-scaling and probability sampling models of attention implies a similar equivalence with the class of *fixed sample size* models.

Increment-scaling

I begin by defining w_k to be a random variable representing the total evidence accumulated in a random walk with *k* increments. For the increment scaling model, the random walks will always accumulate *k* samples in *k* time steps. This is because sampling is perfect and only the size and variability of the samples is affected by attention and set size. Note that for $k > 10$ or so, the distribution of w_k without absorbing barriers is approximately Gaussian (i.e. $N(\mu,\sigma)$):

(III.1)

$$w_k = N\left(kS, \sqrt{k}V\right)$$

For the increment-scaling model the predicted distribution of walk positions as a function of n and ε can similarly be represented by the variable $w_{n,k,\varepsilon}$:

(III.2a)

$$w_{n,k,\varepsilon} = N\left(kSn^{-\varepsilon}, \sqrt{k}Vn^{-\varepsilon/2}\right)$$

such that for $\varepsilon = 1$

(III.2b)

$$w_{n,k,1} = N\left(\frac{kS}{n}, V\sqrt{\frac{k}{n}}\right)$$

Probability-sampling

In the context of a probability-sampling model, I define a new random variable y_k to represent accumulated evidence as a function of time step. In the context of this model there are k time steps, but only some proportion of these provide a sample for accumulation. This proportion, denoted p, is a function of attention and set size:

(III.3)

$$p = n^{-\lambda}$$

Thus, on average there will be kp samples available in k time steps. Accordingly, I define the random variable k^* to be the actual number of samples accumulated in any single random walk given k time steps and a sampling probability of p. The variable k^* is binomially distributed:

(III.4)

$$k^* = B(k,p)$$

where the quantity $B(k,p)$ represents a random deviate from a binomial distribution with parameters k and p. For the probability sampling model the predicted distribution of the y_k's as a function of p are

(III.5)

$$y_{n,k^*,p} = N\left(Sk^*,\ V\sqrt{k^*}\right)$$

Combining the two previous equations gives an expression for the distribution of walk positions at time step k:

(III.6)

$$y_{n,k,p} = N\left(S \cdot B(k,p),\ V \cdot \sqrt{B(k,p)}\right)$$

This equation reveals that the probability-sampling model expresses the position of a random walk after k time steps as a Gaussian random variable whose parameters are scaled by a binomial random variable representing the actual number of samples k^*. Thus, the probability-sampling model has a source of variability that the increment scaling model does not. Despite this addition, the probability sampling model can be equivalently expressed as increment scaling with perfect sampling. A brief proof of this equivalence is given in the next section for the case when there are no absorbing barriers (i.e. $T = \infty$). For now, I simply assert the fact that the distribution of walk positions for k time steps under probability-sampling (eq. III.6) can be simplified as follows:

(III.7)

$$y_{n,k,p} = N\left(kpS,\ V\sqrt{kp}\right)$$

Substituting for p this becomes

(III.8)

$$y_{n,k,\lambda} = N\left(kSn^{-\lambda},\ V\sqrt{kn^{-\lambda}}\right)$$

And setting $\lambda = 1$

(III.9)

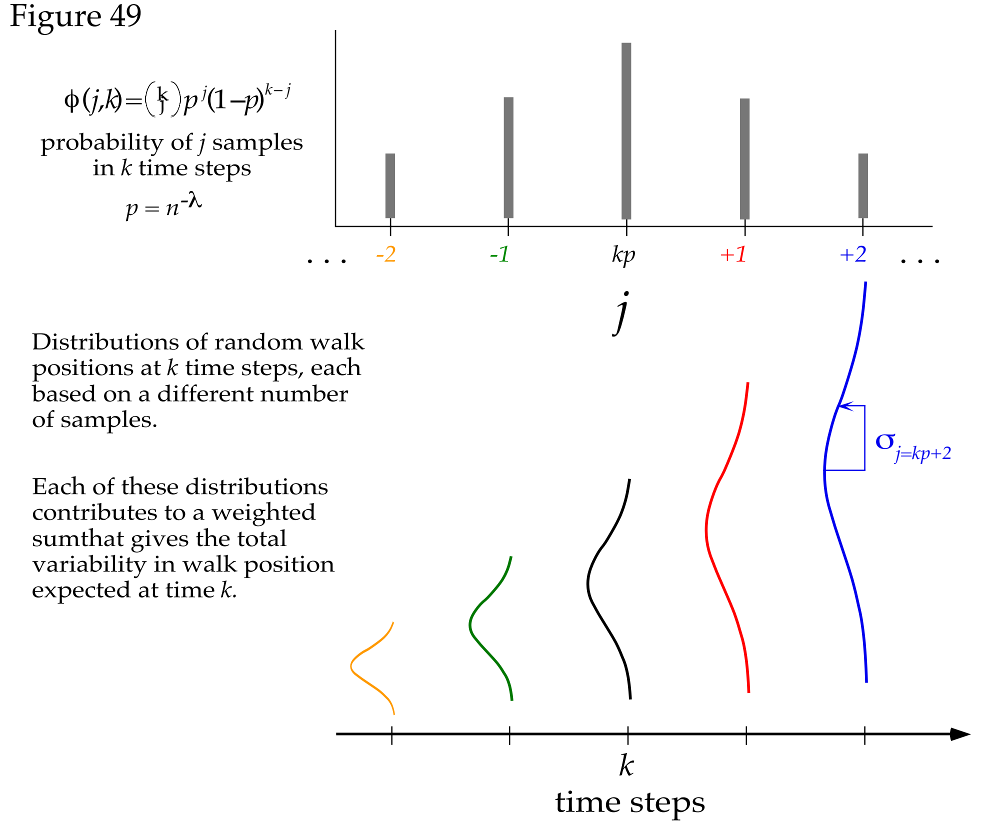
$$y_{n,k,1} = N\left(\frac{kS}{n},\ V\sqrt{\frac{k}{n}}\right)$$

which is equivalent to equation III.2b for the increment-scaling model with $\varepsilon = 1$. Thus, both the increment-scaling and probability-sampling models predict identical effects on the distribution of random walk positions in k time steps (i.e. $w_{n,k,\varepsilon} = y_{n,k,\lambda} \big|_{\varepsilon = \lambda}$).

Proof of the equality of equations III.6 and III.7

Here I show that variability on the number of samples makes no difference in the distribution of walk positions – the random variable k^* can thus be completely replaced by the exact quantity, kp. For reference, the adjacent figure shows the relationship between the important quantities comprising this proof.

Figure 49



The probability of j samples in k time steps is denoted $\phi(j,k)$. This probability derives from the expression for the discrete binomial density function:

(III.10)

$$\phi(j,k) = \binom{k}{j} p^j (1-p)^{k-j}$$

I now define the variability in walk positions associated with *j* samples in *k* time steps as $\sigma_j = Vj^{1/2}$. This is the variability in walk position at time *k* for the subset of random walks having exactly *j* samples in their sums. Similarly, I define the quantity σ_k to be the combined variability across all the random walk positions at *k* time steps. This is effectively the width of the envelope of random walks given variation in *j* and it is obtained by taking a weighted sum of the σ_js over *j*:

(III.11)

$$\sigma_k^2 = \sum_{j=1}^{\infty} V^2 j \ \phi(j,k)$$

[Note that this equation is expressed as a sum and not an integral because the number of samples is always integer-valued]. By factoring out V^2, the remaining sum is simply equal to the expectation of *j*:

(III.12)

$$\mathrm{E}(j) = \sum_{j=1}^{\infty} j \cdot \phi(j,k) = kp$$

and by combining this result with eq. III.11 and taking the square root it then follows that

$$\sigma_k = V\sqrt{kp}$$

In words, I have shown that the variability in walk position at k time steps is determined by a weighted sum of the variabilities associated with each j. Because $\phi(kp\text{-}1,k)$ equals $\phi(kp+1,k)$, the variability gets averaged over, as if all the walks had exactly kp samples. Thus, the binomial variability in the actual number of samples (eq. III.6) can be replaced by an exact quantity given by the expected number of samples in k steps (kp). This substitution yields equation III.7 and completes the proof. The equivalence was also verified via an explicit simulation of the increment-scaling and probability-sampling models (n = 1e7).

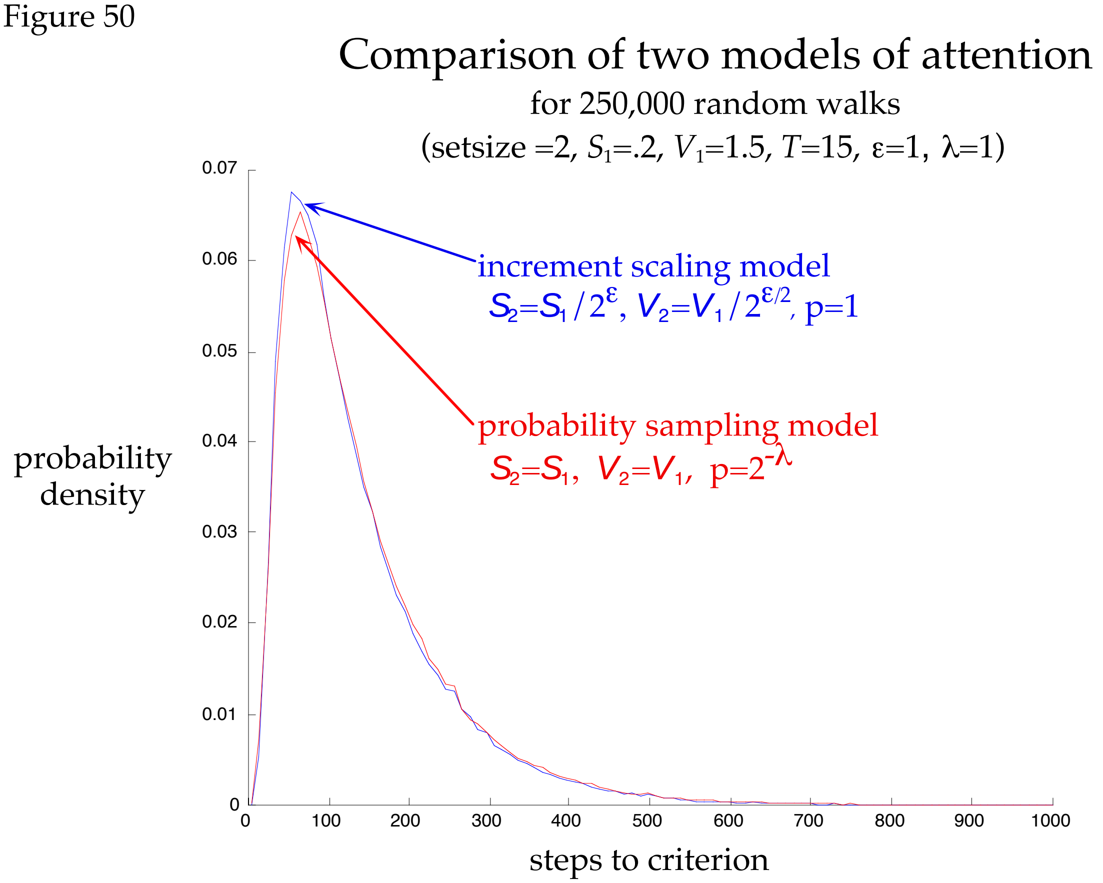
Walks with absorbing barriers

It is important to reiterate that the equivalence of the increment-scaling and probability-sampling models holds exactly only when the random walks are unconstrained, without absorbing barriers. This case is appropriate for modelling *threshold search* performance in which displays are necessarily brief and accuracy is the sole dependent variable of interest (e.g. Palmer et. al, 1993). The unconstrained random walk is also relevant in the simulations of the Bayesian model which provides accuracy by time-step functions reflective of optimal response selection (see Appendix II).

Unfortunately, the equivalence I have just shown is broken when absorbing barriers are introduced. Recall, that the random walk between two absorbing barriers forms the primary architecture for generating both accuracy and RT functions in the MTS models I use to explain my data. To see why the equivalence breaks down when absorbing barriers are introduced into the architecture, consider that in this case the distribution of random walk positions after *k* time steps is no longer Gaussian. Instead we have a distribution of random walk positions with the previously absorbed walks removed (i.e. walks that first reached the barrier for times < *k*). There are expressions for the distribution of *non*-terminated walks (Ratcliff, 1988), and in general it becomes highly non-Gaussian with time because the processes yet to "finish" cluster near the barrier. The non-Gaussian nature of this distribution has the consequence that equation III.6 can no longer be replaced by the simpler expression in equation III.7. This is because the average variability in walk position at time *k* (see equation III.11) is now biased to be slightly lower. This decrease in the width of the distribution of walk positions is a direct result of the fact that those random walks containing an above-average number of samples (and hence higher variability in accumulated position) are more likely to have terminated, and accordingly are no longer part of the average at time *k*. This small decrease in the variance of the distribution of walk positions under probability-sampling predicts that 1) the times-to-absorption will be slightly longer, and 2) the error rates will be slightly lower. As will be shown below, this is in fact what happens and it implies that probability-sampling is no longer exactly the same as increment scaling.

The good news is that even though the distribution of walk positions becomes non-Gaussian with the addition of absorbing barriers, the two

Figure 50

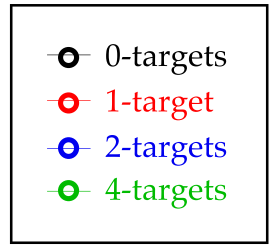


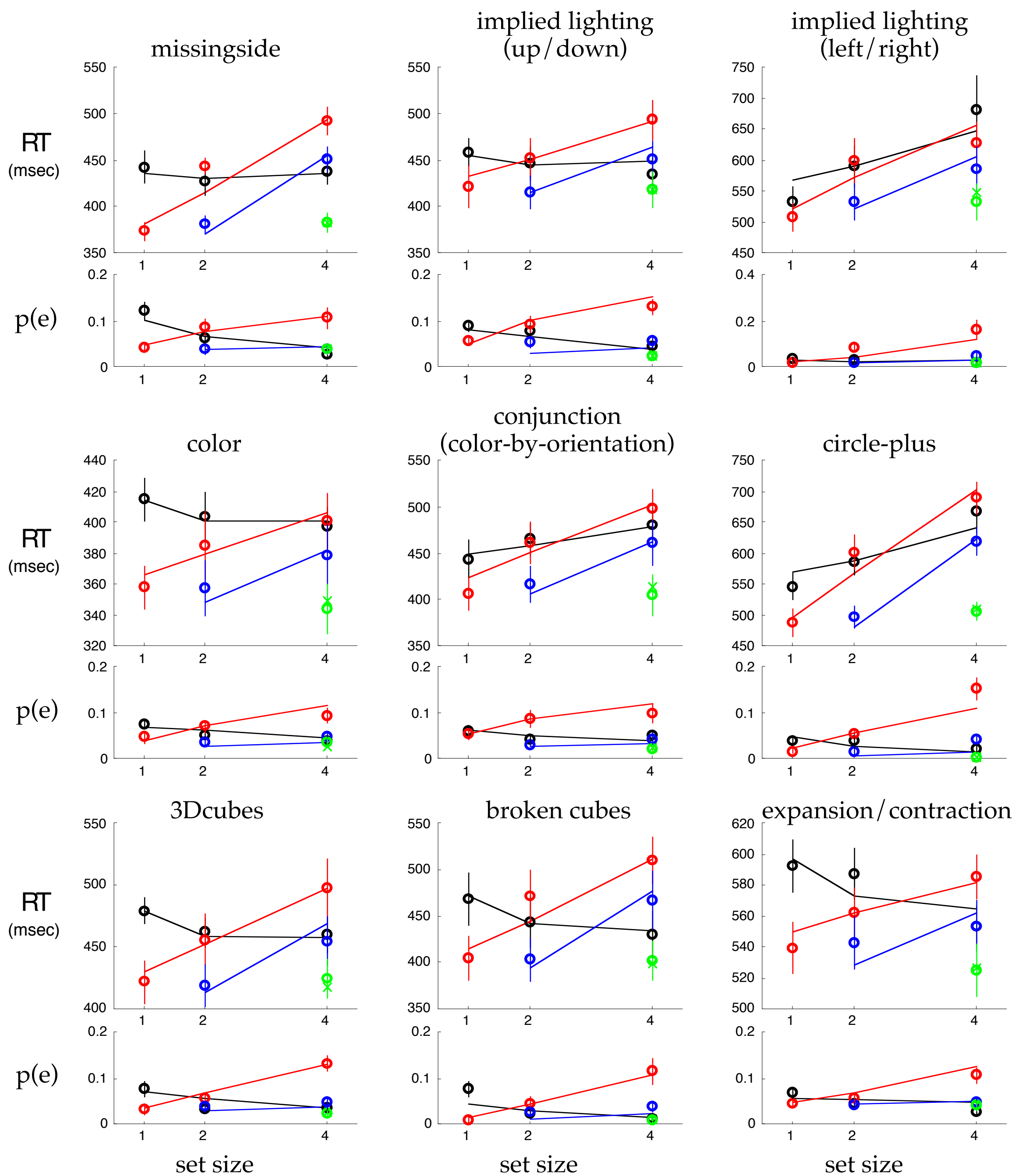
models are *near*-equivalent, such that their predicted RTs and error rates are virtually indistinguishable. The only way to show this is via explicit simulation with matched parameters which I have done. In the adjacent figure I show the distributions of time-to-absorption (RT) for the increment-scaling (blue) and probability-sampling (red) models. These results are based on simulating each model of attention to generate predicted RTs (number of trials = 250,000) for the *set size*=2 condition (attentional limitation is maximal and equal to 1 for both models). The resultant curves are estimates of the density functions and are typical of random walk models in that they are positively skewed by virtue of the geometry of the *diffusion* process. More importantly, the two functions are seen to lie almost completely on top of each other implying that the models are virtually identical at the level of RT. The major difference between the models is most evident near the mode of each density function, such that the increment-scaling model is slightly more likely to produce fast RTs relative to probability-sampling. Direct comparison of the primary moments and median-statistics [median, median-absolute-deviation (MAD)] from each distribution of simulated RT reveals a consistent, but small 2-3 millisecond advantage for increment-scaling (*scaling model*: μ=138.03, σ=100.22, median=110, MAD=50; *sampling model*: μ=141.60, σ=102.84, median=113, MAD=52). A small speed-accuracy tradeoff was also observed such that the *faster* increment-scaling model had a slightly higher error-rate (p(e)=.059), relative to the probability-sampling model (p(e)=.057). The direction of these differences in speed and accuracy is consistent with the intuition that the probability-sampling model generates a distribution of non-terminated walks at time *k* whose mean and variability are slightly lower.

I have compared the performance of each model via simulation across the relevant range of parameters and nowhere do I find discrepancies larger than those reported for these simulations. In fact, as attentional limitation is decreased from its maximum, the increment-scaling and probability-sampling models converge. This is because decreasing limitation means a higher sampling probability, which in turn implies a reduction in the additional source of binomial variability associated with random sampling. It is precisely the magnitude of variability in the number of samples at time k that generates the model differences seen in the previous figure.

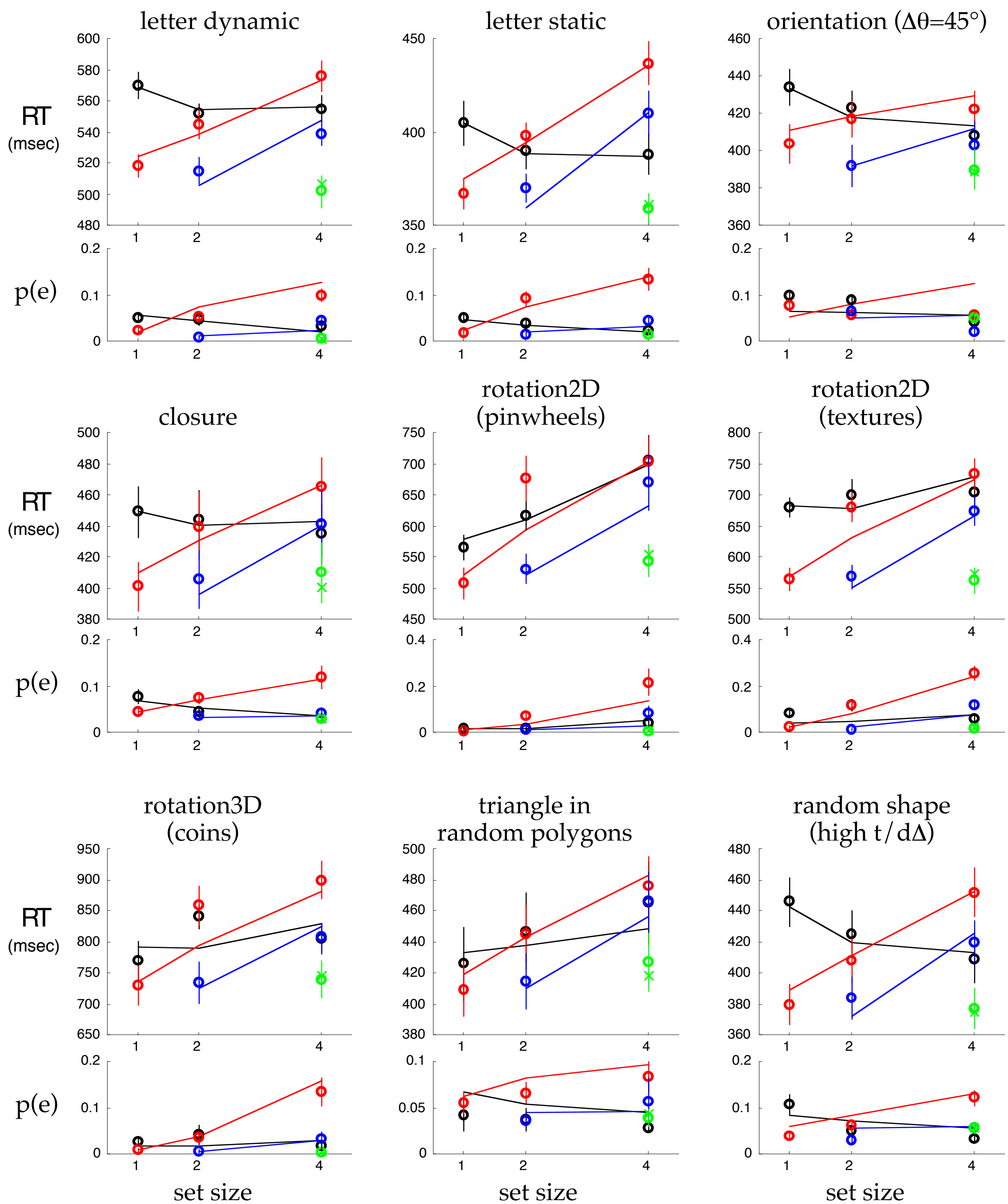
In section II.B.3 I show that the probability-sampling model accounts equally well for patterns of data fit best by an *A-type* model with internal noise. Because of the near equivalence between sampling and scaling models, I chose to generate the predicted RTs and error rates used in that analysis by simulating an increment-scaling model. I used increment-scaling for the simulations instead of a sampling process because 1) it was relatively easier to implement in the context of the existing models, and 2) was computationally cheaper. As this appendix hopefully makes clear, it matters little as to which of these two models is implemented.

Appendix IV *Ensemble data and model fits*
[solid points are data, lines are model fits]

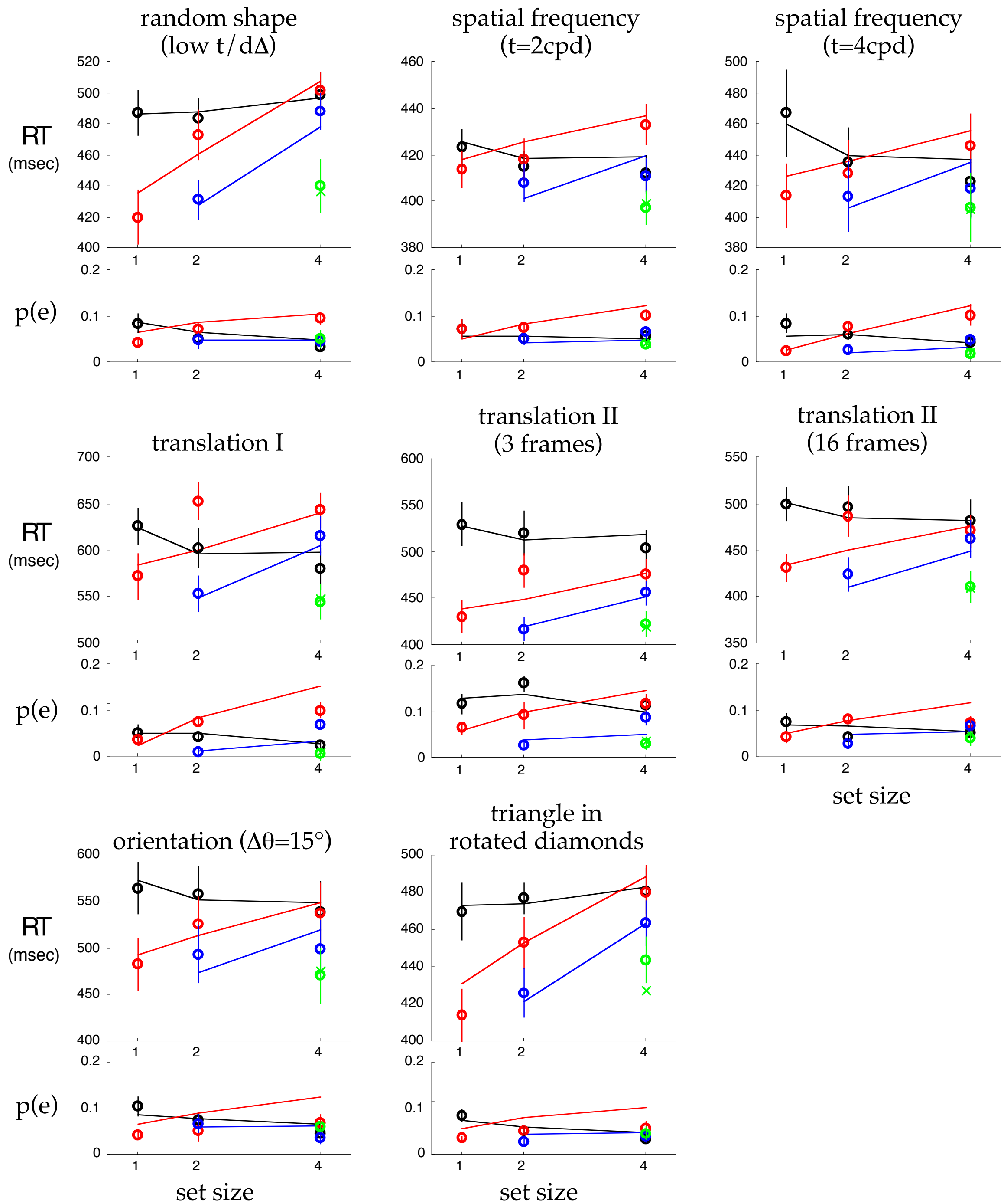




Appendix IV. continued



Appendix IV. continued



Appendix V. Parameters used to account for the MTS ensemble

Experiment	***S***	***V***	***D***	***C***	ε	M_R	***Z***	σ_{INT}	RT *r*	Err *r*
spatial frequency (t=2cpd)	.293 ±.004	1.47 ±.026	1.09 ±.036	2.10 ±.040	.270 ±.016	365 ±1.6	3.2 ±1.0	0	.925	.863
spatial frequency (t=4cpd)	.277 ±.010	1.63 ±.053	1.58 ±.095	2.52 ±.129	.369 ±.037	368 ±4.2	4.5 ±1.8	0	.891	.907
orientation (Δθ=45˚)	.290 ±.004	1.43 ±.033	1.34 ±.052	2.35 ±.061	.271 ±.016	356 ± 1.8	7.1 ±1.7	0	.944	~0
orientation (Δθ=15˚)	.256 ±.014	1.49 ±.052	1.76 ±.118	2.71 ±.136	.504 ±.039	429 ±7.9	39.5 ±7.2	0	.933	.398
color	.268 ±.01	1.64 ±.042	1.53 ±.081	2.40 ±.078	.424 ±.024	309 ± 4.0	18.7 ± 3.2	0	.973	.939
3D cubes	.281 ±.009	1.65 ±.04	1.78 ±.095	3.04 ±.094	.647 ±.039	372 ±2.9	11.1 ± 3.0	0	.971	.952
closure	.289 ±.008	1.64 ±.040	1.46 ±.052	2.72 ±.137	.568 ±.029	357 ±3.2	14.7 ±3.6	0	.953	.985
random shape (high t/dΔ)	.289 ±.007	1.56 ±.048	1.68 ±.094	2.99 ±.093	.614 ±.026	334 ±2.9	17.4 ±3.8	0	.972	.869
random shape (low t/dΔ)	.298 ±.002	1.75 ±.031	1.40 ±.078	2.44 ±.085	.657 ±.021	384 ±2.7	36.7 ±4.1	0	.962	.918
letter static	.293 ±.004	1.60 ±.034	1.40 ±.051	3.20 ±.086	.603 ±.021	323 ± 1.2	11.1 ± 1.9	0	.979	.981
letter dynamic	.256 ±.012	1.64 ±.049	1.45 ±.049	2.70 ±.075	.501 ±.024	463 ± 3.3	21.4 ± 2.6	0	.972	.944
implied lighting (up/down)	.252 ±.013	1.67 ±.059	1.34 ±.068	2.67 ±.097	.568 ±.035	368 ± 4.5	6.0 ± 1.8	0	.955	.964
translation I	.173 ±.016	1.29 ±.066	1.38 ±.070	2.77 ±.150	.437 ±.039	482 ±8.2	10.1 ±3.5	0	.859	.866
translation II (16 frames)	.237 ±.017	1.33 ±.075	1.38 ±.100	2.42 ±.097	.370 ±.031	364 ±6.6	44.8 ±4.8	0	.924	.622
translation II (3 frames)	.244 ±.012	1.84 ±.058	1.84 ±.083	2.43 ±.097	.436 ±.027	376 ±3.5	44.1 ±6.2	0	.955	.881
expansion/contraction	.269 ±.012	1.35 ±.058	1.45 ±.115	2.71 ±.119	.372 ±.028	487 ±4.4	22.9 ± 4.2	0	.911	.927
rotation 2D (textures)	.149 ±.007	1.07 ±.041	1.36 ±.054	2.92 ±.088	.862 ±.019	462 ±5.3	78 ±7.5	.5	.950	.912
rotation 2D (pinwheels)	.156 ±.010	.99 ±.044	1.19 ±.057	2.16 ±.062	.868 ±.014	420 ±6.6	35.5 ±5.7	.5	.921	.938
rotation 3D (coins)	.194 ±.007	1.16 ±.065	1.42 ±.084	2.69 ±.084	.818 ±.034	655 ±5.8	20.8 ±5.9	.5	.840	.971
triangle in rotated diamonds	.288 ±.004	1.69 ±.043	1.31 ±.061	2.22 ±.066	.587 ±.036	375 ±3.5	29.0 ±4.3	0	.930	.617
triangle in random polygons	.277 ±.006	1.53 ±.046	1.03 ±.021	2.08 ±.060	.582 ±.032	364 ±3.5	12.0 ±2.7	0	.924	.846
Conjunction (color-by-orientation)	.184 ±.013	1.33 ±.05	1.10 ±.025	2.11 ±.041	.528 ±.032	337 ± 6.9	20.4 ± 3.9	.25	.962	.962
missing side	.270 ±.007	1.76 ±.027	1.68 ±.066	2.90 ±.075	.823 ±.017	323 ±2.1	21.8 ± 3.7	0	.962	.972
implied lighting(left/right)	.230 ±.014	1.21 ±.057	1.08 ±.029	2.07 ±.043	.837 ± .02	451 ± 6.4	38.6 ± 5.1	.5	.899	.951
broken cube	.254 ±.03	1.51 ±.053	1.77 ±.085	3.53 ±.09	.721 ±.03	351 ±6.2	12.2 ± 4.1	0	.959	.942
circle-plus	.118 ±.007	1.09 ±.038	1.30 ±.039	2.50 ±.113	.723 ±.032	366 ± 8.0	43.6 ± 5.2	.25	.967	.945

Bibliography

Ashby, F. G. (1983). A biased random walk model for two choice reaction times. *Journal of Mathematical Psychology, 27*, 277-297.

Baddeley, A. D., & Hitch, G. J. (1974). Working Memory. In *Recent Advances in learning and motivation Vol. VIII* (G. Bower, ed.), 47-90. Academic Press, New York.

Ball, K., & Sekuler, R. (1980). Human vision favors centrifugal motion. *Perception, 9,* 317-325.

Beck, J. (1966). Perceptual grouping produced by changes in orientation and shape. *Science, 154.* 538-540.

Beck, J. (1982). Textural segmentation. In J. Beck (ed.), *Organization and Representation in Perception,* 285-317. Hillsdale, NJ: Erlbaum.

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

Burgess, A. E. (1985). Visual signal detection. III. On Bayesian use of prior knowledge and cross correlation. *Journal of the Optical Society of America, 2,* 1498-1507.

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

Carrasco, M., McLean, T. L., Katz, S. M., & Frieder, K. S. (1998). Feature asymmetries in visual search: Effects of display duration, target eccentricity, orientation and spatial frequency. *Vision Research, 38,* 347-374.

Carrasco, M., & Yeshurun, Y. (1998). The contribution of covert attention to the set size and eccentricity effects in visual search. *Journal of Experimental Psychology: Human Perception and Performance 24,* 673-692.

Clark, H. H., & Chase, W. G. (1972). On the process of comparing sentences against pictures. *Cognitive Psychology, 3,* 472-517.

Diederich, A., & Colonius, H. (1991). A further test of the superposition model for the redundant-signals effect in bimodal detection. *Perception and Psychophysics, 50,* 83-85.

Dosher, B. A., & Lu, Z. (1998). Perceptual learning reflects external noise filtering and internal noise reduction through channel reweighting. *Proceedings of the National Academy of Sciences, USA, 95,* 13988-13993,

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

Eckstein, M. P. (1998). The lower efficiency for conjunctions is due to noise and not serial attentional processing. *Psychological Science, 2*, 111-118l

Eckstein, M. P., Thomas, J. P., Palmer, J., & Shimozaki, S. S. (2000). A signal detection model predicts the effects of set size on visual search accuracy for feature, conjunction, triple conjunction, and disjunction displays. *Perception and Psychophysics, 62*, 425-451.

Edwards, M., & Badcock, D. R. (1993). Asymmetries in the sensitivity to motion in depth: A centripetal bias. *Perception, 22*, 1013-1023.

Egeth, H. W., Folk, C. L., & Mullin, P. A. (1988). Spatial parallelism in the processing of lines, letters, and lexicality. In B. E. Shepp & S. Ballesteros (Eds.), *Object perception: Structure and process*. Hillsdale, NJ: Erlbaum.

Egeth, H. E., & Mordkoff, J. T. (1991). Redundancy gain revisited: evidence for parallel processing of separable dimensions. In G. R. Lockhead & J. R. Pomerantz (Eds.), *The perception of structure: essays in honor of Wendell R. Garner*. Washington, DC: American psychological association.

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

Fournier, L. R., & Eriksen, C. W. (1990). Coactivation in the perception of redundant targets. *Journal of Experimental Psychology: Human Perception and Performance, 16*, 538-550.

Geisler, W. S., & Chou, K. (1995). Separation of low-level and high-level factors in complex tasks: Visual search. *Psychological Review, 102*, 356-378.

Geisler, W. S., Stern, L., Thornton, T., Kuyel, T., & Ghosh, J. (1998). Bayesian Ideal-observer analysis of texture segregation: Evidence for structure-based models. *Investigative Ophthalmology & Visual Science Supplement.* (ARVO) 39/4, S649.

Gilden, D. L., & Kaiser, M. K. (1992). Kinematic specification of depth and boundary. Presented at the 1992 meeting of the Association for Research in Vision and Ophthalmology.

Gold, J., Bennett, P. J., & Sekuler, A. B. (1999). Signal but not noise changes with perceptual learning. *Nature, 42*, 176-178.

Graziano, M. S. A., Anderson, R. A., & Snowden, R. J. (1994). Tuning of MST neurons to spiral motions. *Journal of Neuroscience, 14*, 54-67.

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

Gurnsey, R., & Browse, R. A. (1987). Micropattern properties and presentation conditions influencing visual texture discrimination. *Perception and Psychophysics, 41*, 239-252.

Harris, L. R., Morgan, M. J. & Still, A. W. (1981). Moving and the motion after-effect. *Nature, 293*, 139-141.

Heathcote, A., & Mewhort, D. J. K. (1993). Representation and selection of relative position. *Journal of Experimental Psychology: Human Perception and Performance, 19*, 488-516.

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

Joseph, J. S., Chun, M. M., & Nakayama, K. (1997). Attentional requirements in a “preattentive” feature search task. *Nature, 387*, 805-807.

Julesz, B. (1981). Textons, the elements of texture perception, and their interactions. *Nature (290)*, 91-97.

Julesz, B., & Hesse, R. I. (1970). Inability to perceive the direction of rotation movement of line segments. *Nature, 225*, 243-244.

Karlin, S., & Taylor, H. M. (1975). *A first course in stochastic processes* (2nd ed.). New York: Academic Press.

Kristjansson, A., & Tse, P. U. (2001). Curvature discontinuities are cues for rapid shape analysis. *Perception and Psychophysics, 63*, 390-403.

Kosslyn, S. M. (1975). Information representation in visual images. *Cognitive Psychology, 7,* 341-370.

Laming, D. (1968). *Information theory of choice reaction times.* New York: Academic Press.

Lewis, C. H., & Anderson, J. R. (1976). Interference with real world knowledge. *Cognitive Psychology, 7,* 311-335.

Link, S. W. (1975). The relative judgment theory of two-choice response time. *Journal of Mathematical Psychology, 12,* 114-135.

Logan, G. D. (1994). Spatial attention and the apprehension of spatial relations. *Journal of Experimental Psychology: Human Perception and Performance, 20,* 1015-1036.

Luce, R. D. (1986). *Response Times: Their role in inferring elementary mental organization*. New York: Oxford University Press.

Maddox, W. T., & Ashby, F. G. (1996). Perceptual separability, decisional separability, and the identification-speeded classification relationship. *Journal of Experimental Psychology: Human perception and performance, 22,* 795-817.

Malik, J. & Perona, P. (1990). Preattentive texture discrimination with early vision mechanisms. *Journal of the Optical Society of America, A7,* 923-932.

McLean, J. E. (1999). Processing Capacity of Visual Perception and Memory encoding. Unpublished dissertation.

Meyer, D. E., Irwin, D. E., Osman, A. M., & Kounios, J. (1988). The Dynamics of Cognition and Action: Mental Processes inferred From Speed-Accuracy Decomposition. *Psychological Review, 95,* 183-237.

Miller, J. (1982). Divided attention: Evidence for coactivation with redundant signals. *Cognitive Psychology, 14,* 247-279.

Miller, J. (1986). Timecourse of coactivation in bimodal divided attention. *Perception and Psychophysics, 40,* 331-343.

Miller, J. (1991). Channel interaction and the redundant-targets effect in bimodal divided attention. *Journal of Experimental Psychology: Human Perception and Performance, 17,* 160-169.

Moore, C. M., Egeth, H., Berglan, L. R., & Luck, S. J. (1996). Are attentional dwell times inconsistent with serial visual search? *Psychonomics Bulletin and Review, 3,* 360-365.

Moore, C. M., Elsinger, C. L., & Lleras, A. (2001). Visual attention and the apprehension of spatial relations: The case of depth. *Perception and Psychophysics, 63*, 595-606.

Mordkoff, J. T., & Yantis, S. (1991). An interactive race model of divided attention. *Journal of Experimental Psychology: Human Perception and Performance, 17*, 520-538.

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

Nakayama, K., Wang, D., & Kristjansson, A. (2000). Efficient visual search not explained by local feature mechanisms or top-down guidance. *Investigative Ophthalmology & Visual Science Supplement*. (ARVO) 41/4, S579.

Neisser, U. (1967). *Cognitive psychology.* New York: Appleton, century, Crofts.

Nothdurft, H. C. (1991). Texture segmentation and pop-out from orientation contrast. *Vision Research*, *31*, 1073-1078.

Nothdurft, H. C. (1993). The conspicuousness of orientation and motion constrast. *Spatial Vision*, *7*, 341-363.

Pachella, R. G. (1974). The interpretation of reaction time in information processing research. In: B. Kantowitz (Ed.), *Human information processing.* 41-82.

Palmer, J. (1994). Set size effects in visual search: The effect of attention is independent of the stimulus for simple tasks. *Vision Research, 34,* 1703-1721.

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., Ames, C. T., & Lindsey, D. T. (1993). Measuring the effect of attention on simple visual search. *Journal of Experimental Psychology: Human Perception and Performance, 19*, 108-130.

Palmer, J., & McLean, J. (1995). Imperfect, unlimited-capacity, parallel search yields large set size effects. Paper presented at the annual meeting of the Society of Mathematical Psychology, Irvine, CA (transcript available from the authors).

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

Pashler, H. (1988). Cross-dimensional interaction and texture segregation. *Perception & Psychophysics, 43,* 307-318.

Pashler, H. (Ed). (1998). *Attention*. Hove, England: Psychology Press/Erlbaum (UK) Taylor & Francis.

Pelli, D. G. (1990). The quantum efficiency of vision. In C. Blakemore (Ed.), *Vision: Coding and efficiency* (pp. 3-24). Cambridge, UK: Cambridge U. Press.

Pike, R. (1973). Response latency models for signal detection. *Psychological Review, 80*, 53-68.

Poder, E. (1999). Search for feature and for relative position: measurement of capacity limitations. *Vision Research, 39,* 1321-1327.

Pomerantz, J. R., & Pristach, E. A. (1989). Emergent features, attention, and perceptual glue in visual form perception. *Journal of Experimental Psychology: Human Perception and Performance, 15*, 635-649.

Raab, D. (1962). Statistical facilitation of simple reaction time. *Transactions of the New York Academy of Sciences, 43,* 574-590.

Ramachandran, V. S. (1988). Perceiving shape from shading. *Scientific American, 259*, 76-83.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85,* 59-108.

Ratcliff, R. (1988). Continuous versus discrete information processing: Modeling accumulation of partial information. *Psychological Review, 95,* 238-255.

Ratcliff, R., & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science, 9*, 347-356.

Reinhardt-Rutland, A. H. (1994). Perception of motion in depth from luminous rotating spirals: Directional asymmetries during and after rotation. *Perception, 23,* 763-769.

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

Rentschler, I., Hubner, M., & Caelli, T. (1988). On the discrimination of compound Gabor signals and textures. *Vision Research, 28,* 279-291.

Saarinen, J. (1996). Visual search for positional relationships between pattern elements. *Scandinavian Journal of Psychology, 37,* 294-301.

Sagi, D. (1995). The Psychophysics of Texture Segmentation. In T. Papathomas, C. Chubb, E. Kowler, & A. Gorea (Eds.), *Early vision and beyond*, 17-25. Cambridge, MA: MIT Press.

Schwarz, W. (1994). Diffusion, superposition, and the redundant-targets effect. *Journal of Mathematical Psychology, 38,* 504-520.

Sekuler, A. B. (1992). Simple-pooling of unidirectional motion predicts speed discrimination for looming stimuli. *Vision Research, 12,* 2277-2288.

Shaw, M. L. (1980). Identifying attentional and decision-making components in information processing. In R. S. Nickerson (Ed.), *Attention & Performance, 8I (106-121)*. Hillsdale, NJ: Erlbaum.

Shaw, M. L. (1984). Division of attention among spatial locations: A fundamental difference between detection of letters and detection of luminance increments. In H. Bouma & D.G. Bouwhuis (Eds.), *Attention & performance, Vol. 10: Control of language processes.* Hillsdale, NJ: Erlbaum.

Shiffrin, R. M., & Gardner, G. T. (1972). Visual processing capacity and attentional control. *Journal of Experimental Psychology: Human Perception and Performance, 22,* 758-779.

Smith, D. G., & Mewhort, D. J. K. (1998). The distribution of latencies constrains theories of decision time: A test of the Random-walk model using numeric comparison. *Australian Journal of Psychology, 50,* 149-156.

Snowden, R. J. (1998). Texture segregation and visual search: a comparison of the effects of random variations along irrelevant dimensions.

Journal of Experimental Psychology: Human Perception & Performance, 24, 1354-1367.

Sperling, G. (1960). The information available in brief visual presentation. *Psychological Monographs, 74*.

Sternberg, S. (1969). Memory scanning: Mental processes revealed by reaction time experiments. *American Scientist, 57*, 421-457.

Stone, M. (1960). Models for choice reaction time. *Psychometrika, 25,* 251-260.

Sun, J., & Perona, P. (1996). Early computation of shape and reflectance in the visual system. *Nature, 379*, 165-168.

Takeuchi, T. (1997). Visual search of expansion and contraction. *Vision Research, 37*, 2083-2090.

Theeuwes, J. (1992). Perceptual selectivity for color and form. *Perception and Psychophysics, 51,* 599-606.

Thornton, T., & Gilden, D. L. (2000). Modeling reaction time and accuracy in multiple target visual search. *Investigative Ophthalmology & Visual Science Supplement*. (ARVO) 41/4, S424.

Thornton, T., & Gilden, D. L. (2001). Attentional Limitations in the sensing of motion direction. *Cognitive Psychology*, *43*, 23-52.

Townsend, J. T. (1972). Some results concerning the identifiability of parallel and serial processes. *British Journal of Mathematical and Statistical Psychology, 25,* 168-199.

Townsend, J. T. (1974). Issues and models concerning the processing of a finite number of inputs. In B. H. Kantowitz (Ed.), *Human information processing: Tutorials in performance and cognition,* 132-185. Hillsdale, NJ: Erlbaum.

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

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

Townsend, J. T., & Nozawa, G. (1995). Spatio-temporal properties of elementary perception: An investigation of parallel, serial, and coactive theories. *Journal of Mathematical Psychology, 39*, 321-359.

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

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

Treisman, A., & Souther, J. (1985). Search asymmetry: A diagnostic for preattentive processing of separable features. *Journal of Experimental Psychology: General, 114*, 285-310.

van der Heijden, A. H. C., La Heij, W., & Boer, J. P. A. (1983). Parallel processing of redundant targets in simple visual search tasks. *Psychological Research, 45*, 235-254.

Van Zandt, T., Colonius, H., & Proctor, R. W. (2000). A comparison of two response time models applied to perceptual matching. *Psychonomic Bulletin & Review, 7*, 208-256.

Verghese, P., & Nakayama, K. (1994). Stimulus discriminability in visual search. *Vision Research, 34*, 2453-2467.

Wang, Q., Cavanaugh, P., & Green, M. (1994). Familiarity and pop-out in visual search. *Perception and Psychophysics, 56*, 495-500.

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

Weiss, Y., Simoncelli, E., & Adelson, E. H., (2002). Motion illusions as optimal percepts. *Nature:Neuroscience, 5*, 598-604.

Wickelgren, W. A. (1977). Speed-Accuracy Tradeoff and Information Processing Dynamics. *Acta Psychologica, 41*, 67-85.

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

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

Wolfe, J. M. (1998b). Visual Search. In H. Pashler (Ed.), *Attention*, 13-73. Hove, England UK: Psychology Press/Erlbaum (UK).

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 and Performance, 15,* 419-433.

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

VITA

Thomas Lafayette Thornton Jr. was born in Austin, Texas on February 13, 1963, the son of Judith Lawrence Thornton and Thomas L. Thornton. After completing his work at McKinney High School, McKinney, Texas in 1981, he entered the University of Texas at Austin. After two years of study in Physics and English, he left the University to pursue a career in music. From 1986 to 1993 he was employed as a teacher, and later as an assistant director at the Phoenix Preschool in Austin, Texas. In the spring of 1988 he married, and shortly thereafter began work on a family of three boys (Jack L., age 13, Jerod G., age 7, and Drew C., age 4). In the fall of 1993 he re-entered the University of Texas at Austin to pursue undergraduate training in Psychology, and in 1994 received a Bachelor of Arts in Psychology with a minor in English and Mathematics. In the fall of 1995 he entered the Graduate School of the University of Texas at Austin to study Cognitive and Perceptual Psychology under the tutelage of David L. Gilden, Ph. D.

Permanent Address: 1603 Cliffwood drive, Austin, Texas 78733

This dissertation was typed by the author.