DECISION MAKING IN VISUAL SEARCH: A DUAL-MODELLING APPROACH TO EXAMINE THE INFLUENCES OF ATTENTIONAL TEMPLATES ON RESPONSE TIME DISTRIBUTIONS

by

YI-SHIN LIN

A thesis submitted to the University of Birmingham for the degree of

DOCTOR OF PHILOSOPHY

School of Psychology

College of Life and Environment Sciences

University of Birmingham

January 2015

Abstract

The remarkable ability of human search inspires computer vision algorithms that have influenced our life (e.g., Itti, Koch, & Niebur, 1998). The algorithms often automatically reach an optimal decision upon finishing analysing stimulus information. This is not how human search works. One missing puzzle psychophysical studies have yet to answer is how a search decision is reached. This thesis addressed this question, using descriptive models to examine response time (RT) distributions and cognitive process models to reveal the hidden information. The thesis focused on those search paradigms, guided (Wolfe, 2007) by an attentional template (Duncan & Humphreys, 1989), because not until recent years, relevant techniques become more accessible to apply the pioneering approach of dual-modelling. The thesis compared RT distributions and cognitive processes when observers were guided by different attentional templates to search for a target. The first search paradigm was to discriminate 2 from 5 (Study 1, Chapter 4). Study 2 (Chapter 5) varied the working memory strengths in templates, and they were represented differently (null, abstract vs. concrete) in Study 3 (Chapter 6). The findings suggest attentional templates selectively influence different parts of a search decision and RT distributions, depending on how a template is represented, whether it is strengthened or weakened by the conditions, and whether it is concrete or abstract.

Table of Contents

Table of Figures

Table of Tables

List of Equations

List of Abbreviations

2

2 alternative forced choice, 2AFC 41

3

3-dimension, 3-D 15

A

Accessory Memory Item, AMI 125

Akikie Information Criterion, AIC 138

Asynchronous Diffusion Model, ADM 18

B

Bayesian inference Using Gibbs Sampling, BUGS 251

Bayesian Informative Criterion, BIC 138

C

Categorical Template, CT 183

D

Deviance information criterion, DIC 66

Drift-diffusion Model, DDM 17

F

Feature Integration Theory, FIT 11

Frontal Eye Field, FEF 229

G

Guided Search, GS 12

H

Hierarchical Bayesian Model, HBM 53

I

Inter-Stimulus Interval, ISI 55

J

Just Another Gibbs Sampler, JAGS 87

L

latent response, *IR* 145

Lateral Intraparietal, LIP 229

Leaky Competing Accumulator, LCA 42

Acknowledgements

I owe a debt of gratitude to my supervisor, Prof. Glyn Humphreys for giving me the opportunity to pursue the Ph.D., especially at the time when he was moving away to Oxford and taking on another great responsibility himself. I also thank Dr. Dietmar Heinke, who took on, the unexpected supervision work in the midway of my Birmingham days. I appreciate his patience with the ever-changing, tedious administration work for managing PhD students and am very grateful for his high tolerance towards my idiosyncratic research approach and my obsession with research ethics.

My thanks should go also to my important collaborators, Prof. Andrew Heathcote and Dr. Scott Brown, whose brilliant mathematical works not only inspire me to take on the challenge of the formidable mathematical psychology, but also undoubtedly make this thesis possible. The collaboration won't be possible without Scott's kind financial support for my visit to their lab 10,548 miles away. The visit would not be possible without Pia's helpful hands, writing me a referee letter in such a short notice when Glyn was too busy to remember the deadline. I shall also thank Dr. Casmir Ludwig and Pia for acting as my thesis examiners. Their meticulous examination and comments help me to improve the thesis quality greatly, and importantly to correct numerous mistakes I made. I shall also thank the colleagues in Birmingham who are always readily to give generous helps.

Finally, I wish to thank my families and friends, giving me unconditional, both financially emotionally, supports during my PhD journey.

Chapter 1 Top-down Guidance in the Cognitive Information Processing

1.1 Visual Search

Visual search describes the situation when an observer uses his/her visual system – the external ocular system and internal visuo-cognitive neural network – to look for some specific objects in the environment. In addition to early perceptual processes, this behaviour recruits multiple cognitive processes, including attention, object recognition, working memory and cognitive control. Finding an item, such as a car or a door key may seem trivial. Finding a hazard in a workplace, a dangerous item in a luggage check point, an adverse sign in an X-ray or MRI image or a missing airplane in an open sea, is however obviously critical to a person, a family, or perhaps hundreds and thousands of people. This explains why understanding visual search is important and why it is crucial to understand its main driving mechanism, attention.

Attention, though simple might it sounds, can refer 'to several different processes, even in the context of visual search' (Wolfe, 2007). In an inefficient visual search where focused attention is required to identify an object, attention often refers to *selective attention* operating at the cognitive bottleneck of information processing. This type of attention selects a subgroup of items seemingly in a serial fashion. Particularly, *selective attention* may lead to the selection of an object (i.e., object recognition). This differs from the attention processes examined in, for instance, an 'attentional blink' paradigm (Shapiro, 1994), whereby search items are presented in a rapid temporary sequence and an item appearing right after an imperative target fails to be registered at a

conscious level (though see Koch & Tsuchiya, 2007, 2008; Lamme, 2003; arguing for treating attention and consciousness differently). This failure to register items in the attentional blink paradigm can be attributed to a lack of attentional capacity, so attention, like the eyes, 'blinks'.

This thesis focuses on *selective attention*. This by no mean implies the other types of attention play a less important role in visual search. On the contrary, they relate to at least two different, but relatively less explored questions, (1) whether attention necessarily leads to awareness and (2) how top-down goals (& reentrant processes) increase attentional sensitivity. It is too ambitious to address all types of attention that may relate to visual search. The thesis chooses to focus on one cognitive process, working memory (WM), which, as recent evidence shows, may interweave with *selective attention* in visual search (see a review in Soto, Hodsoll, Rotshtein, & Humphreys, 2008). The interaction of WM and selective attention determines importantly how an observer, after his/her attention is guided to a target/distractor, reaches a search decision. This thesis proposes a new analytical framework, an integrated model of sequential-sampling and response time (RT) distributions, to examine how different WM representations, in terms of their strength and perceptual concreteness, influence search decisions (see an early review regarding search decision in J. Palmer, Verghese, & Pavel, 2000) and how these influences might reflect on different parts of an RT distribution.

1.1.1 The Architecture of Information Processing

Visual search can be conceptualised as a general manifestation of cognitive information processing. This is for instance illustrated in Townsend and

Ashby's (1983) general framework of memory-scanning and visual-search paradigms. This framework unifies both paradigms as involving the operations of search and stimulus comparison. Both paradigms involve retrieving information from transient memories and comparing this information against a list of items. The difference lies in the temporal order of stimuli. The memory-scanning paradigm (Sternberg, 1966) presents a target after a search list, in which a visual target coming later is compared with an early search list stored in transient memory, whereas the latter paradigm does the reverse. That is, a visual search list is compared with a memorised target. A basic form of this model comprises of four processes: stimulus encoding, stimulus comparison, response selection, response execution (Figure 1-1). The stage of stimulus comparison plays a critical role in a search decision and is the focus of the thesis.

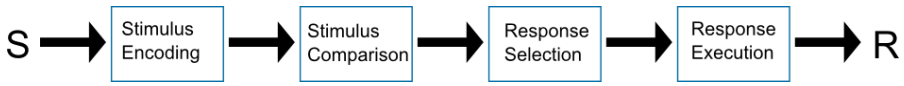


Figure 1-1. A unified discrete model of memory scanning and visual search paradigms. S and R stand for stimuli and responses, respectively. For illustrative purpose, the diagram is presented with arrows pointing from a preceding stage to a succeeding stage. This does not imply a cognitive system processes information sequentially. The issue of processing architecture remains an open question, requiring more data to gain further insights. The diagram is partly adapted from Townsend and Ashby (1983).

1.1.2 The Issues of Cognitive Information Processing

One long-time debate on cognitive information processing is how multiple stimuli or elementary features of a stimulus are processed[1]. Are they subjected to simultaneous processes, are they just sampled one after another and serially

[1] Another equally important issue of parallel vs. serial processing is how separate cognitive subsystems (or channels) operate. For example, one visual stimulus can possess multiple features, such as line orientation and contrast. The channel processes line orientation and that processes the contrast could operate simultaneously, sequentially and the mixture of the two.

fed into the cognitive system, or are there other more complex ways to process stimuli (Townsend & Ashby, 1983)? The processing architecture question, such as parallel vs. serial, is interwoven with three other related issues – when does a cognitive system reach a decision, does a cognitive system process information with a limited or an unlimited capacity, and whether the processing of multiple stimuli is stochastically independent (Fific, Townsend, & Eidels, 2008)?

The first question – whether multiple stimuli are processed serially or concurrently – refers to the processing architecture problem, because it involves the data structure (*architecture*) and how it may be handled (*processing*). Take processing multiple elementary features as an example[2]. A serial structure is a typical example of a simple processing architecture. Take a conjunction visual search task as an example. In this paradigm, a target is usually defined conjointly by 2 (or 3) elementary features, such as colour and orientation. At the outset, the observer may be probed with the identity of a target, such as a black vertical bar (▮). A target-present search display comprises of one target (same as the probe, ▮) and a mixture of distractors (also called non-targets in literature), such as black horizontal (▬) and white vertical bars (▯). To execute a correct response (e.g., pressing either target present or absent key), the observer needs to find the item in the search display that matches both colour and orientation with the probe.

In a serial architecture, the observer may randomly decide first to investigate colour and find that using only colour information is insufficient to tell

[2] The classic debate regarding to serial vs. parallel processing centred on serially or concurrently processing multiple display items, which is different from the feature example here.

what the target is with high confidence. She may then go on to examine the orientation information and thereby gains enough confidence to make a response. This type of architecture is shown in the upper part of Figure 1-2, illustrating which features are investigated one after another. A second type of architecture is to sample simultaneously both colour and orientation information in two parallel processing channels, as shown in the lower part of Figure 1-2. The results from both channels then merge to inform a search decision.

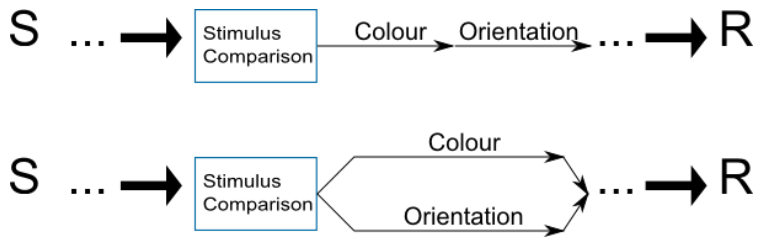


Figure 1-2. The parallel and serial processing architecture. S and R represent stimulus inputs and response outputs, respectively. Features (colour, orientation, etc.) matching happens at the stage of stimulus comparison. The dots refer to other processes, such as stimulus encoding, response selection, motor responses, etc.

The second question asks when and how a cognitive system may finish processing and reach a conclusion. This is usually dubbed the stopping, quitting, or exiting rule. This appears nonobvious in the aforementioned example, whereby the observer requires both the colour and the orientation information to make a correct response. Suppose now both types of distractors are changed to white horizontal (▭) and vertical bars (▯; as in feature search paradigm). In this situation, it is possible for an observer to reach a correct decision, when she compares only the colour information. The exiting strategy can be seemingly 'self-terminating'. Take the hypothetical paradigm as an example. If the cognitive system randomly samples colour feature first and compares it with the probe

colour, the system can soon enter the stimulus comparison stage and proceed to next stage of information processing (i.e., initialising a response). An early exit can also happen in a parallel architecture. Because the processing time may differ, a self-terminating cognitive system will finish the stimulus comparison when, for example, the colour processing channel finishes first and returns its result, indicating enough information to make a correct response. Thus, the system quits the circuit, even though the orientation processing channel is still operating.

Another exiting process often been contrasted to the self-terminating process is that a cognitive system exhausts all possible information. This type of system exits only when all possibilities have been thoroughly examined. When within a parallel architecture feature search task is conducted with an exhaustive process, the system will wait until the orientation information has been thoroughly investigated, even though the early output from the colour channel may be enough to inform a correct response. The exhaustive process however might exist only in theory, because an exhaustive examination of search space is no guarantee of a correct decision. A cognitive system may suffer from memory decay, imperfection of sensory information, the limitation of cognitive resources etc. The exhaustive process might just be a specific example of self-termination process when a seemingly thorough examination of search space is required to reach a high confidence for decision-making. Later in Section 1.5, I will discuss the decision threshold view that considers the exiting rule as an adjustment of a threshold associated with the decision confidence.

The third elementary question in the cognitive information processing is the capacity issue. The capacity of a cognitive system refers to particularly how many stimuli or elementary features of a stimulus a system is able to handle[3]. An elementary processing unit could be, for example, vertical bar (▯), the vertical line of the bar (ı) or the turning corner of the bar (⌋). An early common finding supporting *limited capacity* is a positive (and negative) correlation of the response latency (and accuracy) with the number of elementary units. The limited capacity implies that the system cannot process all stimuli at once. This is frequently observed in an inefficient search task, such as conjunction search.

Nonetheless, a cognitive system may appear to operate with unlimited capacity. This is what has been found in efficient search paradigms, such as feature search [e.g., looking for a target (▮) with a distinct feature dimension from all the other distractors (▭, ▯)]. In the case of efficient searches, response latencies and accuracies are seemingly independent of the number of stimuli. The capacity issue is associated with the processing architecture problem. One reason is that the predicted processing times of a stimulus differ amongst the four systems: capacity limited serial, capacity unlimited serial, capacity limited parallel, and capacity unlimited parallel. With an identical number of stimuli waiting to be processed, a serial system with an unlimited capacity will still prolong response latencies because a late processing stage has to wait until a

[3] The capacity issue is usually understood in the context of the finest cognitive unit, as atoms in classic physics. As physicists now realise Higgs boson, an even more elementary particle in the Standard Model of particle physics (Aad et al., 2012; Barate et al., 2000), experimental psychologists and neuroscientists are also continuously working on searching for a finer cognitive processing unit, such as a communication amongst multiple synapses (Kandel, Schwartz, & Jessell, 2000). At a macro-functional level, it is plausible to model a cognitive system with an elementary unit of one stimulus.

preceding stage is finished, even when processing an identical number of stimuli as a parallel system does. A further issue associated with the interaction of the processing architecture and the system capacity is the unobservable operation of *capacity reallocation* (Townsend & Ashby, 1983). An observer may speed up to an asymptotic level when she gets more experience at handling similar stimuli (i.e., the effect of *priming and/or practice*). She may slow down also to a certain asymptotic level when she gets tired or bored at responding to a large number of monotonous stimuli (i.e., the effect of *tiring/attentional lapse*). The priming effect may result from when an observer realises, with experience, that certain aspects of a stimulus are less critical and thus reallocates resources to other aspects of the stimulus. The attentional lapse may result from when, after numerous trials, only a small amount of capacity is available to correctly process the stimuli. Both effects likely are caused by reallocating resources, thereby leading to changes in the response performance.

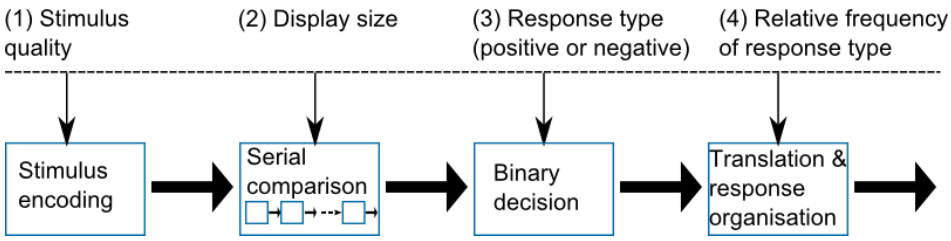


Figure 1-3. Serial processing architecture in a two-alternative forced choice (2AFC) paradigm. Adapted from Sternberg (1969).

The last question concerns whether multiple processing channels influence one another (*stochastic dependence*) before concluding a final output, regardless of parallel or serial processing. This question relates to how a psychophysicist designs an experiment to measure an observer's response, which is frequently quantified as RTs, as a final output from the multiple preceding

processes. In a typical experiment, the psychophysicist usually probes an unobservable cognitive process by measuring the duration from when an observer perceives an imperative stimulus to when she commits an observable response. The duration may be contrasted, for example, between searching for a target in a one-item and in a two-item search display. If the channel processing the additional item in the two-item display depend on that for the first item, a simple subtraction method is able to inform how long one may process the additional item and to imply that an additional channel is dedicated to this process. The stochastic dependence accounts for that the second channel only starts to work until the first channel delivers its outputs to the second channel. If those channels are stochastically dependent, one can apply Sternberg's *additive factor method* (1969) to the data collected from a simple factorial design to determine whether an experimental factor influences a given stage of a presumed serial processing system. This gives rise to the concept of *selective influence* (Dzhafarov, 2003; Dzhafarov & Gluhovsky, 2006). That is, an experimental manipulation influences selectively one processing channel, but not others.

Table 1-1. A summary of architecture issues of the cognitive information processing.

	Mode 1		Mode 2	Other related operations
Processing architecture	Serial	vs.	Parallel	Hybrid
Exiting rule	Exhaustive	vs.	Self-terminating	Resource reallocation
Capacity	Limited	vs.	Unlimited	Supercapacity
Processing independence	Dependent	vs.	Independent	Selective influence, guidance

1.1.3 Early Theories of Information Processing in Visual Search

Early theorists proposed that visual search composes of two fundamental stages: preattentive and attentive processes (J. E. Hoffman, 1979; Neisser, 1967). The preattentive stage processes all perceptual inputs within a visual field simultaneously, and the attentive stage focuses attention to serially analyse the outputs from the preceding stage. The latter was frequently conceptualised as a serial, self-terminating process (Treisman & Gelade, 1980). The two-stage model was based on the common finding that feature search shows a near zero search slope in the function, $f(display\ size) = RT$, and conjunction search typically shows a roughly linear function (Quinlan & Humphreys, 1987; Sternberg, 1966, 1975; Treisman & Gelade, 1980). One explanation to the typical conjunction searches is that the more items a search display has, the more serial stages are needed before responding. That is, adding one more conjunctive item will take up one more processing step, which linearly increases RTs. The self-terminating rule is suggested from the evidence showing that the ratio of the search slopes, $\frac{target\ absent\ trials}{target\ present\ trials}$, is slightly higher than 2. A large corpus of visual search data indicates that an observer spends about 20-30 ms to examine an item in a target (present) trial and about 40-60 ms in a blank (target absent) trial (Wolfe, 1998)[4]. This evidence is consistent with self-terminating rule, because observers seem to search, on average, half of the items in a target trial. Not all items are examined

[4] This is a meta-analysis finding mining from a large corpus of visual search experiments (1 million trials) conducted in Wolfe's laboratory. As Wolfe's GS4 (2007a) used their carwash metaphor arguing adequately that the slope of RT × display size function is a measure of the rate of processing, instead of a per-item search time.

as the exhaustive terminating rule predicts, and some mechanisms determine an early exit. The self-terminating rule becomes naturally one of the candidates.

One prominent two-stage model is Feature Integration Theory (FIT) (Treisman, 1986; Treisman & Gelade, 1980; see also Bergen & Julesz, 1983; Julesz, 1984; Julesz & Bergen, 1983; who proposed the Texton Theory, another equally important and similar early model). FIT claimed that the preattentive stage is composed of multiple parallel processors, the feature maps. These maps compute the level of visual distinctiveness (saliency[5]) for all low-level features (e.g., red, yellow, blue colours, horizontal, vertical orientations, etc.) at each location of a visual field. The saliencies are summed up (integrated) across parallel feature maps and form a master map that determines the final level of saliency at the preattentive stage. The higher the saliency of a location, the earlier it catches the observer's attention. In this case, the highest salient item 'pops-out' from a visual scene. Put it in slightly different words, the saliency affects the observer's bottom-up attention. This suggests that, in feature searches, no matter how many items a search display contains all that is needed to find a target is a highly discriminable feature. No feature binding is required. In contrast to the preattentive stage, the attentive stage recruits focus attention to analyse stimuli when binding multiple features (e.g., inefficient conjunction searches) and when detailed analyses are required with a low discriminability target (e.g., spatial configuration searches; one of this types is searching T

[5] The term, saliency, was coined by Itti, Koch, & Niebur (1998).

amongst rotated Ls). The simple idea of a two-stage process accounts for a number of early findings and thus rendered FIT an attractive hypothesis.

The simplicity of early FIT served a very good working model stimulating further expansion to various visual search paradigms examining more complicated features. The studies expanded the coverage of FIT beyond traditional visual features (i.e., colour & orientation) to others, such as item size and stereoscopic depth. However, the early FIT did not readily account for those features. Neither did it account for some conjunction searches that give flat RT × display size functions (Nakayama & Silverman, 1986a, 1986b; Wolfe, 1998a). Such cases suggest that the preattentive stage can process feature binding relations in a conjunctive search (though see Treisman & Sato, 1990, for an alternative account). In addition to the conflicting findings, the early FIT did not take individual differences into account. In conjunction searches for example, some observers search as efficiently as in feature search, some show moderate search slopes, and the others a steep slope (Wolfe, Cave, & Franzel, 1989). FIT was revised to account for these by introducing several new processes such as inhibiting distractor features if they differed sufficiently from the features of target (Treisman & Sato, 1990). Its main appeal of generality and simplicity no longer stands. The revised FIT became cumbersome without a clear quantitative description or implementation.

1.2 Guided Search

To accommodate the conflicting data and to provide a thorough quantitative description, Cave and Wolfe (1990) proposed Guided Search (GS) model as a successor of the early FIT. GS maintained the two-stage architecture

and introduced, most importantly, the idea of *guidance*. It evolved from the early theories, including Treisman's FIT (Treisman, 1986; Treisman & Gelade, 1980) and Julesz and Bergen's Texton Theory (Bergen & Julesz, 1983; Julesz, 1984; Julesz & Bergen, 1983), with an algorithmic framework to implement a visual search simulation that demonstrated its ability to accommodate conflicting data. GS's computational implementation however simplified many aspects of the search behaviour and thus left room for later 'upgrades'. The theory was upgraded to GS2 in 1994 (Wolfe, 1994), re-implementing an enhanced GS, to GS3 in 1996 (Wolfe & Gancarz, 1996), adding the implementation of eye movements and a saccade map, and to GS4 in 2007 (Wolfe, 2007). In the latest form, the theory is starting to account for decision making (object recognition in Wolfe's term) in search and response distributions. This section briefly reviews the evolution of GS and concludes with a decision to choose GS4 as the working model to examine the role of decision-making and WM in visual search.

1.2.1 Bottom-up Activation Map

On the one hand, GS, similar to FIT, claimed that the preattentive process gathers low-level features simultaneously across all locations in a visual field and the attentive process analyses search items serially. On the other, GS, contrasting to FIT and the account of single, multidimensional map (Duncan & Humphreys, 1989), defined feature maps 'categorically'. It treated different feature dimensions, (colour, orientation, motion, depth etc.), rather than different feature attributes, as FIT does, within a dimension (red & yellow within colour dimension; 15°, 30°, & 45° tiled lines with orientation dimension), as separate maps. The feature maps give different activation values for each location in a

visual field, depending on feature saliencies. For example, when a visual field contains a horizontal bar (▭) and numerous identical vertical bars (▯), the location with the horizontal bar will elicit a higher activation value than the rest of locations, because the horizontal bar is the only item with a unique orientation feature. Mathematically, the bottom-up activation values ($a_{bottom\text{-}up}$) are computed via an exponential function that takes the summed absolute differences of an item's and its neighbours' feature values (*f*), $a_{bottom-up} = e^{\left(\frac{\sum_{j=0}^{n}|f_i - f_j|}{n-1}\right)}$. The *i* indexes the investigated item, and *j* indexes all its neighbouring items (i ≠ j). The resulting summed absolute difference is scaled by the total number of search items (in a visual field) minus 1 (*n-1*) (Wolfe, 1994). The feature maps are weighted according to feature saliencies, and summed across, for example, colour, motion, orientation, and size maps, to form a single activation map.

The activation map charts the saliency value on each location in a search display. This was later revised in GS2 as a terrain map with hills, indicating high probability and valleys, indicating low probability locations (Cave & Wolfe, 1990; Wolfe, 1994). It influences the sequences of attentional deployment. The map guides the attentive process, under the influences of noise, to shift attention serially from one location to a next until it finds a location with an activation value surpassing a terminating threshold or until there is no location with an activation value surpassing a pre-set response threshold.

Guidance is GS's key argument that search decisions are made after the attentive stage. This is in contrast to early FIT, which assumes that both stages are able to initiate a decision. In FIT, when a target is identified by the preattentive

process, a search decision is initiated, bypassing the attentive process. GS asserted that the preattentive stage feeds a noisy activation map to the attentive stage. When the target possesses only a unique salient feature, this map gives the location with the target a high probability, so it unambiguously attracts the first few instances of attention. The attentive stage, thus, is guided rapidly toward the target and afterwards initiates a response. This allows GS to account for the 'standard' (Wolfe, 1994), 2- (or 3-) feature conjunction search pattern. In the conjunction search with a salient target, the target immediately attracts attention because preattentive guidance indicates the target as a high likelihood location. In a similar vein, in the conjunction search with a less discriminable target, attention is guided probabilistically and serially to examine about half of the items. Thus, in this type of conjunction search, RTs increase linearly with display size. As a consequence, the guidance argument accounts for the performance gradient of conjunction searches, from almost 0 to the roughly linear slope of the typical RT × display size function.

Instead of representing each item only by a feature value and, GS2 went further, taking inter-item distance and density into account. In this view, the bottom-up activation value for a stimulus is no longer merely an exponential function that takes the summed absolute differences. Specifically, GS2 computes the differences in feature space[6] between an investigated item and its neighbours (not all distractors). The differences are divided by the distance between the investigated item and its neighbours, so close neighbours result in

[6] As explained previously, GS2, differing from GS, uses a 3-dimension (3-D) activation map. Individual feature dimension is represented by multiple categorical feature channels, such as red, yellow, green and blue in colour dimension.

stronger activation than far neighbours. This is an implementation of the idea of local contrast (Figure 1-4). The values represent how dissimilar an item is with regard to its neighbouring items. This is akin to Duncan and Humphreys's idea (1989) of the target-distractor and distractor-distractor similarity and naturally results in grouping of homogeneous distractors (Humphreys & Müller, 1993). Concisely speaking, the more similar target and distractors are and the less dissimilar distractors are, the harder search will be and vice versa.

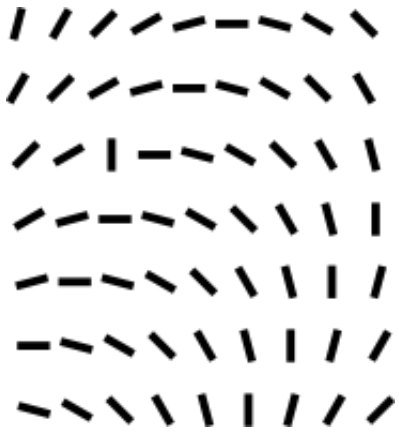


Figure 1-4. The display illustrates the idea of local contrast. The only salient, but not unique, item is the vertical bar in the upper left corner. Three other vertical bars are at the lower-right part of the display. The figure is adapted from Wolfe (2007).

1.2.2 Top-down Activation Map

GS implemented top-down influences on search via another activation map, which contributes also to preattentive guidance. Mathematically, the top-down activation map was computed based on the equation, $a_{top-down} = d|f_i - t|$ at all the locations in a visual field. It reflects a feature dimension-weighted (*d*), which multiplied the absolute difference of the feature values between the investigated item (f_i) and the target (*t*). Note that in GS the top-down activation map computed similarities, instead of differences as the bottom-up activation

map does. Therefore, $|f_i - t|$ measures how close the feature value of the investigated item to the (pre-set) target, presuming that a memorised target representation elicits identical feature values as the visual stimuli in a display. The *d* determines the relative effectiveness of a target feature. For example, looking for a coloured target is faster than looking for an orientation target (Egeth, Virzi, & Garbart, 1984), so the d_{colour} is usually larger than the $d_{orientation}$. An intuitive assumption based on the equation determining top-down effects in search, is that an observer creates, perhaps following the task instruction, a visual image identical to the one been presented in a search display. This renders f_i and *t*, if memory is perfect, sharing an equal hypothetical mathematical unit, so it makes sense to subtract them directly.

As many recent studies have shown, and one of the studies reported in the thesis will demonstrate, there are a number of ways to set up a top-down search goal. A conceptual, abstract top-down goal, for instance, influences differently from a visual preview (see Chapter 6). GS, as well as many other models of computer vision, emphasised more on the bottom-up aspect of search and implemented a simplified 'top-down' (in strict sense, instruction-weighted bottom-up), activation map. The thesis will show later how I incorporated Duncan and Humphreys's (1989) attentional template idea and the decision-making model to address the question of search decision-making, an area that previous GS models left unexplained. GS4 has started to propose its version of the drift-diffusion model (DDM, Ratcliff, 1978)[7], called the asynchronous diffusion model

[7] DDM is one of the most prominent decision-making models introduced in psychological literature by Ratcliff (1978). Later in Section 1.5, I will discuss a number of similar models and how they account for visual search, most importantly, RT distribution data.

(ADM) to address the search decision issue (Wolfe, 2007). One aim of this thesis is to use a wide array of decision-making and hierarchical Bayesian models to address the issue of search decisions. The thesis used this approach to examine a series of visual search paradigms designed to test the role of attention templates when the top-down goal is made clear.

GS2, formulated still from a bottom-up perspective, changed the way it computes the top-down activation map. Instead of using one arbitrary parameter (*d*) as GS does, it selectively boosts the effectiveness of specific categories in feature channels. For example, the top-down activation map may boost the steep category in the orientation map. The selection of which categories to boost is determined by two rules (Wolfe, 1994, pp. 207-208). First, it defines channels as broad categories in a feature dimension. For example, the orientation dimension contains 'steep', 'shallow', 'left', and 'right' categories (Wolfe, Yee, & Friedman-Hill, 1992). These categories are examined and given higher weights in the top-down activation map, if they are found only in a target. Second, the specific feature value belonging to the weighted target channel is compared with the corresponding channel averaged across distractors. Thus, the channel with the greatest positive difference between the target and distractors is selected.

1.2.3 Guided Search 3.0 and 4.0

Aiming to expand from searching laboratory-based images to searching natural images as envisaged by Enoch (Enoch, 1959, 1960), GS3 incorporated the influences of eye movements and viewing eccentricity (Wolfe & Gancarz, 1996). The eccentricity effect reflects that visual search is faster for central than peripheral items (Carrasco, Evert, Chang, & Katz, 1995). GS3 added a saccade

map to mimic the operation of the superior colliculus in overt eye movements in additional to effects on covert attentional shifts. The saccade map used spatial blurring to over-represent the central portion of a visual field (i.e., multiple saccadic movements are around central visual field, because the over-representation of this visual area).

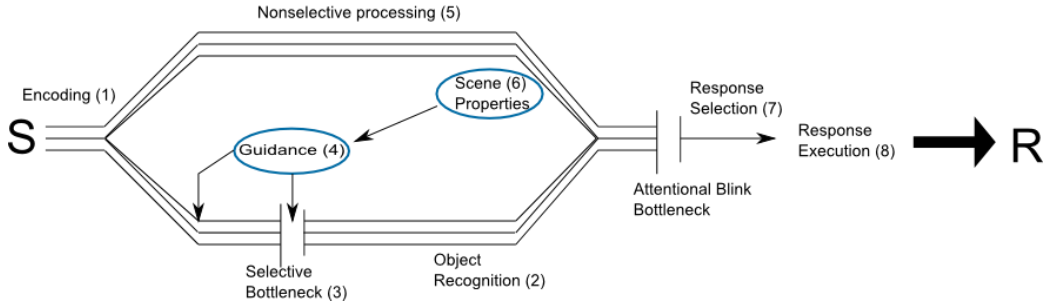


Figure 1-5. The structure of Guided Search 4.0. The oval circles indicate the sources of guidance and the arrow heads point at the guided processes (adapted from Wolfe, 2007). First arrow indicates the preattentive guidance that gathers information from the bottom-up and the top-down activation maps. The second arrow indicates indirect guidance from scene properties, such as pre-knowledge about the type of a search scene and scene statistics. One example for the pre-knowledge guidance is that the knowledge of a typical kitchen will guide (and hence facilitate) searching for a mug, a flying pan or a chopping board, but not a television or a desktop lamp (Wolfe, Võ, Evans, & Greene, 2011). An example for the scene statistics guidance is that when searching for a friend in a beach, the attentive operation will be guided towards the areas of sea or beach, but not the area of sky, because statistically speaking one is less likely to engage in an activity, such skydiving, near a beach.

GS4 (Wolfe, 2007) divorced the process of preattentive guidance from early vision and modified it as a separate control signal (Figure 1-5). The control signal guides attention prior to the stage of selection based on a serial ranking of the activation values from the most to the least probable locations. Each attended location is then passed through the selection, and checked for a match to the target representation (i.e., attentional template).

In addition to repositioning the preattentive process, GS4 started to take the *object recognition* (or in the terminology in this thesis, the *matching* of template-target/template-distractor) into account. It borrowed Ratcliff's well-known DDM (Ratcliff, 1978) to account for the object recognition (OR; i.e., decision-making) process. Suffice to say that the OR process starts to accumulate sensory information and initiates a decision when information is strong enough for identifying either a target or a distractor. GS4 modified DDM as an asynchronous diffusion process, taking a hybrid approach. The important difference is that GS4's accumulators enter the diffusion process serially and accrue information in parallel. As a consequence, an accumulator entering the process later than other early accumulators may leave earlier.

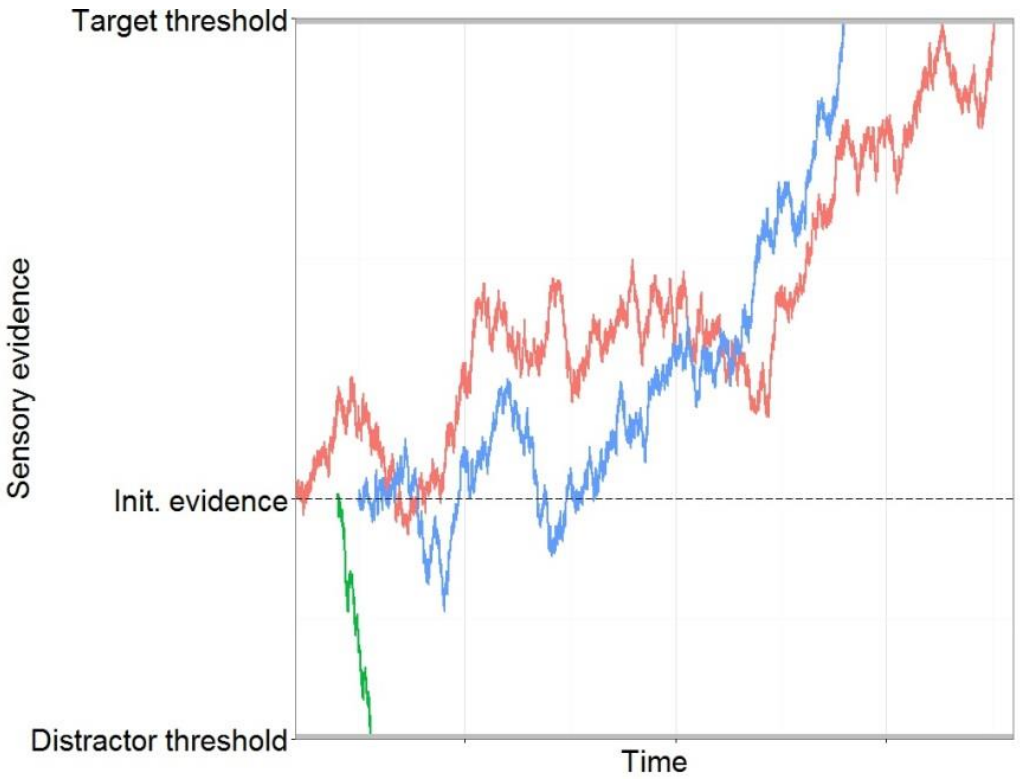


Figure 1-6. An illustration of the asynchronous diffusion model (Wolfe, 2007). The coloured lines represent separate diffusion processes (diffusors in Wolfe's term) for each individual accumulator. Each accumulator starts separately at a different time (i.e., asynchronously), as demonstrated with the blue accumulator starting after the green one, which enters the process after the red one. Also illustrated is that the blue accumulator leaves the process earlier than the red accumulator (see the 'Target threshold' line), because it accrues information on average faster than the red accumulator. The green accumulator leaves the process from the distractor threshold, indicating that an item is identified as a distractor. The different distances between the initial amounts of evidence (Init. evidence) to the target and to the distractor thresholds illustrate that identifying a target requires different amounts of evidence from identifying a distractor.

1.3 Attentional Template

Less clear in the bottom-up search theories is how a top-down goal influences search. Although this may not be an issue in feature (i.e., *odd-one-out*) search due to its use of a salient target, it is a prerequisite for demanding searches, such as conjunction searches, where the target is less discriminable. The formation of a top-down goal is a critical part of a successful search, because its representation differs fundamentally from a physical stimulus. Specifically,

there is a timing asynchrony of the perceptual input of a top-down goal and the physical target in a display. The former is usually encoded during an observer is studying a task instruction, but the latter appears only when the observer starts to engage a search.

The idea of an advanced internal representation, the dubbed *attentional template* (or originally attentional trace, used by Näätänen, 1985) was from Duncan and Humphreys's visual selection (VS) theory (1989). The attentional template is thought to be held as a *perceptual description* formed from the task instruction. One crucial concept in VS theory is to 'make contact' with nonvisual properties in memory, and the selection operates in a way of competition for limited resources to enter visual short-term memory (VSTM). It follows that the attentional template will result in a matched response, if a target item is selected into VSTM; otherwise, mismatched responses are made. As to how a response is decided how the 'make contract' operates, VS did not specify and this is one of the focuses of the thesis.

The other representation in VS theory is the search display, which is typical composed of several items, each a collection of visual features. In VS theory, a search display is also a perceptual description, which is hierarchically organised as *structural units* (analogy to Kahneman & Treisman's, object files, 1984; and Marr & Nishihara's 3-D models, 1978). It results from an early parallel, capacity-unlimited operation that forms a preconscious perceptual organisation process. The VS hierarchy specifies how a visual input is perceived. One example VS theory used is a word (a whole), constructed often by multiple letters (parts). A part can be further viewed in finer scope, such as a horizontal line, a concave, or

a tiled line, so a perceptual description is formed by the hierarchical coding of structural units.

A structural unit can be an entire search display, part of a display, an item in a display, or a colour attribute, such as red, within a search item. The determination of a unit depends on how perceptual descriptions are grouped, by early parallel processes (this was dubbed *perceptual grouping* by Duncan & Humphreys, 1989). Consider two visual principles – proximity and similarity – that contribute to perceptual grouping. Firstly, the closer the perceptual descriptions, the more probable they are to be grouped together as a structural unit. It can be illustrated readily by the word example where two letters within a word tends to be grouped together (probably also with other letters in an identical word) when compared with two letters in two separate words. The second principle states that the elements with the same colour, motion, or even semantic meaning, tend to be grouped together. For example, an observer tends to perceive dots moving toward more or less same direction as more homogeneous than those moving randomly.

In summary, VS theory claims that both the search goal and the search display are perceptual descriptions encoded by a parallel, capacity-unlimited process. The search goal is represented as an attentional template, possibly held in VSTM for the upcoming search display (Duncan, Humphreys, & Ward, 1997). The stimuli in the search display are grouped as a number of size-varied structural units, following principles such as inter-item proximity and perceptual similarity, similar to the ideas as what GS has formulated in its bottom-up activation equations that take item density and distances into account. Those

structural units compete for subsequent operations by accessing the resource-limited VSTM, a key point I will develop in the next section.

1.3.1 Working Memory

One key notion in VS theory is that, before a response (target selection or distractor rejection) is generated, a structural unit must make it into VSTM (i.e., visual working memory, VWM[8]). VWM admits stimuli based on their 'attentional weights', which can be viewed similarly though not exactly the same as GS's activation values. The VWM idea sets VS theory apart from other bottom-up perspective theories, because it gives a clear role of top-down goal. Visual selection is interpreted as a competing process for entering VWM, which only holds maximally around 4 units of representation at a time (Davis & Holmes, 2005; Luck & Vogel, 1997; Sperling, 1967).

Although both VS theory and GS4 use a similar idea – weights in the structural unit or values in the activation map (also see Bundesen, 1990 for a similar idea), to determine the importance (i.e., likelihood of been selected) of a processed unit, they differ in terms of processing architecture. In contrast to GS4's hybrid architecture, which holds that processed units enter the selective bottleneck sequentially like a car wash, VS theory follows Sperling's parallel architecture, proposing that attentional resources are divided into varying proportions, each engaging in separate structural units. The unit receiving more resources becomes more competitive and accesses VWM quicker and more easily than those that have few resources.

[8] In this thesis, I treat VSTM as the same construct as VWM. It is, nevertheless, disputable in strict sense whether VSTM and VWM are identical.

On the basis of the WM competition idea, Olivers and colleagues (2011) proposed a WM 'offloading' hypothesis, stating that an attentional template might be partially removed from VWM when an experimental procedure called consistent mapping is used (Schneider & Shiffrin, 1977). This procedure refers to, in a cueing-search paradigm, an identical target probe is presented before a search display and is reiterated in every trial as oppose to the varied-mapping procedure, where different probes are used. The offloading account proposes the two slightly different procedures render a search template receives different cognitive resources and thereby how it interacts with search target and distractors. Chapter 5 (Study 2) will examine this hypothesis.

1.3.2 Which Architecture?

The aim of this thesis is not to resolve the architecture question concerning attention, though it should be noted that there is mounting evidence, associated with the attentional operation in visual search, pointing to the dominant role of parallel architecture (Thornton & Gilden, 2007). As explained earlier in sub-section 1.1.2, together with the notion of resource reallocation, divided attention is also able to distribute limited resources to process all items in a visual field, and simultaneously check every one of them. In this way, a parallel architecture can also account for most of the visual search findings based on mean RTs, accuracy and the function of RT × display size (Humphreys & Müller, 1993; Townsend, 1971, 1990; Townsend & Wenger, 2004). For example, a recent work (Fific et al., 2008) found evidence favouring a parallel architecture with interacting (dependent) channels by comparing the contrasts of survivor functions, using the new analytic method based on RT distributional analyses. However, rather than

concentrating on the architecture question, the thesis focuses on how a search decision relates to template representations and how these factors affect RT distributions.

As a working hypothesis, the thesis starts with the hybrid GS4, which has both parallel and serial processing properties to examine specifically what happens during the stimulus comparison process. The thesis focuses on paradigms designed to probe the influence of top-down guidance and the role of VWM in search. It examines how WM affects a search decision, reflecting on the mean latencies, the accuracy and most importantly, the RT distributions. Importantly, GS4 is computationally delineated model that is testable, so refutable – particularly when distributional analyses are taken into account.

1.3.3 Guided Search 4.1

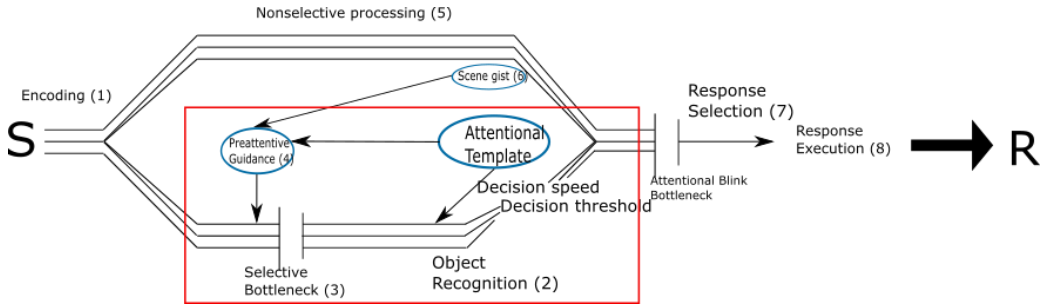


Figure 1-7. A further refinement of GS4. The red rectangle highlights the main difference from GS4.

The thesis assumed an explicit source of guidance from a top-down template, represented in the VWM. The template is a representation that influences the preattentive guidance by giving weights to modify the activation values in the saliency map and a representation that probabilistically determines whether a selective item is a target (i.e., feature comparison). The thesis focuses on how different representations and operations of the attentional template

influences two crucial decision-making processes: (1) the *decision rate* and (2) the *decision threshold*. At this early stage of investigating the decision aspect of visual search, I did not immediately model the asynchrony diffusion structure as GS4 envisions, although it undoubtedly is likely to underlie the true process of a search decision. Instead I used the original structure of the DDM (Ratcliff, 1978), assuming that each accumulator represents one unobserved parallel process of decision-making and manifests what happened when an observer is searching for a target in a display. That is, all possible features within the visual field are taken in at once and presumably, while in the stimulus comparison stage, a selected item (or group of items) is compared against an attentional template encoded earlier and residing in VWM.

Thus, the thesis presumes the *template matching* process influences two visual search aspects. Firstly, the template information correlates with investigated items and thereby generates correlational (response) values that influences the saliency map (Beutter, Eckstein, & Stone, 2003; Najemnik & Geisler, 2005). This pre-attentive influences modify the sequence of attentional deployment (as well as eye saccades). The second aspect of the attentional matching process influences the stimulus comparison. This post-attentional comparison influences the rate of evidence accumulation. Due to the simplified application of the decision-making model, the thesis did not specify how the evidence accumulation process operates and only assumes it occurs concurrently across all features as the time passes.

One reason that I cautiously claim that the attentional template also affects the early preattentive stage is that there exists physiological evidence arguing for

the early modulation of top-down attention on neural structures, including the lateral geniculate nucleus of the thalamus, the striate and extrastriate cortices (Kastner, Pinsk, De Weerd, Desimone, & Ungerleider, 1999; J. Moran & Desimone, 1985; O'Connor, Fukui, Pinsk, & Kastner, 2002). The early guidance due to a top-down memory representation likely influences search before the bottleneck stage, which possibly relevant more to the perceptual, rather than decision, processes. This type of guidance is largely discussed in the literature applying signal detection theory (Green & Swets, 1966) and the Ideal Observer analysis (see a review in Geisler, 2011) when eye saccades are also recorded. Because the thesis focused only on manual responses without simultaneously recording eye saccades, the template influences on early guidance perhaps is better to leave for future modelling efforts and is outside the scope of the thesis.

My version of visual search model branches out GS4 to incorporate the ideas of working memory and attentional templates, as sources of top-down guidance, which GS4 had not made explicit. Because this is not developed by Wolfe and colleagues, the thesis dubs it GS4.1, signifying that it branches out from their main developmental tree and is a beta version[9].

1.3.4 SEarch via Recursive Rejection

In addition to GS4.1, the thesis applied another hybrid search model, SEarch via Recursive Rejection (*SERR*) (Humphreys & Müller, 1993) to account for *item grouping* and *distractor rejection* (see also Heinke & Humphreys, 2004). SERR introduces two important search mechanisms: *group segmentation* and

[9] In the tradition of software development, a beta version is denoted, after the decimal place, with an odd number, comparing to an even number that signifies a stable version of the software.

distractor tagging. The former is an idea built on VS theory's perceptual grouping. This mechanism accounts for a search strategy that segments a large display into several parallel processing groups, instead of analysing item-by-item. The latter mechanism proposes that distractors, instead of completely being irrelevant, serve important functions. This reflects especially in inefficient search, when observers usually conduct few, rather than just one, attentional sweeps to locate a target. SERR hypothesises that distractors are tagged in those unsuccessful sweeps and the tagged distractors, which are represented in memory, facilitate search via serving as distractor templates (see a review in Watson, Humphreys, & Olivers, 2003). This tagging process is recursive, because distractor templates are iteratively reused as long as search is ongoing.

The tagging and recursive processes account for the distractor function, being a facilitator, rather than merely being given less saliency weight. Specifically, the tagging processing enhances the grouping effect for similar distractors by passing tagged features in, for instance, first search sweep to a next, when an observer has not identified the target in the first sweep. The tagging goes on as long as the target has not identified or the observer has not made a guess. The recursive tagging in turn helps offload the cognitive resources, thereby increasing the likelihood of finding a target.

SERR plays an important role especially in this thesis, because the paradigms reported here presented relatively small search items in a confined visible region rendering high density of similar items appeared often when a display contains several of them. Further, except the feature search paradigm in Chapter 4, all other paradigms were designed to require focus attention (i.e.,

inefficient search). SERR provides a good account for the search benefits that are difficult to be explained by purely guidance.

1.4 Response Time Distribution

The analysis of RT distributions, in addition to mean RTs, can permit further insights into different cognitive processes, a point originally raised as early as 1950s in works such as those of Christie and Luce (1956), Hohle (1967), and Sternberg (1969). In essence, one early RT distribution account postulated that the RT is a functional output summing across a decision component that distributes exponentially and a residual component that distributes symmetrically (Hohle, 1965). This postulation later developed into the well-known ex-Gaussian distribution that convolves mathematically the exponential and Gaussian components. The ex-Gaussian function becomes popular mainly because of its capacity of accommodating positively skewed distributions, an observation commonly found in RT data before a representative values, such as mean, are averaged across several observers. A usual practice for analysing RT data is to average multiple observations for different experimental conditions in an observer, and the averaged values in each condition are averaged again across several observers. This practice presumes that the first-level averaged values catch the general shapes of RT distributions, thereby representing well the majority of RT data in a condition. This assumption conflicts with the observation of the skewed RT distributions, which cast doubt on the data generated from the (individual-level) mean RTs, because they may not represent some observations when distributions are skewed.

Further doubt on mean RTs comes from their ambiguity when answering the architecture question. Specifically, the finding of a display size effect and that of the slope ratio of target trial to blank trials calculated from mean RT in fact cannot determine how the cognitive architecture operates. This has been repeatedly demonstrated by Townsend and colleagues, showing that both serial and parallel models are able to predict the data from mean RTs (Townsend, 1971, 1990; see Townsend & Ashby, 1983 for a review; Townsend & Wenger, 2004).

1.4.1 Descriptive Models

The drawback of analysing only mean RTs appears to an alternative approach for analysing RT data, such as the distributional analyses (Lin, Heinke, & Humphreys, 2015; Loft, Bowden, Ball, & Brewer, 2014; Payne & Stine-Morrow, 2014; Toeroek, Kolozsvari, Viragh, Honbolygo, & Csepe, 2014). The ex-Gaussian function, for example, breaks down a distribution into two mathematically and psychologically seemingly separable components: the Gaussian and exponential parts. The latter accounts mathematically for that why an RT distribution skews towards the short latency side. The former keeps the original symmetrical Gaussian part of a distribution. Although the data do not necessarily collaborate the dichotomy of decision and residual components into the Gaussian and the exponential parts of an RT distribution (Gholson & Hohle, 1968a, 1968b; see a recent review, Matzke & Wagenmakers, 2009), recent works show that the value of adopting an ex-Gaussian is that it provides a plausible model to describe positively skewed RT distributions (e.g., Matzke, Dolan, Logan, Brown, & Wagenmakers, 2013). This is in contrast to assuming the Gaussian distribution as the underlying function that generates RTs. The advance is crucial

because the Gaussian function may be only appropriate to account for the mean RTs across multiple observers, rather than the mean RTs across multiple observations within a condition in an observer (see Chapter 4 for the data supporting this point).

The initial attempt to conceptualise RTs, using the ex-Gaussian framework, distinguished perception and decision components and the components involved in the organization and execution of the motor responses (Hohle, 1965). For instance, Hohle's original interpretation of the exponential component was that it reflects perception and decision processes – opposite to McGill and Gibbon's interpretation of a residual motor latency (1965). The early conflicting interpretations and numerous succeeding works (Balota & Spieler, 1999; see a recent review in Matzke & Wagenmakers, 2009; Rohrer & Wixted, 1994) indicates that the mapping of ex-Gaussian components onto cognitive processes only results in paradigm-dependent interpretations. That is, the separation of a Gaussian component and an exponential component is meaningful only at the mathematical, but not cognitive, level; the resultant interpretations of exponential and Gaussian components varied with tested factors and experimental designs. As a consequence of the null finding, it is suggested that the ex-Gaussian model is most useful when been treated 'as a descriptive first-order account of response latency' (Heathcote, Popiel, & Mewhort, D. J., 1991). This view is echoed later in Schwarz's work (2001), where he used an ex-Wald function as a quantitative model, taking advantage of its Wald component to approximate a Wiener diffusion process, a link to the underlying cognitive processes. This was demonstrated in Schwarz's go/nogo

digit comparison experiment, in which observers pressed a button upon detecting a go digit (6 or 9) and withheld button-press upon detecting a nogo digit (5). The results showed that (1) the a priori probabilities of the appearance of go digit (50% vs. 75%) selectively affected only the evidence criterion (a statistical Wald parameter), (2) the numerical distance (6 vs. 9, comparing to 5) selectively affected the drift rate (a second statistical Wald parameter), and (3) the exponential parameter, γ, appeared insensitive to the two aforementioned experimental factors (Schwarz, 2001).

1.4.2 Non-Gaussian Distribution

The modest success by using the ex-Wald function to describe RT distributions lends support to researchers exploring other insights from analysing RT distributions, by comparing different experimental manipulations. Converging evidence supporting the descriptive approach of distributional analyses comes from Ashby and colleagues' work (Ashby, Tein, & Balakrishnan, 1993). They showed, in a Sternberg memory-scanning task, a number of distribution-level predictions, such as the variance RTs and the shape of RT distributions, are inconsistent with the serial and the unlimited capacity parallel models. The descriptive approach to examining RT distributions by comparing different experimental conditions appears promising for understanding other cognitive processes, although not quite as much as the ambitious researchers originally envisioned (e.g., Hohle, 1965).

The success of descriptive distributional analyses using different probability functions and higher distributional moments suggests that the advantage of adopting the ex-Gaussian (or ex-Wald) function does not lie in the

probability function per se, but in its capacity to accommodate empirical RT distributions. The fact is that numerous other probability functions are capable of doing this (Dolan, van der Maas, & Molenaar, 2002; Feige et al., 2013; Heathcote, Brown, & Cousineau, 2004; E. M. Palmer, Horowitz, Torralba, & Wolfe, 2011; Rouder, Lu, Speckman, Sun, & Jiang, 2005). In addition to the probability functions convolving with an exponential function, there are other generic functions, such as the 3-parameter Weibull function, gamma, log-normal, and Wald functions that are flexible enough to accommodate skewed distributions without convolving with an exponential function. These other plausible probability functions allow researchers to describe and compare the shape of RT distributions across different experimental conditions, with a strategy differing from the convolving functions. One of the strategies is to describe RT distributions using the location-scale form of the probability density[10].

1.4.3 The Application of Descriptive Distribution Models to Visual Search

The descriptive approach using plausible probability functions to model RT distributions seems promising to tackle also the problems in visual search paradigms. Specifically, the parallel, limited capacity model can produce linear function for the RT × display size relation, by dividing attention into multiple processing channels when search items increase. The increment of each search item dilutes the limited resources, rendering multiple processing channels sharing increasingly few resources, so lengthening response latencies.

[10] The location-scale form refers to, in the case of the Gaussian function, mean and variance. Further details specific to the probability functions adopted in this thesis will be explained in the later chapters.

One simple application of descriptive distributional analyses is to examine the spreads of RT distributions. RT standard deviations in fact correlate with RT means linearly, as has been showed recently by a meta-analysis for 3 visual search experiments. The finding was dubbed as 'the linear law' of RT (Wagenmakers & Brown, 2007). A more complicated application involves contrasting an entire distribution (i.e., the probability density function) and further its derivatives across different experimental conditions and/or participants. The distributional derivatives include, but are not limited to, the cumulative density function, the survivor function, the hazard function and different applications of these distributional functions based on specific experimental designs. The survivor function, for example, has been used to examine the effect of target-distractor similarity (VS theory originally called it *interalternative similarities* in Duncan & Humphreys, 1989) in a simple visual search paradigm (Fific et al., 2008). Fific and colleagues used the linguistic (Cyrillic) and nonlinguistic meaningless symbols in a simple two-item search task and found evidence favouring a parallel cognitive architecture with positive interacting channels against a serial architecture. This was supported by the data showing different patterns of *survivor interaction contrast function*, which calculated the difference between differences of the survivor functions of the four factorial conditions in a 2 by 2 factorial design experiment (Townsend & Nozawa, 1995) .

In summary, the descriptive distributional analyses suggest a new approach to examining visual search, and the thesis will explore one possible distributional analysis method, the 3-parameter Weibull function, to analyse visual search data.

1.5 Decision-making Models

One drawback of the descriptive distributional analyses is the difficulty in associating the patterns of distributions related to experimental manipulations that are designed to understand cognitive processes. This was addressed in Schwarz's work (2001). He assumed the cognitive process in his go/nogo task is akin to the process of sensory evidence accumulation and this can be approximated by the ex-Wald function, which is a rough approximation of the process of sensory evidence accumulation. A more direct approach is also possible. This direct modelling approach has been studied extensively for simple choice reactions, generally, under the name of sequential sampling models (see a short review in Ratcliff & Smith, 2004). In this section, the thesis discusses two classes of the model: random walk models and race models. They have been applied successfully in a few recent papers (Matzke, Dolan, et al., 2013; Matzke & Wagenmakers, 2009) to complement and validate the interpretations based on descriptive distributional analyses. My discussion focuses on the three sub-classes of random walk model that I have applied on the visual search data presented in the thesis: the DDM (Ratcliff & Tuerlinckx, 2002), the EZ2 diffusion (Wagenmakers, van der Maas, Dolan, & Grasman, 2008), and the linear ballistic accumulator models (LBA; Brown & Heathcote, 2008). Because there exist only few empirical works applying a similar approach to visual search (Purcell, Schall, Logan, & Palmeri, 2012; Ward & McClelland, 1989), I described firstly an early simplified random walk model and then its two specific implementations, the DDM and EZ2 and after that the LBA model.

1.5.1 Random Walk Models

Random walk models account for a discrete time-series process that a variable takes a random step away from its previous value. The value in each step can be sampled from a distribution. In the original random work formulation, the distribution is assumed independently and identically distributed in value. The random walk models are appropriate for choice RT data, because they account for RTs between different choices in a single framework. For example, the noisy operator theory, a subclass of random walk model, accounts for a perceptual matching task as a comparison process (Krueger, 1978). The matching task was to determine if a pair of stimuli is *same* or *different* on the basis of their perceptual features. The noisy operator theory assumed a process that registers differences between two to-be-matched stimuli. Because noise constantly perturbs the cognitive system, the identities of stimuli are not always perfect. Hence, even identical stimuli sometimes elicit different neural responses depending on the noise level. The matching process, explained by the noisy operator theory, is a decision-making process along discrete time steps while an operator checks and counts the numbers of matched features. When the number of matches or mismatches is not enough to conclude *same* or *different,* a rechecking process then takes place. This exemplifies the usefulness of the random work models to explain simultaneously the choice of a *same* or a *different* response and the choice latency.

1.5.2 Drift Diffusion Model

The first application of random walk models with an explicit expression of RT distributions was found in Ratcliff's work (1978; see a recent review in Ratcliff

& McKoon, 2007; see also a neurobiological account of the model in Smith & Ratcliff, 2004). The DDM was applied originally to explain the data on recognition memory (Ratcliff, 1978). Since its first appearance in psychological literature, it has been used in a number of cognitive tasks, including typing (Heath & Willcox, 1990), detection (Diederich, 1995; P. L. Smith, 1995), perceptual matching/discrimination (Ratcliff, 1981; Voss, Rothermund, & Brandtstädter, 2008), lexical decision (Ratcliff, Gomez, & McKoon, 2004; Wagenmakers, Ratcliff, Gomez, & McKoon, 2008), semantic priming (Balota, Yap, Cortese, & Watson, 2008; Voss, Rothermund, Gast, & Wentura, 2013), and visual search (Ward & McClelland, 1989). The wide and successful application lends support to the DDM for modelling cognitive performance, when it is crucial to consider simultaneously both accuracy and response latency.

The DDM posits an instance (time step) of neural computation is to sample sensory information from the environment, which in the case of performing a visual search task, is the search display or GS4's preattentive outputs. The sensory information drives an evidence accumulator (i.e., the diffusor in Wolfe's ADM) starting from an initial amount of sensory information (*z*; influenced by foreknowledge or by a task instruction) and moving towards either a matched (set at a value, *a,* on the y axis of sensory information) or a mismatched decision threshold, while the accumulator drifts randomly (up or down; see Figure 1-8). In contrast to the multiple accumulators modelled by the LBA model[11], the DDM presumes only one accumulator. Depending on the stimulus type, the DDM accumulator is biased, moving towards one threshold, with an average drift rate,

[11] I will discuss it in details in Section 1.5.4

v. For example, in a target trial of a typical 2-feature conjunction search when an instruction has informed some target features (i.e., an attentional template has been set up), the accumulator tends to drift towards a target-present decision, because a presented target is usually given higher attentional weight than a nontarget; thereby the sensory evidence sampled from a target display favouring a present decision. The rate of accumulator drift is theorised as the speed of information accumulation, implying the speed of reaching a decision. The decision speed can be modelled as a random variable sampled from two Gaussian distributions, corresponding to matched and mismatched decisions, with the mean, v^+ *and* v^-, and a common intertrial variance η^2. The within-trial variance of drift rate, namely the scale factor or the drift constant, is usually fixed at a constant s^2 , which some set it at 0.1 (Ratcliff, 1978; Van Zandt, Colonius, & Proctor, 2000), and others set it at 1 (Voss & Voss, 2008). Likewise, the initial amount of sensory information (*z*) is modelled as a random variable sampled from a uniform distribution with an intertrial variance s_z^2. In summary, the limiting forms of the first-passage time distribution for the DDM, respectively for the match and mismatch decisions, are described by the two equations (Feller, 1971, pp 359; Ratcliff, 1978):

$$g^+(t,z,a,v^+) = \frac{\exp[(a-z)v - 0.5v^2t]}{\sqrt{2\pi t^3}} \sum_{n=-\infty}^{\infty} \exp\left\{-\frac{[(1+2n)(a-z)]^2}{2t}\right\}[(1+2n)(a-z)] \quad \text{Eq (1)}$$

$$g^-(t,z,a,v^-) = \frac{\exp[-zv - 0.5v^2t]}{\sqrt{2\pi t^3}} \sum_{n=-\infty}^{\infty} \exp\left[-\frac{(2na+z)^2}{2t}\right](2na+z) \quad \text{Eq (2)}$$

The distributions of the matched and mismatched decision times are described, respectively, by g^+ and g^- functions, each determined by a set of DDM parameters (*z*, *a* & *v*) and the decision time (t). The measured, empirical RT is

then composed of the decision time plus the non-decision time (denoted t_0 or sometimes t_{er}), the time to encode sensory information and to execute motor responses. The drift rate and the initial amount of information are separately modelled by a Gaussian and uniform distributions, as described before.

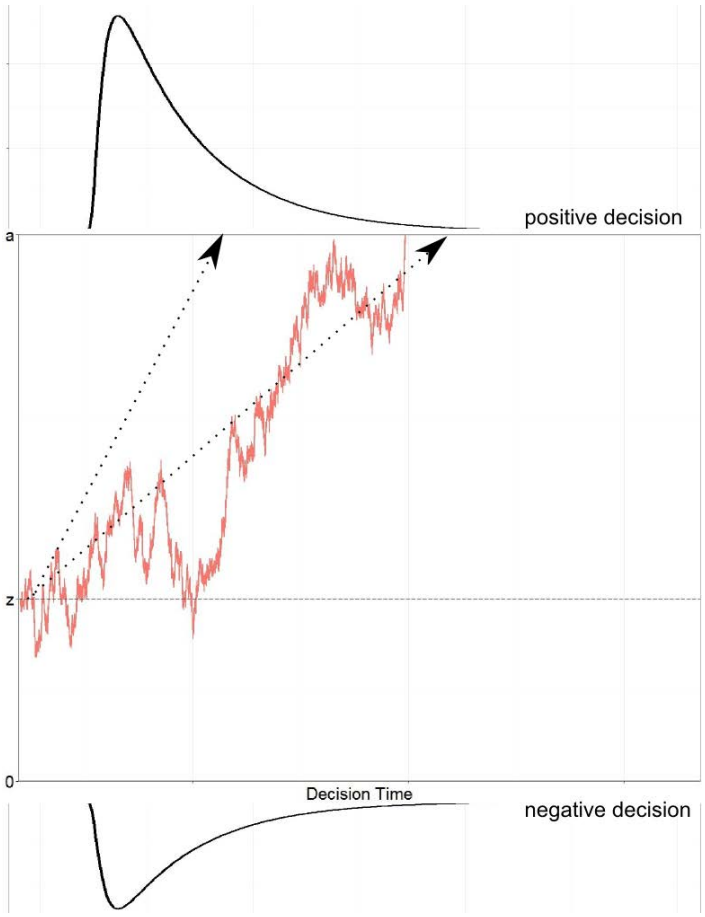


Figure 1-8. Drift-diffusion model. Each arrow represents one instance of decision-making process. Only the second arrow is illustrated with a random-walk line (in red).

The success of DDM comes with its constraints. Firstly, in contrast to the noisy operator theory (Krueger, 1978), the time step of DDM is assumed to be very small. In practice, this is achieved by setting a high precision of time step, such as one DDM implementation set its default time step to minus 3 power below the decimal place (Voss & Voss, 2007). This implies a millisecond resolution and

thereby increases the burden of computation. Secondly, the DDM strictly accounts only for 2 alternative forced choice (2AFC) tasks. As is clear in the DDM density functions [Eq (1) and Eq (2)] as well as Figure 1-8, the initial amount of information is modelled as a relative position to the positive boundary (*a*) and to the negative boundary (*0*), so does the term, *a-z*, in Eq (1) becomes *z* in Eq (2). The third constraint is that the DDM is appropriate only for the situations in which a 'single stage' of decision making governs performance. That is, it accounts only for the paradigms that observers are able to complete a response with one sweep of decision-making process (Ratcliff & McKoon, 2007). Empirically this implies the uncontaminated response times should be less than 1 s (Ratcliff & Rouder, 1998). One additional possible constraint on the DDM is that it fits accuracy data poorly in some cases (e.g., Van Zandt et al., 2000; but see Ratcliff & Smith, 2004), relative to the Poisson race/counter model (Townsend & Ashby, 1983). As will be discussed in Chapter 5, I also found a similar model fitting pattern when the DDM was pitted against the LBA model, which is a subclass of the race model.

1.5.3 EZ2 Diffusion Model

EZ2 model is a closed form variant of the DDM (Grasman, Wagenmakers, & van der Maas, 2009). It models only the averages of *v*, a and t_{er} , presumes no intertrial variability and sets an unbiased *z* (i.e., *z* = *a/2*). EZ, an early version of the estimation method, takes three observed quantities, RT means, RT variances from correct responses and accuracies as inputs. Each of them are estimated from each condition/individual and serves as inputs to calculate analytically *v*, *a* and t_{er} (Wagenmakers, Maas, & Grasman, 2007; Wagenmakers, van der Maas,

et al., 2008)[12]. EZ2 relaxes the EZ assumption of unbiased z and uses a numerical algorithm to estimate the three EZ parameters.

The simplification of the EZ2 method unavoidably sacrifices some merits of the DDM, such as its ability to describe the shape of distributions both for matched and mismatched decisions (Ratcliff, 2008; Wagenmakers, van der Maas, et al., 2008). Even though these disadvantages suggest that EZ2 cannot fit empirical data (as goodness-of-fit criteria cannot be calculated), it serves as a useful approximation to understand the unobservable decision-making process when the per-condition observation is limited and accuracy is high (Ratcliff, 2008). I applied it on the first study (Chapter 4), which tested the benchmark visual search paradigms (Wolfe, Palmer, & Horowitz, 2010). Below I discussed the LBA model, a race model (Townsend & Ashby, 1983) claiming to provide a process (substantive) model, instead of merely descriptive model, to account for the decision-making process.

1.5.4 Linear Ballistic Accumulator Model

The LBA model is adapted from Usher and McClelland's (2001) leaky competing accumulator (LCA) model. The LCA model describes multiple accumulators which gather moment-to-moment information for their respective decision types. The decision process is modelled by a race amongst these accumulators towards a common decision threshold. The decision type and latency is determined by the first accumulator attaining the common threshold. The LCA, similar to the DDM, samples the drift rate and initial amount of

[12] This is in contrast to the full DDM, which uses numerical optimisation algorithms to approximate the parameters.

information from two probabilistic distributions. It further models two nonlinear processes: one is a passive decay of accumulated information and the other is the choice competition process. In contrast to the DDM, the LCA model accounts for the competition amongst accumulators (choices) explicitly, allowing individual accumulators influence on one another and subjecting the accumulated information to probabilistic leakage.

The LBA model aims to present a parsimonious LCA account and to maintain good fits to both correct and error responses. The model simplifies the within-trial and intertrial randomness from the information accumulation process. The early version of the ballistic model still retained nonlinear leakage and between-accumulator competition (Brown & Heathcote, 2005). However, the LBA model argues that a further simplification, replacing linear independent accumulators with the nonlinear dependent accumulators, still accounts for all important empirical phenomena, such as the shape of the RT distribution, speed-accuracy trade-offs, and more importantly the relative speed of correct vs. incorrect responses. Although it simplifies a great deal of process details, the LBA model is able to make accurate predictions for both correct and incorrect responses where numerous previous models did not fit well to the proportion of incorrect responses and to RT distributions. It has argued that one critical reason is, in contrast to the LBA model, the model with only one accumulator (e.g., Reeves, Santhi, and DeCaro's random-ray model 2005), presumes a Gaussian distribution governs the drift rates. A negative sample from the Gaussian distribution drives the accumulator moving towards the lower threshold, and resulting in an incorrect response. This way to model the cognitive process,

namely the drift rate, results in predicted error RT distributions becoming negatively skewed, a contradiction against the empirical data (Luce, 1986). The separate accumulators, allowing the properties of multiple responses distribute similarly, thus solve the pitfall of one-accumulator random walk models.

1.6 Hierarchical Bayesian Model

One major challenge of the distributional analyses is the necessity of collecting large number of observations. This challenge confronts both the descriptive and the process approaches of distributional analyses. A large number of observations are required to make distributional analyses effective, irrespective of whether it is a descriptive statistical model or a cognitive process model. This is best illustrated by some incredible early endeavours, such as, Ashby and colleagues' memory scanning work (1993), which collected about 1,500 trials per participant per study set size and Ratcliff and Rouder's model fitting work on the perceptual discrimination RTs (1998), which collected about 1,000 per-participant trials (see Cousineau & Shiffrin, 2004 for perhaps the most extreme case for near 6000 trials). The quantity of the observations is far beyond a typical psychophysical experiment would collect and difficult, although not impossible, to achieve.

In addition to the model simplification efforts that may detour the overarching goal from simultaneously fitting the data, estimating processing parameters and describing the distributions, to describing the unobservable processes (e.g., Grasman et al., 2009), there exists another way to trim down the required number of observations. The relevant techniques, developed during the last decade alongside the simulation-based algorithms (e.g., Monte Carlo

integration foresaw in the 1940s by two nuclear physicists, Metropolis & Ulam, 1949) and intelligent probabilistic sampling methods (e.g., the NUTS sampler, Hoffman & Gelman, 2011; the Gibbs sampler, Gelfand & Smith, 1990), make feasible two powerful, statistical methods: hierarchical modelling and Bayesian inference (Bayes, 1970).

1.6.1 Hierarchical Modelling

The canonical form of the hierarchical model [13] is similar with the generalised linear model (Draper & Smith, 1998). That is, a functional output (e.g., an RT) is fully determined by one or more regressors plus with a residual term modelled by a stochastic function, for example in matrix form:

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \qquad \text{Eq (3)}$$

$\boldsymbol{Y}$ is a column vector, representing a collection of outputs (y_1, y_2, … y_i), such as 100 RTs collected from an observer's button-pressing responses while viewing a feature search display, repeatedly for 100 times. $\boldsymbol{X}$ is i × j matrix, with j representing the number of explanatory variables, such as the number of items in a display, different spatial frequency contrast, etc. $\boldsymbol{\beta}$ is a j × 1 row vector, with j numbers of coefficient, accounting for the influence magnitudes of the explanatory variables. In a typical ANOVA-design experiment assuming a Gaussian distribution underlying across-participant variability, the $\boldsymbol{\beta}$ term

[13] Hierarchical models are sometimes called multilevel models. The term, 'multilevel' omits the information of a structural hierarchy amongst various levels of dependent variables. Take questionnaire data collected from pupils in a number of classes in several schools, across a wide range of geographical regions (4 hierarchies) as an example. The regions sit on the top hierarchy, followed by the schools, the classes and then the pupils. During the development of this technique, it was called random-effect, or in a broader sense mixed-effect, model (Kirk, 1995; Pinheiro & Bates, 2000). See an argument in page 245, Gelman and Hill (2006) why the term, random-effect/mixed-effect model, is confusing and should be avoided.

represents means for different level of an experimental factor. The last term is a column vector with i elements, derived from a stochastic distribution. In this canonical form, the stochastic component of a function output is solely determined by the residual term, ϵ, which is a collection of all possible sources of variability that affect random variable (i.e., $\boldsymbol{Y}$). When using the Gaussian assumption, the ϵ term is the Gaussian variance parameter (Figure 1-9).

The hierarchical framework relaxes the determinant $\boldsymbol{\beta}$ and the variability ϵ terms, allowing them to be modelled by different stochastic functions, subjected to the nature of data. For example, one can choose a probability function, such as ex-Gaussian, to fit RTs, because the empirical data indicate RT distributions skew towards the positive side, showing a long tail (Luce, 1986; Van Zandt, 2000). The three ex-Gaussian parameters can be mapped on different part of the distribution data, so they derives their own $\boldsymbol{\beta}$ matrices. The hierarchical concept comes in when, for instance, the three ex-Gaussian parameters are also random variables presumably sampled from other probability distributions. Again depending on the nature of data and how complex an analyst wishes, further hierarchies can then be built on top of the lower level functions. For example, if an analyst believes the rate parameter of the exponential component in an ex-Gaussian function selectively correlates with the spatial frequency of a visual display (which is often modelled by a Gaussian function), she/he can then considers the rate parameter as a random variable sampled from a Gaussian distribution. In summary, the hierarchical framework allows an analyst to use any appropriate probability functions and to build model hierarchically to fit the data. In fact, this is akin to the above-mentioned cognitive process models, with the

building blocks arranged closely to one's favourite cognitive processes (either one accumulator diffuses towards either a matched or a mismatched threshold or many accumulators race towards a common threshold).

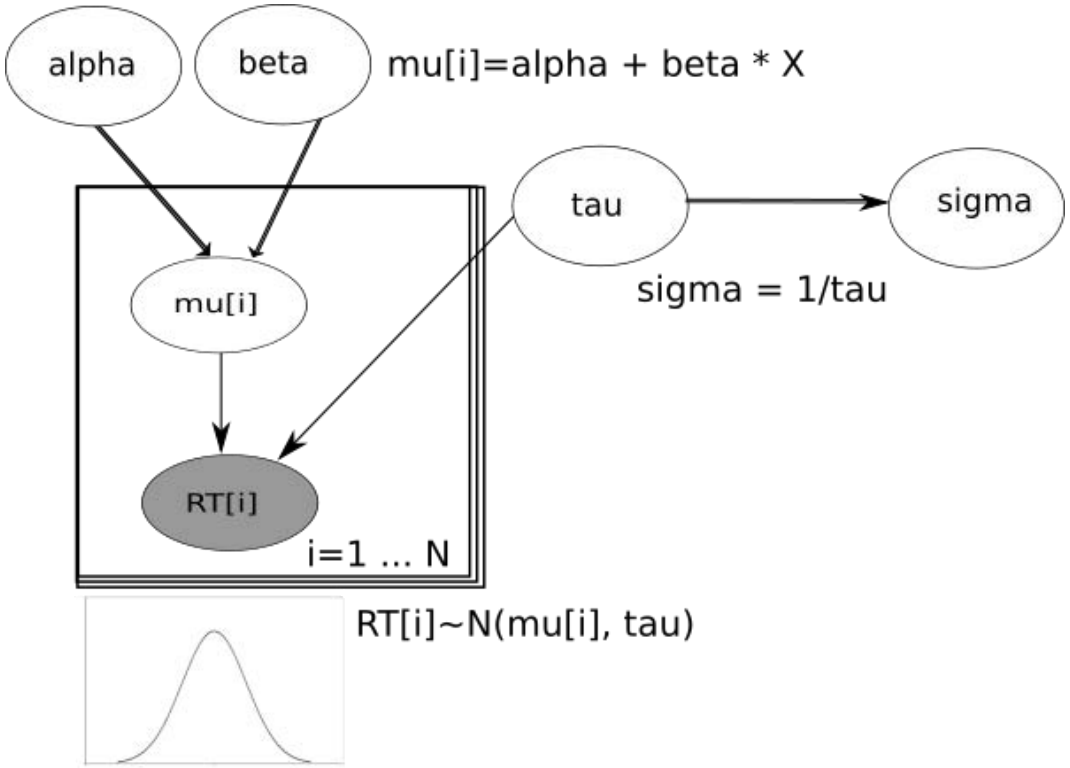


Figure 1-9. An illustration of one-level Gaussian model, using the BUGS-styled directed acyclic graph. The shaded oval represented the observed data. The double-line arrows indicate parameters deterministically associates with another parameter. As the equation on the right of the beta suggests, the mu's can be calculated analytically, using alpha, beta and the explanation variables. In contrast, the single-line arrows indicate parameters stochastically associate with another parameter. For example, RTs are sampled probabilistically from a Gaussian function with the mean, mu[i] and the precision (the inverse of the standard deviation) tau. The line-shaded rectangle indicates multiple observations (i= 1, …, N) are collected, so the model samples estimates iteratively from the probabilistic function.

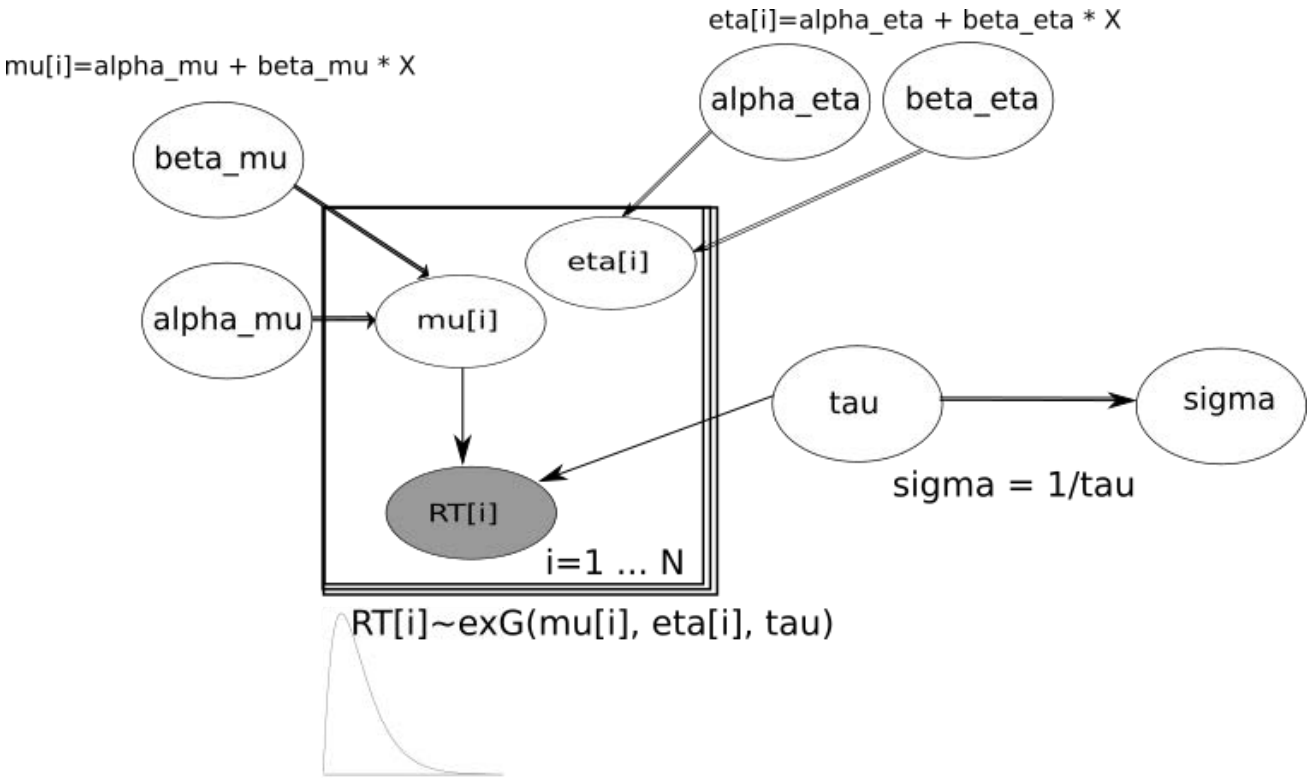


Figure 1-10. An illustration of one-level ex-Gaussian model. This figure illustrates an example to form a one-level ex-Gaussian model when one is willing to relax the Gaussian assumption and to replace it with the ex-Gaussian function.

1.6.2 Bayesian Inference

A natural side-effect of the hierarchical structure is an increase of free parameters. More free parameters to be estimated means more observations are needed. Typically, an analyst calculates a mean value from 25-50 observations per condition per participant. This calculation presumes the underpinning distributions both in the population of participants and in the population of one participant's responses are Gaussian distributions. Because a Gaussian function is fully described by two parameters, mean and variance, the typical practice uses four free parameters in a condition: the mean and variance describing the across- participant Gaussian function and the mean and variance describing the within-participant Gaussian function. A frequent practice to pursue high (statistical) power is to increase the number of participants, which raises the reliability for the across-participant Gaussian estimates; however, the within-participant Gaussian estimates are presumed as direct observations.

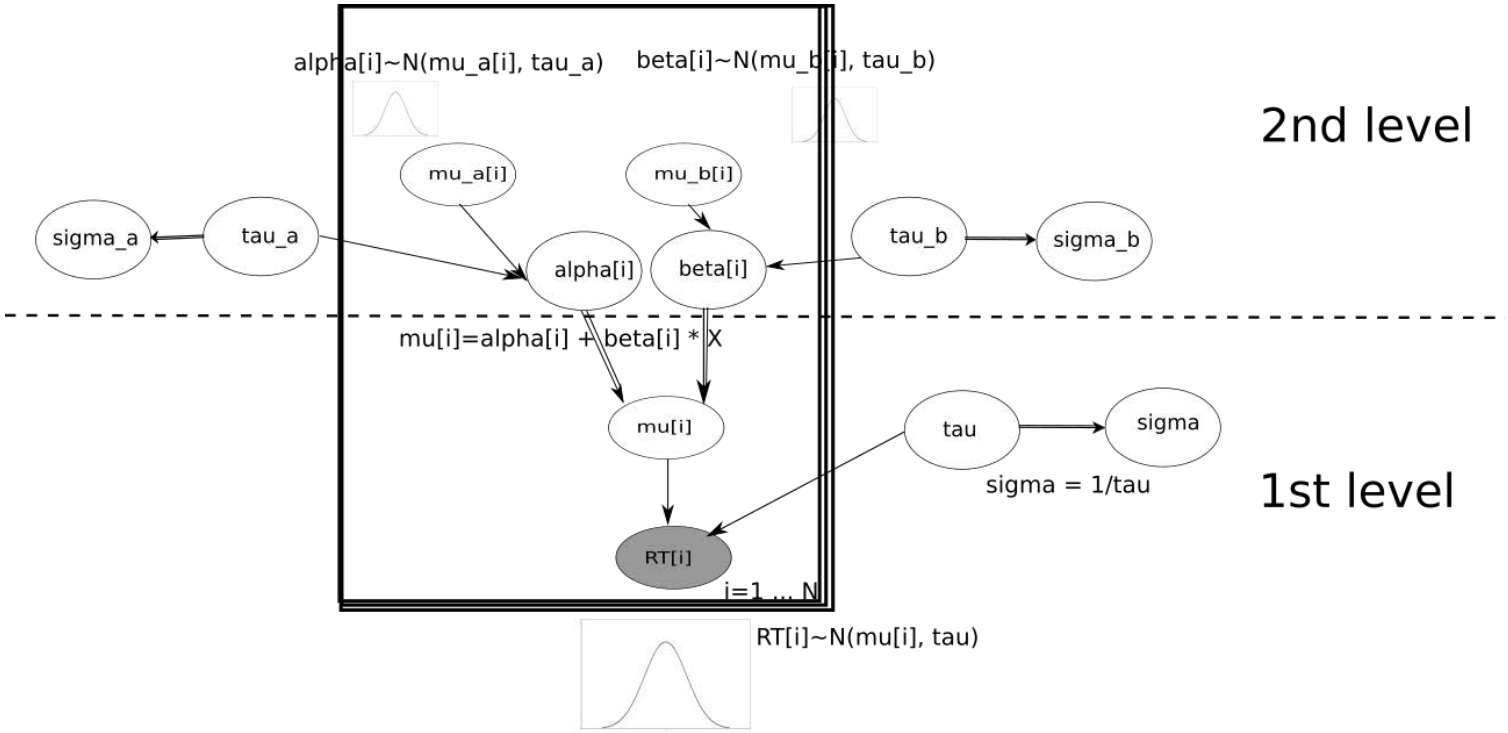


Figure 1-11. An illustration of two-level Gaussian model. This two-level model illustrates that using a hierarchical structure, one can also sample alpha and beta (the original deterministic parameters for the one-level Gaussian mean) from other probability functions. For the simplicity reason, this illustration uses Gaussian functions in both hierarchies. By using, for instance, a two-level structure, one can simultaneously model within-participant and across-participant variabilities on the basis of data. For example, one can use the ex-Gaussian function to fit the within-participant data and the Gaussian function to fit across-participant data.

The hierarchical framework permits an analyst to relax the direct observation assumption and account for the within-participant distribution. This suggests it is also crucial to consider the number of observations in each condition in a participant so as to estimate the parameters reliably.

Although the hierarchical model permits more thorough analyses at each level of data structure, the within-participant parameters now require also enough observations. To ease the burden of observation numbers, the Bayesian approach takes an entire different perspective on the parameter estimation, basically relying on the idea of the conditional probability.

$$P(\theta|Data) \propto P(\theta) \times P(Data|\theta) \qquad \text{Eq (4)}$$

The left-hand side of the equation is a posterior probability function which is determined by a set of parameters (θ; e.g., mean and variance in the case of Gaussian function) and is conditioned on the available data [P(θ|Data)]. According to Bayes rules, the posterior function can be derived from the right-hand side of the equation, which shows the multiplication of a prior probability function [P(θ)] and the likelihood function of data [P(Data|θ)]. Eq (4) states that one's belief (hence estimation) about the parameter set (θ) is conditioned on the observed data. Importantly, this belief can be updated iteratively by reusing a posterior function (in n^{th} step) as a new prior function in the $(n+1)^{th}$ step in light of new data. The iteration thus helps gradually improve the precision of parameter estimation on the basis of an educated guess for first prior function and renewing the prior belief with accumulated data. The advantage of iterative updating of the

posterior function only becomes clear with the help of modern computers which process a larger number of monotonous calculations with tremendous speed.

1.6.3 Hierarchical Bayesian Model

Although, the Bayesian inference suggests a promising solution to lighten the burden of the parameter estimation, only a handful of comparative studies pit the hierarchical Bayesian model (HBM; Rouder, Sun, Speckman, Lu, & Zhou, 2003) against other methods for estimating RT distributions (Farrell & Ludwig, 2008; Rouder et al., 2005), showing the advantage of HBM with small numbers (20 & 80) of simulated observations. It is unclear whether the results also apply to empirical data on visual search. Therefore, I conducted a simulation study to explore how far I can trim down the number of observations per condition (Chapter 2).

1.7 Thesis Outline

1.7.1 The Aim of the Thesis

The thesis aimed to investigate broadly the association of higher-order RT distributional cumulants with visual search findings, and specifically how these cumulants associate with search decisions underpinning the stages of *stimulus comparison*. This strategy was hinted earliest by Sternberg in his classic work of *additive-factor method* (1969) as well as being suggested by many early mathematical psychologists (e.g., Ashby & Townsend, 1980). They foresaw the pitfalls and suggested the potential values of examining higher-order cumulants of RT distributions. This has been pioneered by those mathematical psychologists, such as Townsend (1971), Link (1975), Ratcliff (1978), and Luce (1986). More recently, the estimation methods become more accessible,

because of the efforts made by many of the researchers (e.g., Brown & Heathcote, 2008; Grasman et al., 2009; Heathcote et al., 2004; Lacouture & Cousineau, 2008; Lee & Wagenmakers, 2013; Matzke, Love, et al., 2013; Rouder et al., 2005; Van Zandt, 2000; Voss & Voss, 2007; Wiecki, Sofer, & Frank, 2013).

The thesis starts from Wolfe's GS4 model and proposed one route to incorporate both correct RTs and more importantly, error RTs. There are potentially several approaches to accomplish this. Here I explored the dual-modelling approach with the descriptive probability model complemented with the decision-making models. The thesis, built on the stepping stones amassed by mathematical psychologists, made a small step forwards to explore the association between RT distributions and the decision-making process during search when a top-down goal is explicitly represented in VWM (Duncan & Humphreys, 1989).

1.7.2 Thesis Plan

The thesis begins with a simulation study testing whether the dual-modelling method I developed is able to estimate RT distributional parameters with acceptable precision when per-condition trial numbers are limited. Based on the results of the simulation studies, I then went on to reanalyse a benchmark visual search paradigm (Wolfe et al., 2010), using this new method. I then conducted a replication study, with the aims being to first confirm that the model is able to provide an appropriate description for RT distributions and second, to reveal an effect due to a specific display layout that was used. These were examined against a plausible, but simple, cognitive model, the EZ2 model

(Grasman et al., 2009), to understand how each distributional parameter might associate with the decision-making processes.

After validating the descriptive model of the RT distribution with the benchmark paradigms in the first study, I next conducted a second study with three experiments to examine the association amongst the VWM strength of attentional template, RT distributions and search decision. I designed a simple search paradigm with a within-block trial-by-trial updated template probe, comparing performance to that with a within-block trial-by-trial constant probe. The former manipulation was to re-strengthen the template representation in VWM, so it should exert a stronger influence from WM than the latter manipulation. The first experiment was contrasted to a second and a third experiments, which followed closely the first and introduced two inter-stimulus intervals (ISI; 50 vs. 400 ms) randomly distributed within a block. This rendered the appearance of a search display highly uncertainly temporally. The third experiment simply doubled the per-condition trial in the second experiment, aiming firstly to replicate the second experiment and secondly to test how the large trial number might affect the ISI and the cue factors. One important prediction in the second study is the manipulation of WM strength should selectively influence the decision threshold.

The third study investigated a different aspect of template representation in VWM. I designed an odd-one-out search paradigm in which a pre-search probe was used to fine-tune search performance. In this paradigm, a 3-level probe was set in order to elicit different template representations/WM operations, with a fixation cross (a null representation), a symbol (U/l for upper- or lower-case

letter; a part representation) signifying part of a target's features, and an exact replicate of the oncoming target image (a complete representation). The three levels of the template representation are assessed for their differences on the common average measures (mean RT and error rates) and on the descriptive and process model parameters. One main prediction in the third study is the manipulation of the template perceptual quality should selectively influence the decision rate, with the complete perceptual representation leading to a higher drift rate, comparing to other template representations.

Overall the work presented in this thesis shows the value of taking a detailed mathematical analysis of visual search functions and how this analysis can give new insights particularly into the role and nature of the templates that guide search.

Chapter 2 Estimation of the Minimal Sample Size Using the HBM and Its Correlation with the DDM

A recent surge of simulation studies is mainly driven by the advance in cognitive modelling, computational capability, and psychophysics searching for solutions to control noisy behavioural/neural responses. Typical biological observations consist of large extent of noise, and an ideal observation, similar to those highly controlled experiments in physics, rarely occurs. This problem has challenged researchers since the inception of psychophysics, which aims at studying body and mind in a sound scientific method (Fechner, 1860).

Simulation studies aim to mimic human behaviours in a highly controlled environment. When confined to a few specific behavioural responses, simulation studies are able to achieve the aim of controlling simulated human responses in a predictable way. This is because, rather than directly measuring from a human or animal observer, simulated responses and their noise signals are generated by prescribed cognitive mechanisms designed and fully controlled by researchers.

So far simulation studies are mainly used, but not limited, to (1) comparing different parameter estimation methods (e.g., Farrell & Ludwig, 2008; and as I did in this chapter) (2) fitting empirical data with different cognitive models (e.g., to test model adequacy) (3) remedying problems in traditional analyses (e.g., unbalanced sample sizes & missing data) by imputing simulation data back to empirical data (Schafer, 2010), and (4) more recently, applying the mixture of the

latter two methods to mitigate some obstacles of parameter estimations (Turner & Sederberg, 2014; Turner, Sederberg, Brown, & Steyvers, 2013).

Only a handful of simulation studies have shown the benefits of HBM over maximum likelihood estimation (MLE) method to estimate RT distributions (Farrell & Ludwig, 2008; Rouder et al., 2005, 2003). However, few, if there is any, investigate visual search data and how they may correlate with decision-making parameters. The chapter thus reported two simulation studies to address this issue.

2.1 The Association of Decision and Distributional Parameters

The second simulation study was to examine how the decision parameters estimated by EZ2 model correlate with the distributional parameters. An account (Rouder et al., 2005) presumed that the three distributional parameters–the shift, shape and scale–reflect respectively the minimal response times, the central processing time, and the speed of response execution. As one of the reviewers (see Chapter 4) indicated that this account is only preliminary and perhaps needed further investigation. Furthermore, because the three Weibull parameters jointly determine the spatial contour of a distribution, and the shape parameter aims to describe the distributional shape, rather than exactly manifests the 'shape' of a distribution, the RT distributional shape may change due to various factors, that the shape parameter may not reflect. The simulation study was to investigate this issue.

2.1.1 Method

This simulation conducted three case studies, investigating three

different scenarios. Each of them doubled one distributional parameter and fixed the others. Each case study simulated 200 RT observations, assuming that they were generated from 20 homogeneous observers. The study controlled distributional parameters and used the Weibull functions to generate simulated data. The data were then submitted to EZ2 model to estimate the condition-averaged drift rates, boundary separations and non-decision times. The estimations of the decision parameters were taken as observations to understand how they reacted to the change in each distributional parameter.

2.1.2 Result

The results indicated that firstly doubling the shift parameter resulted in a near two-fold increase in the non-decision time from 0.4 to 0.8 s. The increases in the drift rate (from 0.012 to. 0.013) and the decrease in the boundary separation (from 4.89 to 4.57) were minuscule. Second, doubling the scale parameter resulted in a decrease in the drift rate from 0.013 to 0.009, an increase in boundary separation from 4.70 to 6.57, and a negligible 10-ms increase in the non-decision time (407 vs. 417 ms). Finally, doubling the shape parameter resulted in an increase in the drift rate from 0.013 to 0.018 and a decrease in the boundary separation from 4.57 to 3.39. The increase in the non-decision time (410 vs. 507) is small, although its increase was larger than that of doubling the scale parameter. Figure 2-1 shows a comparison across the three case studies.

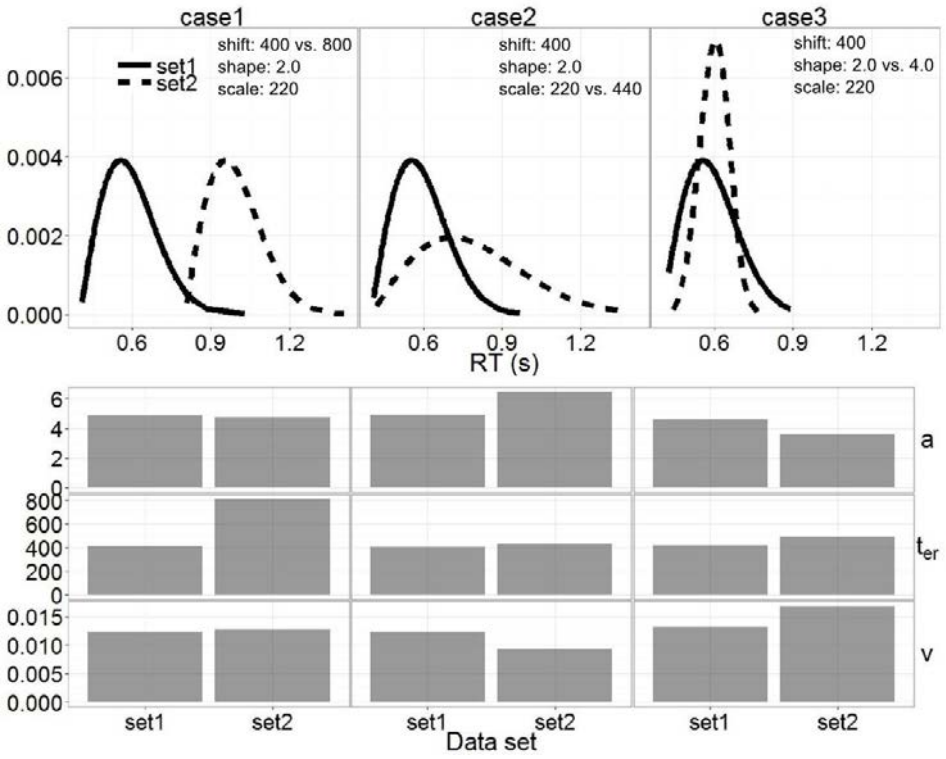


Figure 2-1. The figure shows the corresponding changes in the EZ2 parameter, when doubling the Weibull parameters. t_{er}, a and v stand for the non-decision time, the boundary separation and the drift rate, respectively.

2.2 The Minimal Sample Size: HBM vs. MLE

The second simulation study was to examine the estimation biases on three distributional parameters– mean, variance, and skewness–when various probability functions were fitted with different sample sizes per experimental condition and when the true distributions generating RTs were known. The study investigated minimal per-condition trials and aimed to answer whether a relatively small sample size (e.g., 100 trials) was sufficient to estimate distributional parameters reliably using the HBM.

2.2.1 Method

The simulation study used four R routines, *rnorm*, *rexGAUS*, *rinvgauss*,

and *rweibull3* to generate the simulated data[14]. The study examined four scenarios, respectively assuming that the true distribution followed a normal (*rnorm*), an ex-Gaussian (*rexGAUS*), a Wald (*rinvgauss*) or a Weibull function (*rweibull3*). The true distributions adopted the parameter values listed in Table 3 in Cousineau, Brown and Heathcote's report (2004) to generate simulated data. The true distributions generated twenty homogeneous participants, each contributing RT observations in 10 different sample sizes ranging from 20 to 470 with a step size of 50. The data were then submitted separately to the HBM and the MLE, estimating the three distributional parameters: shift, scale, and shape. Because HBM and MLE are parametric methods, both assumed the data were random variables generated by a Weibull probability function when estimating the parameters. The parameters – shift, scale, and shape – were then analytically converted to the mean, variance and skewness to evaluate the performance of the estimation methods. The estimates were then compared with the true values (Table 3 in Cousineau et al., 2004). The mean of the differences and the standard error of the differences were termed *bias* and *precision*, respectively. The two statistics were summarised in the following 3 figures.

[14] The R routines, *rexGAUS*, *rinvgauss* and *rweibull3*, are not included in the standard packages. To implement them, the user needs to install additional packages: *gamlss*, *statmod*, and *FAdist*.

2.2.2 Result

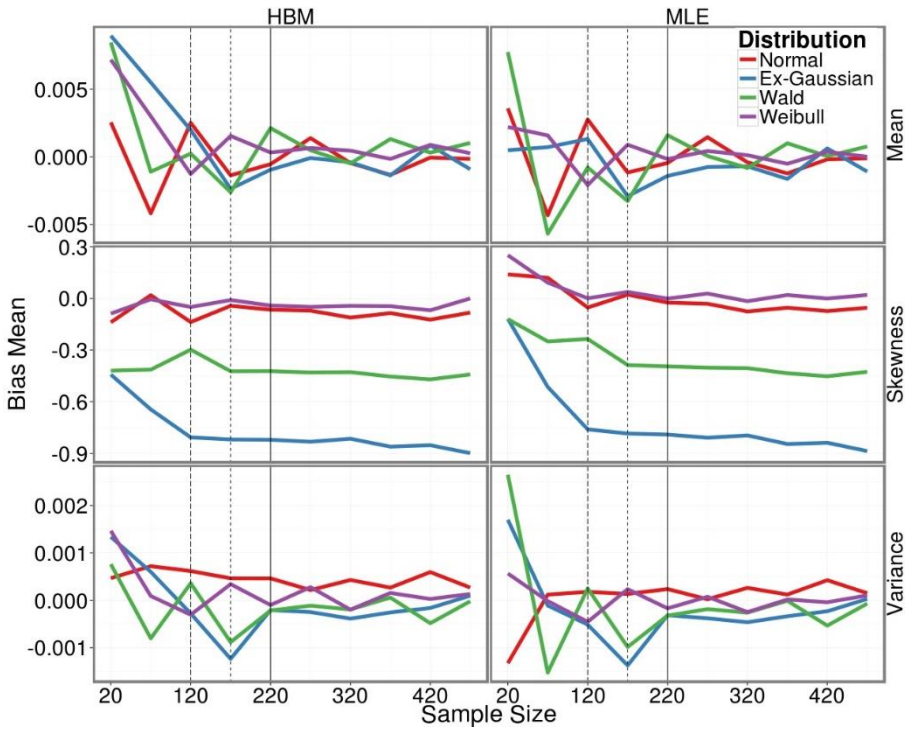


Figure 2-2. This figure compares the average (across participants) of mean, variance and skewness to the true values that generated the simulated data. The mean and variance are on the units of seconds and square of seconds, respectively. The skewness was calculated by the equation, [$m_3 = (RT - RT_{mean})^3/N$] (Crawley, 2002). The reference dashed lines are drawn at the sample sizes, 120, 170, and 220.

Figure 2-2 shows the *bias* for the estimation of means, variances and skewness for the four distributions. Overall, no differences were observed between the two methods when estimating the means. The only factor improving the estimation was the sample size, $F(9, 1520) = 4.90$, $p = 1.69 \times 10^{-6}$. As the figure showed, the more observations were in a condition, the less the estimation error (i.e., the lines become closer to the 0). The bias dropped rapidly when the sample size surpassed 100, from 17 ms at 20 observations to 7 ms at 120 observations, and it decreased at a slower rate when the sample size was well over 120 observations about 5 ms around the 0-difference line. The specification of the true distribution did not alter the bias when the sample

size exceeded 120, even though both estimation methods assumed an underlying Weibull function to account for the data. Similar pattern was observed in the precision (Figure 2-3).

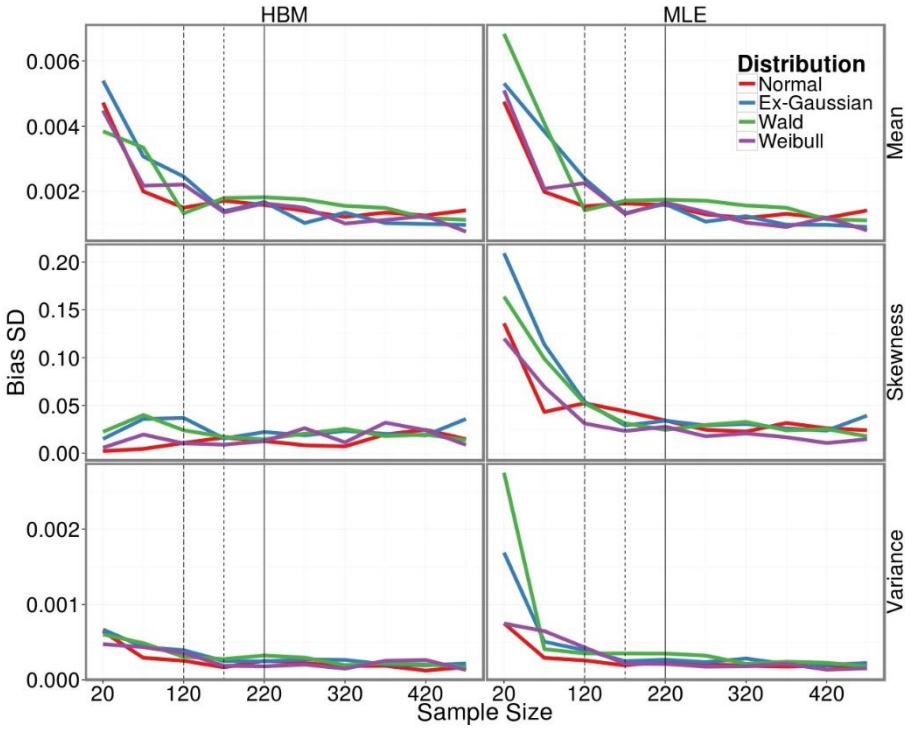


Figure 2-3. The figure showed the average standard errors across participants for the mean, skewness and variance.

In contrast to estimating the mean, the HBM demonstrated an advantage over the MLE when recovering the variances and skewness. Except for the skewness bias score, large sample sizes improved the estimation. The advantage of the HBM showed especially at the small sample sizes, although the disadvantage of MLE resolved gradually when the sample sizes exceeded 120 (see the first dotted line in Figure 2-2 & Figure 2-3). The misspecification of the underlying distribution resulted in different estimations for variance. This showed only at the estimation error (bias) for the variance. Both methods needed a sample size larger than 170 to resolve this problem. That is, when

the sample sizes were 220, 270, 320, 370, 420 and 470, no difference at estimating variances was observed between the two methods. The parameter recovery was better when the true distribution followed the Weibull function. The HBM showed a specific advantage at the precision for both estimating variance and skewness at the small sample size (Figure 2-3 & Figure 2-4).

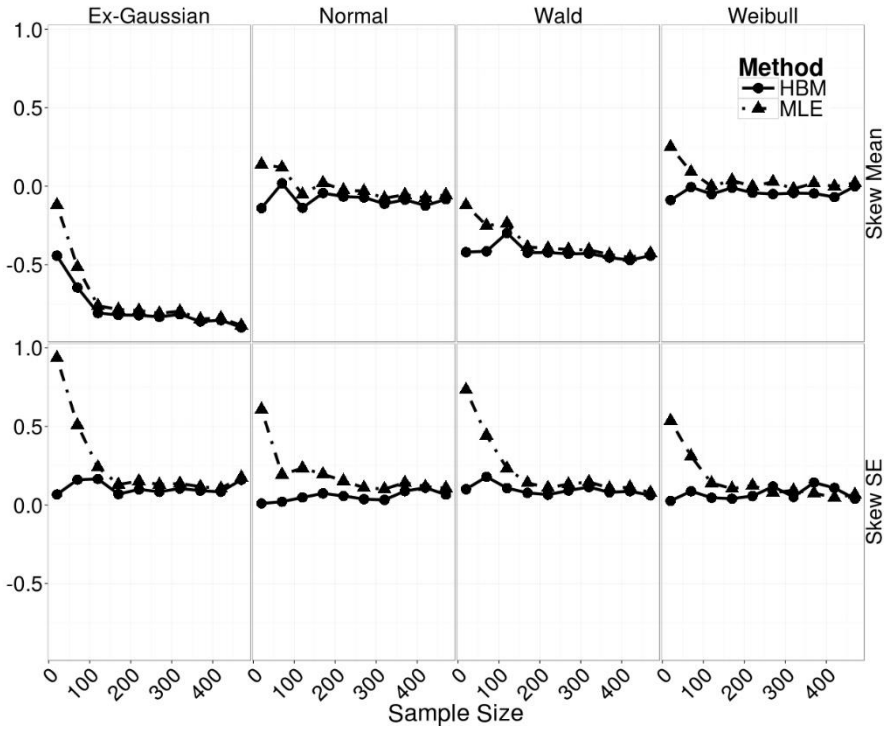


Figure 2-4. The estimation of skewness. The figure shows the difference between the HBM and MLE along different sample sizes.

2.2.3 Summary

By and large, the results suggested that (1) there was no difference between HBM and MLE when the sample size was larger than 120, (2) the HBM performed better than the MLE both at estimating variance and skewness when the sample size was small, (3) the HBM advantage was showed especially at the precision and (4) the specification of equally plausible probability functions was crucial only when it matched the true distribution that

generated the RT data.

2.3 Why Use the Weibull Function?

The Weibull function is one of the many plausible probability functions that can accommodate positively skewed RT distributions. It was chosen as a main vehicle to describe RT distributions because its parametric characteristics enable an intuitive understanding for the shape of RT distributions. Nevertheless, there are other alternatives. These include, but not limited to, gamma, log-normal, and Wald functions. With appropriate parameterisation, all are capable of accommodating skewed RT distributions with the same descriptive parameters. This section justified the reasons to use the Weibull function to fit RT data.

First, the Weibull function summarises concisely the shape of RT distributions. As illustrated in Section 2.1, changes in the distributional parameters – shift, scale and shape – associate with increases/decreases in different parts of RT distribution, and thereby affect the accumulation of RT densities. These corresponding changes inform how an experimental factor may alter different parts of an RT distribution. Selective influences in different distributional parameters thus suggest distinct changes in cognitive processes. Secondly, although the three-parameter gamma function fitted RTs better the Weibull function when both functions were estimated by the MLE method (E. M. Palmer et al., 2011), the gamma function did not converge when fitted with HBM. The gamma function showed signs of non-convergence and perhaps because of this, it fitted the data slightly worse than other plausible functions, supported by

the deviance information criteria (DIC)[15]. Third, the other plausible functions – Wald and log normal – fitted similarly with the Weibull function to both ours and the benchmark paradigm (Wolfe et al., 2010), with only slightly better DICs than the Weibull function. Because amongst the four 3-parameter functions, the Weibull function has been tested in literature extensively to fit visual search data (e.g., Cousineau et al., 2004). Further all four functions fit similarly to the data when examined separately for the tasks, display sizes, target types, and data sets in the benchmark paradigms and my replication (Chapter 4). As a result, the thesis chose the Weibull function as the main vehicle to describe RT distributions. To test whether the four plausible functions fit the positively skewed RT distributions adequately, their DICs are compared with the DIC of a Gaussian function fit (~ -3150). The DIC is far worse than the four plausible functions. See for Appendix C to see fitting the gamma function in HBM.

[15] DIC is a goodness-of-fit index, used in Bayesian parameter estimations.

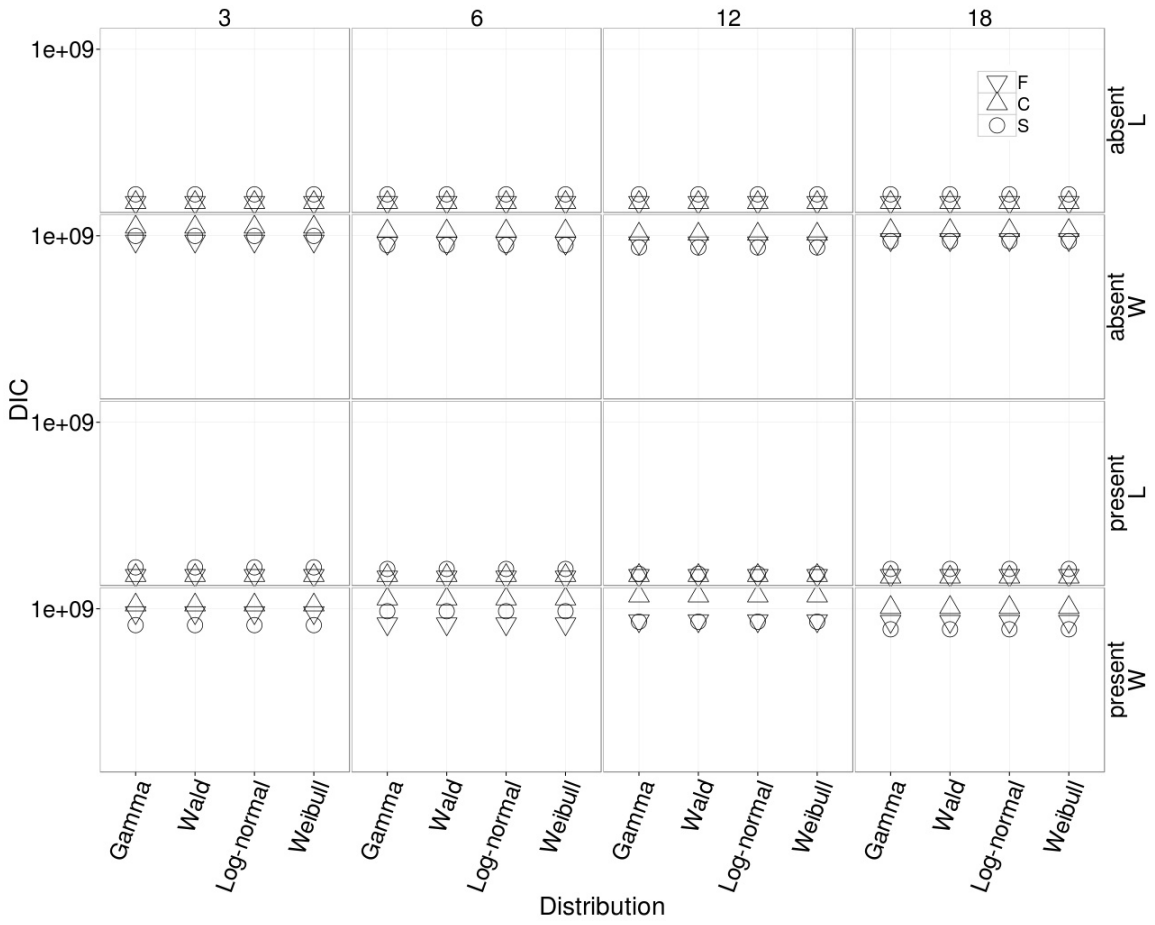


Figure 2-5. The figure compares the Bayesian DICs for the fitted 3-parameter probability functions across the data sets, search tasks, target types, and display sizes. In general, the smaller the DICs, the better the fit. L and W stands for my and Wolfe et al.'s (2010) data sets. F, C, and S stand for feature, conjunction, and spatial configuration searches. Y axis is on log scale.

2.4 Discussion

The results from the two simulation and function fit studies suggest (1) the HBM estimates distributional parameters with good precision when per-condition sample size is around 120, (2) the non-decision time associates with the Weibull shift parameter in a predictable way, but it is not clear how the scale and shape parameters may associate with the decision-making process and (3) the Weibull function is able to describe distributional shape appropriately in the HBM. Informed by the simulation studies, the psychophysical studies in the following chapters collected around 100 observations per condition (per participant), estimated RT distributions via the HBM Weibull function, and interpreted the influences of key experimental manipulations on the shapes of RT distributions, using the decision-making models.

Chapter 3 General Method

This chapter summarised the common methods used in the following three chapters. The differences specific to each study were documented separately in the Method sub-section in the respective chapter.

3.1 Participants

Table 3-1 presented the participant demographic data. Study 1 (Chapter 4) excluded one participant from the analysis because of chance-level responses. Study 2 (Chapter 5) conducted three experiments. The second experiment in Study 2 excluded two participants because one participant performed inadvertently the varied cue condition twice and another dropped out without attending the second visit. Thus, the second experiment in Study 2 analysed 19 valid participants. The third experiment recruited 26 volunteers. Six of them dropped out before the second visit and one completed, relative to other participants, the experiment with less per-condition trials (70). Thus, the third experiment analysed also 19 valid participants. All volunteers took part in exchange of course credits or cash, reported normal or corrected-to-normal vision and signed a consent form before carrying out in the study. The procedure was reviewed and granted permission to proceed by the Ethics Review Committee at the University of Birmingham.

Table 3-1. Participant demographic data. Exp: experiment; P: participant; S: study; F/M: female to male ratio; L/R: left hander to right hander ratio.

		P number	Age (years; Mean ± SE)	F/M	L/R
S1		40	18-22 (18.9 ± 1.01)	33/7	5/35
S2	Exp 1	20	18-31 (23.27 ± 3.85)	18/2	5/15
	Exp 2	22	18-32 (20.18 ± 3.33)	13/9	5/17
	Exp 3	26	18-27 (20.42 ± 2.50)	14/12	0/26
S3		10	19-20 (19.4 ± 0.52)	8/2	0/10

3.1.1 Summary

A clear association was observed between the non-decision time and the shift parameter when the other distributional parameters were kept constant. As for the other two distributional parameters, the simulation study do not indicate a clear distinction for the scale and shape parameters with the decision-making process. Specifically, an Increase in the scale parameter raised the boundary separation and slowed down the drift rate, suggesting a negative correlation between the scale parameter and RTs. In contrast to the scale parameter, an increase in the shape parameter suggests it correlates positively with RTs, because of the decrease in the boundary separation and the increase in the drift rate, although the increase in the shape parameter appears also to increase the non-decision time.

Overall, the current evidence supports a clear link between the shift and the non-decision time but seems to suggest that an overarching influences of the scale and shape parameters on the decision-making process. Even though the detailed examination of the latter two parameters indicates that an increase in the

scale parameter affects more the decision components and less the non-decision component and that an increase in the shape parameter affects both, it is far from clear-cut to assert the scale or the shape parameter reflects a distinctive cognitive process.

3.2 Apparatus and Stimuli

In Study 1 and Study 2, the order and timing of the paradigms were controlled by PsyToolkit (Stoet, 2010), which is a GNU C library designed to implement cognitive paradigms. The paradigms were carried out on a Linux PC, using a kernel specifically tweaked to a hard real-time system (Linux kernel 2.6.31-11-rt). A hard real time system will treat a designated computer programme as system critical, thereby responding immediately when the programme is called. I designated cognitive paradigms as critical programmes when they were running. In contrast to general operating systems, such as Windows 7 and Windows 8, the user is not allowed to tweak their kernel. This type of operating systems prioritises system-related programmes, but not the programmes running cognitive paradigms. In some situations, the programming running cognitive paradigms may be interrupted by, for example, system updates. This is not good news, because the programmes executing system updates mostly own a higher priority than those running cognitive paradigms. Further the system-related programmes are controlled by IT staffs (maybe remotely by the staffs in software companies), suggesting that the time the programmes interrupt experiments is not completely random.

The graphic card, NVidia GeForce 8500 GT, controlled visual displays in Study 1 and Study 2. Study 3 used E-Prime 2.0 on a Windows 7 personal

computer, equipped with an NVidia GeForce GT430 graphic card to control the timing of stimulus presentation. All participant responses were made using, a Cedrus RB-830 response pad. All experiments presented stimuli on a Sony CPD-G420 CRT monitor set at the resolution of 1152 × 864 pixels and 100 Hz refresh rate, except the second experiment in Study 2, which presented stimuli on a Dell P991 CRT monitor at the 1024 × 768 resolution and 85-Hz refresh rate. Participants sat about 60 cm in front of the monitor in a well-lit cubicle and were asked to make speeded responses without compromising their accuracy.

In Study 1 and Study 2, the visual stimuli were presented on a 2.526° invisible circle in black and grey (white in Study 2) colours onto a grey background (RGB, 190, 190, 190). The visible area contained the entire screen, but the relevant stimuli were all drawn within the viewing area of 7.59° × 7.59°. Study 1 used the visual stimuli similar to the benchmark paradigm (Wolfe et al., 2010). Study 2 used 13 English uppercase letters, A, B, D, E, F, G, H, J, M, N, Q, R, & T, sized 0.63° × 0.63°. In Study 3, the search items (A, B, D, E, F, G, H, M, N, R, & T) were scaled to 0.32° × 0.46° in black colour, presented on E-prime's default grey colour background. The search items were randomly allocated to 10 possible locations on an invisible circle (see Figure 4-1 for an illustration). While viewing an imperative stimulus, participants indicated whether the target was present or absent (Study 1) or whether the target was on the left or the right side of the invisible circle (Study 2 and Study 3).

3.3 Design

To minimise one of the experimenter biases related to the analysis of null hypothesis significance testing (NHST; Kruschke, 2010), the studies set a fixed

target sample size (20 participants in Study 1 and Study 2; 10 participants in Study 3) before collecting data. The target sample size was determined based on commonly used sample sizes (approximately 5 to 20 participants) in visual search literature. The data from participants who withdrew and completed only part of the tasks were not analysed; these participants were replaced with other individuals.

3.4 Decision-making Models

Three decision-making models will be applied in the thesis in separate chapters. Chapter 4 applied the simplified decision-making model, EZ2 to estimate the decision parameters in the benchmark search parameter. The EZ2 model differs slightly from the other two decision-making models, applying in Chapter 5 (DDM and LBA) and Chapter 6 (the fast-dm version of DDM). EZ2, because of the mathematical simplification, estimated the drift rate, non-decision time and boundary separation. Specifically, the boundary separation merges the decision threshold and the initial bias (as estimated separately in DDM and LBA) as one parameter. Study 2 (Chapter 5) fited data with LBA and DDM and used the model selection procedure described in Donkin, Brown and Heathcote (2011) to balance the good fits and parsimonious factors to fit model. Study 3 (Chapter 6) allowed all experimental factors to depend on most DDM parameters, thereby fitting data with a saturation model.

Chapter 4 Modelling Visual Search Using Three-parameter Probability Functions in a Hierarchical Bayesian Framework

4.1 Introduction

Distributional analyses are becoming an increasingly popular method of analysing performance in cognitive tasks (e.g., Balota & Yap, 2011; Heathcote et al., 1991; Hockley & Corballis, 1982; Ratcliff & Murdock, 1976; Sui & Humphreys, 2013; Tse & Altarriba, 2012)[16]. When compared with analyses based on mean performance, distributional analyses potentially allow a more detailed assessment of the underlying processes that lead to a final decision. In particular it has long been noted that RT data, before being averaged across multiple participants, frequently show a positively skewed, unimodal distribution (Luce, 1986; Van Zandt, 2000). Distributional analyses begin to allow us to decompose such skewed data and to address the processes that contribute to different parts of the RT function. One approach to this is through HBM, a method that blends Bayesian statistics and hierarchical modelling. The latter uses separate regressors to assess variations across trial RTs collected from a participant by estimating regression coefficients, contrary to conventional single-level ANOVA models which directly use RT means as dependent variables. The hierarchical modelling then carries on assessing the coefficient variations across participants at the second level, accounting for individual differences. One direct advantage of the hierarchical method is that variation across trials can be described by a positively skewed distribution (or other

[16] This chapter has been accepted for publication by the journal, Attention, Perception, & Psychophysics.

distributions, as analysts wish), in contrast to the Gaussian distribution implicitly adopted by a single-level ANOVA model (which works directly on the second level of the hierarchical method). The flexibility to choose an underlying distribution liberates analysts from using statistics derived from the Gaussian distribution to represent each participant's performance in an experimental condition, since a Gaussian assumption may not be appropriate given positively skewed RT distributions.

Hierarchical modelling typically relies on point estimation, which itself depends on the critical assumption of independence of random sampling – making performance highly sensitive to the sample size. Hierarchical modelling may perform less than optimally when, relative to the number of estimated parameters, trial numbers are too few to account for the parameter uncertainties at each hierarchical level (Gelman & Hill, 2006). This is possible when a non-Gaussian distribution is used to estimate parameters for each participant separately in a hierarchical manner. For example, a data set with ten participants, when using an ex-Gaussian distribution (fully described by three parameters), estimates simultaneously at least 30 (3 × 10) parameters, each of which should be derived from a distribution with an appropriate uncertainty description (i.e., parameters for variability). This is assuming that only one experimental condition is tested. It follows that small trial numbers within an experimental condition may result in biased uncertainty estimates, which render the effort of adapting hierarchical modelling in vain[17]. Bayesian

[17] The requirement for a reasonable large sample size relates to the increase of estimated parameter. This also applies to other methods, when the number of estimated parameter

statistics is one of the solutions to the problem of point estimation inherent in the conventional approach. Building on the nature of the hierarchical structure of parameter estimations, Bayesian statistics conceptualize each parameter at one level as an estimate from a prior distribution. Based on Bayes' theorem, the outputs of prior distributions can then be used to calculate posterior distributions, which are conceptualized as the underlying functions for the parameters in the next level. By virtue of Monte Carlo methods, HBM is able to estimate appropriately the uncertainty at each level of the hierarchy, even when trial numbers are limited (Farrell & Ludwig, 2008; Rouder et al., 2005; Shiffrin, Lee, Kim, & Wagenmakers, 2008). Note that Bayesian statistics here are used to link variations in the trial RTs within an observer with the variations at aggregated RTs between observers. This differs from applying Bayesian statistics to account for how an observer identifies a search target by conceptualizing that her prior experiences (e.g., search history; modelled the RTs in N^{th}-1 trial as prior distributions) influence the current search performance (modelled the RTs in N^{th} trial as posterior distributions).

HBM has been used previously in cognitive psychology to examine, for example, the symbolic distance effect – reflecting the influence of analogue distance on number processing (Rouder et al., 2005; other examples see, Matzke & Wagenmakers, 2009; Rouder, Lu, Morey, Sun, & Speckman, 2008;). In symbolic distance studies observers may be asked to decide if a randomly chosen number is greater or less than 5. Observers tend to respond more

increases. Hierarchical model helps to trim down sample sizes by constraining outliers, when estimating the same number of parameters.

slowly when the number is close to the boundary (5), compared to when the number is far from it. One interpretation based on mean RTs is that an additional process of mental rechecking is required when numbers are close to 5. The result from HBM however suggests a further refinement for this interpretation by showing that the locus of effect resides in the scale (rate) of RT distributions. A scale effect, interpreted together with other symbolic distance findings using a diffusion process or a random walk, implies broadly a change in drift rates or decision boundary, as opposed to a change in (cognitively) functional architecture, such as mental rechecking process (Rouder et al., 2005).

4.1.1 Application to Visual Search

The present study applied HBM and distributional analyses to account for the RT distributions generated as participants carried out visual search. To do this, I compared participants' performances under 3 search conditions varying in their task demands: a feature search task, a conjunction search task, and a spatial configuration search task. A typical visual search paradigm requires an observer to look for a specific target. The 'template' (Duncan & Humphreys, 1989) set-up for the target can act to guide attention to stimuli whose features match those of the expected target. Depending on the relations between the target and the distractors, and also the relations between the distractors themselves (Duncan & Humphreys, 1989), performance is affected by several key factors, including the presence or absence of the target, and the similarity between the target and the distractor and the similarity between distractors (for a computational implementation of these effects based on

stimulus grouping see, Heinke & Backhaus, 2011; Heinke & Humphreys, 2003).

The display size effect relates to how performance is affected by the number of distractors in the display. Effects of display size are frequently observed in tasks where target-distractor similarity is high and distractor-distractor similarity low (conjunction search being a prototypical example; Duncan & Humphreys, 1989). In addition, the display size × RTs function shows a slope ratio of absent trials to present trials slightly greater than 2, which varies systematically with the types of search task, from efficient to inefficient (Wolfe, 1998b).

To date these effects have mostly been studied by examining mean RTs across trials, with the variability across trials considered as uncorrelated random noise (though see, for example, Ward & McClelland, 1989, who used across-participant variation to examine how search might be terminated). The assumption of across trial random noise unavoidably sacrifices the information carried by response distributions, which may help to clarify underlying mechanisms (e.g., the influence of top-down processing on search). In contrast to this, hierarchical distributional analyses set out to use the variability at each possible level of analyses as well as the mean tendency across responses, and through this, they relax the assumption of an identical, independent Gaussian distribution underlying trial RTs. This then permits trial RTs to be accounted for by a positively skewed function. The reasons I adopt HBM (Rouder & Lu, 2005; Rouder et al., 2005, 2003) in the present study are: (1) it harnesses the strength of Bayesian statistics which take into account the evolution of the entire response distributions from trial RTs in one participant to aggregated RTs

across all participants, (2) it uses the dependencies between each level of response as crucial information for identifying possible differences between the experimental manipulations, (3) it takes into account the differences between individual performances, and (4) it allows the higher-level parameters to constraint the lower-level parameters, thereby preventing potential outliers from over-influencing the estimates. Notably, the response variability across different trials is no longer assumed to constitute random noise but rather is treated as crucial information that must be modelled.

My study examined the effectiveness of distributional analyses and the HBM approach for understanding performance in 3 benchmark visual search tasks, which were modified from Wolfe, Palmer and Horowitz's paradigm (2010; a different set of analyses was reported also in Palmer et al, 2011; also see a computational model aiming at clarifying the mechanism of search termination in Moran, Zehetleitner, Müller, & Usher, 2013). In their paradigm, an observer searched for an identical target throughout one task - either a red vertical bar in the feature and conjunction tasks or a white digital number 2 in the spatial configuration task. The distractors, either a group of homogeneous green vertical bars or a mixture of green vertical and red horizontal bars, set the feature and configuration tasks apart. In the feature task, the homogeneous distractors enabled the target's colour to act as the guiding attribute (Wolfe & Horowitz, 2008) making search efficient. In the conjunction task, and possibly also in the spatial configuration task, a further stage of processing might be required in order to find the target amongst the distractors as no simple feature then suffices. All search items were randomly presented on an invisible 5 by 5

grid. One of the crucial contributions derived from previous work using RT distributions is that observers set a threshold of search termination depending not only on prior knowledge but also on the outcome of prior search trials (see Lamy & Kristjánsson, 2013, for a review). As a consequence, instead of always exhaustively searching every item in a display, an observer may adapt the termination threshold dynamically (Chun & Wolfe, 1996). A second contribution has been to show that variations in the display size can have relatively little impact on the shape of the RT distribution (Palmer et al., 2011; Wolfe et al., 2010) and effects on the shape of the distribution only emerge at the large display sizes (i.e., 18 items) when the task difficulty is high (i.e., on target absent trials in the spatial configuration task; Palmer et al., 2011; though see Rouder, Yue, Speckman, Pratte, & Province, 2010, for a contrasting result).

4.1.2 The 3-parameter Probability Functions

My study adopted four three-parameter probability – lognormal, Wald, Weibull and gamma[18] – functions (Johnson, Kotz, & Balakrishnan, 1994) to estimate RT distributions using the HBM. Differing from the frequently used ex-Gaussian function, the 3-parameter probability functions describe an RT distribution with the parameters, shift, scale and shape that characterise the pattern of a distribution. An increase of scale parameters reduces the height of a distribution, thereby lengthening its tail. This implies that some responses originally accumulated around the central part become slower, thereby being

[18] The functions describe a distribution with the same set of parameters, shape, scale and shift. Because comparing to other functions the previous analysis (Palmer et al., 2011) reported a worse χ^2 fit of Weibull function, I constructed the comparable 3-parameter HBM to test if other functions gain a substantial better fit using hierarchical Bayesian approach than the Weibull function. I thank Evan Palmer for this suggestion.

moved to the tail side. An increase in the shape parameter, on the other hand, elevate the height of a distribution, implying some original very slow and fast responses become moderate. Hence the increase of the shape parameter not only changes the kurtosis, skewness, and variation, but also likely moves the measures of the central location. An increase in the shift parameter preserves the general pattern of a distribution. That is, an identical curve is moved rightwards (see Figure 2-1 or an illustration).

The study assumed that changes in RT distributions reflect unobservable cognitive processes (a similar argument also made by Heathcote et al., 1991). The visual search processes that may change RT distributions include, but not exclusively, the clustering process of homogeneous distractors, the matching process of a search template with a target and distractors, and the process of response selection (see Duncan & Humphreys, 1989; Heinke & Humphreys, 2003; Heinke & Backhaus, 2011; Palmer, 1995). Some previous work (e.g., Rouder et al., 2005) suggests interpreting Weibull-based analyses as reflecting psychologically meaningful processes. For example, the shift, scale and shape parameters of an RT distribution have been suggested to link respectively with the irreducible minimum response latency (Dzhafarov, 1992), the speed of processing, and high-level cognition (e.g., decision making). This is similar to some reports applying distributional analyses on RT data, attempting to link distributional parameters with psychological processes directly (e.g., Gu, Gau, Tzang, & Hsu, 2013; Rohrer & Wixted, 1994). Although it is ambitious to posit links between distribution parameters and underlying psychological processes, a better strategy is to take advantage of the descriptive nature of distributional

parameters (Schwarz, 2001), which permit a concise summary of how a distribution varies in response to a particular experimental manipulation. The distributional parameters describe how an RT distribution changes in three different separable aspects (shift, scale & shape). This enables researchers to examine RT data as an entirety, building on top of what can be provided by an analysis of mean RTs. However, one potential pitfall is how the distributional parameters can be understood with regard to unobservable psychological mechanisms (e.g., the visual search processes I investigated here). I explored a possible avenue to resolve this issue by applying a plausible computational model to understand the same set of RT data (a similar strategy was reported recently in Matzke, Dolan, et al., 2013; and suggested also in Rouder et al., 2005).

To understand how my distribution-based HBM correlates with underlying cognitive processes, I compared the HBM parameters with those estimated from EZ2 diffusion model (Grasman et al., 2009; Wagenmakers et al., 2007; Wagenmakers, van der Maas, et al., 2008) which is a closed-form and simplified variant of Ratcliff's diffusion model (1978). The diffusion model conceptualizes decision-making in a 2AFC task as a process of sensory evidence accumulation. The accumulation process is described through an analogy in which a particle oscillates randomly on a decision plane where the x axis represents the lapse of time and the y axis represents the amount of sensory evidence. When the amount of the evidence surpasses either the positive or negative decision boundaries of the y axis, a decision is reached and the time the process takes is the decision RT. The merits of the diffusion model

are that it directly estimates three main cognitively-interpretable processes – the drift rate, the boundary separation, and the non-decision component – three parameters that turn the random oscillation into a noisy deterministic process. The drift rate is associated with the speed to reach a decision threshold (Ratcliff & McKoon, 2007), which is determined by the correspondence between the stimuli (search items) and the memory set (search template). In the case of template-based visual search, the drift rate correlates with the matching of the template to the search items; thus, it is conceivable that the shape of an RT distribution will correlate with the drift rate, if the processing of template matching influences an RT shape. The boundary separation, on the other hand, may reflect how conservative a participant is. Liberal observers may reach a conclusion earlier than conservative observers on the basis of the same amount of evidence if their decision criterion is set lower. The non-decision component is a residual time, calculated by subtracting the decision time (estimated by the diffusion model) from the total (recorded) RT; this may reflects the time to encode stimuli (perceptual times) together with the time to produce a response output (motoric times) (Ratcliff & McKoon, 2007).

The diffusion model has been used on various 2AFC paradigms and so far both psychophysics and neurophysiological studies indicate its usefulness to probe the two latent decision-making processes and the decision-unrelated times (e.g., Cavanagh et al., 2011; Towal, Mormann, & Koch, 2013). The EZ2 model is one of the simplification types (Grasman, Wagenmakers, & van der Maas, 2009; though see a review for more complicated statistical decision models of visual search in Smith & Sewell, 2013), which provides a coarse and

efficient estimation for the two important aspects of search decision: decision rate and decision criterion. By dissecting the joint data of RT and accuracy into the part that is influenced by decision-related processes and that is influenced by the non-decision-related process, the EZ2 model is able to account for the changes in RT distributions in a psychologically meaningful way. For instance, the factor that affects the non-decision process should reflect on the shift parameter that hardly changes the general pattern of an RT distribution, because its effect would be on all ranges of a distribution. If most responses in a distribution are delayed equally, the shift parameter will also increase selectively. On the other hand, the factor that delays the decision-related processes may consistently delay only the responses from the quick to the central band of a RT distribution, so it will result in an increase of the scale parameter. That is, as Figure 2-1*Figure 2*- showed, a scale increase reduces the height of a distribution. Alternatively, if a decision-related factor delays the quick to central band of a RT distribution, but speeds up the very slow band of responses, it will result in a shape increase.

The diffusion model was used to complement the distributional analysis. The three model parameters – the evidence accumulation, the boundary separation, and the non-decision process – are operated at the stage of stimulus comparison in a search trial. I used the EZ2 model to estimate the means across trials of the diffusion parameters in each condition. The Weibull HBM on the other hand summarises the shapes of RT distributions in each condition. The RT distributions thus are the aggregated outputs from the diffusion processes. Therefore, the dual-modelling approach, at one end,

assumes one search response is driven by the diffusion process, and at the other, all the responses in one experimental condition aggregate to form an RT distribution, described by the Weibull parameters. Even though the Weibull model takes only correct responses into account, the EZ2 estimations will still be able to account for the descriptive model, because the benchmark paradigms produce high accuracy responses.

In summary, this study examined three questions related to the perceptual decision making during visual search. The first question is whether the demands of search task affect the drift rate of sensory evidence accumulation related to decision speed and how this influence manifests in an RT distribution with regard to its shift and shape. The three benchmark search tasks here likely required various high-level cognitive processes, such as focusing attention to improve the quality of sensory evidence and binding multiple features to match a search template. Particularly, the spatial configuration search task has been showed highly inefficient (Bricolo, Gianesini, Fanini, Bundesen, & Chelazzi, 2002; Kwak, Dagenbach, & Egeth, 1991; Woodman & Luck, 2003). It is reasonable to expect this particular search task changes the shape of the RT distribution drastically. The second question examined whether the display size affects the shape of the RT distribution. As the stage model of information processing (Rouder et al., 2005) presumes, the shape of an RT distribution is likely affected specifically by late-stage cognitive process. If the increase of search item in a display merely adds burden on early perceptual process, I should expect no influences from the display size on any decision parameters and thus the RT shape. The third question examined the hypothesis of group segmentation and recursive rejection

processes in search (Humphreys & Müller, 1993). Specifically, segmentation and distractor rejection may involve both late-stage cognitive processes (binding multiple search items as a group), and early-stage perceptual processes (recursively encoding sensory information). This may in turn affect the decision and non-decision parameters and therefore, manifest as an interaction effect in the shape of the RT distributions.

4.2 Method

4.2.1 Design

The study used a similar design to Wolfe et al. (2010) with a slight modification. Specifically, I used a circular display layout with a viewing area of 7.59° × 7.59°, which allocates 25 locations to hold search items. Wolfe and colleagues (2010) used a viewing area of about 22.5° × 22.5° (also with 25 search locations) and each search item subtended around 3.5° to 4.1°. Relative to Wolfe et al.'s study, my setting (i.e., using a similar number of search items presented in a smaller viewing area) rendered a high density of homogeneous distractors more likely when display sizes were large.

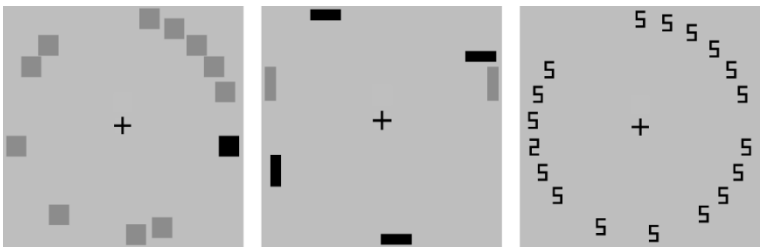


Figure 4-1. A schematic representation of the tasks; in each there was a target present [black item (feature); black vertical (colour-form conjunction) and the number 2 (spatial configuration)].

The study investigated two factors, the display size (3, 6, 12, & 18 items) and whether the target was present or absent, using a repeated-measures,

within-participant design. One group of participants (N = 20) took part in the feature and conjunction search tasks, and a second group took part in the spatial configuration search task (N = 20).

In the feature search task, each observer looked for a dark square amongst varying numbers of grey squares (both were 0.69° × 0.69°). In the conjunction search task, observers looked for a vertical, dark bar (0.33° × 0.96°) amongst two types of distractors, vertical grey bars (0.33° × 0.96°) and horizontal dark bars (0.96° × 0.33°). In the spatial configuration search task, each observer looked for the digit 2 amongst digit 5s (both are 0.33° × 0.58°). Before the search display was presented, a 500-ms fixation cross appeared at the centre of the screen, followed by a 200-ms blank duration. A trial was terminated when the observer pressed the response key.

4.2.2 HBM

The framework of the HBM is based on Rouder and Lu's R code (2005), which used a Markov Chain Monte Carlo (MCMC) algorithm to implement hierarchical data analysis assuming a three-parameter Weibull function. I modified Rouder and Lu's code into an OpenBUGS-based R program by adapting Merkle and van Zandt's (2005) WinBUGS code to run a Weibull hierarchical BUGS model (Lunn, Spiegelhalter, Thomas, & Best, 2009), which was linked with R codes by R2jags (Sturtz, Ligges, & Gelman, 2005) and Just Another Gibbs Sampler (JAGS) (Plummer, 2003)[19].

The Weibull function was used to model the individual RT observations,

[19] The readers who are interested in the programming details could visit the authors' GitHub at https://github.com/yxlin/HBM-Approach-Visual-Search

assuming that each of them was a random variable generated by the Weibull function. The function comprises three parameters, the shape (i.e., β, describing the shape of a RT distribution), the scale (i.e., θ, describing the general enhancement of the magnitude and variability in a RT distribution), and the shift (i.e., ψ, describing the possible minimal response time of a RT distribution). The β parameter was then modelled by a γ distribution with two hyper-parameters, η1 and η2, and the θ and ψ parameters were modelled by two uniform distributions. The former (θ) was initialized as an un-informative distribution, whereas the latter (ψ) was set to the range of zero to minimal RTs in the respective condition and participant, because the ψ parameter assumed a role as the non-decision component. The hyper-parameters underlying the γ distributions were then modelled by other γ distributions with designated parameters, following Rouder and Lu (2005). Likewise, I replaced the Weibull function with the 3-parameter gamma, lognormal, and Wald functions (Johnson et al., 1994), keeping similar prior parameter setting.

In the HBM, correct RTs were modelled for each participant separately in each condition. The HBM ran 3 simultaneous iteration chains. Each of them iterated 105000 times and sampled once every 4 iterations to alleviate possible auto-correlation problems. The first 5000 samples were considered to be arbitrary and discarded (i.e., burn-in length). The same setting was applied both to my data and to Wolfe et al.'s data (2010) to help a direct comparison.

4.2.3 EZ2 Diffusion Model

The analyses also used Grasman, Wagenmakers and van der Mass's (2009) EZ diffusion model, implemented in R's EZ2 package, to estimate the drift

rate, boundary separation and non-decision component separately for each participant in each condition. Following the assumption of the EZ diffusion model (Wagenmakers, van der Maas, et al., 2008), the across-trial variability associated with the drift rate, boundary separation and non-decision components was held constant. Due to the high accuracy rate, the analyses applied the edge correction procedure [20] following Wagenmakers et al. (2008; see also other possible solutions in Macmillan & Creelman, 1991) for the conditions where an observer committed no error. Present or absent responses were modelled separately, using the Simplex algorithm (Nelder & Mead, 1965) to approach a converging estimation. The initial input values to the EZ2 model was set according to the paradigm and literature: (1) the paradigm permitted only two response options, either the target was present or absent and (2) the search slope for present-to-absent ratio was slightly greater than 2 (Wolfe, 1998b). Accordingly, the initial values of the drift rates for present and absent responses, were respectively set at 0.5 and 0.25. The non-decision component and the boundary separation were arbitrarily, but reasonably, set at 0.05 and 0.09. The initial values are simply educated guesses provided for the algorithm approaches reasonable estimations.

Both for the HBM and the diffusion model, the parameters were estimated as per-condition per-participant basis, so data from each participant contributed 24 (3 × 2 × 4) data points for each parameter. The analyses assessed the variability across individuals in visually-weighted regression lines, using a non-

[20] When an observer make no error response (i.e., 100% accuracy, P_c), the accuracy is replaced with a value that corresponds to one half of an error, following the formula, $P_c = 1 – (1/2n)$.

parametric bootstrapping procedure, implemented by Schönbrodt (2012) for Hsiang's visually-weighted regression (VWR) method (2013)[21]. The VWR is a data visualisation technique that attempts to give a visually intuitive impression for what the data may inform. It applies lowess to estimate regression functions and the bootstrap method to estimate confidence interval. The principle to represent the confidence interval in VWR is to use the bootstrapped line density and its colour saturation to inform uncertainty intuitively. That is, the denser the bootstrapped lines, and the more saturated its colour is, the more probable the data suggest.

4.3 Result

I report the data in four sections. Firstly, I report standard search analyses, using mean measures of performance for individuals across trials. Next, I present the distributional analyses, using box-and-whisker plots, probability density plots with quantile-quantile subplots, and empirical cumulative density plots, to recover the RT distributions. The distributions from each condition were then compared. Thirdly, the standard search analyses and the distributional analyses were then contrasted with previous findings reported by Wolfe et al. (2010) and by Palmer et al., (2011)[22]. In the last section, I report the analyses, using the HBM and the EZ2 diffusion model. These include the data for the Weibull and the diffusion model parameters, presented separately, with visually-weighted non-parameter regression plots. From this I go on to discuss the factors contributing to the RT shape, shift and scale parameters,

[21] The technique was discussed and implemented in the blogsphere before it was formally published in the 2013 technical report.
[22] I thank Jeremy Wolfe and Evan Palmer for their permission.

based on how these parameters change across the different search conditions and contrast them with the decision parameters from the diffusion model. Chapter 2 presents a simulation study to examine if Weibull HBM estimates of distributional parameters are reliable with a small sample size and that Bayesian diagnostics verify the reliability of MCMC procedure.

I focus on the data from target present trials because absent trials likely involve a different set of decision processes (one possibility is an adaptive termination rule, suggested by Chun & Wolfe, 1996; alternatively see a recent computational model in R. Moran et al., 2013). A decision in an absent trial is reached, possibly based on, for example, a termination rule that an observer deems the collected sensory evidence is strong enough to refute the presence of a target. Although it is likely an observer, in a present trial, may also adopt an identical termination rule to infer the likelihood of the target presence, he/she would rely on the stronger sensory evidence extracted from a target than those from non-targets. This is likely when a target image is physically available in a present trial and target foreknowledge is set up in an attentional template. Thus, the main aim of report is to examine the role of factors such as target-distractor grouping effect on the distribution of target present responses in search. I nevertheless append also standard analyses for absent trials in all the figures.

4.3.1 Mean RTs and Error Rates

As is typically done for the aggregation RT analyses, I trimmed outliers by defining them as (1) incorrect responses or correct responses outside the range of 200 ms to 4000 ms for feature and conjunction searches and 200 ms to 8000 ms for spatial the configuration search (though see, Heathcote et al., 1991, for

the downside of trimming RT data). The trimming scheme was the same as in Wolfe et al. (2010). This outlier trimming resulted in a rejection rate of 9.2%, 12%, and 7.2%, of the responses respectively for the three tasks. After excluding the outliers, the data were then averaged across the trials within each condition, resulting in 76 averaged observations for the feature and conjunction searches and 80 observations for the spatial configuration search. All outliers were defined as error responses.

The mean RT were modulated by all experimental factors and their interaction, as indicated by the two-way ANOVA[23], showing the display size effect, $F(3, 165) = 176.11$, $\eta^2_p = .76$, $p = 1 \times 10^{-13}$, the task effect, $F(2, 55) = 108.39$, $\eta^2_p = .80$, $p = 1 \times 10^{-13}$, and the interaction effect, $F(6, 165) = 68.63$, $\eta^2_p = .71$, $p = 1 \times 10^{-13}$. The spatial configuration search (RT_{mean} = 913 ms) required reliably longer response times than the conjunction search task (mean difference = 327 ms, 95% CI, [244, 411] ms, $p = 5.89 \times 10^{-13}$), which in turn had longer mean RTs (586 ms) than the feature search task (428 ms; mean difference = 158 ms, 95% CI, [74, 243] ms, $p = 6.68 \times 10^{-5}$). Separate ANOVAs support the display size effect in all search tasks, [feature, $F(3, 54) = 7.49$, $\eta^2_p = .29$, $p = 2.78 \times 10^{-4}$; conjunction, $F(3, 54) = 103.15$, $\eta^2_p = .85$, $p = 1 \times 10^{-13}$; spatial configuration, $F(3, 57) = 113.80$, $\eta^2_p = .86$, $p = 1 \times 10^{-13}$].

The error rates showed a similar pattern as the mean RT with a display size, effect $F(3, 165) = 38.09$, $\eta^2_p = .41$, $p = 1 \times 10^{-13}$, a task effect, $F(2, 55) =$

[23] The three task levels were treated as a between-participant factor for straight-forward presentation, although the levels of feature and of conjunction search are within-participant factor. Even under this calculation (leaving more variation unexplained), the RT_{mean} amongst three tasks still showed reliable differences.

5.75, η^2_p = .17, p = .005 and the interaction effect, F(6, 165) = 10.867, η^2_p = .283, p = 3.52×10^{-10}. This is consistent with there being no trade-off between the speed and accuracy of responses. The spatial configuration search (error $rate_{mean}$ = 11 %) was more difficult than the conjunction search task (8 %), but the difference did not exceed significant level after Bonferroni correction (the difference of mean error rate = 3 %, 95% CI, [1.774, 8.134], p = .356). The conjunction search task in turn was more difficult than the feature search task (5%; the difference of mean error rate = 4 %, 95% CI, [-1.396, 8.628], p = .241). The only reliable difference of error rates was between the spatial configuration search and the feature search tasks (the difference of mean error rate = 7 %, 95% CI, [1.847, 11.755], p = .004)

For the feature search, the effect of display size was not reliable, F(3, 54) = 1.52, η^2_p = .08, p = .22, while there was a reliable effect of display size for both the conjunction task, F(3, 54) = 6.08, η^2_p = .25, p = .001, and the spatial configuration task, F(3, 57) = 41.43, η^2_p= .69, p = 1.24×10^{-13} (lower panel in *Figure 4-3*). Post-hoc t tests indicated that in the conjunction search task participants committed more errors at display size 18 (13 %) than at display sizes 12 (9 %; p = .028) and at 6 (7 %; p = .043, Bonferroni correction for multiple comparisons). In the spatial configuration search, there were differences across all display size pairings, p = 5.90×10^{-5}, 9.85×10^{-6}, 3.58×10^{-4}, 6.80×10^{-6}, & 1.21×10^{-5} (3 vs. 12, 3 vs. 18, 6 vs. 12, 6 vs. 18, & 12 vs. 18; Bonferroni correction for multiple comparisons), except for display sizes 3 and 6 (p = .16).

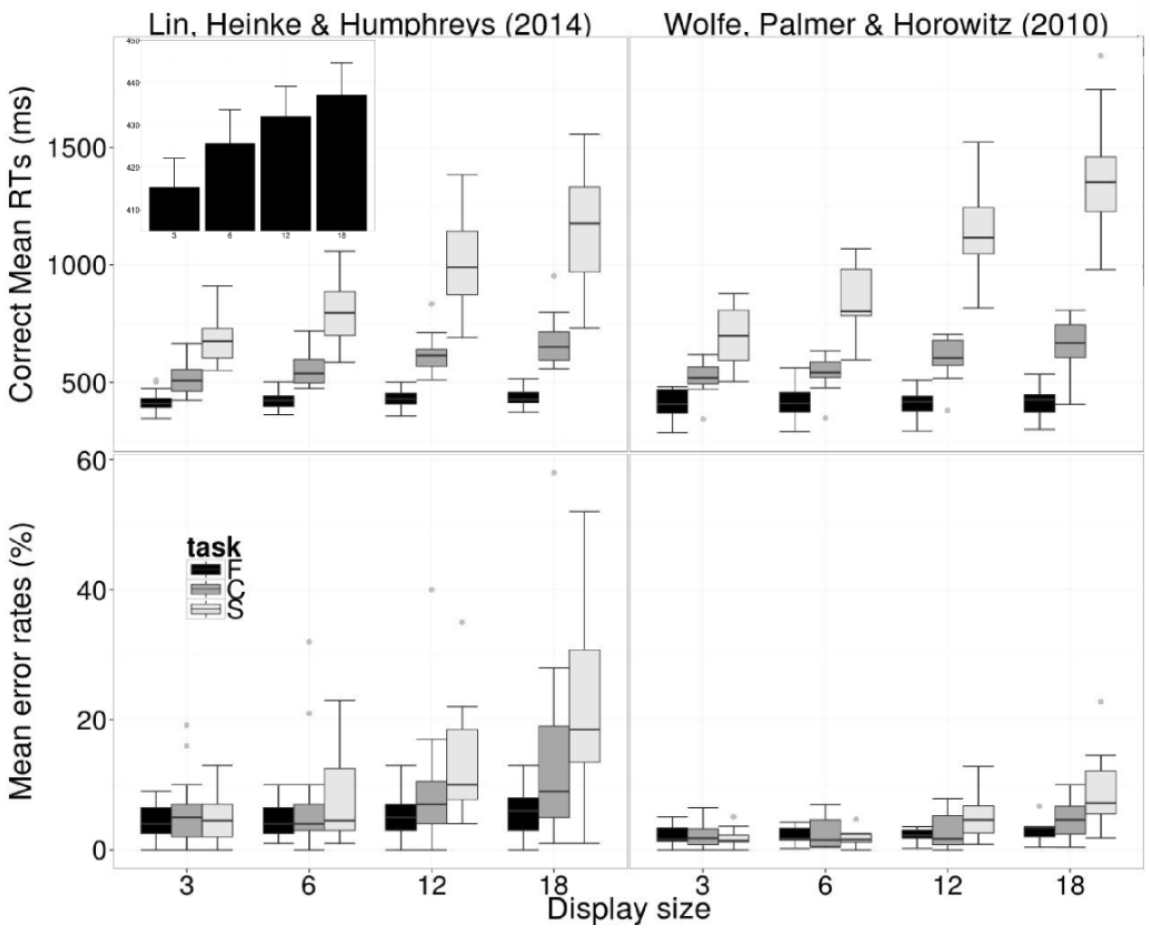


Figure 4-2. The box-and-whisker plots. The upper and lower panels show the means of RTs and error rates, respectively. The subplot in the upper-left panel shows a zoom-in view of the bar-plot of the feature search task (y axis ranging between 405 to

450 ms; x axis labelling four display sizes). The left and right panels present the analyses from the current and Wolfe et al.'s (2010) data sets, respectively.

4.3.2 Error Analysis

To test if the shape changes in an RT distribution are due to an increase of miss errors (Wolfe et al., 2010), I also analysed two types of error, miss (participants pressed the absent key in target present trials) and false alarm (participants pressed the present key in target absent trials).

The miss error rate showed a reliable display size effect, $F(3, 165) = 38.08$, $\eta^2_p = .41$, $p = 1 \times 10^{-13}$,a search task effect, $F(2, 55) = 5.75$, $\eta^2_p = .17$, $p = .005$ and an interaction effect, $F(6, 165) = 10.85$, $\eta^2_p = .28$, $p = 3.62 \times 10^{-10}$. Both the spatial configuration, $F(3, 57) = 41.37$, $\eta^2_p = .69$, $p = 1.25 \times 10^{-13}$, and the conjunction search task, $F(3, 54) = 6.08$, $\eta^2_p = .25$, $p = .001$, showed increasing miss errors as the display size increased, but not the feature search task, $F(3, 54) = 1.52$, $\eta^2_p = .08$, $p = .221$. The false alarm rate showed only a display size effect, $F(3, 165) = 3.94$, $\eta^2_p = .07$, $p = .010$. The reliable effect of false alarm errors was observed in both feature, $F(3, 54) = 2.81$, $\eta^2_p = .14$, $p = .048$ and conjunction search, $F(3, 54) = 2.96$, $\eta^2_p = .14$, $p = .04$, but not in spatial configuration search, $F(3, 57) = 1.14$, $\eta^2_p = .06$, $p = .34$ (Figure 4-3).

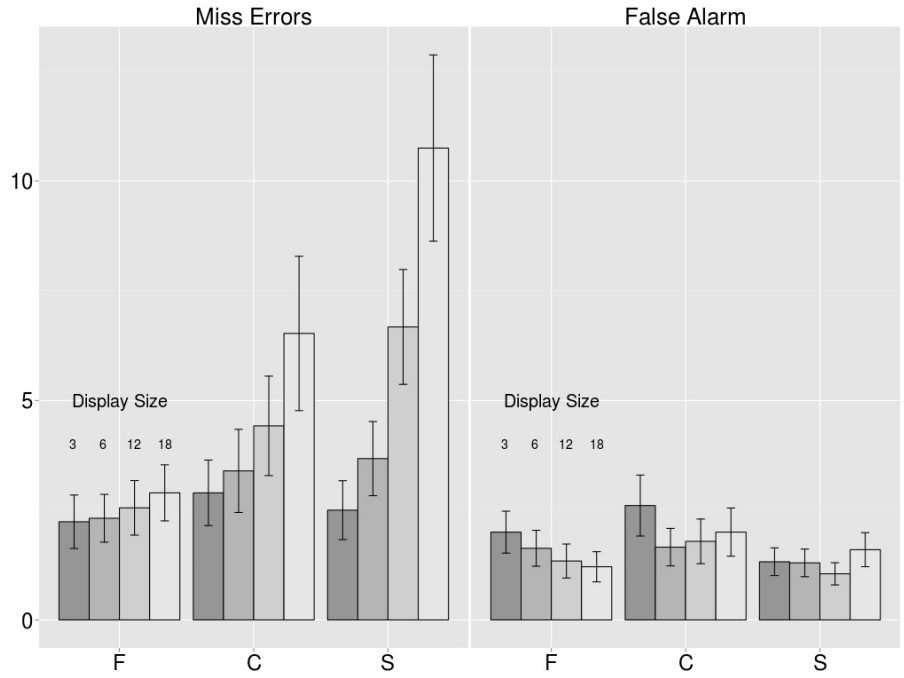


Figure 4-3. Mean rates of miss and false alarm errors. The error bars show one standard error of the mean. The y axis shows percentage of errors. F, C and S stand for feature, conjunction and spatial configuration searches.

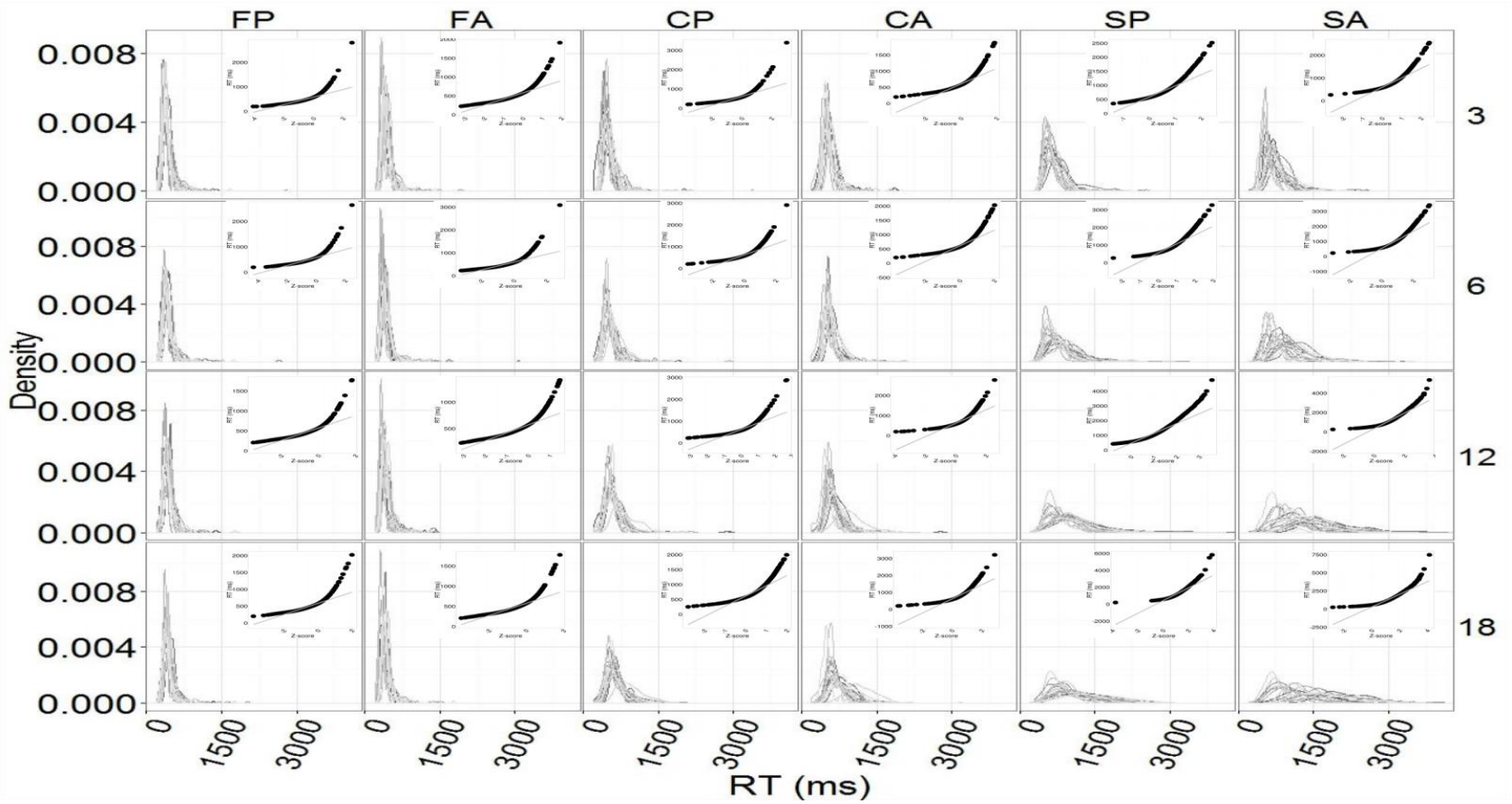


Figure 4-4. The trial-RT distributions. The subplots are the normalised quantile-quantile (Q-Q) plot for the corresponding density plots. The y axis of the Q-Q plots compares trial RTs [y axis labelled, RT (ms)] with normalised z scores [x axis labelled, (Z-score)]. The black dots in the Q-Q plots represent data points. The more black dots deviate from the normal reference (grey diagonal) line, the stronger it suggests the data are drawn from a non-normal distribution. The plot is re-scaled to make the small print easier to read. P and A stand for present and absent trials, respectively.

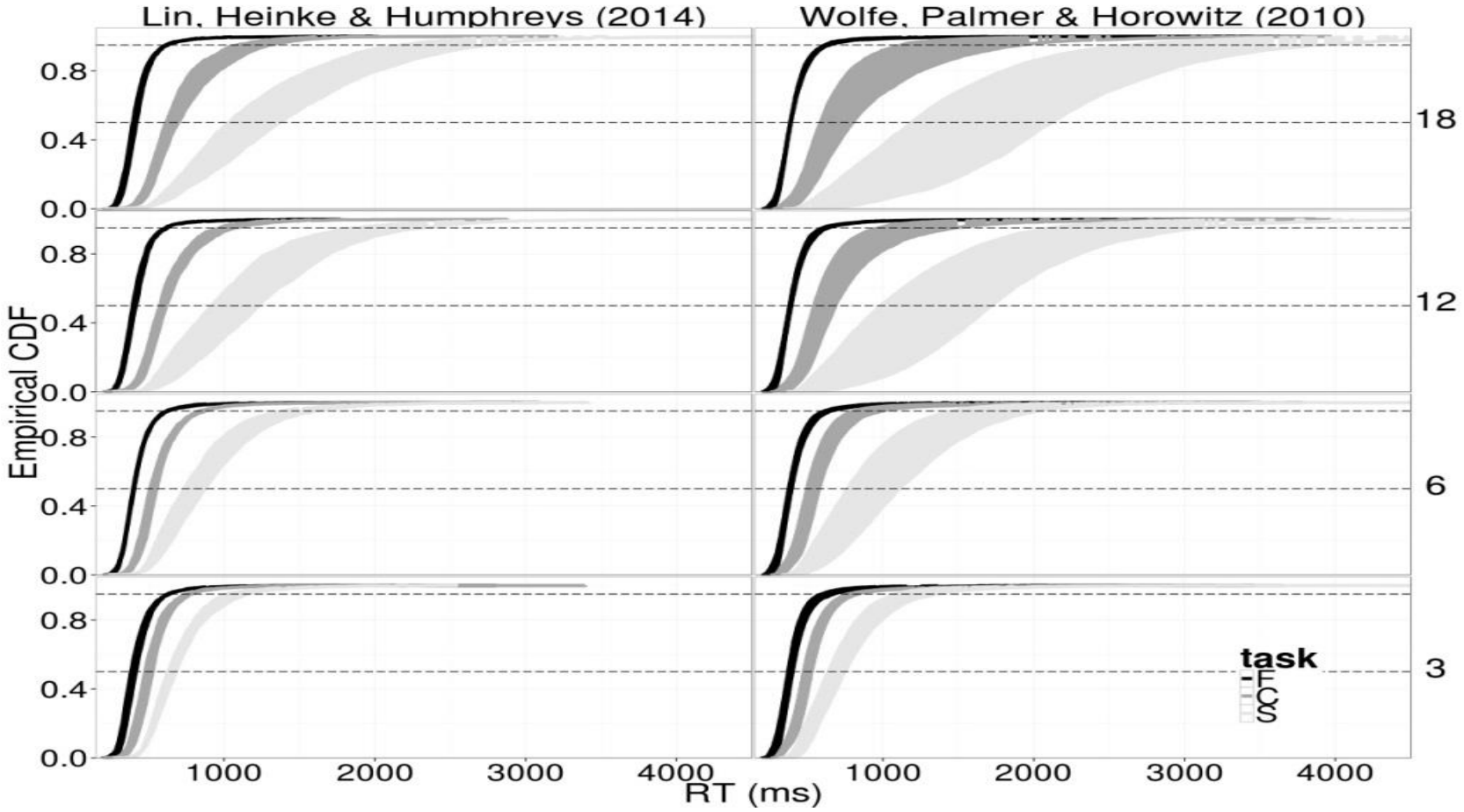


Figure 4-5. The empirical cumulative RT density curves drawn based on the trial RTs. The areas within each envelope represent the differences between target present and target absent trials for each task. The two dashed lines show the positions of the 50% and 95% cumulative densities. Long latencies (right border of envelopes) were consistently observed on target absent trials. The plot is re-scaled to make the small print easier to read.

4.3.3 Distributional Analysis

Figure 4-4 shows the RT distributions and quantile-quantile plots. The distributions were constructed based on the trial RTs (43485 data points). Each density line represents the data from one participant. Evidently, the normality assumption was untenable across all the conditions. All sub-plots showed that the data clearly deviated from the theoretical normal lines. It is also apparent that individual differences play a more important role for the conjunction and spatial configuration tasks than for the feature task, judging by the diversity of the density lines in the two difficult search tasks.

Figure 4-5 shows the empirical cumulative distributions, drawn based on trial RTs (43485 and 109036 data points in my and Wolfe et al.'s data sets, respectively). The contrasting RTs across the display sizes confirm Wagenmakers and Brown's (2007) analysis that, in inefficient relative to efficient search tasks, the RT standard deviation, together with the RT mean, play crucial roles in describing visual search performance. Specifically, the elongated cumulative distributions suggest that the more items are present, the more likely an observer will produce a response that falls in the right tail of the RT distribution. This observation again cautions against a reliance solely on using the measurement of the central location when investigating visual search performance.

4.3.4 Contrasts with Prior Data

The data here was compared with those of Wolfe et al.'s (2010) data. The mean RTs and error rates indicated similar patterns across the studies. The data for the trial RTs revealed skewed distribution in both studies. Even though the

general pattern is similar, the two data sets showed some differences. One difference is the variability. The experiment in Wolfe et al. (2010) recruited less participants, and each of them contributed far more per-condition observations than the current study. Although it is unclear if the strict experimental procedures, such as the real-time machine and high-precision respond pad, helped also for decreasing the variability in the current dataset, the distributional plots suggest that decreased variability may also contribute to the different, especially in the Weibull parameter, findings in the two studies.

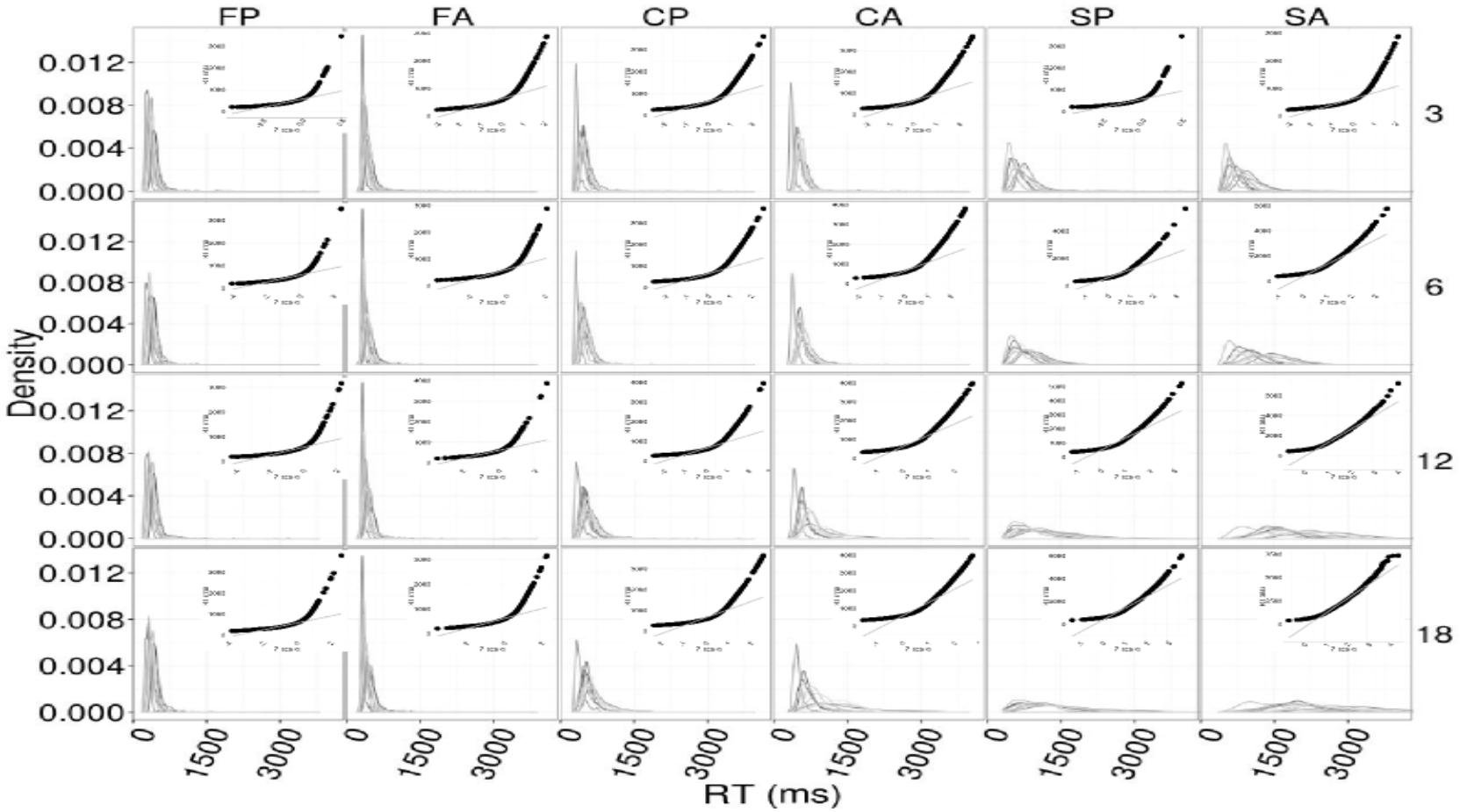


Figure 4-6. The trial-RT distributions. The y axis of the Q-Q normalised plots compares trial RTs [y axis label, RT (ms)] with normalised z scores [x axis label, (Z-score)]. Data are from Wolfe et al (2010). The plot is re-scaled to make the small print easier to read.

4.3.5 The HBM Estimates

In this section, I firstly presented each parameter separately for the respective ANOVA results, and I compared the data for the three search tasks at the different display sizes, modelled by the HBM. Next, I conducted a non-parametric bootstrap regression to assess the relationship between the display size and the difficulty of the search task. The analysis focused on target-present trials. I used the DIC to evaluate the function fit to the data. In general, the small the DIC, the better fit (Lunn, Jackson, Best, Thomas, & Spiegelhalter, 2013). Although the lognormal function showed the smallest DIC[24], the DICs across the four fitted functions were close. Although the lognormal and Wald functions showed the smaller DICs than the Weibull function, the DICs across the four fitted functions are close (Table 4-1Table 4-). Moreover, the diagnostics of the gamma HBM suggest its posterior distributions did not converge. Excluding the non-converged gamma function, I reported arbitrarily the estimates from the Weibull HBM, given that prior work shows this provides a highly robust account, not strongly moderated by noise in the data (see a specific pathology of the Weibull function in Rouder & Speckman, 2004, pp 424-425; and how HBM resolves this problem in Rouder et al., 2005, pp. 203).

[24] DIC suggests which model may fit better, but is not a definitive theoretical index, as its usage depends on the context (e.g., how a model is constructed). I used it as a convenient proxy to understand if Weibull function fits substantively different from the other comparable functions.

Table 4-1. The DICs of the 4 fitted functions. They are averaged across the absent and present trials, tasks and display sizes. The smaller the DICs, the better the fit.

	Lin, Heinke, & Humphreys (2015)	Wolfe, Palmer, & Horowitz (2010)
Gamma	385,348,342	975,871,147
Log normal	385,348,002	975,870,279
Wald	385,348,026	975,870,358
Weibull	385,348,139	975,871,078

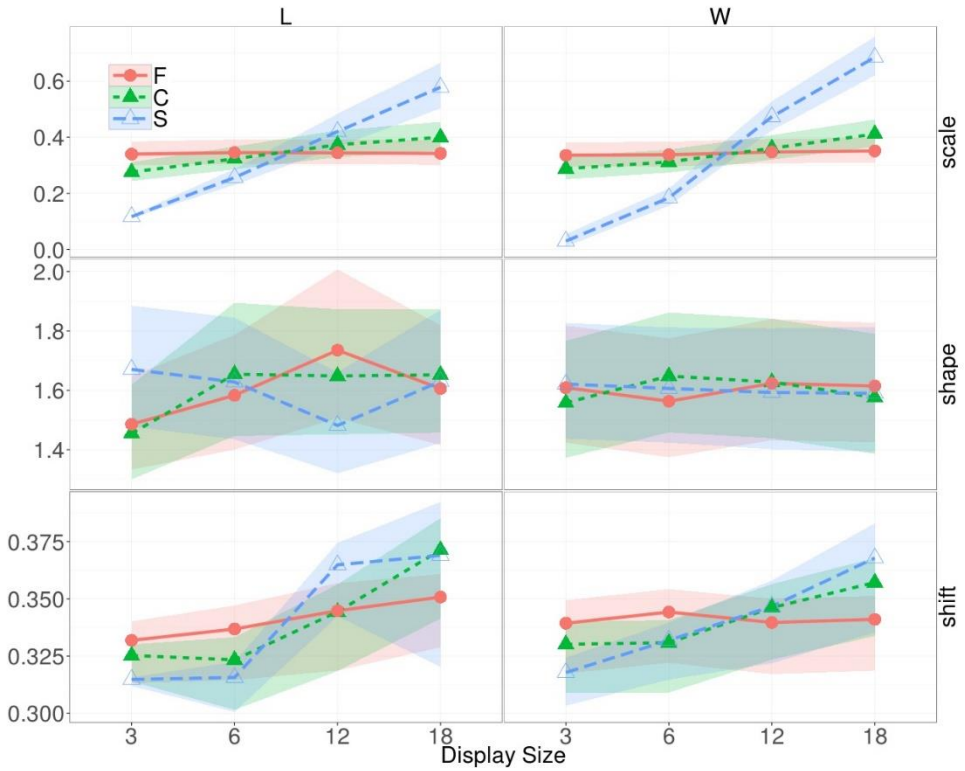


Figure 4-7. The line plots for the Weibull parameters. L and W stand for my and Wolfe et al.'s (2010) datasets. The error ribbons were drawn based on the credible intervals estimated by the BUGS model.

4.3.5.1 Shift

The shift parameter were influenced by the task, $F(2, 55) = 129.75$, $p = 1.0 \times 10^{-13}$, $\eta^2_p = .83$, and the display size factors, $F(3, 165) = 9.03$, $p = 1.43 \times 10^{-5}$, $\eta^2_p = .14$ (two-way ANOVA). The shift parameters increased gradually from the feature, and the conjunction search to the spatial configuration search (post-hoc t test; 246, 342, vs. 436 ms; $p = 2.37 \times 10^{-10}$ & 2.83×10^{-10}). The shift plot in Figure 4-8 collaborated with the ANOVA results, showing that each task demonstrates very different magnitudes. The display size effect is due to the contrast between the largest and the smallest display size. Comparing to other Weibull parameters, the gradient of display size change in the shift parameter is relatively small.

4.3.5.2 Shape

The shape parameter were influenced by the task factor, $F(2, 55) = 23.50$, $p = 4.21 \times 10^{-8}$, $\eta^2_p = .46$, and the task × displays size interaction, $F(6, 165) = 3.45$, $p = .003$, $\eta^2_p = .11$. The display size influenced the shape parameter only marginally, $F(3, 165) = 2.44$, $p = .067$, $\eta^2_p = .04$. The marginal display size effect was supported by separate ANOVAs, showing the display size influences on the conjunction search, $F(3, 54) = 4.21$, $p = .01$, $\eta^2_p = .19$ (1496, 1731, 1695 vs. 1702 ms) and spatial configuration search, $F(3, 57) = 4.45$, $p = .007$, $\eta^2_p = .19$ (1573, 1541, 1397 vs. 1529 ms).

The shape plot in Figure 4-8 suggests that the search task differences appears to occur at large display sizes (i.e., 6, 12 and 18). Secondly, there is a

U-shaped function for the spatial configuration task – both for the magnitude and variability of the shape parameter. This U-shape pattern is not observed in Wolfe et al.'s (2010) data. The emergent decreases in the mean shape parameter and the associated increase in the variability suggest there might be an emergent mechanism influencing search at the larger display sizes. The SERR group segmentation account for the emergent observation will be discussed in the Discussion.

4.3.5.3 Scale

Similar to the shift parameter, the scale parameter is modulated by the two experimental factors [task, $F(2, 55) = 161.70$, $p = 1.0 \times 10^{-13}$, $\eta^2_p = .86$; display size, $F(3, 165) = 39.75$, $p = 1.0 \times 10^{-13}$, $\eta^2_p = .420$] and their interaction, $F(6, 165) = 19.31$, $p = 1.0 \times 10^{-13}$, $\eta^2_p = .413$. The display size main effect reflects from the conjunction search, $F(3, 54) = 10.00$, $p = 2.42 \times 10^{-5}$, $\eta^2_p = .36$ (206, 257, 301 & 334 ms) and the spatial configuration search, $F(3, 57) = 33.47$, $p = 1.42 \times 10^{-12}$, $\eta^2_p = .64$ (302, 444, 607 & 760 ms). The VWR plot showed the shape parameter is more variable than the other two parameters and this is especially the case for the feature search.

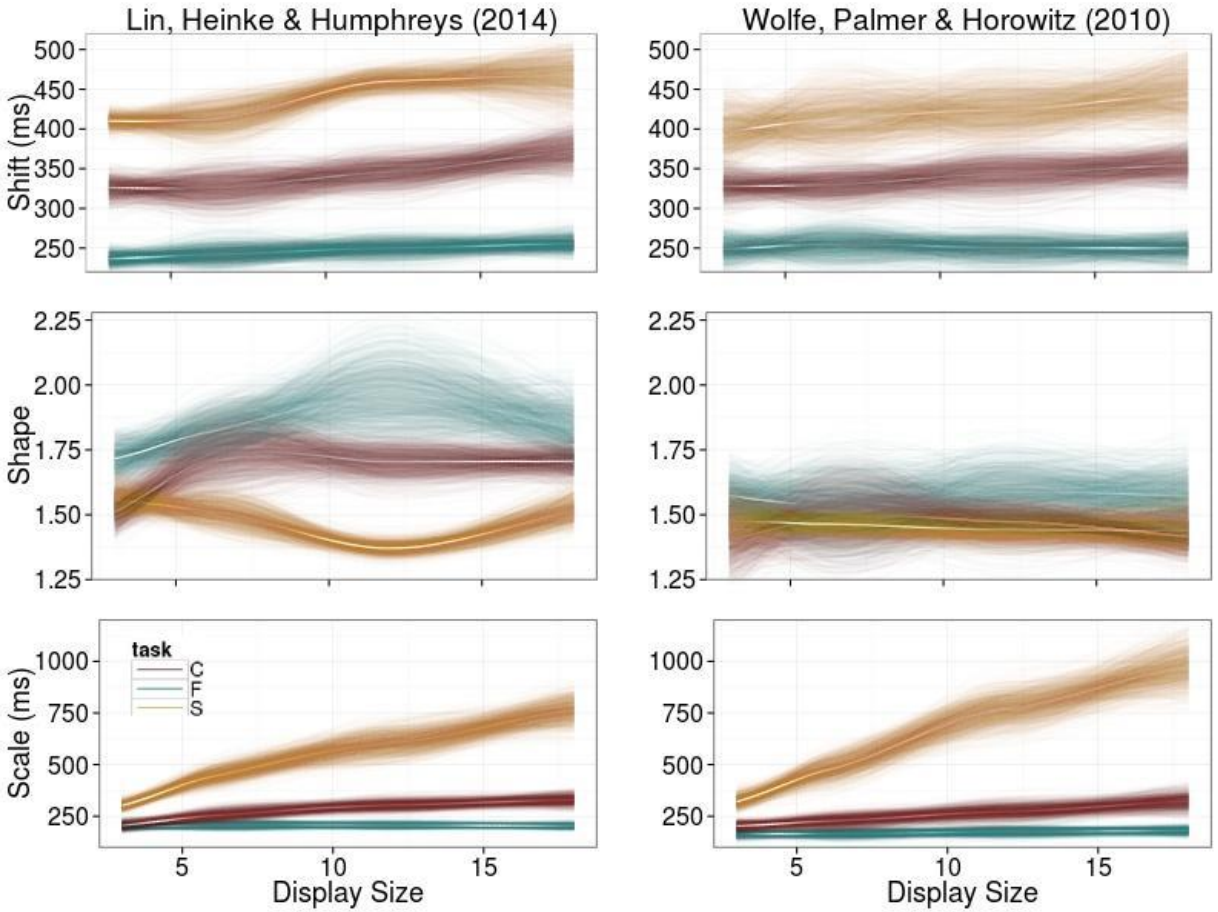


Figure 4-8. Visually-weighted regression (VWR) plots (Hsiang, 2012) for the three Weibull parameters. VWR performs regressions using display size as the continuous independent variable and Weibull parameters as the predicted variables separately for the three search tasks. The white lines in the middle of each ribbon show the predicted regression lines. To show

differences across the conditions (display sizes and tasks), the uncertainty, which usually error bars aim to communicate, is estimated via bootstrapping nonparametric regression lines. Here I used locally weighted smoothing (loess; Cleveland, Grosse, & Shyu, 1992). The density of lines and its colour saturation were drawn in a way to reflect the extent of uncertainty. The denser and more saturated a ribbon is, the less between-participant variation it shows.

4.3.6 Diffusion Model

The section I present the three diffusion model parameters, using an identical analysis protocol as in previous section. Again, the analyses focused on target-present trials.

4.3.6.1 Drift rate

The drift rate was influenced by the task factor, $F(2, 55) = 9.47$, $p = 2.92 \times 10^{-4}$, $\eta^2_p = .26$, but not the display size factor, nor their interaction. Further tests a larger drift rate for the feature search (0.323) than that for the conjunction search (0.265; marginally significant, 95 % CI, [-0.12, .001], $p = .057$) and that for the spatial configuration search (0.220; 95 % CI, [0.04, 0.16], $p = 1.81 \times 10^{-4}$). No difference was found between the conjunction and spatial configuration searches.

The drift rate in Figure 4-9 maintained at a stable rate across the display sizes in the feature and the conjunction searches. On the other hand, the three tasks showed noticeably different drift rates. There was a tendency also for the drift rate to rise at the large display size in the spatial configuration task, suggesting that there was an emergent factor, although the variability across observers suggested that this was not universally the case for all participants. This was not evident in absent trials[25]. This upward trend was also not present in the data of Wolfe et al. (2010).

[25] See https://github.com/yxlin/HBM-Approach-Visual-Search for absent trial data.

4.3.6.2 Non-decision time

The non-decision time differed reliably among the search task, $F(2, 55) = 5.64$, $p = .006$, $\eta^2_p = .170$, which interacted with the display size factor, $F(6, 165) = 4.16$, $p = .001$, $\eta^2_p = .131$. But no display size difference was observation in the non-decision time. Post-hoc t tests showed that spatial configuration search (79 ms) was associated with a longer non-decision time than feature search, (57 ms; 95 % CI, [4.53, 38.1], $p = .008$) and conjunction search (61 ms; 95 % CI, [0.71, 34.2], $p = .038$). The interaction effect is due to a reliable display size influence on the spatial configuration task, $F(3, 57) = 6.89$, $p = 4.89 \times 10^{-4}$, $\eta^2_p = .27$ (60.59, 80.54, 89.50 vs. 84.23 ms), but not other tasks.

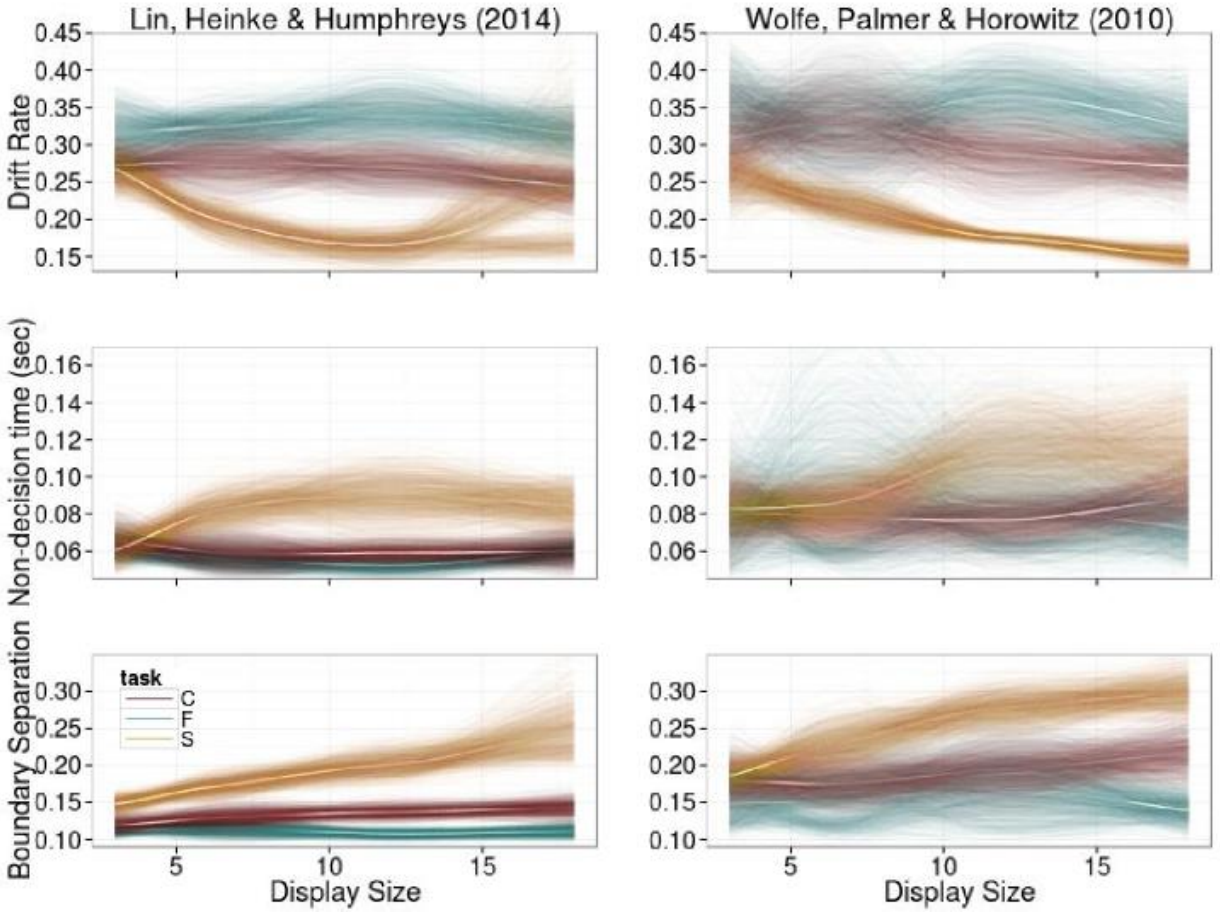


Figure 4-9. The VWR plot for the EZ2 diffusion model

4.3.6.3 Boundary separation

Boundary separation, similar with the mean RT, accuracy rate and the Weibull scale parameter, showed reliable main and interaction effects [the task, $F(2, 55) = 31.75$, $p = 6.81 \times 10^{-10}$, $\eta^2_p = .54$; the display size, $F(3,165) = 7.6$, $p = 8.61 \times 10^{-5}$, $\eta^2_p = .12$; the task × display size interaction, $F(6, 165) = 4.76$, $p = 1.69 \times 10^{-4}$, $\eta^2_p = .15$]. The task main effect is due to that the observers responded to the feature search, using a more liberal criterion than the spatial configuration search (0.11 vs. 0.19, $p = 1.01 \times 10^{-9}$, 95% CI, [0.06, 0.11]). Although the response criterion for the conjunction search was in-between the above-mentioned two tasks, it differed reliably only from the spatial configuration condition ($p = 1.49 \times 10^{-6}$, 95% CI, [0.03, 0.09]).

Table 4-2. Summary table for the significance of two-way ANOVAs for all tested parameters

	Mean RT	Error rate	RT Shift	RT shape	RT scale	Drift rate	Non-decision time	Boundary separation
Display Size	V	V	V		V			V
Task	V	V	V	V	V	V	V	V
Interaction	V	V		V	V		V	V

4.4 Discussion

The study applied an integrated approach to modelling visual search data. It examined the data not only using standard aggregation approaches, but also using distributional approaches to extract cognitive-related parameters from the trial RTs. This approach reveals the possible accounts of the three distributional

parameters – shift, shape and scale – associating them with non-decision time, drift rate and boundary separation estimated from the diffusion model. This study goes further than most previous studies (Balota & Yap, 2011; Heathcote et al., 1991; Sui & Humphreys, 2013; Tse & Altarriba, 2012) that have applied distributional analysis to RT data. It used conventional distributional analyses to examine empirical RT distributions and the associated parameters were complemented with Bayesian-based hierarchical modelling to optimise estimates. Moreover, the study examined those distributional parameters against a plausible computational model – the EZ2 model – to link the distributional parameters to underlying psychological processes. The former statistical model associates the experimental factors with the changes in RT distributions and the latter EZ2 model associates the experimental factors with the EZ2 parameters, the drift rate and the response boundary. A cross-comparison of the EZ2 parameters and the RT distributions could suggest how the changes in the cognitive processes affects the RT distributions.

Replicating many previous findings in the search literature the data show efficient search for feature targets and inefficient search when targets can only be distinguished from non-targets by conjoining multiple features (shape and color, or shape only, see Chelazzi, 1999; Chun & Wolfe, 2001, for a review). The display size effect presented in the feature search (415, 426, 432 & 437 ms) suggests some limitations on selecting feature targets but the analyses based on mean RTs do not differentiate if the effect (η^2_p = .294) is due to post-selection reporting (Duncan, 1985; Riddoch & Humphreys, 1987) or an involvement of focal attention in feature search. This question is addressed by examining the

estimates from the HBM together with the EZ2 model. The lack of display size effects in the non-decision time suggest that the increasing trend in the mean RTs is unlikely due to a delay in the peripheral processes, such as motor or early perceptual times. Neither drift rate showed a reliable effect in the display size at the feature search. The only possible difference is an unreliable display size effect ($p = .106$) together with an increase of variation at the shape parameter in the condition of display size 18. This result appears to favour the explanation of focal attention.

Though previous results have indicated that search is often inefficient for conjunction and configuration-based stimuli, my findings indicated that spatial configuration search was particularly difficult (Bricolo et al., 2002; Kwak et al., 1991; Woodman & Luck, 2003). This could reflect either a reduction in the guidance of search from spatial configuration compared with simple orientation and colour information, or in the speed for identifying each item after it had been attended. Interestingly, although when compared with the standard error of the conjunction search (9.68 ms), the configuration search generally showed a larger value across participants (24.54 ms), the standard errors within the configuration search decreased as the display sizes increased (35.17, 27.12, 15.38, vs. 20.49 ms). This last result suggests high density homogeneous configurations of distractors rendered search less variable, a point I return to below (Bergen & Julesz, 1983; Chelazzi, 1999; Duncan & Humphreys, 1989; Heinke & Backhaus, 2011; Heinke & Humphreys, 2003).

4.4.1 RTs Between and Within Participants

The analyses for the (between-participant) mean RTs do not always

accord with the analyses of (within-participant) trial RTs. The analyses of the trial RTs reveal clear skewed RT distributions, suggesting that the data are distributed symmetrically when taking average across participants and skewed positively and distributed towards long latencies when examining at the within participants (Luce, 1986).

In additional to the analyses of mean performance, the integration of hierarchical Bayesian and EZ2 model threw new light on search. Following Rouder and colleagues (2005), HBM dissects an RT distribution into three parameters, shift, scale and shape. The shift parameter has been linked to residual RTs, the scale parameter linking with the response rate and the shape parameter with post-attentive response selection (Wolfe et al., 2011). The EZ2 model estimates directly three parameters: (1) the drift rate, reflecting the quality of the match between a memory template and a search display (the *goodness-of-match*, in Ratcliff & Smith's term, 2004), (2) the boundary separation reflecting the response criterion (Wagenmakers et al., 2007), and (3) the non-decision time reflecting the time an observer encodes stimuli and executes a motor response. This conceptualisation can help articulate the correlation between the descriptive parameters from the RT distribution and those estimated by the diffusion model. For example, the role of shift in a Weibull function is to set directly a minimal threshold for responses and rules out the possibility of negative responses. This suggests an association between the RT shift and non-decision time parameters.

4.4.2 Model-based Analysis

The EZ2 model and the HBM suggest that distributional parameters reflect different aspects of search processes. The current data from the two models

inform some modification for the hypotheses posited by the stage model of information processing, which claims that different aspects of peripheral and central processing might associate with different Weibull parameters (Rouder et al, 2005). Firstly, the shift parameter was associated with the speed of peripheral processes (i.e., irreducible minimum response latency, Dzhafarov, 1992).

First, the shift parameter varied across the search tasks and the display size, a pattern that might attribute to non-decision time changes. The different tasks require very different perceptual burdens, so does the increase in the display size. These factors should reflect on the psychological processes influencing evenly all responses in a distribution, which the shift parameter measures (Chapter 2). However, a direct comparison for the ANOVA result (Table 4-2) did not show a clear one-to-one mapping in the experimental factors with the non-decision time. One possible reason for the ambiguous finding is that the EZ2 model, although enables rough estimates for the drift rate and the boundary separation, might give contaminated non-decision time. That is, the current EZ2 estimate for the non-decision time might carry the influences also from some cognitive processes. This possibility is suggested by the interaction effect. As showed in the VWR plots, the shift parameter increases monotonously (Figure 4-8) with the task and the display size factors, but this is not observed in the non-decision time (Figure 4-9). A second possible account is that the processes estimated by the non-decision time might reflect the perceptual operation been influenced by the decision-related processes. For example, in the spatial configuration search, observers engage in the cognitive operation of rotating the mental image of 2/5 digit and matching the digits against the search

template. If this cognitive operation might affect the non-decision processes, such as encoding the digits and perceptually grouping them together, the EZ2 estimate for the non-decision time might include also the task factor related to higher cognitive operations. In general, the current data support the view that the shift parameter reflects unambiguously the component in the factors that influences evenly all responses in a distribution, but whether the non-decision time also reflects the same process remind to be clarify.

To sum up, the current study do not support a direct association of the shift parameter with the speed of peripheral processes, assuming that the EZ2 non-decision time parameter is a good estimate for the peripheral processes. Nevertheless, if the EZ2 non-decision time estimate is also contaminated by non-peripheral processes, such as the perceptual grouping that rendering a drastic change at the display size 7, the shift parameter might still be a good candidate for estimating the peripheral processes. A further modelling effort may help to clarify this point.

In contrast to the shift parameter, the shape parameter showed marginal effect of the display size, a reliable effect at the task, and an interaction between these factors. The magnitude of this parameter increased monotonically with the display size for the feature and conjunction searchers but there was a U-shaped function for the spatial configuration search. This last result is consistent with there being a contribution from an emergent property of the larger configuration displays, such as the presence of grouping between the multiple homogeneous distractors leading to a change in perceptual grouping (see also Levi, 2008, for a similar argument concerning *visual crowding*). This

change in the shape parameter in the large configuration search display is in line with a sudden increase of the drift rate standard deviation (.080, .050, .054, .344). This might be just a coincidental observation across the two separate models and is ready accountable by the SERR's spatial grouping hypothesis. It might also be a true visual search phenomenon been captured by my exploratory analysis, as the Rouder et al's (2005) hypothesis predicted the association between the distributional shape and the higher cognitive processes. Presuming the results from the memory recognition paradigm is applicable to the visual search paradigm and the matching process of a template to search items is a WM operation, the current study suggests the drift rate changes might be due to the matching process of template to a search item and to the group segmentation process. Both processes underlie the change in the shape parameter.

In addition, I observed a general increase in the values of the shape parameter across the display sizes at absent trials in the spatial configuration task, $F(3, 57) = 6.13$, $p = .001$, $\eta^2_p = .244$ (1.73, 1.86, 2.05, & 1.96; 3, 6, 12, & 18). The target absent-induced shape change in the spatial configuration task was observed also in Palmer and colleagues' analysis (2011). However, their data showed no reliable shape change across display sizes for present trials (E. M. Palmer et al., 2011). Following Wolfe et al.'s (2010) suggestion, Palmer and colleagues (2011) speculated that the display size effect for the shape parameter might result from the premature abandoning of search, a view that is supported by their data showing high rate of miss errors in the spatial configuration task (Wolfe et al., 2010). The high rate of miss errors might reflect

when an observer prematurely decides to give an absent response on a target present trial. This will in turn reduce the overall number of slow responses leading to an RT distribution with low skew. This indicates that in the conditions with high miss errors, participants tended to set a low decision threshold for the target absent response. The tendency might also appear in the absent trials, resulting in correct rejection by luck, a result leading to RT distributions in the absent trials with an increase in the shape parameter. I, applying a more sensitive method under the constraint of limited trial numbers, show reliable display size effects on the RT shape in the present trials of the spatial configuration and the conjunction searches. Together with the miss error data, my data do indicate that a link between the miss errors and the shape of the RT distribution is plausible. In addition to the explanation of participants abandoning search prematurely (i.e., a dynamic changes of boundary separation), I propose another explanation that, relative to the feature search, the factor that changes the RT shape in the spatial configuration search is the *goodness-of-match* between the search template and the search items (thereby the change in the drift rate). This implies the factors contributing a change at the rate of evidence accumulation will result in shape changes for the RT distributions.

Taken together the data in the simulation study (Chapter 2), the data suggest that shape parameter does associate higher complex cognitive operations, because (1) the boundary separation is associated with the miss-error account; (2) the increase in the drift rate is associated with the *goodness-of-match* account; and (3) the emergent grouping effect is associated with the

view that an additional stage is inserted into a cognitive operation (Rouder et al., 2005). The association of a myriad cognitive operations with the shape parameter suggests also that rather than reflecting a specific cognitive operation, the shape changes depend on the paradigms.

Among the three Weibull parameters, the scale parameter showed the highest correlation with mean RTs (Pearson $r = .78$, $p = 2.20 \times 10^{-16}$), a result replicating Palmer et al.'s (2011) analysis. The high correlation should not be surprising, considering that both the RT scale and the mean RTs capture the change in the central location of RT distributions. The scale parameter estimates an overall enhancement (or reduction) of response latency as well as response variance, so do the mean and variance RTs (see a review in Wagenmakers & Brown, 2007). Unlike the mean RTs, however, the scale parameter in my dataset was not sensitive to the display size in the feature search task. A cross-examination with the boundary separation in the diffusion model appears to indicate that the scale parameter might reflect the influence of response criteria, with only the inefficient tasks showing the display size effect. This should not be taken as evidence indicating that the scale parameter is a direct index of the response criteria however; rather changes in the scale parameter are a consequence of altering the response criteria. An observer with a conservative criterion, for example, might show a general change of response latency and variance (the more reluctant to make a decision, the more variable a response will be), so the scale parameter reflects this change.

The scale-criterion account is however just one possibility, which is based on the similar ANONA finding (i.e., the significant effects for the two factors and

the interaction for the scale and the boundary separation). Because the simulation study showed the scale parameter links to two EZ2 parameters. The current study does not exclude the possibility that the scale effect is due to the change in the drift rate. A further direct test is required to explore this possibility.

4.4.3 Limitation

The analytic approach I adopted assumes that individual RTs are generated by the 3-parameter probability functions. The selection of the Weibull function is determined, on the one hand, by its probing three important aspects, the shift, scale and shape, of an RT distribution, differing from what the ex-Gaussian function describes (mu, tau, & sigma). On the other hand, I selected the Weibull function, because it permits a reliably converged posterior distribution, and has broad applications to memory and to visual search (Hsu & Chen, 2009; Logan, 1988) when being modeled in the hierarchical Bayesian framework (Farrell & Ludwig, 2008; Rouder et al., 2005).

4.5 Conclusion

In conclusion, my study shows how the HBM-based distributional analysis, complemented with the EZ2 diffusion model, can help to clarify visual search processes. The data indicate that different effects of search difficulty contribute to performance, with the effects of the search condition being distinct from effects of display size in some cases (on the drift rate and shift parameters) but not in others (effects on non-decision factors and the separation of decision boundaries). I have linked this dissociation to the involvement of distractor grouping and rejection (on the one hand) and serial selection of the target and

the setting of a response criterion (on the other). The approach goes beyond what can be done using standard analyses based on mean RTs.

Chapter 5 Updating the Template Changes the Response Threshold

5.1 Introduction

Our ability to conduct efficient visual search for a target varies greatly according to the similarity relations between targets and distractors (Duncan & Humphreys, 1989). Search efficiency can vary from minimal or even decreasing RT functions as the display size of distractors increases (in 'feature search'), to highly inefficient search when the target is defined by a conjunction of features or a spatial configuration of form elements. In some cases, though, conjunction targets can be searched efficiently – for example when the feature values defining the conjunction target and the distractors are sufficiently different – when the features are thought to 'guide search' (Quinlan & Humphreys, 1987). Here we consider the mechanisms by which top-down guidance of search to targets operates.

5.1.1 Attentional Template

Many models of visual search assume the involvement of an attentional template held in WM (Bundesen, 1990; Desimone & Duncan, 1995; Duncan & Humphreys, 1989; Wolfe, 1994), which acts to direct search to task relevant (target) items. There is good evidence that this memory representation for the target modulates visual selection. For example, Pashler (1987) instructed participants to search for either letter C or E (in Experiment 2) in a display, so participants anticipated both target letters but only one of them appeared in a search display. The distractors consisted of the salient letters X and N and the

confusable letters G (relative to C) and F (relative to E). . The former group of letters were salient, because they were less similar to the targets than the latter group of confusable letters (van der Heijden, Malhas, & van den Roovaart, 1984). In this experiment, Pashler observed a specific type of display size effect relative to the number of confusable distractors where a search display always contained 6 items. Importantly, the confusable distractors disrupted search even when it was similar to the target letter represented only in memory (i.e., been anticipated before search display, but not appear in it). At this condition, the search performance was still affected in trials with target C/E when the confusable distractors were F/G (e.g., the distractors were the letter Gs and the target was an E). Such data suggest that similarity to the memory template modulates search efficiency. Other studies showed further evidence suggesting that giving participants a memory template for the target is critical for generating efficient search. Hodsoll and Humphreys (2001), for example, examined search for a large target amongst medium and small-sized distractors. The large target 'popped out' from the search display, but only when participants had an expectation for a large target in a trial. In this instance, pop out was determined by the presence of a target template and not simply by bottom-up differences between the items. There is also evidence that a template held in WM can direct search in an involuntary fashion. Soto et al. (2005) had participants hold an irrelevant item in WM for a later matching task whilst undertaking a different search task. They found that the irrelevant item in WM still directed attention to a matched non-target item in a display – even when that item was always a distractor. More recently there has been evidence that a secondary memory

representation (i.e., accessory memory items, AMIs) can modulate the effects of the search template in WM (see reviews in Olivers, Peters, Houtkamp, & Roelfsema, 2011; Soto et al., 2008). However, the finding of the AMI influence is not entirely clear (Dowd, Kiyonaga, Egner, & Mitroff, 2015; Downing, 2000; Woodman & Luck, 2007; Woodman, Luck, & Schall, 2007).

Olivers and colleagues (2011) have argued that one key factor that determines the different findings is the status of AMIs in WM. Because in a dual-search and memory recognition-task an observer engages or prepares to engage both tasks at once, ideally, she/he needs to maintain two active WM representations. One is the search template relevant to the primary search task, and the other is the AMI, which should be held for the later memory task. Olivers et al. argued that observers can only hold one strong WM representation at a time. If this is the primary template for the search task, then search can be directed efficiently to the search target without interference from the other item in WM. This can happen, for example, when the primary search target changes from trial-to-trial (Schneider & Shiffrin, 1977), so the search target is prioritised in WM. Similarly, when several search templates are held in WM, the AMI's WM representation may degrade (Soto & Humphreys, 2008). Sometimes, the AMI may be prioritised for example when an identical search target is tested in one block of trials (Schneider & Shiffrin, 1977) or when it is no longer needs to be highly active in WM. Under these circumstances, the AMI distractor may catch attention during search. This was exactly what Olivers found in his Experiment 6 (2009). This experiment showed only in the consistent mapping condition, the AMI-related distractor prolonged search time with an 80-ms (mean) RT

difference. By contrast, the varied mapping condition showed an opposite non-significant difference (-27 ms).

This finding nonetheless appears trivial, because the error rate data suggested (1) Olivers's (2009) participants might trade speed with accuracy and (2) in his consistent mapping condition, the AMI-related condition showed fewer (insignificantly) errors than the AMI-unrelated condition. A recent correlation study even suggested that the mixed results might be due to the two seemingly identical paradigms (Soto et al., 2005; Woodman & Luck, 2007) tap into two different - visual working memory and attentional control - cognitive abilities (Dowd et al., 2015). Although both Olivers et al's (2011) and Dowd et al.'s (2015) hypotheses are derived from credible data, their experiments did not lend strong supports for the arguments. One important reason is that those studies tested only small number of trials per condition. As explained in our previously Chapter 4) and numerous works investigating RT distribution (see a review in Balota & Yap, 2011), the data of mean RT are sometimes biased and miss important information hidden in the distribution (P. L. Smith & Ratcliff, 2004). Furthermore, a seemingly opposite finding demonstrating the AMI-related interference showed higher search slope only in the varying-mapping participants (Woodman et al., 2007). This, although in my opinion incomparable finding, contradicted Olivers's data (2009). He had argued the key factors resulting in the contradictory findings are the relatedness of the AMI to the search template and their activation states. In Woodman et al.'s varying mapping participants, the AMI interfered with search performance, because it was less similar with the search template. Thus, when the WM load is high, the AMI pushed out the template (not yet fully activated) and

dominated the WM operation. As opposed to this, in Olivers's varying mapping condition, the AMI ceded the attentional control to the template, because they were related and due to the relatedness, both fully activated.

Nevertheless, current interpretations do not address the key question about search decision, but only purposed hypotheses (Dowd et al., 2015; Olivers et al., 2011; Soto et al., 2008; Woodman & Luck, 2007). Woodman et al.'s (2007) search slope data suggested the AMI interference in the varying-mapping participants associated with search efficiency, whereas Olivers's mean RT data suggested the AMI interference in the consistent-mapping condition associated with search strategy. As will be discussed later in Section 5.1.2, to decide whether a target is found, an observer collects sensory evidence with a decision rate (i.e., drift rate) and declares a target is found (e.g., by pressing a response button) when the amount of sensory evidence has reached a response criterion. The search slope data suggested changes in the drift rate (associated with efficiency), whereas Olivers's mean RT (2009), together with the reversed error rate, data suggested changes in the drift rate and/or the response criterion. Olivers's activation-state hypothesis (2009), though plausible, does not address whether changes in the activation state for a search template affects search efficiency or search strategy and which part of a search decision associates with the WM state of a template.

In the present study we set out to test the activation hypothesis (Olivers, 2009), using a search-alone paradigm to compare the consistent mapping and varying mapping conditions. Specifically, I conducted three experiments to examine how the memory consolidation of an attentional template might affect

different aspects of a search decision when (1) the template is held constant versus when it is varied from trial to trial and when (2) the duration between the presence of an template probe and that of a search display is 50 ms versus 400 ms. To attain the aim of disentangle different aspects of a search decision, I go beyond previous work in the field by using mathematical modelling to characterise specific aspects of the search process. This is achieved by using the decision-making models (e.g., the DDM, Ratcliff & McKoon, 2008), which incorporates response latency and accuracy jointly to account for two main aspects of decision making: the decision (drift) rate and the decision (response) threshold. The former suggests how fast a decision is reached as a function of sensory evidence and time. The decision rate describes the instantaneous slope of the function. That is, under the linearity assumption, the slope equals the amount of sensory evidence divided by time. The more the evidence accumulates, the higher the probability that an observer reaches a reliable decision. Thus, if search conditions generate differences in perceptual sensitivity and/or in the guidance of search to target features, there should be a corresponding change in the decision rate. The decision threshold, on the other hand, measures the (internal) response criteria, with a response triggered when sufficient sensory evidence has been accumulated to exceed the threshold. Mathematical models of decision making have been developed to formally separate effects of decision rate and decision threshold on memory recognition (Ratcliff & McKoon, 2008; Ratcliff & Smith, 2004) as well as been successfully applied to visual search (Purcell et al., 2012). Here we used such models to examine how attentional templates modulate search.

I examined visual search under two main conditions: when the identity of the target was held constant across trials and when the identity of the target varied in each trial (in both cases the target identity was pre-cued, but in the constant condition the cue was always the same letter). Targets and distractors were randomly selected letters, making search relatively difficult and likely dependent on serial selection of items (Czerwinski, Lightfoot, & Shiffrin, 1992; Fecteau & Enns, 2005; Malinowski & Hübner, 2001; see a review about whether letter identities guide search in Wolfe & Horowitz, 2004). I also varied the time from the cue to the search display (in Experiment 2 and 3), to examine the timing influence on consolidating a new search template in WM. Experiment 2 was to assess how the consolidation process of a template modulated the parameters in a decision-making model, compared with the baseline condition, in which the target and cue were constant. Experiment 3 increased the per-condition sample size, aiming to replicate the finding in Experiment 2. Details of the modelling processes are described below.

5.1.2 Linear Ballistic Accumulator

The LBA model is a simplified decision-making model which uses choice RTs (Luce, 1986) to estimate the decision rate and the decision threshold. It was adapted from a model describing multiple independent 'leaky competing accumulators', each starting from and racing towards respective initial points and response thresholds (Usher & McClelland, 2001). When one of the accumulators surpasses its threshold, a decision is made. Each accumulator underlies one decision type, for instance deciding a target is present or absent. Accumulators gather sensory evidence gradually along a temporary scale;

hence decision rates and times can be assessed. When an accumulator, inconsistent with a correct response, wins a race, an error response is committed. Take an error response in a target present trial as an example. This will result from a target-absent accumulator reaching its threshold earlier than a target-present accumulator. Accounting for choice RTs simultaneously in a single framework, the LBA model can assess the decision rate, threshold, and time, both for correct and incorrect responses.

The LBA model is a further simplification from an early ballistic model of decision making (Brown & Heathcote, 2005). The ballistic model eliminated the within-trial randomness from the model of leaky competing accumulators (Usher & McClelland, 2001) and assumed an across-trial constant decision rate. The LBA model further eliminates the nonlinearity of the evidence accumulation process (Brown & Heathcote, 2008). Although sacrificing the complexity of some models, the LBA can accommodate choice RT data collected from lexical decision and the brightness discrimination tasks (Ratcliff et al., 2004; Ratcliff & Rouder, 1998), providing excellent fits in both cases. Therefore, the LBA model appears to provide a good fit for empirical data with minimal complexity.

5.1.3 Response Time Distributions

A further advantage of modelling choice RT in a decision-making framework is that it accounts for variations in entire RT distributions across the experimental factors. In this study, I used quantile analysis, calculated from the empirical data and predicted by the two decision-making models. The quantile descriptive results were analysed together with the decision parameters

assessed by the LBA model and the DDM to understand the search decision processes associated with RT distributions.

5.1.4 Study Aims

The current study was to investigate three main questions related to the guided search. First, it tested whether an automated template influences the search decision comparing with a non-automated template. Second, it examined the effect of inter-stimulus interval (ISI) on the consolidation of WM template and third how the ISI effect influenced search efficiency as well as decision. Particularly, the study tested whether a prolonged consolidation time (400 ms vs. 50 ms) prior search will strength a non-automated template, rendering it been converted into a similar memory status as an automated template. To ameliorate the pitfall in previous studies and to account for the search decision, this study collected relatively large number of trials in each condition. This strategy may make the data viable to be processed more reliably by decision-making models, so the search decision can be assessed.

5.2 Method

5.2.1 Design and Procedure

5.2.1.1 Experiment 1

The task started with a white fixation square centred on the screen for 500 ms, followed by a 200-ms cue. The search display appeared after a 50-ms blank. A trial was terminated when a participant made a response to indicate whether a target letter was located on the left or on the right side of the display, relative to the fixation square. A message was then presented to inform the

participant if a correct response was made. If an incorrect response was made, the target location was highlighted by a yellow square circled around it. Each participant visited the lab twice to complete the fixed and varied cue conditions in two separate days in a counterbalanced sequence. Participants first completed an 8-trial warm-up block before commencing ten 48-trial testing blocks. In the fixed cue condition, all participants always searched for an identical letter (H). This target letter was pre-determined for all participants and kept constant in this condition. In the varied cue condition, a letter was selected randomly from the 13-letter pool in each trial as the target letter. This randomly selected letter was used as a cue prior a search display appeared. A schematic screen sequence is shown in Figure 5-1.

The experiment used a two-factor, display sizes (3, 5, 7 & 9) and (fixed vs. varied) cues, repeated-measures, within-participant design. The fixed cue condition presented an identical probe letter in every trial, whereas, the varied cue condition randomly changed the probe letter in every trial. Participants contributed 130 trials in each condition. In total, one participant contributed 1040 trials.

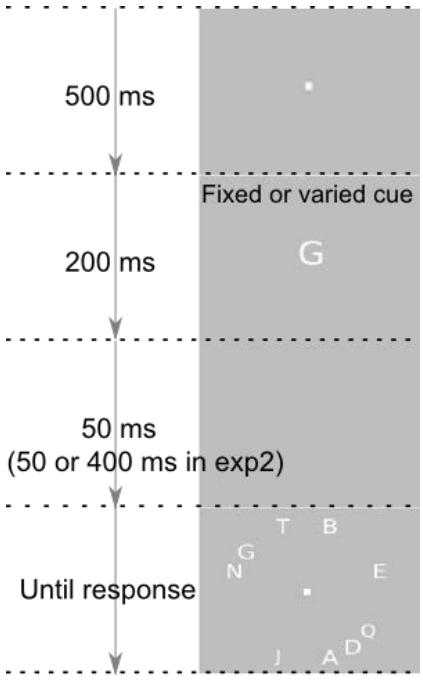


Figure 5-1. An illustration of the search paradigm. Experiment 2 tested one additional 2-level ISI factor. Except for doubling per-condition trial numbers, Experiment 3 replicated exactly Experiment 2. The search goal is to decide whether a cued letter is on the left or the right side, relative to the fixation point.

5.2.1.2 Experiment 2 and Experiment 3

In Experiment 2, an additional factor was introduced to examine the timing influence of memory consolidation (50 vs. 400 ms) on visual search. Two ISIs were randomly assigned to each trial within a block. Participants were not informed the two variants of ISIs. Experiment 2 and 3 tested only three display sizes (3, 7 & 9). Participants performed respectively 100 and 200 trials per condition (cue × display size × ISI) in Experiment 2 (1200 in total) and in Experiment 3 (2400 in total)

5.2.2 LBA Model and DDM

The LBA model describes a two-choice search decision with two accumulators, corresponding to deciding a left/right target (internally), and to initiating a left/right button press response (externally). Each accumulator

begins at a starting amount of evidence (i.e., 'starting point', *A*) and accumulates evidence at a speed which is described by a decision rate (i.e., 'drift rate', *v*). The first accumulator to reach a common response threshold (*B*) determines the overt response. The decision time is then assessed by simple algebra, $(B - A)/v$. An RT reflects the decision time plus a residual time (i.e., the 'non-decision time', t_{er}). The latter non-decision time, broadly speaking, accounts for the early perceptual and late motoric processes (Brown & Heathcote, 2008) and is presumed to be independent of the decision-making process. Accuracy is determined by the first accumulator corresponding to one of the answers for a trial. For instance, if an accumulator for the left target reaches its response threshold first in a search display containing a right target, the measured RT reflects an error response.

The measured RTs were cut at 0.2s and 2s and then fit with two specific versions of decision-making models, the DDM (Ratcliff & Tuerlinckx, 2002) and the LBA models (Heathcote & Love, 2012). Both models were assessed via the MLE method (Myung, 2003) and used the same notation, t_{er} and *v*, to refer to the parameters of mean non-decision time and of mean drift rate. As is standard practice, the diffusion model uniformly distributed variability in non-decision time with width s_t. The LBA model accumulates sensory evidence deterministically, with a drift rate that varies from trial-to-trial according to a normal distribution with mean *v* and standard deviation s_v. The diffusion model specifies these same two parameters (*v* & s_v) with a normal distribution that gives rise to a random sample of drift rate for each trial. Within each trial, evidence accumulates on average at the speed given by the drift rate sample, and a moment-to-moment random

variability. This variability sets a standard deviation *s*, I conventionally fix at 0.1, which ensures model identifiability (this can also be fixed at 1, see Voss & Voss, 2008).

As the drift-diffusion and LBA models differ in their accumulator structure, they describe the parameters with different symbols. The DDM represents the distance between the (positive & negative) thresholds, with parameter *a*. The starting point for accumulation, sometimes denoted *z*, was estimated by its relative position between the thresholds, denoted $Z = z/a$. Both models assume uniformly distributed variability in the accumulation starting point. Thus, the implementation of a starting point in the DDM varies around *z*, with the variability, *SZ*. In the case of the shorter of the distances from *z* to the threshold, $SZ = \min(z, a-z)/a$. The starting points of the LBA are uniformly distributed between zero and an upper bound denoted *A*. The distance from *A* to the threshold is denoted as *B* (see Figure 1 in Donkin, Brown, & Heathcote, 2009 for a graphic comparison for the evidence accumulation structure between the two models).

Table 5-1. The top-level LBA and the drift-diffusion models. See text in the next paragraph for the meanings of each mathematical symbol.

	LBA model	Drift-diffusion model
Experiment 1	$\begin{cases} B = lR \times Q \\ v = S \times Q \times N \times M \\ sv = M \\ A = 1 \\ t_{er} = N \\ pc = 1 \end{cases}$	$\begin{cases} a = Q \\ v = S \times Q \times N \\ sv = 1 \\ Z = Q \\ SZ = 1 \\ t_{er} = N \\ st = 1 \\ pc = 1 \end{cases}$
Experiment 2 & 3	$\begin{cases} B = lR \times Q \times I \\ v = S \times Q \times N \times M \\ sv = M \\ A = 1 \\ t_{er} = I \times N \\ pc = 1 \end{cases}$	$\begin{cases} a = Q \times I \\ v = S \times Q \times N \\ sv = 1 \\ Z = Q \times I \\ SZ = 1 \\ t_{er} = I \times N \\ st = 1 \\ pc = 1 \end{cases}$

The drift-diffusion model was fit to all variants of the factor combinations, from the most to the least complex (where all parameters were equal across all conditions), resulting in 64 models to be analysed for each participant's data. The LBA model, likewise, resulted in 128 models per participant. It did not consider variants in which the M factor (matching vs. mismatching accumulator) was dropped for the *v* parameter (without this, the model is forced to predict chance accuracy). Model variants were fitted starting from the simplest, with the best fits of simpler models providing starting points for fitting more complex models (Donkin et al., 2011).

For all three experiments, I began with a complex, top-level model to fit the data as specified in Table 5-1. The decision threshold (*B* in the LBA and *a* in the

drift-diffusion models) varies with the cue factor (*Q*). The LBA model, because of modelling two separate accumulators, fits also the latent response (*IR*) factor, which corresponds to the two accumulators. *IR* acts like a two-level experimental factor. The drift rate (*v*) varies with all possible factors: stimulus (target) location (*S),* cue (*Q*), display size (*N),* and a scaling factor reflecting the matching/mismatching of an accumulator with correct answer (*M*). That is, when the left-sided target accumulator (lR-left) reaches the threshold first and a trial contains a right-sided target, the M factor will indicate a mismatched accumulator (hence its drift rate) is chosen; likewise, for the case of matched accumulator. The between-trial standard deviation of the drift rate (s_v) varies only with the scaling factor (*M)*, presuming constant variation across the two accumulators (Brown & Heathcote, 2005). That is, the accumulators' *sv* does not depend on other experimental factors, such Q, N or S. This is achieved in the DDM with an intercept. The starting point of an accumulator (*A*), residual time (t_{er}) and the contamination factor (*pc*), presumed to be invariant with the experimental factors (*Q*, *N*, & *S*) and the accumulator matches (*M*), were modelled as intercepts. Due to the different accumulator structure in the DDM, its starting point (*a*) was modelled with the cue factor. Note that for both models the ensemble of the parameters forms one probability/cumulative density function, so the equations were fitted as one function to the data).

For Experiment 2 and 3 I added an ISI factor to a similar complex model. The ISI factor (*I*) was added on the decision threshold and the residual time, but not the drift rate. This decision was based on an assumption that the long ISI increases the likelihood of response preparation and memory consolidation of

the target template, but not the quality of the match between a template and a search display (associated with the decision rate). A similar rationale was also applied to the DDM.

The complex models were subjected to a model selection process. The minus log-likelihoods of all possible models (different combinations of experimental factors) were calculated, using the MLE. The most probable model was then acquired via the MLE, based on it accounting for the highest variation with minimal factors, with selection based on the lowest Akikie Information Criterion (AIC) and Bayesian Informative Criterion (BIC) values aggregated over participants. This method of model selection provides a good trade-off between goodness-and-fit and model complexity, as measured by number of parameters. The trade-off is measured differently by the two criteria (see Burnham & Anderson, 2004), with AIC tending to select more complex models than BIC.

5.2.3 HBM

In addition to applying the decision-making models, the study used a Weibull probability function to describe RT distributions. Mathematically and theoretically, the approach, compared to using the Gaussian probability function, provides a more liberal and realistic perspective to describe RT distributions. It also exploits the descriptive nature of a probability function, permitting an intuitive impression about how RT distributions changes associated with experimental manipulations. See previous Chapter 2 and Chapter 4 for more details.

5.3 Result

The result section presents the traditional analyses for the mean correct RTs and accuracy rate, which were followed by the analyses of model fitting and RT distributions.

5.3.1 Traditional Analysis

5.3.1.1 Experiment 1

The data-trimming scheme resulted in a 2% rejection rate, removing those trials with RTs less than 0.2s and greater 2s RTs. The RTs were then averaged across trials within a condition, resulting in 320 mean RTs (20 × 4 × 2 × 2; participants × display sizes × fixed/varied cues × left/right stimulus). The mean RTs then were subjected to ANOVAs, which reported the factors showing significant differences. Only when the sphericity assumption was untenable did I report p values corrected with the Greenhouse-Geisser method. The degrees of freedoms are reported as their original values before the correction.

For the mean RTs, the three-way ANOVA showed reliable effects of display size, $F(3, 57) = 183.7$, $\eta^2_p = .906$, $p = 3.23 \times 10^{-12}$, (GG-corrected) and cue, $F(1, 19) = 8.27$, $\eta^2_p = .303$, $p = 9.67 \times 10^{-3}$ (fixed vs. varied, 524 ms vs. 552 ms), and two interactions: cueing × display size, $F(3, 57) = 5.19$, $\eta^2_p = .215$, $p = 3.05 \times 10^{-3}$ and a stimulus × display size interaction, $F(3, 57) = 3.04$, $\eta^2_p = .138$, $p = 3.64 \times 10^{-2}$. The more items a search display contained (3, 5, 7, vs. 9), the slower participants responded (438, 502, 564, vs. 649 ms). The cue × display size interaction was due to a stronger display size effect for the varied cue than the fixed cue condition. The RT differences between the two cue levels were 12, 24, 40, and 38 ms, going from the small to the large display sizes (Figure 5-2). Post-

hoc t tests indicated that the fixed cue resulted in quicker RTs than the varied cue at display sizes 5, 7 and 9, t(19) = 2.76, 3.23, 2.65, ps = .013, .004, 016, but not at display size 3, t(19) = 1.61, p = .125. Participants responded slightly faster towards the right target – though the effect was marginal, t(19) = 1.96, p = .065 (657 ms vs. 689 ms) in the varied cue condition when the display size was large (9).

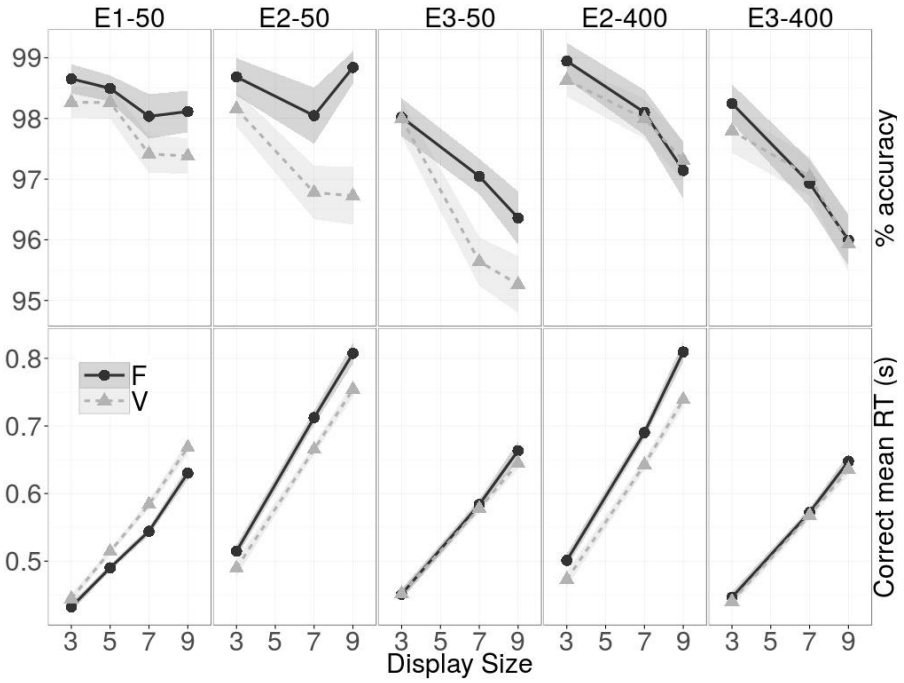


Figure 5-2. Mean correct RTs and percent accuracy in Experiments 1, 2 and 3. No display size 5 was tested in Experiments 2 and 3. Error ribbons are drawn at ±1 SE.

For the accuracy rate data, the ANOVA showed a reliable display size effect, $F(3, 57) = 5.53$, $\eta^2_p = .225$, $p = .005$ (3, 5, 7, vs. 9; .985, .984, .977, vs. .977) and a marginal cue effect, $F(1, 19) = .355$, $\eta^2_p = .158$, $p = .07$ (F vs. V; .983 vs. .978). Two marginal interactions were also observed: display size × stimulus, $F(3, 57) = 2.95$, $\eta^2_p = .134$, $p = .06$ and the three-way interaction, $F(3, 57) = 2.54$, $p = .073$. Participants tended to respond more accurately towards a right target, when the cue was unchanged. The average data showed no obvious speed-

accuracy trade-off.

5.3.1.2 Experiment 2

Experiment 2 used a similar protocol to analyse the average data. The RTs were averaged across trials, resulting in 456 mean RTs (19 × 3 × 2 × 2 × 2; participants × display sizes × cue × ISIs × left/right stimulus).

For the mean RTs, the ANOVA showed a reliable display size effect (3, 7 vs. 9; 495, 678, vs. 777 ms), $F(2, 36) = 148.12$, $\eta^2_p = .892$, $p = 1.65 \times 10^{-10}$ (G-G correction), an effect of the ISI (short vs. long; 657 ms vs. 643 ms), $F(1, 18) = 15.23$, $\eta^2_p = .458$, $p = 1.04 \times 10^{-3}$, and the cue (fixed vs. varied; 673 ms vs. 627 ms), $F(1, 18) = 10.42$, $\eta^2_p = .367$, $p = 4.67 \times 10^{-3}$. Participants responded marginally faster when the target letter was on the left side (639 ms) than on the right side [661 ms; $F(1, 18) = 4.19$, $\eta^2_p = .189$, $p = 0.06$]. There were two two-way interactions, for display size × cue, $F(2, 36) = 5.92$, $\eta^2_p = .247$, $p = .01$ (G-G correction) and display size × stimulus, $F(2, 36) = 4.27$, $\eta^2_p = .192$, $p = .04$ (G-G correction). The former interaction again was due to the increasing differences between the fixed and varied cue conditions as the display size increased. This time, however, a stronger effect of display size was observed in the fixed cue condition. The interaction of display size × ISI only reached a marginal level, $F(2, 36) = 2.82$, $\eta^2_p = .135$, $p = .08$ (Figure 5-2).

Separate ANOVAs at each display size showed a reliable cue effect at all display sizes, $F(1, 18) = 6.05, 10.78, 10.31$; $\eta^2_p = .251, .375, .364$, ps = .02, .004, 005. The RT differences between the fixed cue and varied cue conditions, respectively, at display sizes 3, 7 and 9, were 27, 47, and 62 ms. The ISI effect were observed at the display size 3 and 7, Fs(1, 18) = 29.57, 13.28; η^2_p

= .622, .427 ps = 3.64 × 10-5, .002. When the display sizes were averaged across, separate ANOVAs at each percentile indicated significant cue × ISI effect [until 0.4 percentiles, Fs(1, 18) = 6.78, 11.02, 7.83, & 7.34, ps = 0.02, 0.004, 0.01, & 0.01, η^2_p = 0.27, 0.38, 0.30, & 0.29]. A percentile-by-percentile comparison suggests the interaction is due to significant ISI differences only in the varied cue condition in the leading edge of the RT distribution.

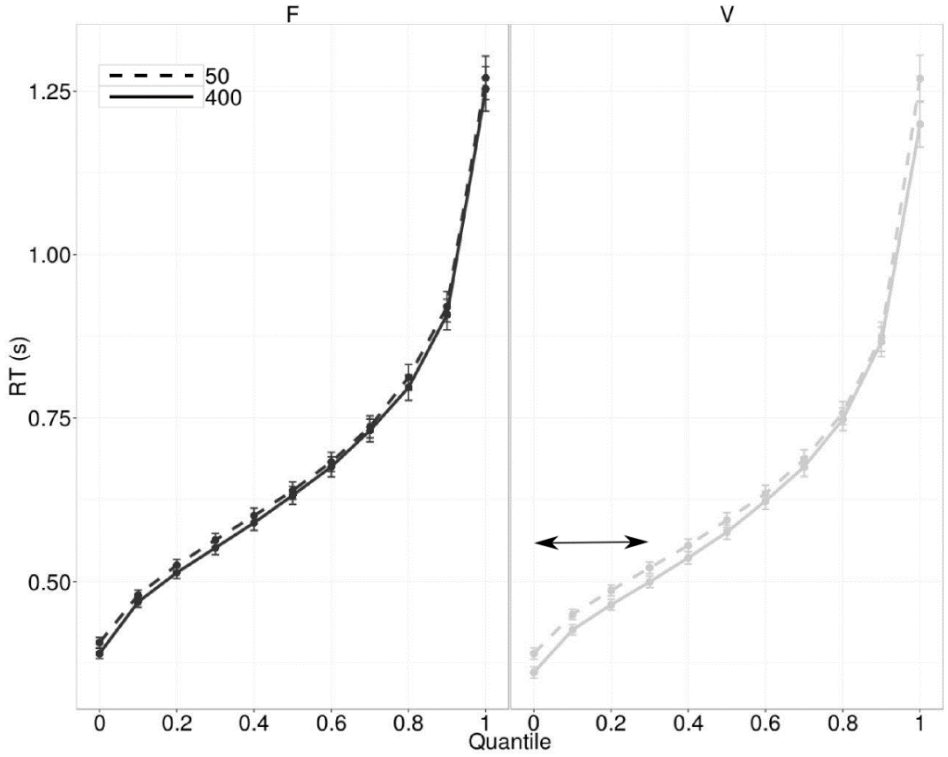


Figure 5-3. The percentile plot shows the small ISI effect in Experiment 2 is due to the RT differences at the early percentiles in the varied cue condition. Error bars were drawn at ±1 SE.

For the accuracy data, the ANOVA showed reliable effects of display size, F(2, 36) = 5.84, η^2_p = .245, p = .015 (G-G correction), cue, F(1, 18) = 7.89, η^2_p = .305, p = .012 and stimulus, F(1, 18) = 16.46, η^2_p = .478, p = .0007. Participants showed a tendency to respond better towards the right targets, comparing to the left targets (.984 vs. .975).

There were two two-way interactions: stimulus × display size, F(2, 36) =

4.73, ,η^2_p = .208, p = .016, and ISI × cue, F(1, 18) = 9.85, η^2_p = .354, p = .0057. The stimulus × display size interaction is due to a larger advantage when responding for a right target in a 7-item display (1.59% increase; 1.16% increase in the marginal effect in Experiment 1), comparing to other display sizes (in average 0.48% increase). The ISI × display size interaction reached only marginal difference, F(2, 36) = 3.52, η^2_p = .164, p = .052.

The ISI × cue interaction stems from a reliable difference in the two cue conditions at the short ISI (fixed vs. varied; .985 vs. .975), but not at the long ISI (fixed vs. varied; .981 vs. .980). Accuracy was higher for the fixed target than for the varied target. The effect was mainly due to significant accuracy differences at display size 9, F(1, 18) = 9.31, η^2_p = .341, p = .007 and at display size 7, F(1, 18) = 6.30, η^2_p = .259, p = .002 when the ISI was 50 ms.

5.3.1.3 Experiment 3

Experiment 3 replicated Experiment 2, thereby using an identical protocol to analyse the average data, with 480 mean RTs (20 × 3 × 2 × 2 × 2; participants × display sizes × cue × ISIs × left/right stimulus).

For the mean RTs, the ANOVA showed a reliable display size effect (3, 7 vs. 9; 447, 575, vs. 648 ms), F(2, 38) = 140.71, η^2_p = .881, p = 1.01 × 10^{-10}, an effect of the ISI (short vs. long; 562 ms vs. 552 ms), F(1, 19) = 7.66, η^2_p = .287, p = 1.22 × 10^{-2}. The cue factor is not significant (fixed vs. varied; 561 vs. 553) nor the interactions was found significant.

For the accuracy data, the ANOVA showed reliable effects of display size, F(2, 38) = 13.84, η^2_p = .42, p = 4.65 × 10^{-4}. The cue factor showed only marginal effect, F(1, 19) = 4.04, η^2_p = .175, p = .06 (fixed vs. varied; .971

vs. .966). The stimulus × display size interaction replicated the finding in Experiment 2, $F(2, 38) = 11.35$, $\eta^2_p = .208$, $p = 5.77 \times 10^{-4}$. The ISI × cue only reached marginal level, $F(1, 19) = 3.48$, $\eta^2_p = .155$, $p = .08$. The marginal ISI × display size interaction found in Experiment 2 was not significant in Experiment 3. One three-way interaction, cue × ISI × display size, was found marginally significant, $F(1, 19) = 3.10$, $\eta^2_p = .14$, $p = .06$. This interaction is due to the display size dependent differences between the varied and fixed cue, which observed only in short ISI.

As for the slope, when fitted using linear regression models, the search slopes across the three experiments did not differentiate fixed and varied cue condition ($p > .3$).

Table 5-2. Summary table for search slope. The slopes were calculated based on simple linear regression model, using display sizes regressed on mean RT. Slope comparisons between the two cue conditions were conducted using a null model, presuming that two lines were parallel, comparing to an alternative model, presuming that the cue factor interacts with the display size factor.

	Cue	50-ms ISI	400-ms ISI
E1	F	32 ms/item	NA
	V	37 ms/item	NA
E2	F	49 ms/item	50 ms/item
	V	44 ms/item	44 ms/item
E3	F	35 ms/item	33 ms/item
	V	32 ms/item	32 ms/item

Overall, the participants in Experiment 1 (538 ms) responded quicker than those in Experiment 2 (650 ms), but relative to Experiment 1, the

participants in Experiment 3 slowed only slightly (557 ms). Participants in Experiment 2 showed a reversed cue effect at mean RTs, though this effect did not suppress significant level in Experiment 3 when per-condition observation was doubled.

The accuracy advantage for the fixed condition consistently emerged across the three experiments, but reduced to non-significant difference in the condition of 400-ms ISI. This suggests that a long ISI might allow one to consolidate the varied, trial-by-trial updated search template and to strengthen it to a similar level as a consistently-mapped template (i.e., fixed condition). Nevertheless, when considering with the data of accuracy rate, one may argue that the reverse pattern in the mean RT suggests that participants might trade speed with accuracy (or vice versa). I suggest that this is a matter of adjusting decision threshold and decision rate (Forstmann et al., 2008). In the following, I used the two decision-making models – LBA and DDM – to approach an optimal fit solution for the data and propose an account for how the WM strength of a search template relates to the decision parameters.

5.3.2 Model Fitting

All possible models with different combinations of factor levels were fitted with MLE and evaluated by their BIC, AIC and minus log likelihood values. The detailed procedure of model selection was described in Donkin, Brown and Heathcote (2011). These models differed in how the various experimental manipulations influenced the parameters. The LBA model allowed for differences between accumulators corresponding to left vs. right responses in the parameters of the starting point (*A)* and decision threshold (*B*). It fitted also a mathematical modelling factor, the latent response (*IR*), which captures response bias (e.g., a bias to respond left by having a lower *B* for the left accumulator). The *IR* effects

on the *A* and *B* parameters in the LBA model are analogous, respectively, to the *Z* and *SZ* parameters in the diffusion model.

The LBA model also varied the drift rates, depending on the match between a stimulus and an accumulator. For example, the mean drift rate for a left stimulus could be higher for the left (matching) accumulator than the right (mismatching) accumulator. Changing parameter values across the matching factor (denoted *M*) allows the LBA model to capture the effect of the stimulus on accuracy depending on the difference between the matching drift rate and the mismatching drift rate. In the diffusion model there is no corresponding factor, as the diffusion drift rate is analogous to the difference between matching and mismatching LBA drift rates. The between-trial standard deviation of drift rate (s_v) in the LBA was set at 1 for the mismatching accumulator, a minimal assumption that makes the model tractable (Donkin et al., 2009).

Table 5-3 presents the different versions of the diffusion and LBA models by indicating the factors that affect each parameter. Where a parameter might be influenced by more than one factor I fit all possible orders of interactions among the factors as well as the main effects.

Table 5-3. The model selection table. The table shows model selection for the three experiments. Q, S, N, and I refer to the cue, the stimulus, the display size, and the ISI factors, respectively. The diffusion parameters, a, v, s_v, Z, SZ, t_{er}, and s_t stand for, respectively, the distance of boundary separation, the mean drift rate, the standard deviation of drift rate, the relative position of the starting point, the variability of the starting point, the non-decision time, and its standard deviation. The contamination factor, pc, is not shown in the table, because it is presumed invariant with the experimental factors. The response bias (IR for latent responses) and the accumulator parameter of match/mismatch (M) are specific to the LBA model. k stands for the total number of model parameters per participant. Dev. is the abbreviation for deviance, a goodness-of-fit

measure. Dash signs (-) signify a common value was estimated for all conditions.

	DDM	*A*	*v*	s_v	*Z*	*SZ*	t_{er}	s_t	k	Dev.	AIC	BIC
Exp. 1	Top-level	Q	S, Q, N	-	Q	-	N	-	27	-28993	-27913	-23625
	AIC	Q	Q, N	-	Q	-	N	-	19	-28742	-27982	-24964
	BIC	Q	N	-	-	-	-	-	11	-28194	-27754	-26007
Exp. 2	Top-level	Q, I	S, Q, N	-	Q, I	-	I, N	-	29	-18244	-17142	-12717
	AIC	Q, I	Q, N	-	Q	-	I, N	-	21	-18032	-17234	-14030
	BIC	Q	N	-	-	-	N	-	12	-17310	-16854	-15023
Exp. 3	Top-level	Q, I	S, Q, N	-	Q, I	-	I, N	-	29	-54701	-53541	-48455
	AIC	Q, I	S, Q, N	-	I	-	I, N	-	27	-54627	-53547	-48812
	BIC	Q	N	-	-	-	I, N	-	15	-53820	-53220	-50589
	LBA	*B*	*v*	s_v	*A*		t_{er}		k	Dev.	AIC	BIC
Exp. 1	Top-level	lR, Q	S, Q, N,M	M	-		N		39	-29292	-27732	-21538
	AIC	lR, Q	N, M	M	-		N		15	-28987	-28387	-26004
	BIC	Q	N, M	M	-		-		13	-28926	-28106	-26042
Exp. 2	Top-level	lR, Q, I	S, Q, N,M	M	-		I, N		36	-19345	-17977	-12484
	AIC	lR, Q, I	N, M	M	-		N		18	-19104	-18420	-15674
	BIC	lR, Q	N, M	M	-		I		14	-18929	-18397	-16261
Exp. 3	Top-level	lR, Q, I	S, Q, N,M	M	-		I, N		40	-57114	-55514	-48499
	AIC	lR, Q, I	Q, N, M	M	-		I, N		28	-56886	-55766	-50855
	BIC	lR, Q	N, M	M	-		I		14	-55586	-55026	-52570

The top-level[26] models that I fit were parameterised to test two questions. First, did the cue factor affect the decision threshold, the drift rate or both? In the diffusion model this implies that both the parameters, *a* and *SZ*, are a function of the Q factor. Likewise, the *B* parameter in the LBA model is a function of the Q and IR factors. Secondly, I asked whether the display size factor affects the mean drift rate, the mean non-decision time, or both. In both models this implies the *v* and t_{er} parameters are a function of the N factor. I also assumed that the stimulus (S) factor affected only the mean drift rate (*v*). In the LBA model I allowed the *v* and s_v parameters to be a function of the match (M) factor that is unique to that model.

5.3.2.1 Experiment 1

Figure 5-4 shows the fit of the top-level models to the accuracy rate and the RT distribution for correct RTs. The RT distribution is represented by the 10th, 50th and 90th percentiles. The figure reflects the deviance of fit reported in the 'Dev.' column in Table 5-3. The smaller a deviance value is, the better a model fit to the data. In general, the LBA model fitted better than for the DDM in terms of the deviance; however, when one considers the BIC column for the top-level models, the DDM is superior to the LBA model. One possible reason for the LBA advantage at the deviance is that, comparing to the DDM, the LBA adds two model factors (*M* and *IR*). More about this point will be discussed in Section 5.4.6. Contrasting the model fits in Figure 5-4, the LBA model performed better than the DDM at fitting the error rate data. That is, the line with the hollow circles (i.e., the prediction of the LBA model) traces closer along the error ribbons than the line

[26] That is, the least constrained or most complex model.

with the hollow triangles (i.e., the prediction of the DDM; see a similar finding in Van Zandt, Colonius, & Proctor, 2000 for the perceptual matching task). Both models provide a very good account of the RT distribution for correct responses, although the LBA model tended to underestimate the 90th percentiles, particularly for large display sizes.

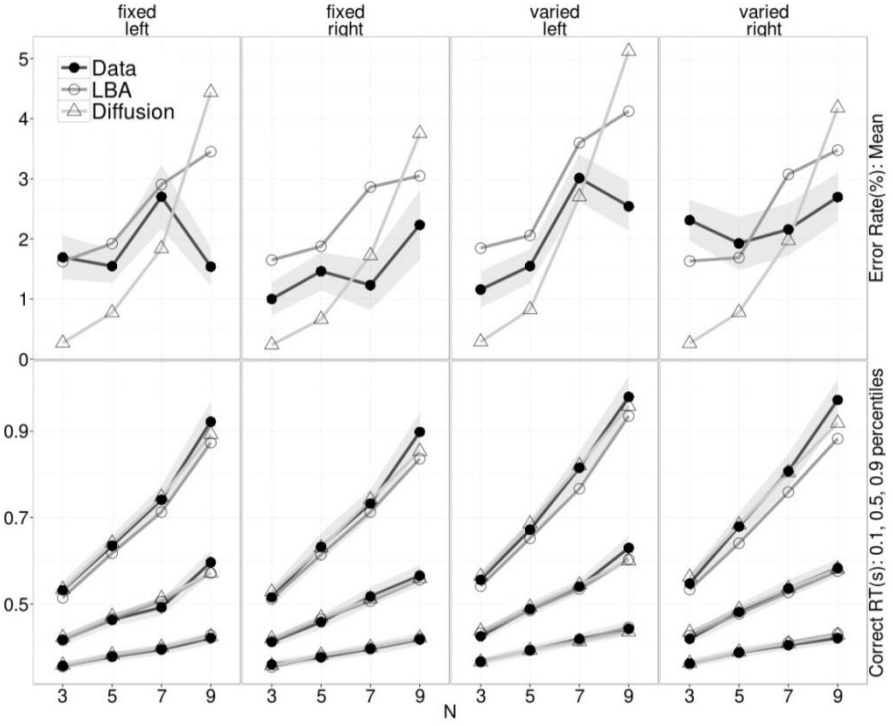


Figure 5-4. Experiment 1 top-level model fits for the correct RTs and error rates. N in the x axis stands for the display size. The figure shows clearly the advantage of the LBA model over the DDM when accounting for the error rates when the per-condition trial number is limited (~200).

The final AIC and BIC models shown in Table 5-3 were selected based on the lowest AIC and BIC values (aggregated over participants) amongst all the tested models (not shown in the tables). The BIC-based model answers unanimously the two questions posed above. That is, the cue factor selectively influences the decision threshold (*B*). To put it differently, updating a template (by cueing a new target on each trial) modulates the amount of information required

to initiate a decision, but not the response bias (*lR*) or the drift rate (*v*). The LBA AIC model only adds response bias to the decision threshold, so does not change interpretation of the first question. An identical interpretation is also acquired in the drift-diffusion BIC model (i.e., *Q* influences *a* and *Z* only).

As regards the second question, the model fits show that the *N* factor affects only the mean drift rate but not the mean of the non-decision time. However, the AIC model found that the display size effect accounted also for the mean of the non-decision time. This influence, at best, was small. In the top-level model, it explains only 11% (0.023s) of the increase in mean RT (0.211s) and the same is true for the LBA AIC model.

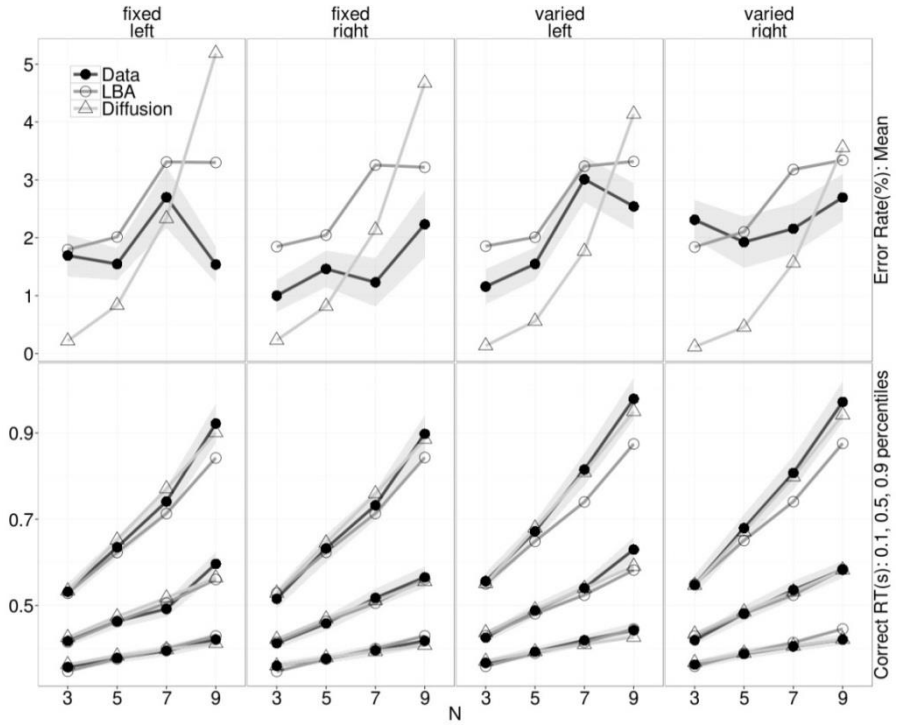


Figure 5-5. BIC model fits to Experiment 1 data.

As shown in Figure 5-5, the DDM drift rate accurately accommodated the display size effect on RT distributions and performed better than the LBA model

especially at the 90% percentile. However, the good performance of the DDM for the RT distributions appears to be at the cost of the account of error rates. It underestimates the error rate data at the small display size and over-predicts it at the large display sizes. Note that this failure of the diffusion model is not much improved in the top-level model, which, because of including more factors, should accommodate the data better than the less-complex BIC model.

In summary, the observers in Experiment 1 responded quicker towards right than left targets and showed a substantial increase in the decision rate at the display 7. Their decision threshold is lower for the fixed cue condition than for the varied cue condition.

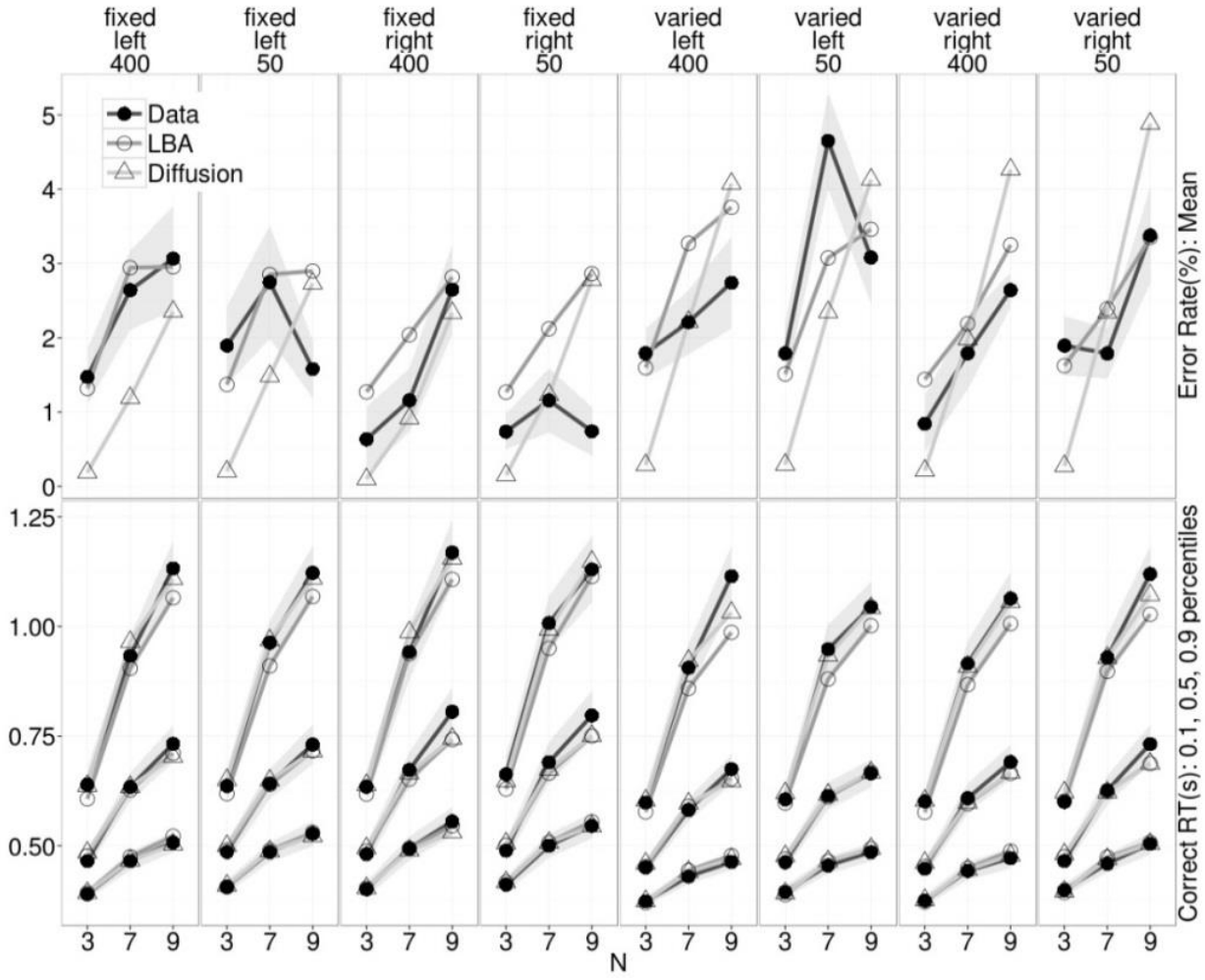


Figure 5-6. Top model fits to Experiment 2 data.

5.3.2.2 Experiment 2

Experiment 2 asked a further question of whether the ISI (I) factor influences the decision threshold (*B*) and the non-decision time (t_{er}). Likewise, for the DDM, whether the I factor affects *a*, *Z* and t_{er}. Model selection confirmed most of the effects in Experiment 1, except for a response bias effect (*IR*) accounting also for the decision threshold in the LBA BIC model. The ISI effect was associated with the decision threshold in the AIC model, whereas in the BIC model, it was associated with the non-decision time (Table 5-3). The deviance in Table 5-3 confirmed that the top-level model accounted for more variations than the BIC and the AIC models. Because the BIC models strike a good balance between parsimonious number of factors and good fits and because the display size effect in t_{er} in the AIC model was rather implausible, with non-decision time (t_{er}) decreasing as N increased, I focus on interpreting the LBA BIC model.

In contrast to Experiment 1, the cue effect shows an opposite relation between the two levels with F > V (0.92 vs. 0.83). Because, relative to Experiment 1, the only experimental difference is the random ISI pattern, the speed-accuracy-tradeoff pattern seen in Figure 5-2 likely reflects an adjustment of decision threshold, attributable to the ISI factor.

As in Experiment 1, the drift rate was affected strongly by the display size (*N*) and the matching factor (*M*). The two factors interact and this interaction explains the small decrease in accuracy with N and the large increase in RT (see Figure 5-2). The N × M interaction at the drift rate reflects also the displays size × stimulus interaction in the data of mean RT and accuracy rate. Specifically, the observers reached a decision faster for a matched accumulator (i.e., a correct

response) than a mismatched accumulator (i.e., an incorrect response). The tendency matches with the faster and less accurate responses towards left-target displays and with the slower and more accurate responses towards right-target displays.

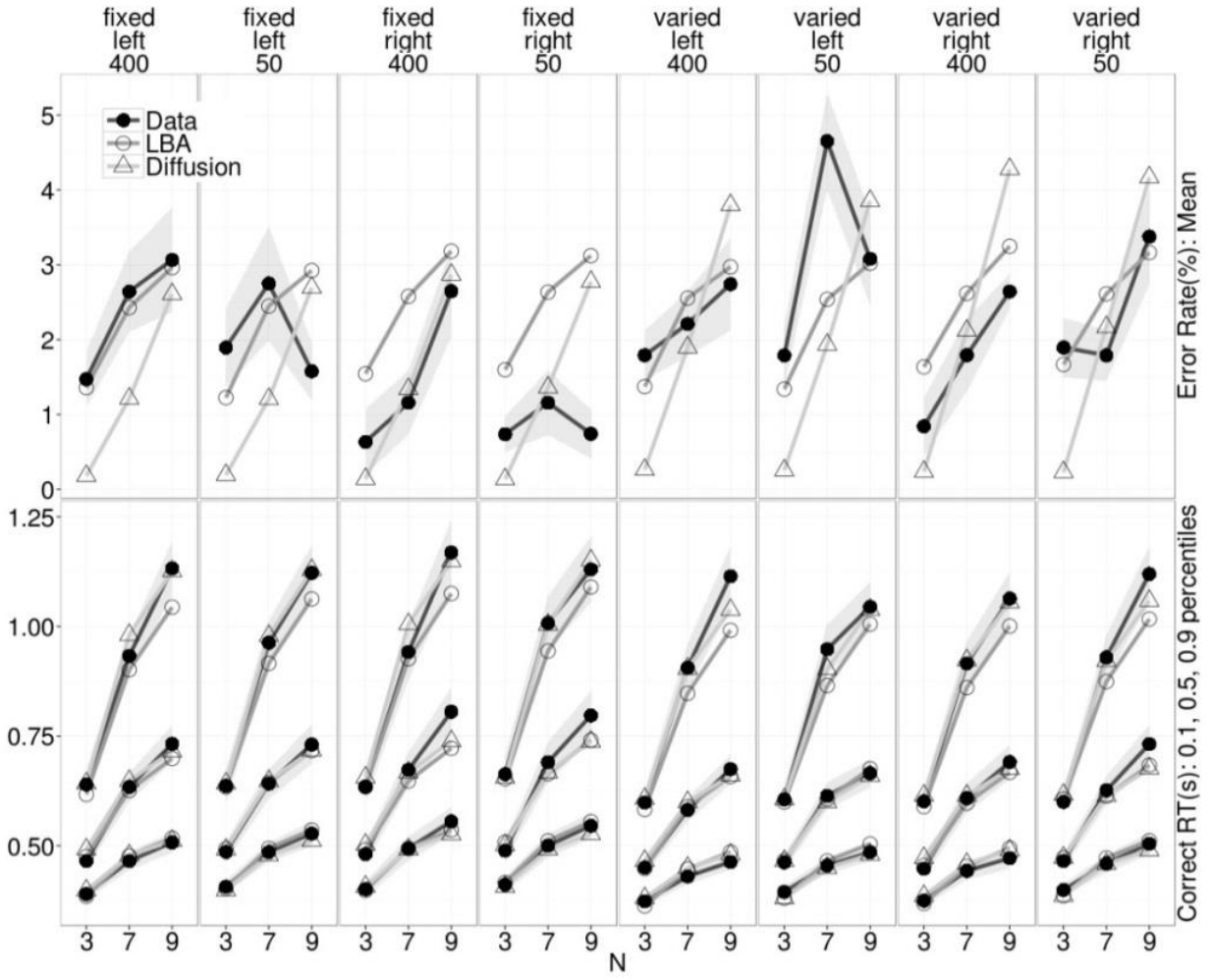


Figure 5-7. BIC model fits to Experiment 2 data.

5.3.2.3 Experiment 3

Experiment 3 aimed to test if the factors in Experiment 2 are reliable when per-condition trial numbers are increased. The deviances in Table 5-3 showed an identical pattern as previous experiments. The top-level model accounted for more variations than the AIC model, which in turn accounts for more variations than the BIC model. This pattern is observed in both LBA and diffusion Models. The BIC model uses the least number of parameters and predicts both the error rate data and RT distribution closer.

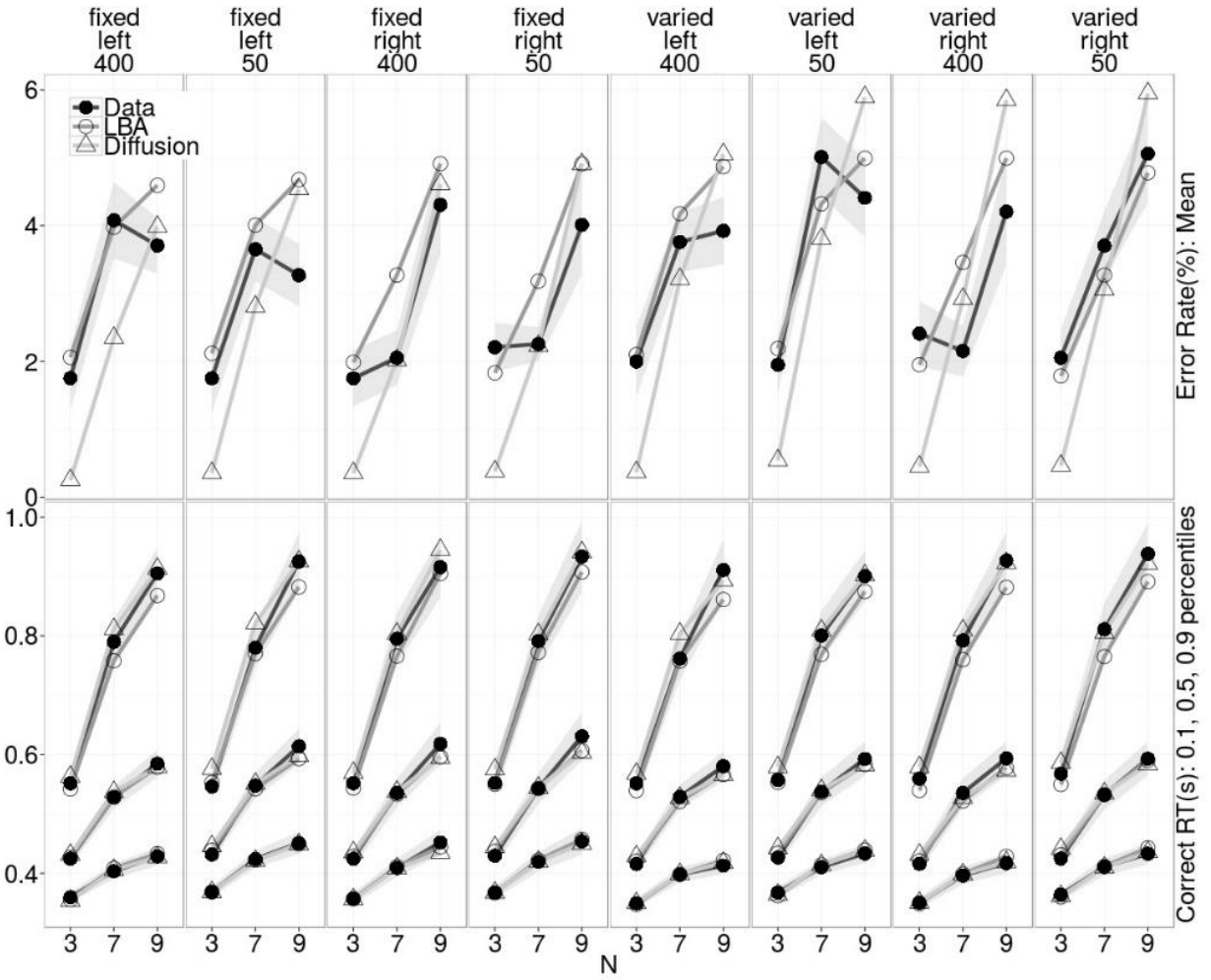


Figure 5-8. Top model fits to Experiment 3 data

The BIC model in essence replicates the BIC result in Experiment 2. The AIC model is slightly different this time. It shows that the cue factor becomes significant in decision rates (v) and that the ISI factor in non-decision time (t_{er}). This likely results from the increase of per-condition trials, making small effects more reliable. As before LBA fits better than DDM both in BIC and AIC models. The following we focused on the BIC ANOVA for the LBA fit.

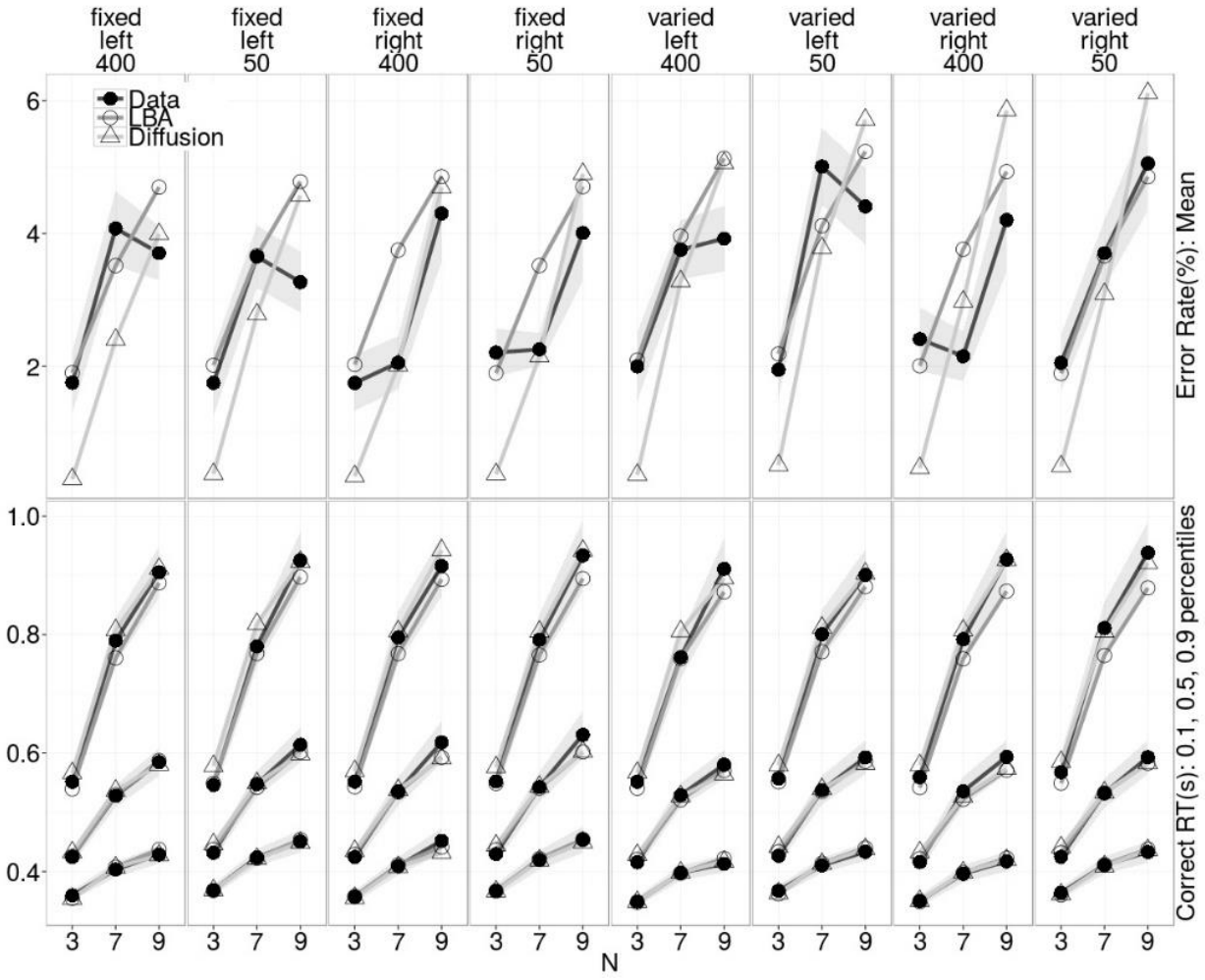


Figure 5-9. BIC model fits to Experiment 3 data.

Experiment 3 showed very similar thresholds for the two cue levels F vs. V (0.55 vs. 0.54). The cue-effect dissipation likely reflects the influence of the increase in per-condition trials, rendering an increase in template strength and thereby the decision confidence. The increases in the template strength and the decision confidence in turn result in the less pronounced threshold adjustment.

5.3.3 Distributional Analysis

In this section, the HB Weibull parameters were contrasted using ANOVAs.

5.3.3.1 Experiment 1

The two-way ANOVA showed a reliable display size effect for both the shift parameter, $F(3, 57) = 4.42$, $\eta^2_p = .189$, $p = .007$, and the scale parameter, $F(3, 57) = 35.24$, $\eta^2_p = .650$, $p = 5.15 \times 10^{-13}$. For the shape parameter, the ANOVA showed a reliable display size effect, $F(3, 57) = 16.04$, $\eta^2_p = .458$, $p = 1.12 \times 10^{-7}$, and a marginal cue effect, $F(1, 19) = 3.35$, $\eta^2_p = .150$, $p = .083$. No interaction was observed.

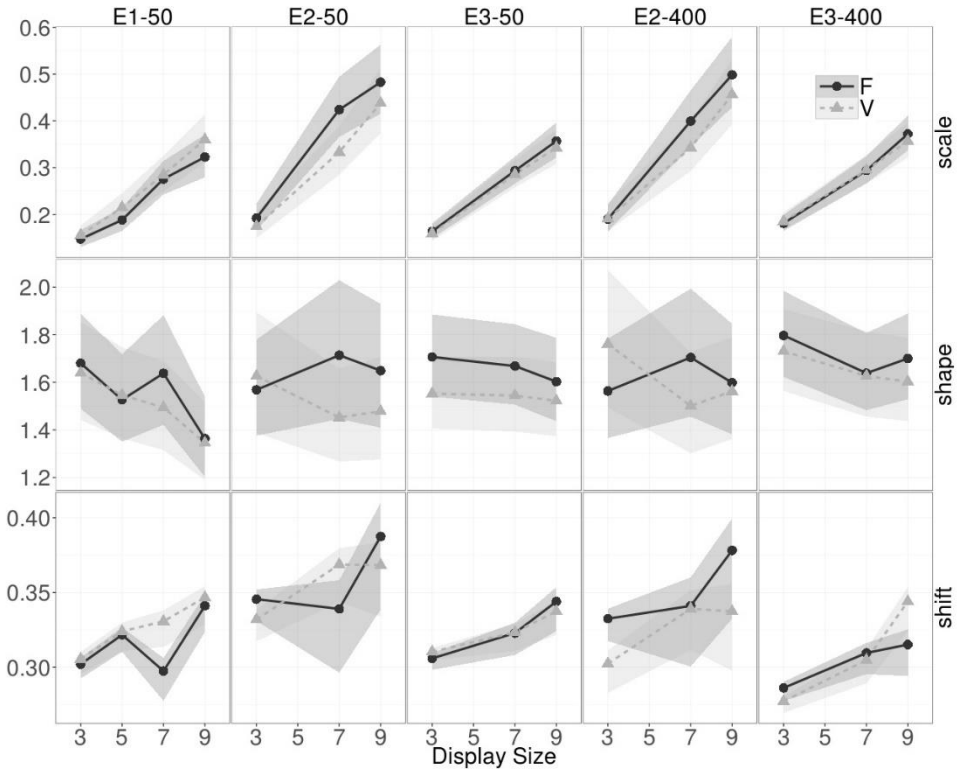


Figure 5-10. The line plots for the Weibull parameters. Note that the error ribbons were drawn based on the credible intervals estimated by the BUGS model.

5.3.3.2 Experiment 2

The two-way ANOVA for the shift parameter showed a reliable display size effect, $F(2, 36) = 4.27$, $\eta^2_p = .192$, $p = .022$ and a marginal display size ×cue interaction, $F(2, 36) = 2.96$, $\eta^2_p = .141$, $p = .06$. For the shape parameter, the ANOVA showed reliable interactions of display size × cue, $F(2, 36) = 7.94$, $\eta^2_p = .306$, $p = .001$, and cue × ISI, $F(1, 18) = 7.44$, $\eta^2_p = .293$, $p = .014$. A marginal cueing effect, $F(1, 18) = 3.89$, $\eta^2_p = .178$, $p = .06$, was also observed. For the scale parameter, the ANOVA showed reliable effects of display size, $F(2, 36) = 85.30$, $\eta^2_p = .826$, $p = 2.19 \times 10^{-14}$, and cue, $F(1, 18) = 7.02$, $\eta^2_p = .281$, $p = .016$.

5.3.3.3 Experiment 3

The two-way ANOVA for the shift parameter showed a reliable display size effect, $F(2, 38) = 13.66$, $\eta^2_p = .418$, $p = 2.92 \times 10^{-4}$ and an ISI effect, $F(1, 19) =$

14.307, η^2_p = .430, p = 1.26 ×10^{-3}. For the shape parameter, all main effects were reliable: the display size, F(2, 38) = 3.53, η^2_p = .157, p = 3.93 ×10^{-2}, the cue, F(1, 19) = 19.05, η^2_p = .501, p = 3.34 ×10^{-4} and the ISI factors, F(1, 19) = 28.48, η^2_p = .600, p = 3.77 ×10^{-5}. The display size × ISI interaction was also reliable, F(2, 38) = 3.769, η^2_p = .165, p = 3.21 ×10^{-2}.

Table 5-4. ANOVA summary for the mean RT, accuracy rate and the Weibull parameters. N, Q, and I stand for display size, cue, and ISI factors, respectively. • p < .09; * p < .05; ** p < .01; *** p < .005; **** p < .001. ; empty cells are either non-significant or not available.

	Shift			Shape			Scale			RT			Accuracy rate		
	E1	E2	E3	E1	E2	E3	E1	E2	E3	E1	E2	E3	E1	E2	E3
N	**	*	****	****		*	****	****	****	****	****	****	***	*	****
Q				•	•	****		*		**	***		•	*	•
I			***			****			•		***	*			
N×Q		•			***					***	*				
N×I					*	*					•			.	
Q×I														**	•
N×Q×I															•

5.4 Discussion

The study investigated the visual search influences of the WM template updated in each individual trial, comparing to an identical template been iterated in every trial. In contrast to most previous studies, I not only summarised data with central tendency measures, such as mean RTs, but also assessed search performance in a multifaceted fashion by jointly analysing the response accuracy and latency – allowing the RT distributions, decision thresholds and decision rates to be assessed. The decision parameters were estimated using the drift–diffusion and the LBA models, both accounting for the non-decision times, decision rates and decision thresholds. The result indicated that, depending on the duration between the offset of the cue and the onset of a search display, template updating does affect search performance. This reflected less clearly at the mean RT, but was suggested in the accuracy of performance, an aspect of the data previous studies mostly had difficulties to discern. Most importantly, via the decision-making models, the study clarified that template updating affects decision thresholds, but not decision rates. The direction of influence depends on the timing and certainty of a search display and the WM strength of an attentional template. In the following, I discuss the role of the attentional templates, its association with the WM and search efficiency, how they affect the RT distributions and finally the differences of the two decision-making models.

5.4.1 The Automated Template

The results from Experiment 1 indicate that repeating the target probe trial-by-trial (Olivers, 2009) does reduce search time and error rates. The

purpose of repeating the probe was to 'offload' a search template 'partially [from the WM] to other systems' (Olivers et al., 2011). Instead of testing the competition of WM resources between an AMI and a search template in a dual (search and memory) paradigm (Olivers, 2009), here I tested the 'offload' idea in a search-only paradigm, comparing the consistent-mapping (fixed cue) to the varied-mapping conditions (varied cue). The 'offload', following Olivers et al.(2011), refers to a process that a consolidated template is removed from WM or becomes weak in terms of WM strength. I focused specifically on how the WM strength of a template affects search performance. Specifically, I asked which part of a decision-making process associates with offloading a template. The results from the varied cue condition in Experiment 1 suggest that a successful guided search is viable even without an active WM template, presuming that the template no longer resides in the WM system. The offloaded template appears to render search more efficient (Figure 5-2) than when a template is updated in every trial. The offloading advantage however was observed only in Experiment 1 when the timing of search display appearance was highly predictable, a condition leading to a decrease in the decision threshold. This result is consistent with the view of the activation-state hypothesis (Olivers, 2009). The automated template, been achieved via repeating an identical target probe in every trial, lowers only the response threshold. The null effect in the decision rate, on the other hand, suggests that the perceptual sensitivities of matching a target to a WM template are similar in the two cue conditions. The explanation of similar perceptual sensitivities in an identical template and an updated template is likely only when an observer

needs not raise vigilance to prepare an upcoming a search display, which either appear immediately (50-ms ISI) or after a short while (400-ms ISI). High certainty about a target's identity (i.e., fixed cue condition) and a display's timing (i.e., a single 50-ms ISI in Experiment 1) can cause an observer to adjust his/her response threshold when an identical template renders similar perceptual quality as an updated template. That is, both forms of memory representation – an automated template and an active WM template – are capable of permitting a successful search with a similar decision rate when a search task is highly predictable. As a consequence, the difference lies in the decision threshold, as shown in Experiment 1.

However, when a 400-ms ISI was randomly inserted into a testing block originally containing only 50-ms ISI, the relation between the fixed and varied cue conditions was reversed, as shown in Experiment 2. This reverse pattern is robust, as Experiment 3 replicated the pattern with more per-condition observations. One explanation of the reversed relation is the certainty of a target's appearance timing, thereby affecting an observer's decision threshold. This is suggested by a significant ISI contribution on the decision threshold in the AIC model, although the BIC model indicates that the ISI factor contributes only to non-decision time.

Two critical differences were introduced in Experiment 2 (& Experiment 3), because of to the ISI factor. Firstly, the timing of a search display appearance became uncertain. Observers could not predict explicitly the when a display might appear neither automate implicitly their key-press actions with respect to the display time. The uncertain timing likely results in increasing in

cautiousness. Secondly, because Experiment 2 used two ISIs and allocated them randomly, observers might respond to several 50-ms (or 400-ms) ISI trials and encounter unexpectedly a 400-ms (or 50-ms) ISI trial, and vice versa. Apparently, the latter sequence (i.e., several 400-ms trial and then 50-ms trial) hinders search more than the former sequence. Further, comparing to the per-condition observation numbers in other studies (Olivers, 2009; Woodman et al., 2007), Experiment 2 and Experiment 3 used relatively large number of trials, so the observers should experience both ISI sequences. The mix of the two ISI sequences, as suggested by the model fitting, may result in an adjustment of response strategy (as AIC model suggested) and this associated with the cue factor. Admittedly, the current data suggest also that the ISI factor affected also the non-decision time (see the discussion in Section 5.4.4).

The fixed cue condition, if as the activation-state hypothesis (Olivers, 2009) presumed, offloads a WM template to other systems. The result associated with the fixed cue condition in Experiment 2 then implies either (1) that to reload an automated template back to the WM takes up additional time when a search display comes up unpredictably early (i.e., 50 ms) and this is a consequence of decision threshold (as well as non-decision time) adjustment or (2) that the uncertain ISI pattern causes an undecided state of WM template offloading. The second possibility might result in an increase in the decision threshold when in some trials observers kept a dissipating WM template.

The result in Experiment 3 further support for the argument of cue-related threshold adjustment, because when an additional route (i.e., double trial number) to automate template is introduced, the accuracy rate in 50-ms ISI

condition becomes the only traditional statistics differentiating the fixed and varied cue condition. This result supports Olivers's (2009; also Olivers et al., 2011) account that whether the template is automated plays a critical role in the search performance. When observers become highly familiar with the task via performing a large number of trials, the advantage of the mean RT and drift rate for the varied cue condition disappears, but its accuracy disadvantage still exists in the 50-ms ISI condition. Apparently, the automated template, though may not alter the drift rate, helps to maintain a strong decision confidence when a hard-to-predict upcoming search task is displayed immediately. This hypothesis however remains to be verified, because only the main effect of the ISI, rather than the cue × ISI interaction, contribution in the AIC model and top-level model variations was observed.

More concretely, I suggest that the varied cue condition enforced a process of template rehearsal in WM. When given a long ISI, observers were given the opportunity to go over the template, thereby maintaining an accuracy rate as the fixed cue condition. The observation of the similar accuracy rate in 400-ms ISI but not in 50-ms ISI in Experiment 2 and 3 suggests that the template strength in the varied cue condition might reach similar level as that in the fixed cue condition. This is in line with the BIC model, showing that Q factor contribute to the decision threshold, but not the drift rate variations. That is, although the varied cue manipulation rendered the target identity less certain, the long ISI strengthened the template and this resolved the accuracy rate, but not mean RT, difference.

When the two uncertain factors - display timing and target identity – were

introduced, observers might become less confident to commit a response. The data for accuracy suggest an increase in the response criterion, but cannot rule out a decrease in the decision rate. This ambiguity is made clear by the drift-diffusion and LBA models, showing in Experiment 2 the cue factor depends only on the decision threshold, with a lower threshold for the varied cue condition (0.83 & 0.54) than for the fixed cue condition (0.92 & 0.55). This is consistent with the argument of an enhanced activation for the memory template in the varied cue condition (Olivers, 2009). In other words, the enhanced WM template results in a decrease in the decision threshold. However, this is in contrast to an insignificant difference of the decision threshold in Experiment 1 (fixed cue = 0.701, varied cue = 0.727), when there was no temporal uncertainty and in Experiment 3 (fixed cue = 0.55, varied cue = 0.54), when there was a drastic increase in trial number. A further evidence for the certainty-related threshold change is the general magnitude of the decision threshold. The decision thresholds in Experiment 2 are generally higher than those in Experiment 1, which then are higher than those found in Experiment 3. The changes in decision thresholds across the three experiments and the null effect of decision rate support the interpretation of task certainty and decision threshold.

In summary, comparing to updating a target probe in every trial, repeating a target probe does influence the decision threshold, reflecting on the patterns of mean latency and accuracy. The random ISI patterns cause changes of cue effect in the decision threshold and manifest as a pattern of speed-accuracy trade-off in the average data.

5.4.2 Search Efficiency

The display size effect, a common efficiency indicator, shows the frequent observed findings: a positive linear relation of the mean RTs and the display size. So does slope data (within-participant comparison) replicate the previous finding (Woodman et al., 2007; between-participant comparison). Both Woodman et al.'s and our data show that the fixed- and the varied-cue conditions result in similar search slopes when the data were fitted with ordinary linear regression. Even though the cue factor influenced the overall slope little, Woodman and colleagues' data indicated that the search slope was affected by the interaction of the cue and the WM load. Specifically, the cue factor only modulated search slopes in the dual task paradigm only when the observers loaded WM with a memory representation for an upcoming (search) task.

Because the current study did not implement a dual-task paradigm, it cannot directly verify the point that a concurrent WM load renders updating template become effectively to decrease search efficiency. Our model fitting results do suggest two possibilities. First possibility is that the BIC model suggests Woodman et al's (2007) interaction finding might result from an increase in non-decision time, because the cue factor does not account for variation of the decision rate, and because the ISI factor, which relates to the memory consolidation, affects the non-decision time. The AIC model in Experiment 3, on the other hand, suggests the cue factor might affect the decision rate while both the ISI and the display factor also affect the non-decision time. This specific result however is not observed in Experiment 2 when the per-condition sample size is half of that in Experiment 3. A clear

answer may require a direct test using the dual-task paradigm when both RT and accuracy data are jointly accounted for by the decision-making model.

5.4.3 The Certainty of Search Display Appearance

The reverse finding of the fixed versus varied cues raises a question as to why the 50-ms ISI in Experiment 2 (as well as in Experiment 3), a condition seemingly replicating the procedure in Experiment 1, showed a drastic difference between the two cue conditions. Examining together with the accuracy data, the data in the left two columns in Figure 5-2 suggest there might be a pattern of speed-accuracy trade-off. That is, the pattern in Experiment 1 replicates what Woodman et al (2007) had reported, but that in Experiment 2 (50-ms ISI) does not. The only difference in term of design in Experiment 2 (& 3) was the randomly mixed 50- and 400-ms ISIs.

The different design might trigger two different cognitive changes: (1) the readiness of response and (2) the WM strength of the search template. In the 400-ms ISI condition, observers were able to prepare the search template during the 400-ms interval, whereas in the 50-ms ISI condition, this readiness operation might only occur when a search display has already been in place. Thus, the additional mean RT observed in the 400-ms ISI condition reflects this operation. In contrast, the observers in the fixed cue condition in Experiment 1 were not affected by the short ISI, because they might have prepared (i.e., automated) a strong search template via exposing to it repetitively without suffering from memory decay. Secondly, the observers might maintain a less consolidated WM template in the 50-ms ISI in the latter experiments than that in the first experiment. To facilitate search effectively, the observers might spend

time to reload the fading template back to the WM (regain confidence), as the activation-state hypothesis predicted (Olivers, 2009). In light of the model-fitting data, the second account seems more likely, because the cue factor contributes only to the threshold parameter (in the BIC model). Nonetheless, the data do not clearly rule out the first readiness account, because the AIC model indicates that the ISI factor contributes also to the threshold parameter and because both the AIC and BIC models show the ISI factor associates with also the non-decision time.

The additional 400-ms ISI condition and the within-block randomization of two ISIs might cause (1) the reversed relation between the fixed and varied cue conditions in the mean RTs, and (2) the increase difference between the two cue levels at mean accuracy (at 50-ms ISI condition). The latter observation was replicated in Experiment 3 when per-condition trials were doubled, but the first pattern became trivial, even the varied cue advantage in mean RT still descriptively exists. Although one may argue this is due purely to the speed-accuracy trade-off, both decision-making models suggest the decision rate is unaffected by the ISI and cue factors. Therefore, the decision rate is not traded, because of the cue factor, for increasing accuracy. The average data, at best, suggest the increase difference in mean accuracy is due to an increase in response caution (between-experiment comparison), and the reverse relation might be caused by the mixture of ISIs.

One explanation for the reversed cue relation at the mean RT is that because, in addition to the fading WM strength for a template, the uncertain appearance timing of a search display made observers become less reluctant

to commit a decision. In Experiment 1, observers were exposed to one ISI and a fixed target probe throughout the entire task, so the task with regards to the display timing and to the target identity were highly predictable. Thus it is reasonable to infer the search template had been automated. As a consequence of the high certainty, observers might commit less cognitive resource, which is associated with a low response threshold. Experiment 1 showed the quality of an automated template was strong enough to ensure a successful search (i.e., null cue effect on the decision rate), so I observed a reduction of response latency and error rate. A similar level of mean RT performance was observed also in Experiment 3, which supports the claim that an automated template permits an effective search in terms of the decision rate, because Experiment 3 used a larger number of per-condition trials. However, the same automated template in Experiment 2 seemed no longer able to ensure a successful search, possibly because it might not strong enough (in terms of WM strength) to guide search when the search display became less unpredictable. When observers were unsure when a search display was going to come up or when their strategy was set to respond to a particular ISI sequence, an automated, but fading WM template increased target uncertainty. This automated, weak WM template makes observers cautious about a decision. On the contrary, an updated WM template helped observers maintain a WM template as long as its strength is strong enough for a successful search. This explanation is supported by the long ISI accuracy data when observers were given opportunity to rehearse the search template and thus kept it active in the WM. That is, the data of the similar accuracy rates in the two cue

conditions in the E1-50, E2-400 and E3-400 columns in Figure 5-2.

5.4.4 Response Threshold and Perceptual Sensitivity

The findings from the ISI factor support further the interpretation that the WM strength of a template influences the response threshold. Although the accuracy data, together the mean RT, cannot rule out whether decision rates was also changed in Experiment 2, they suggest that the observers became more hesitant to respond at the short ISI. That is, the mean RTs of 400-ms ISI is consistently slightly lower than those of 50-ms ISI, except in display size 9, fixed cue condition.

When Experiment 3 increased the trial number to 200, an unambiguous influence of the cue factor on the mean accuracy, but not mean RT, was revealed in 50-ISI condition. Taken together with the data in 400-ms ISI, it is clear that the long ISI resolves the accuracy difference between the two cue conditions. Note that the mean RT between the two cue conditions remains very similar. Although it needs more evidence to understand how the ISI factor interacts with the changing strength of a WM template and how this interaction influences the decision threshold and/or non-decision time, both decision-making models suggest a clear null effect of the cue and ISI factors on the decision rate.

Evidence from the quantile analyses suggest the cue × ISI interaction occurred on in first 4 RT percentiles (0 to 30%), and waned afterwards (Figure 5-3). This pattern favours a difference in the early perceptual or late motoric processes, reflecting in the non-decision time difference (60 ms vs. 77 ms; 400-ms ISI vs. 50-ms ISI) and an ISI factor contribution in the BIC model.

Nevertheless, further investigation is needed to understand why the AIC model suggests that the ISI factor contributes also to the variation of decision threshold.

5.4.5 The RT Distribution and Decision Parameters

One substantial shape change in the distribution is the one associated with the display size increase, a consistent observation found also in other visual search paradigms (Cousineau & Shiffrin, 2004; Palmer et al., 2011; Chapter 4). Because we did not design this study to examine the question of cognitive architecture (serial vs. parallel and/or termination rules) of visual search, we make no claim if the display-size related shape changes of the RT distributions support either parallel or serial search, instead we suggest one underlying reason leading to the shape changes is the different decision rates. This is in line with our previous simulation study, suggesting that the change of the decision rate might reflect also on the scale and the shift parameters (Chapter 4). The more items a search display contains, the lower the average decision rate and the wider an RT distribution spreads (Cousineau & Shiffrin, 2004).

The shape changes associated with the cue factor were less apparent, however. Two clear cue factor-related shape changes are observed in the tail and peak of a distribution. The tail became thickened and peaks became shortened across the four display sizes when comparing fixed cue to varied cue conditions in Experiment 1 and reversed in Experiment 2 (Figure 5-10).

In summary, in the guided search paradigm, the response threshold change might be associated with a shape change in the RT distribution at the

micro level, and the variation change in the decision rate might be associated with a shape change in the RT distribution at the macro level. I showed here how the dual-modelling approach, together with an appropriate experimental design and a strict timing control, can provide evidence to link the three levels of data (mean RT/accuracy, RT distributions, decision parameters) to the cognitive process in the WM. The underlying reasons leading to the different RT distributions and reverse relation between the two cue conditions may be the unpredictable display timing and/or the intertwined ISI pattern. Further studies are needed to clarify how ISI may influence the cognitive process of template updating/offloading.

5.4.6 LBA vs. DDM

The LBA model and the DDM were contrasted in the two experiments. Both models fit RT data closely, although the LBA model performed worse than the DDM in the slow responses (90% percentile) when the display size was 9. For the error rate, the LBA model showed better fits in the two experiments than the DDM, a result also found in Van Zandt and colleagues' (2000) perceptual matching task (but see Ratcliff & Smith, 2004; Ratcliff, Van Zandt, & McKoon, 1999). As already been explained in detail in Van Zandt et al. (2000), our result does not argue against the DDM as being generally worse model to accommodate error rates than the LBA model, but it does suggest that the DDM may perform worse under our specific conditions. One key difference between the class of race models (LBA) and that of random walk models (DDM), thereby their difference associated with accommodating the error data, is how they parameterise the cognitive process. The race models are designed to

accommodate multiple choices using several accumulators, whereas the random walk models restrict the process of evidence accumulation in a single accumulator to accommodate only two choices. This leads to two consequences. Firstly, the mechanisms to determine an error response and its RT differ; thus affecting how they fit error RTs. The random walk models associate an error response with a correct one, whereas the race models determine whether a response is correct or not on the basis of a matching factor (*M* in our modelling nomenclature). Second, due to multiple accumulators, the race model applies more parameters to fit data. In the case of LBA, it allows the matching factor associated with *v* and s_v, and *IR* factor with *B*. Naturally, the more parameters, the higher freedom for it to fit data. These mathematical differences contribute to the fitting difference.

Compared with Van Zandt et al.'s (2000) paradigm, our data showed most substantial deviance of fit to error rates between the two models are in display sizes 7 and 9. This cross-study comparison suggests that the experimental reason of worse fits may not be what Van Zandt et al (2000) hypothesised that the simultaneous presentation of the elements of a letter pair, because I did not observe homogenous worse fits across all display sizes, but considerable advantage of the LBA fit only in display sizes 7 and 9. I suspect this is because multiple occurrences of the comparison between a WM template and a display item (or a grouped unit). Single accumulator models perform optimally when errors are rare and a decision can be reached within the first sweep (1 s) of a visual scan (Ratcliff & Rouder, 1998). In the current study, the former was observed but the latter may only be possible in small displays (3 or

5). This model design may not fit well to the paradigm involving in multiple occurrences of the parallel process, for example the 7- and 9-item search display in my paradigm. This search paradigm likely involved some degrees of parallel processing of the display and serial selection for multiple items in a trial (Cousineau & Shiffrin, 2004; Dosher, Han, & Lu, 2010). As Cousineau and Shiffrin's (2004) distributional analyses showed, given highly trained observers and huge number of response trials, a search process may become an informed serial search (see Cousineau & Shiffrin's data for clear multi-mode RT distributions). The characteristic of serial, multiple parallel processes may render multiple accumulator models accommodate better for the error rate data.

5.5 Conclusion

This study addressed the questions relating to search decision-making. The results suggest informed searches depend on the WM strength of an attentional template (consistent vs. varying mapping) and its association with the timing (ISI) and certainty of a search display (ISI patterns). The data indicate that the decision rate in visual search correlates highly with the display size, and this correlation changes the shape of an RT distribution. Most importantly, the study showed the decision threshold depends not only on how a template is set up, but also depends on the certainty of an upcoming search display. This study provides evidence, showing the effectiveness of harnessing the RT distributions and the decision-making models to account for the WM mechanism for the search template. The results provide evidence suggesting that how to set-up a search template affect decision threshold. Further studies are needed to clarify how the factors that influence the decision threshold

associate with the direction of changes for the threshold (e.g., see Section 7.2 for a discussion).

Chapter 6 Visual Templates Improve Decision Rate

6.1 Introduction

Broadly speaking, visual search is mainly modulated by external stimuli and internal goals. The former, stimulus-driven information is often theorised as visual saliency, a measure reflecting the perceptual distinctiveness of a stimulus. Take the attentional capture theory (Theeuwes, 1992, 2010) as an example. It hypothesises that the visual saliency captures the initial attentional allocation reflecting differential activation values on the saliency map (Itti & Koch, 2000; Li, 2002; Treisman & Sato, 1990). The activation values on the map reflect the distinctiveness of a stimulus in different features dimensions, such as colour, motion, line orientation, and luminance. The search path driven by the attention forms in a display from the highest peak saliency to a next. The notion is captured well by GS4's bottom-up activation map, which demonstrates computationally how visual search may operate on the basis of attentional shifts following the sequence of activation value on the saliency map (Wolfe, 2007).

In addition to the stimulus factor, a search goal may adjust the activation values on the saliency map when an a prior search template is set up (Found & Müller, 1996; Weidner, Pollmann, Müller, & Cramon, 2002; Wolfe, 1994). Compared to when observers are instructed to search for an outstanding item, they tend to pay additional attention to a set of specific features when they are informed with target features/descriptions. Mostly, the target descriptions facilitates search (Bravo & Farid, 2012; Hodsoll & Humphreys, 2005; Maxfield & Zelinsky, 2012; Schmidt & Zelinsky, 2009; Yang & Zelinsky, 2009), but not always (Bravo & Farid, 2014). Generally, goal-driven information seems to

operate differently on visual search from stimulus-driven information.

6.1.1 Attentional Template

The information held about a search target has been referred to as the 'attentional template', and it is assumed, in most search models, (Bundesen, 1990; Desimone & Duncan, 1995; Duncan & Humphreys, 1989; Wolfe, 1994) as a WM representation. A WM template takes around 200 to 400 ms to set up, so it could facilitate optimal guidance (Knapp & Abrams, 2012; Vickery, King, & Jiang, 2005; Wilschut, Theeuwes, & Olivers, 2013, 2014; Wolfe, Horowitz, Kenner, Hyle, & Vasan, 2004). Although the template set-up time depends on various factors, observers go through two main steps to form an effective template: (1) encoding/memorising and (2) setting/maintaining a template in WM. Apparently, the former step could be isolated from the latter step, which might independently influence search performance (Wilschut et al., 2013, 2014). In Wilschut and colleagues' paradigm, observers studied a first pre-search cue, two coloured disks, for 1.5 s, knowing that one of which (but not sure which one) indicated the relevant colour group a target belonged to. After the first pre-search cue disappeared, a second pre-search cue flashed for 50 ms at the one of the locations of the coloured disks. The second cue informed the observers which coloured disk was relevant. As a result, this paradigm was able to separate the stage of template encoding (cue 1) from that of setting-up (cue 2), suggesting that merely setting up a template affected search efficiency. Because a WM template was used during search to match and to verify the target, the observation of an independent template set-up process suggests that the WM template might only start to guides search when the encoding is

finished or at least sufficiently consolidated.

6.1.2 The Properties of Templates

If the memory matching process does play a critical role in guided search, the descriptions coded in the template should affect the extent of guidance; thereby how the template is represented in WM. The nature of the template has been examined in a series of experiments by Anderson and colleagues (2010). They compared the templates for the colour-defined and for the orientation-defined targets. Their data showed that, relative to being set up via an orientation cue, a template set up via a colour cue resulted in a stronger effect on RTs. The colour dominant role of the template was also found when the colour was target-related, comparing to when it was unrelated (Ansorge & Becker, 2014). In addition to colour being a strong primitive feature that easily captures attention (Motter & Belky, 1998), these findings suggest that there is colour-based guidance associated with the WM template.

6.1.3 Top-down Role of Templates

Not only do the physical features carried by templates influence search efficiency, so do templates' semantic properties. The template set up only at a conceptual level, such as using a verbal cue, influenced search performance (e.g., Castelhano, Pollatsek, & Cave, 2008). As another example, Hodsoll and Humphreys (2005) used a verbal cue to inform observers the target features and found that search time were reduced when the cue described a typical member in the category the target belonged to (categorical template, CT), compared to when the cue described an atypical member. The CT benefit however diminished when observers previewed a visual image of a typical

member, compared with when an atypical member was previewed. It appears that the categorical template set up by describing target category works differently from that set up by previewing an image. On the other hand, the CT benefit is not limited to describing one type of feature, as a similar effect was observed too when the category is defined by orientation (J. P. Hodsoll & Humphreys, 2005). Moreover, the CT benefit apparently is evidenced at different search stages. It reduces search time, increases the numbers of initial saccades towards a target, and decreases the numbers of distractor fixations (Schmidt & Zelinsky, 2009; see also Arita, Carlisle, & Woodman, 2012, for cueing distractor colour can enable a template set-up at a conceptual level, by guiding attention away from distractors).

One consistent finding in this series of studies is that no matter how detailed and elaborate a verbal description is, a visual preview often provides stronger guidance. A visual preview ostensibly carries complete target information, whereas a categorical or verbal cue describes only incomplete information. Although a clear mechanism remains to be clarified, the preview advantage in guidance from a visual cue seems to reflect a search optimisation that other types of template are difficult to attain. The preview advantage has also been demonstrated in a large number of cueing paradigms. So far, these paradigms have not identified a condition wherein a verbal or categorical cue permits equivalent guidance as a visual preview (Bravo & Farid, 2012; Knapp & Abrams, 2012; Meyers & Rhoades, 1978; Schmidt & Zelinsky, 2009; Vickery et al., 2005; Wilschut et al., 2013, 2014; Wolfe et al., 2004; Yang & Zelinsky, 2009). Apparently, the sub-optimal guidance provided by the verbal cue cannot

be attributed to an insufficient time to form/encode a template (Wolfe et al., 2004), because the verbal cue disadvantage remains even with a prolonged time window for viewing the description (up to 6 sec; see the RT data in Experiment 2 in Knapp & Abrams, 2012).

However, some recent studies suggest that the template set up via a visual preview does not always generate a search advantage. For example, Anderson, Heinke and Humphreys (2010) identified a condition that a template set up by a visual preview showed only equivalent guidance to a verbal cue (Bravo & Farid, 2014; Maxfield & Zelinsky, 2012; Soto & Humphreys, 2007). This result is perplexing because, on the one hand the contrast between orientation and colour features in templates in Anderson et al.'s data (2010) showed that the feature differences exerted strong differential effects on search; on the other hand, the contrast between concrete (visual) and abstract (verbal) features did not.

Following the SERR model (Humphreys & Müller, 1993) and their group-segmentation account (Heinke & Humphreys, 2003), Anderson and colleagues (2010) interpreted the findings that colour exerts a goal-driven and a stimulus-driven effects on the group segmentation. They argued that the differences in search time should result from increased search efficiency due to the colour-induced group segmentation, rather than to a change in a *response selection*. This argument was supported by the data collected using a 'compound' search task (Duncan, 1985) wherein an observer looked for a colour or orientation pre-cued target (either a coloured vertical or horizontal bar) which carried also a symbol (a plus sign, '+' or a cross sign, '×'), serving as a response indicator

(responding to colour or orientation). Because they found a reliable search advantage for the colour over orientation cue in the compound task, which the response selection bore no direct relation to target identification, they argued that the colour advantage is a result from its strong effect on the template description, rendering effective group segmentation, rather than facilitating response selection.

6.1.4 Response Selection and Attentional Guidance

Anderson and colleagues' study (2010) demonstrated a condition in which visual cues might not provide strong guidance and the condition in which verbal cues might be capable of guiding search as efficiently as visual cues. In addition to the strong verbal-cue finding, another recent study has shown that, compared to when search was based solely on stimulus saliency, observers could not benefit from knowing target descriptions and showed deteriorated RTs (Bravo & Farid, 2014). This finding however is inconsistent with large number of template-based search literature demonstrating a pre-cueing advantage. One explanation is the additional task, *response selection*, embedded in the search task, as in Anderson et al.'s compound task (2010). It is possible that this additional task contributes to the reduced advantage for visual cues over verbal descriptions. As the authors inferred, the additional task differentiated the stage of attentional guidance from the stage of response selection. A similar argument has also been put forward in the target detection paradigm, observing that the foreknowledge of target form and target location could reduce RTs (Bruhn & Bundesen, 2012). These effect became insignificant when a response selection task was separated from target identification

(Theeuwes, 1989). It appears that attentional guidance is a process separable from *response selection*, and thus the response threshold bears no direct relation to the factors altering guidance (see also eye tracking evidence that separated guidance from target verification in Maxfield & Zelinsky, 2012). The cognitive process of *response selection* may associate with a late target verification step, instead of the attentional guidance (Bravo & Farid, 2014). Specifically, after the attention is guided to the most distinctive item in a display, the item is then matched against the template, either a visual image or verbal description, to evaluate how likely this item is the target. If the likelihood exceeds a decision threshold, a response is initiated after a decision is made. Accordingly, a compound search task may measure RTs reflecting the matching, verification, and decision-making, but not response selection. The RT differences in the compound task thus reflect mainly the advantage of visual preview guidance, rather than the difference in response selection.

The current study examined the influences of different template representations on different parts of information-processing, including guidance, memory matching, target verification, decision-making and response selection. In contrast to many previous works, the study evaluated an entire RT distribution via the DDM (Ratcliff, 1978) and the HBM (see Chapter 4). Particularly, the study examined the changes of the DDM decision parameters and the shapes of RT distributions when observes were probed with three types of template: no template, a verbal cue, and a visual preview. The study assumed (1) the guidance and memory matching affect mainly the perceptual quality of stimuli, which is associated with the decision rate, (2) the target

verification affects the decision criterion, and (3) the response selection affects mainly the non-decision times. The first hypothesis was tested by comparing three different template probes. The latter two hypotheses were tested by comparing search items in a display (??).

The hypotheses regarding to RT distributions follow the view of the stage model of information-processing (Rouder et al., 2005), which divides human performance broadly as central process and peripheral process. The central process associates with the cognitive computations, such as attentional guidance and target-template matching, whereas the peripheral process controls the computation for taking early perceptual inputs and the computation for executing late motor responses. On the basis of the stage model, the study assumes that the changes in attentional guidance and the quality of target-template matching will alter the shape of RT distributions and that the change in the performance related to analysing early perception and to executing late motor response will alter RT distributions little.

6.1.5 Study Aims

This study aimed to examine whether the template set up by a visual preview reduces search times and errors, relative to the template set up by a verbal cue and to search without a template. A further question is how a visual preview enables efficient search with respect to its influences on the decision rate, decision threshold and non-decision times. In addition to the search task, a detection task was embedded in the odd-one-out search paradigm. The detection task serves as a control condition, because all pre-search cues were in fact redundant in the sense that a successful response could be made

without knowing the target identity/description.

Specifically, the study investigated questions in two key domains. First, the study investigated questions related to attentional template and its influences on perceptual decision-making. These questions were: (1) Does a visual template improve search efficiency? (2) If so, does a visual template enhance attentional guidance, reduce response threshold, and/or merely reduce non-decision times? (3) Does a verbal template exert an equivalent effect as a visual template in any aspects of search and decision processes? (4) If not, does a verbal template improve the processes and in what way it affects search differently relative to a visual template? (5) Does a redundant cue improve target detection?

Second, the study addressed questions related to RT distributions. These questions were: (1) If a visual or a verbal template does improve search, as might be shown in ANOVA analyses, or more specifically be revealed in the distributional parameters? (2) Does the linear relation between display size and mean RTs manifest as a change in the shape of RT distributions that can be summarised by the three Weibull parameters? (3) If a redundant cue improves target detection, does the effect alters the shape of RT distributions?

6.2 Method

6.2.1 Design and Procedure

The experiment used a two-way repeated-measures, within-participant ANOVA design. The first factor was display size (6 levels: 1, 3, 5, 6, & 7, vs. 9), and second was the cue (3 levels: null, verbal, vs. visual). A trial began with a central fixation cross lasting for 500 ms, and then a 100-ms pre-search cue

informed participants about the target. Immediately after a 1-s inter-stimulus interval, participants searched for a target letter in a search display. The numbers of search items were randomly chosen for each trial. The display lasted only for 100 ms. A trial was terminated 3 seconds after the onset of the search array. In this paradigm, a response resulted in either a hit or a miss. All participants were asked to give speeded responses without compromising their accuracy. Participants visited the lab three times in different sessions and received an identical instruction, except for the nature of the pre-search cue. The cue was either uninformative (the null cue: a fixation cross), a verbal cue (an l for lowercase or a U for uppercase), or a target preview (the visual condition). In the null cue condition, participants were told to find a letter that stood out from the others because it was in a different case. No further information was given. Because the target was always a lowercase (or an uppercase) letter in an all-uppercase (or all-lowercase) distractor array, participants were able to identify the target without prior knowledge. Comparing to the paradigms used in previous chapters, one critical difference in the current paradigm is that the target is always an 'oddball' in a display. Regardless of the template conditions, a search display is randomly drawn from an identical stimulus pool in every trial. Even though in all three conditions the observers were told to use pre-search cues as much as they can to help the search, the likelihood for them to rely on stimulus saliency to find a target increase when display size increase.

Each session was divided into 14 blocks. Participants had chances to rest when finishing a block. Each condition (6 × 3) contained 112 trials. In total, one

participant contributed 2016 responses. Although relatively speaking, this is not a very large sample size (e.g., 500 per-condition trials were collected in Wolfe, Palmer, & Horowitz, 2010), it is approximately ten times, comparing to the per-condition trials used in prior similar studies (e.g., Knapp & Abrams, 2012). Each trial was separated by a 1.5-sec inter-trial interval.

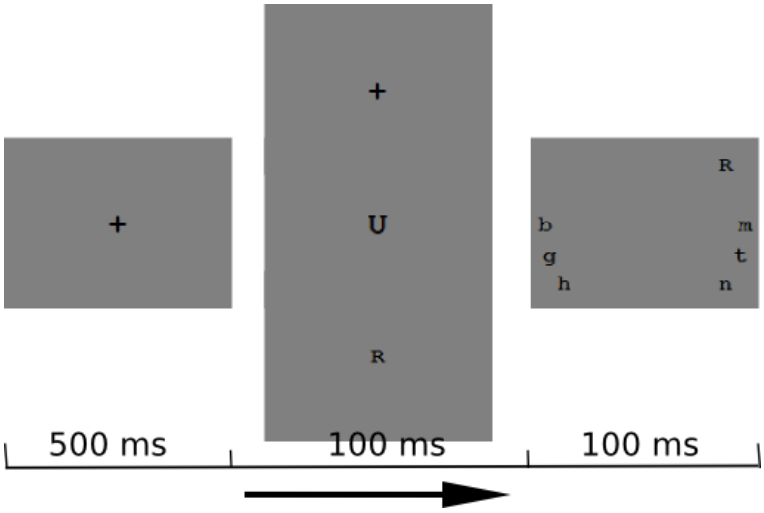


Figure 6-1. An illustration of the paradigm. The images have been re-scaled for the illustration purpose. The exact size of the stimuli allowed only a clear view when a participant sat around 60 cm in frontal of the monitor. The search display was selected so that in this example all cue conditons indicate the same target, R. The actual paradigm randomly selected search items in each display.

6.2.2 DDM

The decision parameters were calculated using the joint data of RTs and response accuracy as inputs to the fast-dm-29 computer programme (Voss & Voss, 2007, 2008), an implementation of the DDM. The fast-dm-29 estimates the three decision parameters, drift rate, decision threshold, and non-decision times and respectively their standard deviations. The programmes fitted the data, using a top-level (saturated) model separately for each participant and condition.

I set the initial point at .5 (i.e., no prior bias towards either positive or

negative boundary), because the likelihood of an identical target appearing in two consecutive trials was low (3.2% mean repeat rate, within-block). Further, because the two alternative responses in my paradigm were to indicate whether a target was on the right or left side of the central fixation, it is reasonable to assume the initial point was in the middle of the matched and mismatched boundaries. The sample size (112 trials) in this study was well above the suggested lower limit in a recent comparative study for different implementations of the DDM (80 trials; van Ravenzwaaij & Oberauer, 2009).

6.3 Results

This section was divided into three subsections. First subsection reported the general search efficiency across the experimental conditions, analysing the RT means, response accuracy and the least-square search slopes. Second subsection reported the DDM parameters. The last subsection analysed the RT distributions using the percentile plots and the Weibull parameters.

6.3.1 Mean RT and Accuracy

ANOVAs were performed on correct RTs between 0.2 to 2 s (0.8% rejection rate) and accuracy. A total of 180 summary observations for median RTs and accuracy were obtained (10 × 6 × 3; participants × display sizes × cues). Because there should be no process involving distractor selection/rejection in display size 1, the analysis of the display size 1 was dubbed the detection task and reported separately from the search conditions.

The detection task showed reliable cue effects both for the RT medain, $F(2, 18) = 6.10$, $\eta^2_p = .30$, $p = .009$ and the accuracy rate, $F(2, 18) = 3.89$, $\eta^2_p = .40$, $p = .04$. Compared with being probed by a fixation cross (the null cue; 390

ms, 98 %; p = .008) and being probed by a target description (the verbal cue; 413 ms, 99 %; p = .012), the observers detected the target quicker when previewing a visual image (the visual cue; 341 ms, 99 %). The cue effect at the accuracy rate is due to the difference between the visual and the null cue, t(9) = 2.59, p = .03.

The result in the detection task suggested the target preview improved performance. Individual analyse showed this performance improvement did not occur homogeneously across all observers. Three of them showed no reliable difference between the visual and null cues.

The search tasks replicated the RT-display size linear function findings in Chapter 4 and Chapter 5, $F(4, 36) = 30.76$, $\eta^2_p = .77$, $p = 3.61 \times 10^{-11}$. In particular, the fewer items a search array presented (3, 5, 6, 7, &. 9), the quicker the observers responded (541, 570, 597, 627, & 634 ms, the last pair, 7 vs. 9, did not exceed the .05 significant level). The search tasks indicated also a cue, $F(2, 18) = 9.07$, $\eta^2_p = .50$, $p = 0.002$, and a cue × display size interaction, $F(8, 72) = 5.62$, $\eta^2_p = .38$, $p = 1.54 \times 10^{-5}$ influences. The cue effect was due to the differences between the visual (467 ms) and the other two cue conditions [verbal, 634 ms, t(9) = 3.62, p = .006; null, 679 ms, t(9) = 4.65, p = .001]. The interaction was due to a decreasing differences between the null and verbal cues when the display sizes were larger than 5 (Figure 6-2)

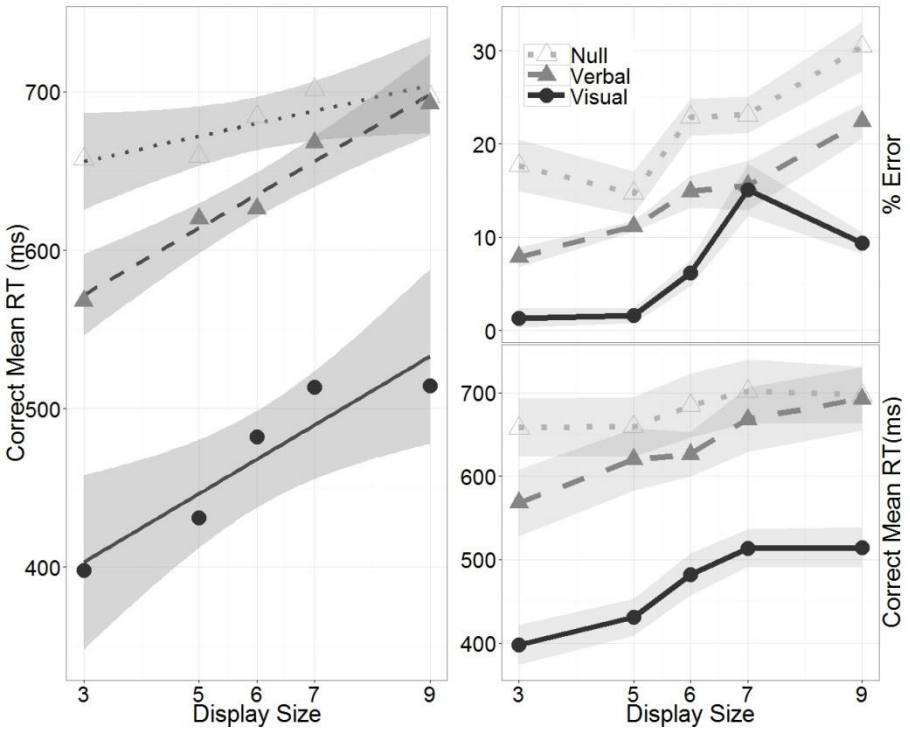


Figure 6-2. Search slope, correct mean RTs and error rates. The right panel drew the error ribbons based on ± 1 standard error corrected for within-participant variation (Morey, 2008). The left panel showed search slopes estimated by the least-square linear regression and drew the error ribbons, using the 95% confidence intervals.

The accuracy rate showed a similar pattern as the RT means. The few items a search array contained, the higher the accuracy participants achieved, $F(4, 36) = 21.12$, $\eta^2_p = .70$, $p = 4.91 \times 10^{-9}$. This manifested as a decreasing trend in accuracy when the display sizes increased (from display size 3 to 9, respectively, 91%, 91%, 85%, 82%, & 79%. The cue factor also affected the response accuracy, $F(2, 18) = 35.84$, $\eta^2_p = .80$, $p = 5.29 \times 10^{-7}$ (from the for the null, verbal to visual cues, 78%, 86%, & 93%). The two factors interacted, $F(8, 72) = 3.30$, $\eta^2_p = .27$, $p = 0.003$. The above-chance accuracy in the null cue condition supports that the observers were able to find the target relying on the oddball strategy.

Averaged across the display sizes, post-hoc t tests indicated reliable

differences between the three cues: the verbal cue vs. null cues, $t(9) = 3.29$, $p = .009$; the visual vs. null cues, $t(9) = 8.18$, $p = 1.86 \times 10^{-5}$; the visual vs. the verbal cue, $t(9) = 7.46$, $p = 3.65 \times 10^{-5}$. Note no difference was found between the verbal and null cues in the data of RT mean.

Further t tests at each display size showed reliable differences across most pairs of the cue conditions, $ps < .05$, except between the verbal and null cues in display size 5, $t(9) = 1.45$, $p = .18$, and display size 7, $t(9) = 2.07$, $p = .07$; between the visual and verbal cues at display size 7, $t(9) = 0.31$, $p = .76$.

A traditional way to estimate search efficiency is by assessing search slopes for the mean RT × display size function. Excluding the data of display size 1, the search slopes respectively were 7.99, 21.07 and 21.66 ms/item for the null, verbal and visual cues. The 95% confidence ribbons in Figure 6-2 give a quick assessment for statistical significance.

6.3.2 Decision Parameters

The visual-cue advantage for detecting a target appeared from the drift-rate difference, $F(2, 18) = 6.13$, $\eta^2_p = .41$, $p = .009$, because no cue effect was found either at the decision threshold or at the non-decision time. The visual cue resulted in a larger drift rate than the null cue [6.57 vs. 5.3; $t(9) = 3.71$, $p = .005$] and the verbal cue [5.46; $t(9) = 2.71$, $p = .024$]. This result seems to suggest a target previewing but not a target description, increased the decision rate.

6.3.2.1 Drift Rate and Non-decision Time

Both the drift rate and non-decision time were influenced by the display size, [$F(4, 36) = 29.29$, $p = 7.10 \times 10^{-11}$, $\eta^2_p = .77$; $F(4, 36) = 26.44$, $p = 2.78 \times 10^{-10}$, $\eta^2_p = .75$], and the cue factors, [$F(2, 18) = 64.05$, $p = 6.54 \times 10^{-9}$, $\eta^2_p = .88$;

$F(2, 18) = 9.77$, $p = .001$, $\eta^2_p = .521$], as well as their interaction, [$F(8, 72) = 11.20$, $p = 3.80 \times 10^{-10}$; $F(8, 72) = 2.94$, $p = .007$, $\eta^2_p = .25$] The interaction was due to the decrease in the cue effect when the display sizes increased.

On average, the drift rate was larger in the visual cue condition than the verbal [3.96 vs. 2.07; $t(9) = 7.58$, $p = 3.38 \times 10^{-5}$] and the null cue condition [1.54; $t(9) = 19.88$, $p = 9.58 \times 10^{-9}$]. No difference was observed between the verbal and null cue conditions [$t(9) = 1.94$, $p = .085$]. Further t tests at each display size showed the visual cue drifted faster than the other two conditions at all display sizes ($ps < .05$), except comparing with the verbal cue condition at display size 7. The verbal cue condition drifted faster than the null condition at display sizes, 3, $t(9) = 2.33$, $p = .045$, and at display size 7, $t(9) = 2.61$, $p = .03$. This result suggested that both the effects of display size and cue influenced the drift rate and the display size seemed to modulate how the visual and verbal cues affect the drift rate.

With regard to the non-decision time, again on average the visual cue condition showed less non-decision time than the other conditions ($ps < .01$). Also no difference was found between the verbal and null cue conditions. Separate t tests at each display sizes showed that the visual cue requires less non-decision time than the other two conditions at all display sizes ($ps < .06$), except when comparing with the verbal cue at display size 6. The only difference between the verbal and null conditions is at display size 5, $t(9) = 2.35$, $p = .04$. Overall, the non-decision time showed a very similar reversed pattern as the drift rate (Figure 6-3).

6.3.2.2 Decision Threshold

In contrast to the other two decision parameters, the decision threshold showed only a reliable effect at the display size, $F(4, 36) = 4.17$, $p = .01$, $\eta^2_p = 0.32$. The effect on the decision threshold was due to the null cue condition, $F(4, 36) = 5.67$, $p = .001$, $\eta^2_p = 0.39$, showing a decreasing trend as the display sizes increased (1.28, 1.17, 1.11, 1.12, & 1.01). The decreasing threshold suggests that when the local contrast was increased and the observers were set to find an odd target, they adjusted the decision threshold, instead of the drift rate. This result suggests an increasing number of display size facilitates attentional capture when the observers were guided only by the location contrast. The lack of display size effect in the verbal and visual cue conditions suggest that when the observers were guided mainly by a WM template, either via a visual preview or a verbal description, the display size does not affect the decision threshold

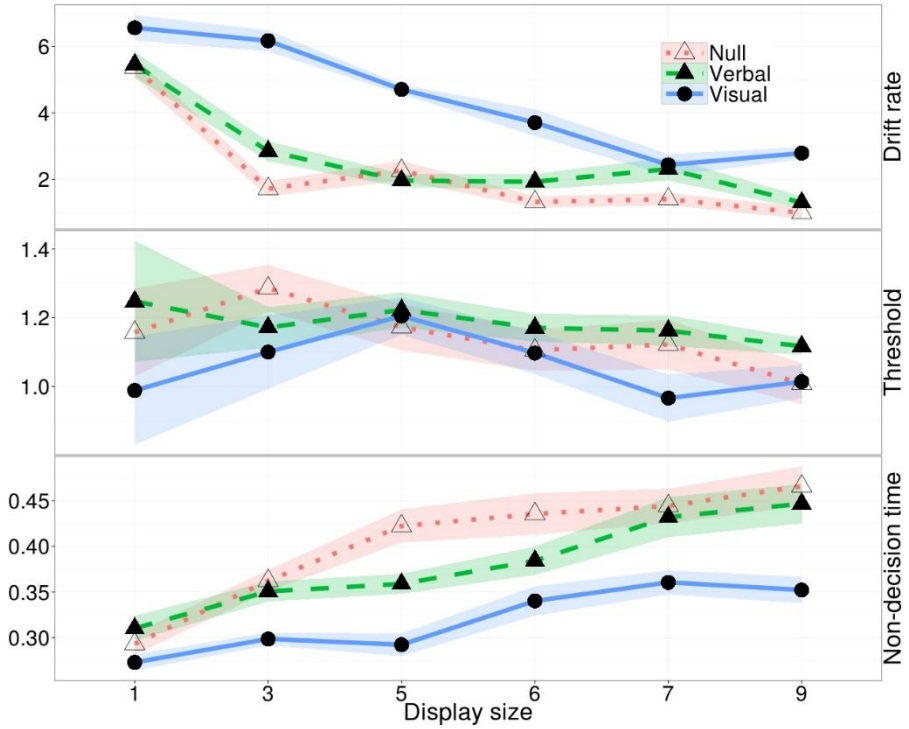


Figure 6-3. Diffusion model parameters. The error ribbons was drawn based on

± 1 standard error corrected for within-participant variation (Morey, 2008). Note the decrease in the decision threshold in the empty triangle (i.e., null cue) when the display size increases. The unit of the non-decision time is second.

6.3.3 Distributional Analyses

This section reported the percentile analyses respectively for error and correct RT distributions to compare the effects of the display size and cue factors from fast to slow RTs. Because the observers gave relatively less error, than correct, responses and the per-condition sample size was small, some error percentiles were estimated based on few than 10 observations (18.67%). Following the percentile analyses, the correct RT distributions were compared across different experimental conditions, using the Weibull parameters estimated by the Weibull HBM (Chapter 4).

6.3.3.1 Percentile Analyses

This percentile analyses clarified where in an RT distribution the cue and display size factors exerted influences. Only .1, .3, .5, .7 and .9 percentiles were presented in Figure 6-5. The procedure to summarise each mean percentile across individual trial was the same as mean RTs. In other words, .5 percentile in correct RT analysis is the median (correct) RT.

The detection task showed reliable visual cue advantages in all percentiles, compared with the verbal and null cue conditions. Interestingly, the null cue condition showed faster RTs than the verbal cue condition from .5 to .9 percentiles, a result suggesting that the top-down guidance from a verbal template is ineffective and that the bottom-up guidance from increasing local contrast became effective when at the tail side of the RT distributions.

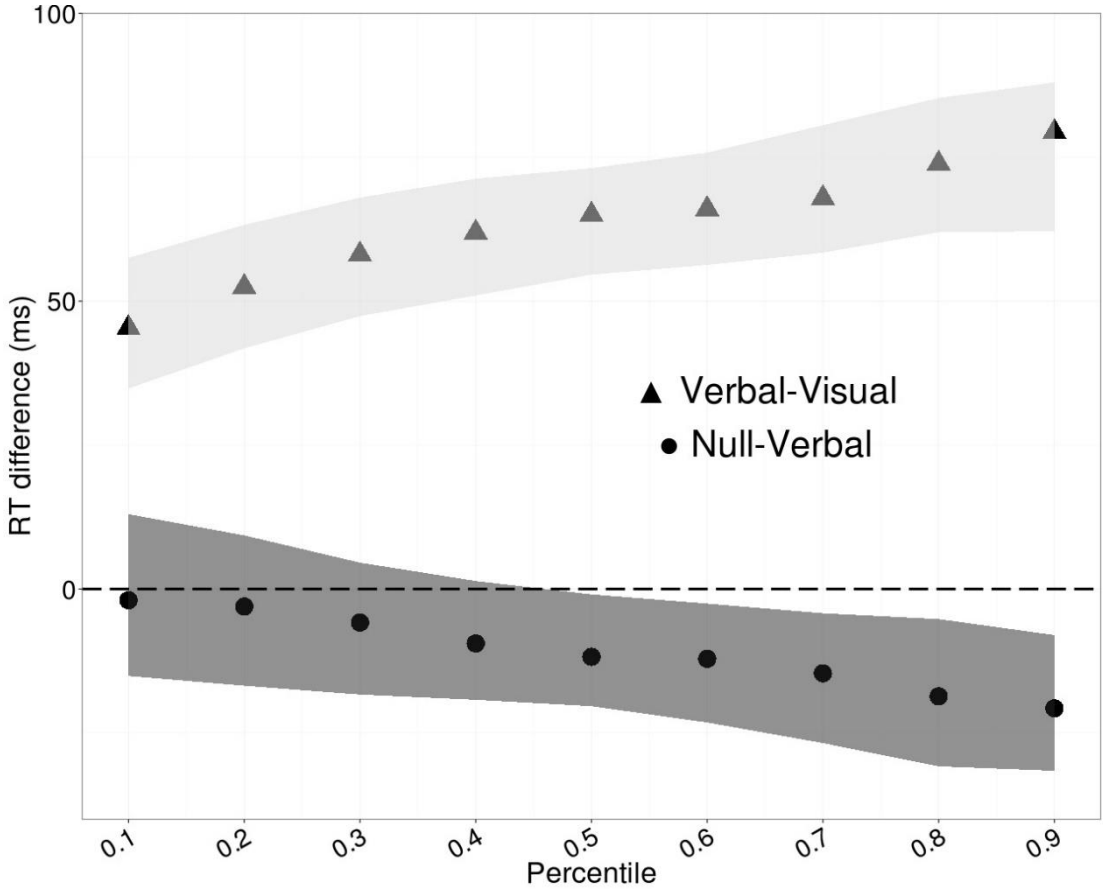


Figure 6-4. The RT-difference plot for display size 1. The figure showed RT differences along 8 percentiles from .1 to .9. The shaded areas indicate 95% confidence intervals. When a shaded area covers 0 difference (the dashed line), it suggests no reliable difference between the two cue conditions. The confidence intervals and statistics were calculated based on a robust procedure (Wilcox, Erceg-Hurn, Clark, & Carlson, 2014).

The correct RT distributions in the search tasks showed three important results. First, the visual cue condition showed unambiguously faster RTs across all display sizes in all percentiles than the other cue conditions (ps < 0.05; Figure 6-5). The slopes of the visual cue condition increased from .1 to .9 percentiles. The increase in the slopes suggests that the display size factor influenced differently on each percentile, from .1 percentile with a weak effect to .9 percentile with a strong effect. The differences between the visual cue condition and the others also showed an increasing trend along the RT distributions.

In contrast to the result in the detection task, the search tasks appeared to suggest that the verbal cue condition might be able to reduce RTs, because it showed consistently faster RTs than the null cue condition along the distributions, even though mostly the differences did not exceed .05 significant criterion. Specifically, amongst the insignificant differences, the verbal cue condition in display size 3 showed significant faster RTs than the null condition after .5 percentile. One way to interpretation this result is that the verbal description did guide search, but when the display size exceeded 3, the increase in the local contrast and the oddball strategy in the null cue condition exert a stronger effect on RTs than the template guidance due to the verbal description. I will return this point in the Discussion.

Third, the null cue condition showed an upward trend in search slopes, a tendency occurred only in the leading edge of the distributions (.1 & .3 percentiles) and diminished after .5 percentile. This result suggests the local contrast may work most effectively in slow RTs (.7 & .9 percentiles).

This observation in the slow RTs specific to the null cue condition suggests the effect of local contrast could be suppressed in an inefficient search task when the observers explicitly engage the verbal template to search. That is, the local contrast became effective only when the observers were explicitly looking for an odd target. The effect of local contrast appeared less effective when the observers set up a search template, because no obvious diminishing upward trend was observed in the visual and verbal cue conditions. The only observation suggesting that the increase in search items reduced RTs was in the visual cue condition when the display size increased from 7 to 9 in .7 and .9 percentiles.

The error RT distributions showed similarly a clear visual cue advantage over the other two. Compared to the quick percentiles (.1, & .3), the cue effect became stronger in the slow percentiles (.7 & .9), although the clear advantage of the visual cue showed only from .5 percentile. In large display sizes (7 & 9) specifically, the increase in variations resulted in insufficient observations. The null and verbal cue conditions did not show any clear differences, nor did the null cue show the decrease in search slopes from quick to slow percentiles. The verbal cue still showed increase in search slopes, suggesting the display size effect became gradually stronger from quick to slow responses, even in error RTs.

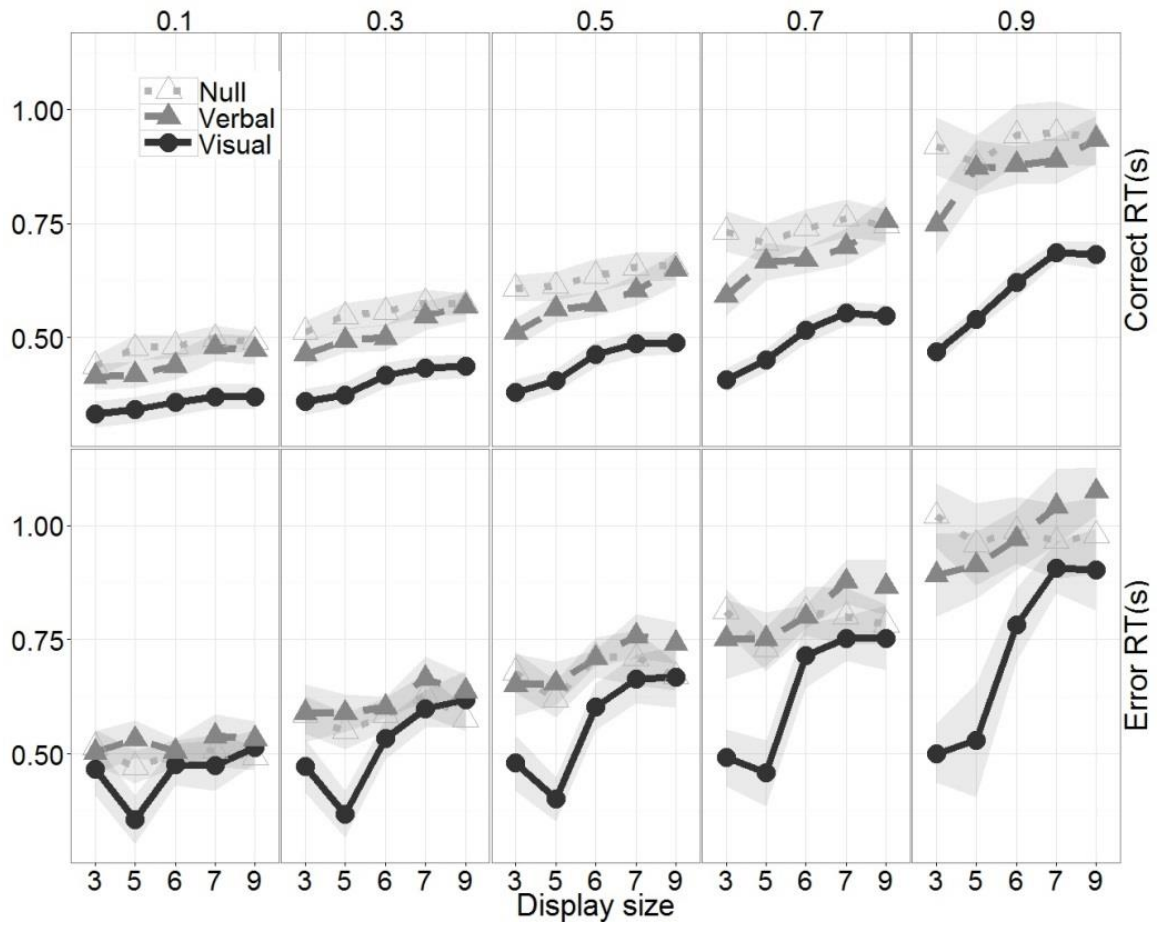


Figure 6-5. Percentile plots. The figure compared the differences of correct and error RT distributions in the three cue conditions along the 5 display sizes. The shaded areas drew ± 1 standard error corrected for within-participant variation (Morey, 2008).

6.3.3.2 Distribution Parameters

The detection task showed a reliable cue effect only in the shift parameter, $F(2, 18) = 4.1$, $\eta^2_p = .31$, $p = .03$. This was due to due to a significant difference between the verbal and visual cues [303 ms vs. 266 ms; $t(9) = 3.00$, $p = .02$]. The null cue condition (290 ms) showed no reliable difference, compared to the other two conditions (Figure 6-6).

6.3.3.2.1 RT Shift

The shift parameter showed only a reliable cue effect, $F(2, 18) = 7.39$, $\eta^2_p = .45$, $p = .01$, which was due to the difference between the visual and null cues, $t(9) = 2.87$, $p = .02$ and between the visual and verbal cues, $t(9) = 4.97$, $p = .001$.

Table 6-1. The Weibull parameters in the search tasks averaged across display sizes in the three cue conditions.

Condition	Shift (ms)	Shape	Scale (ms)
Null	347	1.77	377
Visual	281	1.74	216
Verbal	345	1.60	311

6.3.3.2.2 RT Scale and Shape

The scale and shape parameters show a similar ANOVA pattern, with a reliable display size effect [scale, $F(4, 36) = 5.50$, $\eta^2_p = .38$, $p = .002$; shape, $F(4, 36) = 6.76$, $\eta^2_p = .43$, $p = .0004$], a cue effect [scale, $F(2, 18) = 14.14$, $\eta^2_p = .61$, $p = .0002$; shape, $F(2, 18) = 5.09$, $\eta^2_p = .36$, $p = .02$], and an interaction effect, [scale, $F(8, 72) = 2.35$, $\eta^2_p = .21$, $p = .03$; shape, $F(8, 72) = 1.84$, $\eta^2_p = .17$, $p = .08$].

Post-hoc t tests indicated, when the scales were averaged across the display sizes, the visual cue condition showed a smaller value than the null, $t(9)$

= 4.81, p = .001 and verbal cue conditions, t(9) = 3.22, p = .01. Only marginal difference was found between the null and verbal cue conditions, t(9) = 2.02, p = .08. Further t tests showed reliable larger scale value in the visual cue condition than the others in all display sizes (ps < .05), except that in display sizes 7 and 9, compared with the verbal cue condition. The difference between the verbal and null cue condition was only observed in display size 3, t(9) = 2.38, p = .04.

Post-hoc t tests in the shape parameter found reliable verbal-null cue differences at the display sizes, 5, t(9) = 2.51, p = .03, and 6, t(9) = 4.24, p = .002. The visual-null difference was found only at display size 5, t(9) = 2.31, p = .046. And the display size 3 showed a reliable difference between the verbal and visual cues, t(9) = 3.22, p = .01.

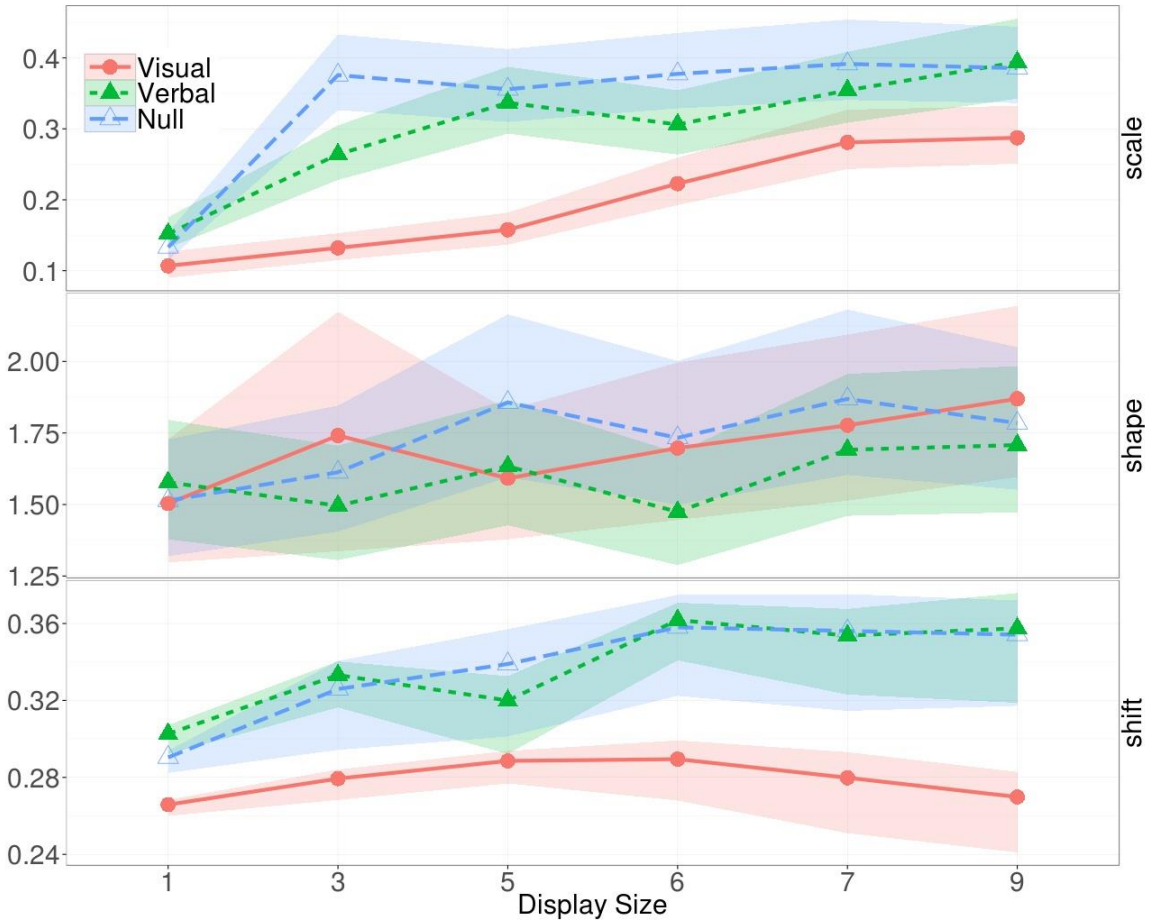


Figure 6-6. The line plots for Weibull parameters. The error ribbons were drawn based on Bayesian 95% credible intervals.

Table 6-2. The ANOVA summary for the mean RT, accuracy rate, the Weibull parameters, the DDM parameters. • p < .09; * p < .05; ** p < .01; *** p < .005; **** p < .001. ; empty cells signify non-significant.

		Mean RT	Acc. rate	Shift	Scale	Shape	Drift rate	Non-decision time	Decision threshold
Detection	Cue	*	*	*			**		
Search	Display Size	****	****		***	****	****	****	**
	Cue	***	****	**	****	*	****	***	
	Interaction	****	***		*	•	****	**	

6.4 Discussion

The study examined three classes of dependent variables, the conventional measures (RT means, accuracy, & search slopes), the decision parameters, and the distributional parameters, when the observers were set in three distinct mental operations, search for an odd letter, a upper-/lower-case letter, and a previewed letter, and were searching for three types of display sizes: a one-item display, a small display and a large display.

The data suggested clear influences of the display size and cue factors on the shape of RT distributions. The distributional changes due to the display size factor can be understood in light of the decision rate, the decision threshold, and the non-decision time. In the current cue-oddball paradigm, the display size factor affects the three main decision parameters as well as two Weibull parameters. This implies the display size factor modulates both early perceptual and late cognitive processes. On the other hand, the cue factor selectively showed no influence on the decision threshold. The condition of the displays size 1 further supports the point that the cue altered the decision rate. The preview cue increased the decision rate, whereas the verbal cue did the opposite. In contrast to the cue-threshold association found in Chapter 5, the finding in the current study suggests the cue factor can also affect the decision rate, suggesting that it is how the template operates in WM affects the decision parameters. In the following, I discussed the implications of different representations of WM template on search efficiency and how they affect search decision and thereby change RT distributions.

6.4.1 Search Efficiency

The analyses of RT means[27] replicated several previous findings (Knapp & Abrams, 2012; Vickery et al., 2005; Wolfe et al., 2004). Firstly, performance correlated positively with the display sizes and showed the visual template reduces search time more than the verbal and null templates (Knapp & Abrams, 2012; Vickery et al., 2005; Wolfe et al., 2004). The search slope however indicates the most efficient condition in this study was the null template, following by a verbal and then a visual templates. Both the visual and verbal cue conditions showed lower intercepts than the null cue condition, suggesting that a WM template, either a target preview or a verbal description, reduces base processing time.

Although the result of search slope appears to suggest the observers search more efficiently when relying only on stimulus features than when using both the stimulus features and a WM template, the superior efficiency in the null condition might be just an epiphenomenon of an increase in the local contrast. Together with the drift rate data, it is clearly that the null cue advantage not only resulted from the increase in its local contrast, but also from the decrease in search efficiency due to large display sizes in the other conditions (Figure 6-3). The increase in the display size appears to affect the null cue condition little, even though all three conditions drew search displays from an identical stimulus pool. This result suggests the instruction, setting observers in a mental operation for looking for an odd item, plays a critical role. This stands in contrast to looking for an uppercase letter such as in a ‘U’ cue trial or to looking

[27] Here I refer to a broad sense of the central location measure of a distribution.

for a letter R in an ‘R’ cue trial. In the latter conditions, the effect of an increase in the local contrast due to a small increase in displays sizes (e.g., from 5 to 6) did not noticeably improve search efficiency possibly because the instruction set observers’ mental operation to analyse and compare stimulus features against a template. In the null cue condition, however, a small increase in search items improved drastically search efficiency and countered the burden for searching more items. This is possibly due to the instruction set observers’ to rely only on the local contrast, so the display size factor also worked in an opposite way. This result, complemented by the drift rate data, demonstrates clear evidence that the search slope suggests two distinct interpretations when the observers were set on different mental operations: relying on a pre-set template or on stimulus features.

The positive correlation between RT and display size is a robust finding in visual search when mean or median performance is analysed (see a review in Wagenmakers & Brown, 2007). One interpretation for this correlation is the selective recursive rejection process (the SERR model, Humphreys & Müller, 1993), which argues attention samples a subset of items when searching through a large display that exceeds the capacity of one parallel processing (i.e., the group segmentation account, Heinke & Backhaus, 2011; Heinke & Humphreys, 2003, 2004). The visual features extracted from a sample are then to match against the features in a pre-set WM template. A selected subset of items is rejected as non-targets, if the amount of sensory evidence assessed is insufficient to reach the decision threshold. Hence, the parallel process of attention sequentially shifts from one subset to a next and matches it against

the WM template. The small display sizes (3 & 5) likely reflected only the segmentation process, and the large display sizes (6, 7, & 9) reflected both the processes of recursive rejection and group segmentation. This interpretation is consistent with a reduction of per-item search time when display items increased. The reduction gradually abates and reaches an asymptotic level.

This asymptotic relationship is in contrast to previous linear findings (e.g., Vickery et al., 2005), wherein tested only few display sizes. Those paradigms allocated search items in a sparse display rendering the segmentation process less likely and the linear relationship inconsequential. The result of mean RTs here instead correlate with the display sizes asymptotically (see a plausible non-linear explanation using a neural-network model in Mavritsaki, Heinke, Allen, Deco, & Humphreys, 2011). Specifically, the differences between consecutive display size (5-3, 6-5, 7-6, & 9-7), when averaged across the cue conditions, showed a decreasing trend (27.38, 29.08, 25.07, & 16.23 ms; see also Figure 6-2). This follows the prediction of the SERR model (cf. the item density hypothesis in Bergen & Julesz, 1983; the visual crowding hypothesis in Levi, 2008; & re-entrant process hypothesis, Di Lollo, Enns, & Rensink, 2000). Because the current paradigm allocated search items in a small viewing area that all displayed items can fall within a visual field, the number of items in one group segmentation increases when the item density increases and it becomes more likely that search items within the same area have similar features (Duncan & Humphreys, 1989). Thus, the increase in item density recruits the segmentation process and this reduces the detrimental effects of large display sizes. These processes seem to interact particularly

with the WM operation relying on a pre-set template, because the study observes a clear reduction of the detrimental effect due to increases in search items (i.e., from 7 to 9) only in the visual template condition (Figure 6-2). Nevertheless, this hypothesis is only a tentative account inferred based on the data in this study and previous study (Chapter 4), so requires more evidence to be consolidated.

The current data also suggest that some extent of pre-attentive guidance, as supported by the intercept data (Figure 6-2). The intercept data suggest the visual template resulted in less baseline time than the verbal template, following by the null cue condition. If this were only observed in the visual cue condition, the interpretation of perceptual priming might account for all the effects. The data of the verbal template, which carries only conceptual features (see also Wolfe et al, 2004; and Anderson, Heinke, & Humphreys, 2010 for prior data) suggest the perceptual priming , although exert strong effect, may not explain all the reduction in the baseline time. The data of the non-decision time and the RT shift support this point.

6.4.2 Search Decision

The detection task indicated the visual cue condition resulted in a quicker decision rate than the null cue condition, which showed an almost identical decision rate as the verbal cue (Figure 6-3). The finding of visual cue advantage suggests there might be a perceptual priming effect due to previewing the target, a possible cognitive mechanism causes the drastic increase in the decision rate. However the void finding from the cue factor in the decision threshold poses questions as to what factors might lead to the

accuracy difference.

In the detection task, because there was no distractors, a visual preview is impossible to improve accuracy by tagging distractors and rejecting them latter. That is, the recursive rejection process is unlikely and it is impossible to adjust the decision threshold by tagging distractors. On the other hand, the abstract foreknowledge of the target appears to benefit little for response accuracy. One explanation is the extent of cognitive processing. In the detection task, observers attained a correct response simply by noticing whether either left or right side on the grey viewing area darkened (Figure 6-1). A reasonable strategy to make sped responses is to process the cue and stimulus as minimal as possible. Because all display sizes were randomly mixed in a block, observers had to take note of the cue, but might discard to further process it as soon as they noticed a display contained only one item. This strategy suggests that a simple visual preview may improve only perceptual quality and may not be harnessed by higher cognition. This shallow preview process can still increase response accuracy without altering the decision threshold. The selective influence on the Weibull shift parameter appears to support the shallow processing account. Nevertheless, a clear mechanism as to how the preview process might benefit preattentive guidance and how the early guidance might increase accuracy is not directly addressed in the study. This remains to be explored.

In the search task, indicated an increase in the decision rate is mainly due to the attentional guidance from the visual and verbal cues. Because when they are compared with the null cue condition with no obvious WM operation,

the cued conditions involved to explicitly set up a template. The WM operation in the cued conditions reduced search baseline time, guided attention towards a probable target, as suggested by the search intercept data. One may argue that the data of search intercept and the drift rate were due to a strong priming effect, as a result of the target preview. However, the account cannot explain the data in the verbal cue condition. Further, the increase in the decision rate cannot entirely be attributed to the priming, because (1) the effect at the search task (η^2_p = .877) was far stronger than that of the detection task (η^2_p = .405), (2) the marginal effect across the display sizes (the post-hoc t findings at display sizes 3 and 7, when comparing the verbal to the null cue conditions), suggest a verbal template did guide search and (3) the interaction effect suggests that the cue alters the search, not merely detecting a target.

The DDM data suggest that template guidance does not affect *response selection,* reflected by the decision threshold. This decision parameter did not vary across the cue conditions. This is consistent with Anderson, Heinke, and Humphreys's (2010) 'compound' task finding, where there was an equivalent effect of the visual and verbal templates (also in Soto & Humphreys, 2007). The decision criterion data here suggest that this is because in the compound task setting, RTs mainly reflected not the process of attentional guidance, but from the response selection at a stage of target verification (Maxfield & Zelinsky, 2012). For this reason, the previous studies showed equivalent RTs when using the compound task to examine response selection.

The finding of the display size influence on the decision threshold is consistent with the argument that more-than-one visual scan occurred in the

current paradigm. Intuitively, this decision parameter should not associate with the display size, which may only set different levels of perceptual burden and thus should correlate positively with the non-decision times, as the linear assumption of analyses of RT-display size functions. In contrast to this, the study found that the decision threshold correlated negatively with the display sizes. The values of the decision threshold gradually decreased from 1.19, 1.20, 1.12, 1.08, to 1.04 when the display sizes increased from 3 to 9 (Pearson $r = -.18$, $p = .02$). This is predicted by the SERR model. As explained earlier, my paradigm renders the segmentation process more likely to contribute to performance when the display size increases. Because the amount of sensory information grows in one attentional parallel process when more items are in a search unit, the decision threshold may decrease with each perceptual sample.

The non-decision times, as expected, correlated with the display size positively (Pearson $r = .41$, $p < .001$). On the other hand, the cue factor showed only a visual cue benefit, but not any advantage from the verbal cue. This is in contrast to the finding measuring the time eyes move towards a target (Maxfield & Zelinsky, 2012). In Maxfield and Zelinsky's study, the data of the time to saccade to a target changed as a function of a cueing categorical hierarchy, from subordinate to superordinate cues with the visual (pictorial) cue showing the strongest effect. I did not, however, observe this dependence of the information hierarchy of template on the non-decision times. It is possible that the time to saccade to a target (Maxfield & Zelinsky, 2012) reflects largely the process of attentional guidance. The guidance is mostly accounted for by the drift rate, as it showed in my study a marginal significance between the verbal

and null cue conditions. Nevertheless, at the two moderate display size (5) the verbal cue showed a reduced non-decision time, compared with the null cue condition. One interpretation for the insignificance between the two cue conditions is the strategy difference. In the small displays, the observers might just detect a darkened light flash, so the fixation cross required less non-decision time than encoding an I or U for template setup. In the large displays, the observer applied the oddball strategy and this might result in a different perception encoding strategy, thereby, reducing non-decision time relative to the verbal cue condition.

6.4.3 RT Distributions

6.4.3.1 Detection Task

Only in the tail side of the distribution (.5 to .9), the observers responded slower when probed by a verbal cue than when by a fixation cross (Figure 6-4). Figure 6-4 shows the disadvantage of processing a verbal cue, even though it was not critical for detecting a target in display size 1, increases gradually towards long-latency RTs. The redundant process suggests two possible modes of target detection. First is a process of target detection without processing the verbal cue, suggested by the equal RTs in the early percentiles. Second mode is to simultaneously detect a target and process a redundant verbal cue. This mode is manifested in the late percentiles. The dual-mode account is plausible because the detection task was randomly mixed within the search task. To maintain good performance, the observers prepared to respond to, detection or search, task. Even though the observers were aware of that the target can be identified without referring to the pre-search cue, the processing

mode was likely determined probabilistically.

6.4.3.2 Search Task

The visual cue condition confirms the results of mean RTs, showing an overwhelming RT advantage over the other cue conditions across all quantiles and display sizes. The benefit of visual cue is due both to the target preview and the attentional guidance. The suggestion of attentional guidance is supported by two pieces of evidence. Firstly, because the verbal cue condition exerted no priming effect and carried only additional (redundant) conceptual information related to a target, an RT advantage due to the verbal cue suggests attentional guidance. This is shown in Figure 6-4. The same robust test as used in Figure 6-5 suggests the observers responded reliably faster when probed by a verbal cue condition than by a fixation in display size 3, 5, and 7, across all percentiles. Secondly, when no search is required (i.e., display size 1), the verbal cue condition resulted in slower RTs in the tail side of distributions. This result suggests simply processing a redundant verbal slowed down RTs when the observers engaged only in target detection. The verbal cue influence on the tail part of RT distributions also suggests that the process of conceptual information may cause only late processing stage.

6.4.4 Distributional Parameters

The detection task indicates that the different attentional templates changed only the shift parameter, suggesting the priming effect due to the visual cue effectively moves an entire distribution leftwards without significantly altering the shape and the scale of the distribution. This observation is inconsistent with Rouder and colleagues' stage model hypothesis (2005), which

assumes that the shift parameter correlates with the non-decision time. The data here indicate that without cross-examining with the decision parameters, one may infer that the perceptual priming due to the target preview alters only the shift parameter, and thus supports the stage model account. However, the effect from perceptual priming increases mostly the drift rate, rather than just reducing motoric times (Figure 6-3). This result suggests either the perceptual priming influences also higher cognitive processes, or the shift parameter reflects more than just the motoric time. Either way, the current data suggests further modification of the stage model. As suggested previously (Chapter 4), the geometric characteristics of the RT distribution provide a useful window to understand, but may not directly reflect, cognitive processes.

The effect of attentional guidance on entire RT distributions is visualised in Figure 6-3, plotted as a trellis graph of RT density curves. Figure 6-3 drew three reference lines at 0.3, 0.5, 0.7 and 1 seconds, enabling a visual comparison across the display sizes and cueing conditions. First important observation from the figure is that compared to the visual cue condition, the distributions move towards long latency less when the display size increases. This is especially clear in the null cue condition. The density curves are almost overlapped in the long latency part. Secondly, in the visual cue condition, the range between the reference lines of 0.5 and 1 seconds covers the RTs beyond the mean (in all display sizes). In contrast, in the verbal and null cue conditions, the same range covers RTs around the mean, although the verbal cue condition shows a clear leftwards shift when the display size decrease.

The scale parameter showed a similar pattern with the mean RTs,

replicating previous findings (Palmer, Horowitz, Torralba, & Wolfe, 2011;Chapter 4) and prediction (Wagenmakers & Brown, 2007). This similarity between the two parameters is not unexpected. As can be seen in the Figure 6-6, the density curves in visual search paradigms reflect most evidently the changes of scale parameter when the experimental factors that may lead to inefficient search are contrasted. The two traditional Gaussian parameters, RT means and RT standard deviations follow the scale changes closely. Each of them informs part of what the scale parameter reflects.

The shape parameter aims to reflect how the general pattern of an RT distribution changes. This parameter indicates that when averaged across the display sizes, the distributional shapes are similar between the visual (1.74) and null cue (1.77) conditions, but different between the null and verbal cue (1.60) conditions. The equivalence of shape parameter between the visual and null cue conditions possibly is because the drastic change in the scale parameters (216 vs. 377 ms). As illustrated earlier, when the scale parameter increases, the density curve becomes thickened in the tail and shortened in the middle. This leads to a very different visual impression in the two conditions, even their shape parameters are similar (their scale parameters differ reliably). The shape parameters do show a difference between the verbal and null cue conditions when their scale parameters are similar, suggesting that the attentional guidance due to a verbal cue may change the distributional shape.

In summary, the distributional shapes can be altered by the mechanisms associated with the display size and cue factors. This is supported by the observation that the three distributional parameters showed significant changes

due to the two experimental factors and their interaction. The selective influence on RT shift due to the cue factor is consistent with the argument that to a target preview enhances performance because of perceptual priming. Varying the information gradient of an attentional template, the study compared the mean RTs, accuracy, the decision and distributional parameter when the target foreknowledge was either absent, conceptual, or visual. The study, to the extent of my knowledge, provides the first empirical evidence accounting for the association amongst the attentional guidance, the attentional template and the decision-making processing in visual search by analysing the joint data of RTs and response accuracy. By using the attentional template paradigm, the current study found that visual templates guide and improve search efficiency by accelerating decision rate and reducing non-decision time, but not by changing response threshold. The result suggests it is instructive to categorise the different information carried by an attentional template.

6.5 Conclusion

This study used the HBM to describe RT distributions and the DDM to assess attentional guidance from the visual and verbal templates. The findings suggest that the attentional template exerts differential guidance, depending on whether how WM-related attentional guidance is operated. The evidence of guidance was found in the improvement of the drift rate when a visual or a verbal cue or was used to set up a template, compared to a null cue. The improvement of the drift rate suggests that the decision time is reduced due to focus attention is drawn towards a target quick and/or stronger, so the rate of information accumulation increase. Importantly, the different templates and induced

guidance alter the decision rates and non-decision times, but not the decision threshold. This evidence further supports the guidance account, instead of the explanation of changes of response selection/threshold. This study also showed the template guidance manifests in RT distributions as the changes in the scale and shape parameters. Both parameters alter drastically the RT density curves. The findings derived from dual-model analyses uncover hitherto the unknown aspects in search.

Chapter 7 Concluding Remarks

Psychophysical findings are directly relevant to our daily lives (Li, 2002; Shi & Yang, 2007). Work on visual search, for example, informs devices such as mobile phones, helping them to recognise and track eyes and faces. The knowledge brought out by previous works indicate that early search (Neisser, 1967; Treisman & Sato, 1990; Wolfe, 2007) and possibly attentional selection simultaneously takes in a large constellation of visual features in a visual field (Humphreys & Müller, 1993). The amount of parallel information an observer captures (Theeuwes, 2010) is under the constraint of acuity within a single eye fixation (Dosher et al., 2010), although parallel search might not always operate across all search tasks (Thornton & Gilden, 2007).

The missing puzzle regarding to how a search decision is reached is only starting to be tackled with the recognition that the psychophysical data – error and correct RT distributions – might hide a great deal of information that previous analytic methods had not noticed. Analysing both RT distributions efficiently in a joint framework is a formidable task. This is best illustrated by Ratcliff and Murdock's early works (1978, 1979; Ratcliff & Murdock, 1976b) on memory retrieval and group reaction time distributions. One of the challenges is to collect a vast number of observations (120,000 in total in the two studies). This is more than one-tenth of the 10-year effort on visual search work (1 million observations) that makes determined the 'a-bit-greater-than-2' slope ratio of target-absent to target-present search (Wolfe, 1998b). The sheer number of observations requires not only adequate experimental designs to keep participants motivated, but also an experimenter's perseverance, let alone the ingenious analytic

methods – such as the drift-diffusion models and the Vincentising procedure (Ratcliff, 1979; Vincent, 1912). These two methods were implemented when high-performance computing cluster was scarce. Thanks to the advances of computation power and algorithms, such as the Gibbs sampler, and the OpenBUGS Bayesian language (Lunn et al., 2009), we are now capable of using hierarchical Bayesian models to reduce the number of observations and meanwhile to retain reliable parameter estimations for RT distributions. Although it still requires more empirical data to make certain whether the hierarchical Bayesian method can apply generally on other probability functions (Ratcliff, 2014) or even process models, such as the DDM or Poisson race/counter model (Vandekerckhove, Tuerlinckx, & Lee, 2011; Wiecki et al., 2013), this thesis has demonstrated that the Bayesian modelling approach does improve the parameter estimation greatly when using Weibull function (Farrell & Ludwig, 2008; Rouder et al., 2005) and it can make a significant contribution to understanding human search processes. Below I summarise the significant contributions the thesis makes and discuss possible avenues to expand the current work to explore the neural correlates of search decisions.

7.1 Scientific Contributions

7.1.1 A New Descriptive Method to Fit RT Distributions

In contrast to the ex-Gaussian and other 3-parameter probability functions, the Weibull function, even though it outperforms the Gaussian function, appeared not to be an ideal candidate for fitting RT distributions with maximum likelihood method and chi-square goodness-of-fit (Chapter 4; E. M. Palmer et al., 2011). This drawback of the Weibull function is due to the shape invariance tendency.

When the Weibull shape parameter has high values, increasing the other parameters have little effect on the distributions (Rouder & Speckman, 2004). This Weibull pathology can be resolved with the HBM, which improves parameter estimation even with small sample sizes (Rouder et al., 2005). The problem of chi-square goodness-of-fit also starts to be tackled by, for example, Voss, Voss and Lerche (2014) whose fast-dm programme (version 30) now allows users to fit the DDM with Kolmogorov-Simirnov, chi-square, or maximum likelihood routines. Thus, users can fit the DDM with flexible optimisation routines tweaked to their specific data types.

Along this line of methodological development on distributional analyses, this thesis contributed a particular Weibull HBM routine to fit RT distributions. This routine is built on the general-purpose, open-source programming language, R (R Core Team & others, 2012) linking with the Bayesian language, OpenBUGS (Lunn et al., 2009), so it frees future users to modify it to suit their data without hindrance. This design is intentional. When similar programmes were built using particular commercial tools or highly technical programming languages, they have tended to fade away (Dawson, 1988) or to become confined to one lab (e.g., the DDM in the early years). The Weibull HBM routine presented here can fit RT data generally (collected from other cognitive paradigms), and be modified to fit with other probability functions that are readily available in OpenBUGS (Lunn et al., 2013).

In addition to the methodological contribution, the thesis has contributed original data and codes of analyses in the public domain, so future replication efforts or meta-analyses can build not only on what has been found here, but also

on the original data, applying a similar hierarchical technique to gain further insights via analysing accumulated data (Curran & Hussong, 2009). This is particularly crucial in cognitive neuroscience, in which a tradition of open and accumulating data in a public database, such as done in biology (Benson et al., 2012), has yet to be established (Grethe et al., 2001; Van Horn & Gazzaniga, 2013).

7.1.2 Scientific Findings on Search Decisions

The three studies reported in the thesis examined the role of attentional template on search decisions and how it is represented in VWM might affect different parts of a decision-making process. The findings are clear: the memory strength, the memory representation of a template, the WM operations as well as search display size selectively influence different parts of decision-making and can be probed via examining RT distributions.

This point was illustrated in Study 1 (Chapter 4) in which, relative to feature search, inefficient searches conjoining multiple features affected the drift rate and interacted with the display size when group segmentation might emerge. The shape of RT distributions revealed this particular search operation, manifesting significant changes in the shift and scale parameters. A critical step in the increase of display sizes underlies an emergent change of segmentation process when the search was to work out the difference between two mirrored numbers (2 & 5). This emergent effect, observed in some individuals, likely reflects an increase in the numbers of item in a search unit that can be processed in parallel. The change in the segmentation size reflects in the drift rate and the shift parameter.

Study 2 (Chapter 5) showed the strength of template representation in WM selectively influenced the decision thresholds, but not the drift rate. The direction of the influence depended on whether the timing of a search display is predictable. When it was unpredictable and too short for an observer to be prepared, a strengthened, as opposed to an automatic, template reduced RTs via lowering the decision threshold. A further finding in Study 2 showed a race (LBA) model might fit error distributions better than the DDM.

Study 3 (Chapter 6) found that a visual, as opposed to an abstract, template selectively increased the drift rate. Study 3 found also that, although the increase in display sizes and thereby in local contrast in an odd-one-out search (the null cue condition) played roles in determining search performance, it was also affected by the top-down goal of searching for an odd target. When contrasted those in Study 1 and Study 2, the finding in Study 3 further suggests that the factors that influence RT distributional shapes can affect different parts of a search decision, with the template strength in WM affecting the decision threshold and with the forms and the mental operations of WM representations affecting the decision rate.

One important simplification in the thesis is that the decision-making models assumed a search display results in a rapid perceptual experience driving an underlying diffusion/race process, determining a decision. There are general visual attention theories proposing a simpler mathematical, but more complicated cognitive, formulation to account for attentional selection, object categorization and thereby decision choices in one framework (Theory of Visual Attention, TVA, Bundesen, 1990; Neural TVA, Bundesen, Habekost, & Kyllingsbæk, 2005). In

this cognitive formulation, a race is amongst multiple attentional selections entering VWM. The selection process is to choose different elements/grouped elements in a search display into VWM (in Bundesen's terms, 'select element *x* to category *i*'). The VWM can represent various, but limited number of categories (i.e., attentional templates). Each category races independently with a rate depending on three parameters: pertinence (perceptual priority of a certain category *i*), perceptual decision bias (towards the category) and the strength of the sensory evidence related to the category. In a one-equation form, the mathematical formulation of TVA can be expressed as (Bundesen, 1990):

$$v(x,i) = \eta(x,i)\beta_i \times \frac{\sum_{i \in R} \eta(x,i)\pi_i}{\sum_{z \in S} w_z} \qquad \text{Eq (5)}$$

The π_i represents the pertinence parameter favouring category *i*, which times the parameter of the strength of the sensory evidence, *η(x, i)* to determine the attentional weight of the category. The strength parameter indicates how strong the sensory evidence an element *x* carries leading to a selection into category *i*. The denominator is summed across all elements ($z \in S$) in a visual field and the nominator across all categories ($i \in R$) that may relate *x*'s features to category *i*. This part of the equation determines the overall attentional weight of element, x. The weight then multiplies the parameter of perceptual decision bias, β_i, and another strength parameter to inform the decision (race) rate, *v(x,i)*. The rate refers to how fast a VWM category reaches a decision type. A final decision is determined by a first VWM category, amongst many others, winning the race.

Take a typical conjunction search as an example to illustrate TVA. Upon being informed a target's identify, an observer set up a VWM category, for example, a memory representation for a vertical dark bar. The correlation between the VWM category and other display items produces the attentional weights. The more the features of a display item overlaps with the VWM category, the higher the correlation (pertinence parameter). Also, the more prominent the features, the higher the sensory strength (strength parameter). For example, dark horizontal bars receive higher values than white vertical bar, because colour is a prominent feature. An unspecified (strength) threshold determines how many display items are selected into VWM category. The perceptual bias parameter may be influenced by for example one's tendency to respond to high contrast (while bars; bias paramter). Therefore, the decision rate is a multiplication of (1) the perceptual bias, (2) the sensory strength values correlated with the VWM category, and (3) the weighted sensory strength values correlated with the VWM category.

A common feature between this cognitive formulation and the models used in the thesis is the (drift) rate parameter, but how it is determined and what cognitive process it represents are different. The decision-making models (or denoted sometimes the evidence accumulation models) in this thesis simplified one perceptual experience as a sweep of sampling sensory evidence leading to a perceptual decision, so it may accommodate a very general class of cognitive tasks that involve only simple perceptual decision-making. However, the models only approximated the cognitive complexities that other cognitive oriented models designed to accommodate. The examples include the leakage (i.e., neuronal

lateral inhibition) in the LCA model (Usher & McClelland, 2001), recursive (attentional) selection and (distractor) rejection in the SERR (Humphreys & Müller, 1993), feature/element selection and WM categorisation in the TVA/NTVA (Bundesen, 1990; Bundesen et al., 2005), and the asynchronous entering of diffusors in the ADM (Wolfe, 2007). The general decision-making models fit the data well in highly simplified paradigms, but may not be enough if the models are to apply on real-world search tasks, such as an algorithm in a robot-assisted search mission. Real-world search scenes are highly complex and a successful model is measured by success rates. Simply attaining good fits of data may be only a first step.

7.2 Outlook

7.2.1 Neural Correlates of Search Decisions

One less explored area in the application of the accumulator model (e.g., DDM, LCA, LBA, NTVA etc.)[28] is whether the brain does work as the models envisage (see a review in Purcell et al., 2010). One neural account is that visual cortices encode sensory evidence, for instance, as a form of neuronal firing rates or synchronous firing frequencies across a network of cerebral cortices. When the quantity or strength of the neural signals exceeds a certain threshold (which favouring one of the decisional choices), it then triggers the premotor/ motor cortex to initiate a motor response (either eye saccades or manual movement). This type of neural hypothesis of accumulator model (e.g., NTVA) raises the question of how the process of evidence accumulation is neurally represented,

[28] Broadly speaking, race and diffusion models all belong to the accumulator model.

and how an above-threshold pre-motor neural signal triggers the motor cortex. These questions, and the relevant techniques to address them, are just starting to be unfolded (Purcell et al., 2012; Ratcliff, Philiastides, & Sajda, 2009; Turner, Forstmann, et al., 2013).

Specific to visual search, a recent study recorded activity from two macaque monkeys in the frontal eye field (FEF), the superior colliculus (SC) and the lateral intraparietal (LIP) cortex, when the animals were performing a T/L conjunction search in which they had to make a saccade to the target (Purcell et al., 2012). The firing rates of neuronal population in the FEF were contrasted to the behavioural data accounted for by the accumulator model. Purcell and colleagues (2012) recorded two populations of FEF neurons: the movement-related and the visually responsive neurons. The former neuronal population converged at a threshold of fixed firing rate immediately before the chosen saccade (suggesting an overt decision) regardless of the display size and RT. This observation suggested that, at least in a small population of movement-related neurons in FEF, there exists a neural mechanism for a decision threshold. The neural evidence on the decision rate comes from the firing rates of visually responsive neurons, which decreased with target-distractor similarity (Duncan & Humphreys, 1989). That is, the visual saliency and local contrast elicited by, for example one target amongst numerous homogeneous distractors, positively correlated with firing rates and this correlation was observed selectively only in the visually responsive, instead of movement-related, neurons in FEF. This study is striking, because it suggests that the brain may works as the way the accumulator model envisages and that why it is so successful to account for a

wide range of experimental paradigms (Brown & Heathcote, 2008; Pleskac & Busemeyer, 2010; Ratcliff, 1978; Ratcliff & McKoon, 2008; Ratcliff & Starns, 2009; Roe, Busemeyer, & Townsend, 2001; Usher & McClelland, 2001; Wagenmakers, 2009).

7.2.2 Integration vs. Summation

The inspiring finding from the above single-unit recording study demonstrates there might be a dissociation in the neural representations for the decision threshold and the decision rate. However, the data cannot explain how the process of sensory information accumulation (Bogacz, Brown, Moehlis, Holmes, & Cohen, 2006) and that of the decision threshold switch are represented. One recent debate is whether the process of accumulating evidence is represented by summing sensory information to a pre-existed informative function or by integrating sensory information to an un-informative function (consistent with Bayesian inference). The former view presumes for different cognitive tasks there exist different ideal response functions. For example, a fisherman knows how to conduct a search effectively for finding a shoal of fish, a mental operation that might be represented as pre-existed informative distribution. On the basis of the informative distribution, the process of information accumulation is to sample sensory evidence that fits into this pre-existed distribution. In a new environment conducting an identical task, the fisherman might have presumed what an ideal searching strategy is and applied this strategy (represented as an informative distribution), which at the outset is uncertain. The fisherman samples evidence to reaffirm that the pre-existed function does work (Berkes, Orbán, Lengyel, & Fiser, 2011). Accordingly, s/he

might adjust the decision rate and threshold, because the pre-existed function is getting evidence to confirm its effectively and predictability.

In the latter Bayesian view, by contrast, the fisherman is presumed to be completely naïve towards the task, representing the mental operation as an un-informative function. The evidence sampling process is to integrate the new data into this prior distribution, which then become a different distribution in light of the new data. As a consequence, each new piece of data is integrated into a previous prior function, which then becomes a posterior distribution. The posterior distribution then becomes a new prior distribution for next step of accumulating process. This process goes on until the first prior distribution evolves into a distribution that might lend the fisherman strong confidence to claim 'fishes found'.

Critically, even though being very successful in predicting a large number of behavioural data, the conventional (simple) accumulator model appears to account only for part of the neural activities. In the search paradigm of speed-accuracy-trade-off (SAT), the activities in the FEF movement neurons suggested, contrary to the predicted direction, a higher threshold in the speed-stressed condition than that in the accuracy-stressed condition (Heitz & Schall, 2012). The question regarding to whether the simple accumulator models are adequate to fit neural data or whether a better account of the accumulation process is required, remain to be explored. Some proposed that the inconsistent finding prompts a requirement for adding new mechanisms (Heitz & Schall, 2013), but other constrained the cognitive model with the neural data, arguing that the existing model might still applicable (Cassey, Heathcote, & Brown, 2014). Further human

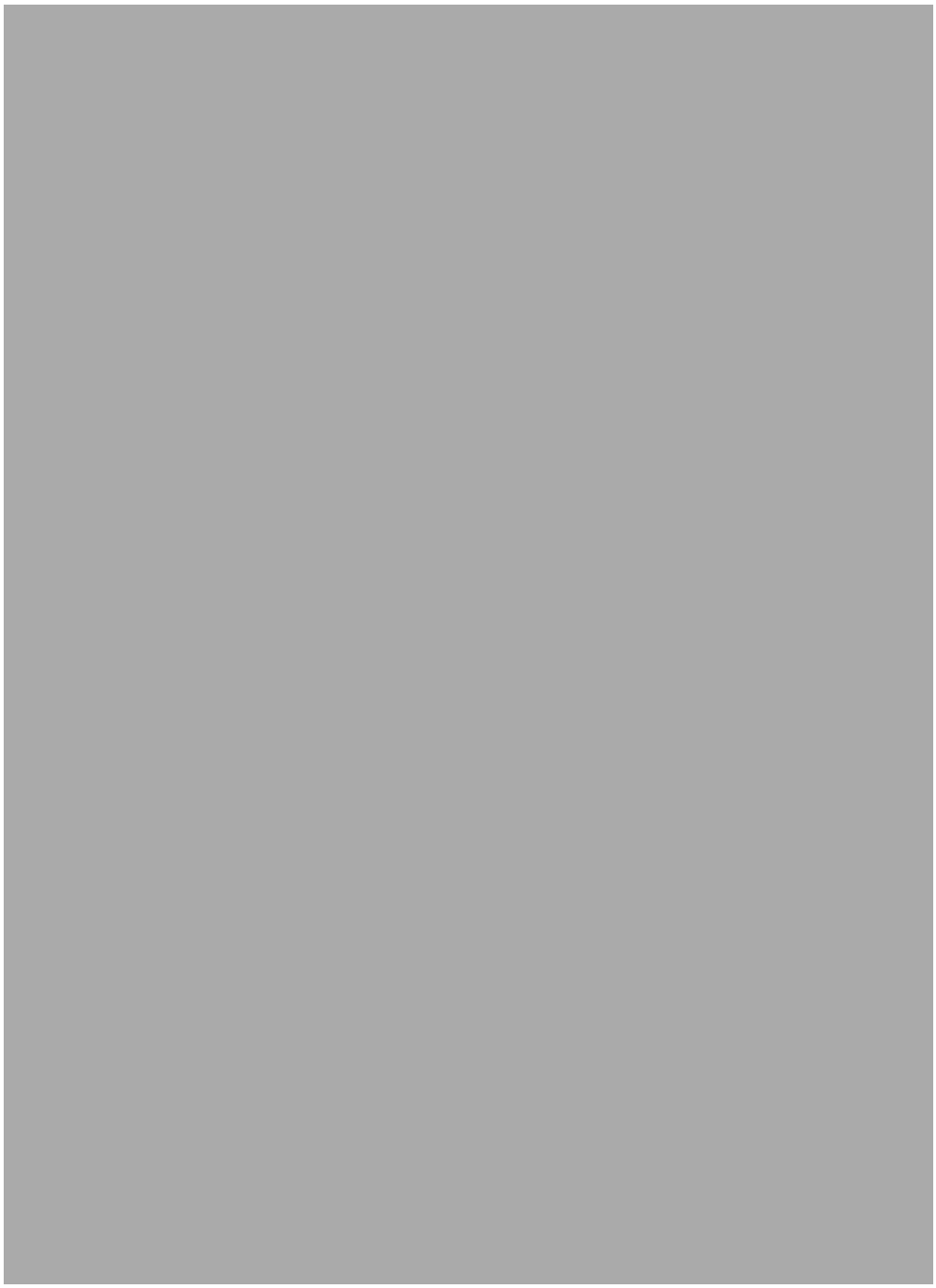
neurophysiological data have suggested that when car/face images were subjected to various degrees of blurring, two ERP components correlate selectively with decision times and non-decision times (Ratcliff et al., 2009). The pioneering work demonstrates one way of investigating human neurophysiological data within the decision-making framework. However, neither the ERP study can provide a clear account to answer how the neural signals (EEG recoding from the scalp) associate with the direction of decision threshold (see also the finding related to this question in Chapter 5). It remains unclear how a mathematical model can link neural data to behavioural data (but see Turner, Forstmann, et al., 2013) and why the saccadic/single-unit recording data sometimes suggested reversed direction of decision thresholds in the SAT search paradigm (Heitz & Schall, 2012).

7.3 General Conclusion

The thesis harnessed both correct and error RT distributions to address the questions related to search decisions when an attentional template is explicitly represented in VWM. The results are overall consistent with psychological theories such as the biased competition theory of selective attention (Desimone & Duncan, 1995; Duncan et al., 1997). That is, at least at the level of decision choices (present vs. absent or left vs. right), the results showed good fits to the data and provided reasonable accounts when modelling two alternative choices, either by one or two accumulator(s), competing for a perceptual decision, which depend on WM strength and representations. The questions related to mutual (neuronal) inhibition or leaky of neuronal/sensory evidence during the competition process remain to be explored.

The notion of competition was implemented in the DDM as an accumulator drifts to either alternative (2AFC) or two accumulators race towards a common threshold (LBA). Instead of directly contrasting two or multiple competing WM representations (designed in paradigms), the thesis asked how an attentional template might affect selective attention when the template strength and the template status were changed. The findings indicate that depending on how an observer operates WM, different template representations affect contrasting parts of a decision-making process and they manifest as different shapes of RT distributions, reflecting in different distributional parameters. A future direction may be further to explicitly model other important neural competition mechanisms, such as mutual inhibition of interneurons and signal leakage in a neuron. These biologically inspired modelling will ultimately lead to a complete account of accumulator model and provide intelligent applications that directly influence our life about how to make an optimal search decision.

Appendix A The Ethics Approval Letter for the Psychophysical Experiments



Appendix B Software Developed in the Project

Section B.1 is the source code, programmed in R, for estimating Weibull distributional parameters for the feature search data. The main function of this programme is to read in raw data and to reformat them be processed by BUGS model. Similar programmes with small tweaks were used in the data collected from other similar searches paradigms. The programme can be modified to fit other types of data or even applied with other probability functions.

Section B.2 is the source code of the BUGS model. In the case of this project, it is a hierarchical Weibull model, designed to fit RTs at the first level. The hyperparameters (higher hierarchical levels) are modelled with other probability functions. I implemented pseudo-Poisson trick to model the 3-parameter (shift) Weibull function, because the available Weibull function in the latest OpenBUGS (version 3.2.3 rev 1012) appears to contain (programming) bugs and other 3-parameter, for example gamma, functions have not implemented in OpenBUGS. Specifically, one bug is its inconsistent variable naming, as I found out in its source codes, using BlackBox Component Builder (version 1.6 built on 11.10.2013).

Section B.3 is the source code of PsyToolkit (Stoet, 2010) for the feature search task. The stimulus table is omitted due to its large volume. The table was generated by another R programme, separately for each participant run.

B.1 Data formatting Source Code

```
# Disclaimer----------------------------------------
# Author: Yishin Lin
```

```
# Last modified Date: 21 May, 2013
# Description: Run HBM
# Load data and functions-----------------------------
# Gelman and Hill's (2007) random imputation function
load('./data/myData/featureT.RData')
source("./functions/random.imp.R")

# Set up sequences of each factor----------------------
size.seq <- sort(unique(featureT$size))
target.seq <- sort(unique(featureT$target))
# I should have 19 (out of 20 participants)
# valid participants
NSubj <- length(unique(featureT$subj))

# Start the for loop-----------------------------------
for(i in seq(along = size.seq)){
  subdata1 <- subset(featureT, size == size.seq[i])
  if (nrow(subdata1)  == 0) next
  for(j in seq(along = target.seq)){
    subdata2 <- subset(subdata1, target %in% target.seq[j] )
    if (nrow(subdata2)  == 0) next

    # Processing the data now-------------------------
    library(plyr)
    dataPerSubj <- ddply(subdata2, .(subj), summarise,
                         N = length(rt.sec),
                         MeanRT = mean(rt.sec),
```

```
                MinRT = min(rt.sec))
minrt  <- dataPerSubj$MinRT

dataPerSubj.ordered <- dataPerSubj[order(dataPerSubj$N,
                                         decreasing=TRUE),]
dataPerSubj.ordered$subj.o <- dataPerSubj.ordered$subj
dataPerSubj.ordered$subj <- 1:nrow(dataPerSubj.ordered)
minrt <- dataPerSubj.ordered$MinRT

rtSample <- NULL
for(sIdx in dataPerSubj.ordered$subj.o) {
  tmpContainer <- subset(subdata2, subj == sIdx,
                         select = 'rt.sec')[,1]

  if(!is.null(rtSample) && ncol(rtSample) >
     length(tmpContainer))
    {
    tmp <- ncol(rtSample) - length(tmpContainer)
    tmpContainerNA <- append(tmpContainer, rep(NA, tmp))
    tmpContainer <- random.imp(tmpContainerNA)
    }
  rtSample <- rbind(rtSample, tmpContainer)
}

# Store data as a list ---------------------------------
dataList <- list(NSubj = NSubj,
                 NTrials = ncol(rtSample),
```

```
            y = structure(.Data = rtSample,
                          .Dim = dim(rtSample)),
            minrt = minrt)

# Save data files as JAG format ------------------
library(BRugs)
fileNameRoot <- paste('data', size.seq[i], target.seq[j],
                  sep='')
BRugsName <- paste('./data/myData/BayesDataF/',
                fileNameRoot, 'BRugs.txt', sep='')
bugsData(dataList, fileName = BRugsName)
detach(package:BRugs)

library(R2jags)
JAGSName <- paste('./data/myData/BayesDataF/',
                fileNameRoot, 'JAGS.txt', sep='')
bugs2jags(BRugsName, JAGSName)

# Set initial prior values-----------------------------
# for each participant, set prior beta=0.9;  lambda=12,
# theta (scale) = lambda^(-1/beta);
# psi = min(each particpant) - min in the group
    thetaInit <-  runif(NSubj, .3, 4)
    betaInit <- runif(NSubj, 0.9, 2)
    psiInit <- dataPerSubj.ordered$Min -
  min(dataPerSubj.ordered$Min)
```

```
# Initialization ---------------------------------------------
# Rouder et al.'s (2005) initial values
# "to simulate analysis of real data, these should be
# reasonable, but far from true values"
#*****************************************************
 initList <- function(){
   list(beta = betaInit,
        psi = psiInit,
        theta = thetaInit,
       eta1 = .1 , eta2 = .1,
       xi1 = .1, xi2 = .1)
 }

# Define parameters to monitor ---------------------------------
parameters <- c('beta', 'psi', 'theta')
nb <-  5000; nbt <-  10
nc <- 3; nct <- 1
ni <-  105000; nit <-  1000
nt <- 4; ntt <- 1

# RUN THE CHAINS (Create, initialize, and adapt the model:)
# Test run ----------------------------------------------------
# jagsfitT <- jags(data=JAGSName,
# model.file='./BayesRuns/model.txt',
# inits=initList, parameters.to.save=parameters,
# n.iter=nit, n.chain = nct, n.burnin = nbt,
# n.thin = ntt)
```

```
# Run ------------------------------------------------------
jagsfit <- jags(data=JAGSName,
                model.file='./BayesRuns/model.txt',
                inits=initList, parameters.to.save=parameters,
                n.iter=ni, n.chain = nc, n.burnin = nb,
                n.thin = nt)


pathWithNameRoot <- paste('./data/myData/BayesDataF/',
                      fileNameRoot, sep="")
imageName <- paste(pathWithNameRoot, '.RData', sep="")
save.image(file=imageName)
  }
}
```

B.2 Model Source Code

```
data {
    for (i in 1:NSubj) {
            for (j in 1:NTrials){
                    zeros[i,j] <- 0
            }
    zero[i] <- 0
    }
}


model {
C <- 100000   # a big constant to certain phi is non-negative
# Likelihood
```

```
for (i in 1:NSubj){
for (j in 1:NTrials){

#-------------------------#
# Weibull density       #
#-------------------------#
term1[i,j] <- beta[i]*log(theta[i]) + pow(y[i,j] - psi[i],beta[i])/pow(theta[i],beta[i])
term2[i,j] <- log(beta[i]) + (beta[i]-1)*log(y[i,j] - psi[i])

# zeros trick
phi[i,j] <- term1[i,j] - term2[i,j] + C
zeros[i,j] ~ dpois(phi[i,j])
}

#--------------------------------#
# beta prior (shape)          #
#--------------------------------#
# to ensure the posterior moments exist for the beta[i]
is.censored[i] ~ dinterval(beta[i], 0.01)
beta[i] ~ dgamma(eta1,eta2)   # I(0.01,), BUGS's old way to truncate

#--------------------------------#
# theta prior (scale)         #
#--------------------------------#
theta[i] ~ dunif(.01, 10000)  # modify from dflat()

#-------------------------------#
```

```
    # pow(theta[i], beta[i])   #

    #------------------------------#

    phip[i] <- (xi1 + 1)*log(pow(theta[i],beta[i])) + xi2/pow(theta[i],beta[i]) + loggam(xi1)
- xi1*log(xi2)

    zero[i] ~ dpois(phip[i])


    # Quote from Rouder et al.(2003, Psychometrika, p593)

    # ..., the hierarchical priors in β and θ^β yield shrinkage in all three

    # parameters.  It is not clear that a hierarchical prior on

    # psi (shift) would yield any additional gain.

    psi[i] ~ dunif(0,minrt[i])

    }


    #----------------------#

    # Hyper-prior       #

    #----------------------#

    # Priors as recommended by Rouder et al (2003).

    # how the shape and scale vary across individuals within the population

    # hyperpriors on xi1,..,eta2. In Rouder et al's original code

    eta1 ~ dgamma(2,0.02)   # c1 = 2, d1 = .02

    eta2 ~ dgamma(2,0.04)   # c2 = 2, d2 = .04

    xi1 ~ dgamma(2,0.1)     # a1 = 2, b1 = .1

    xi2 ~ dgamma(2,2.85)    # a2 = 2, b2 = 2.85

    }
```

B.3 Paradigm Source Code

```
# This programme is modified from Stoet's (2011) visual search task
```

```
# 1152x864 is SONY's resolution with 100 Hz RF.
# Author: Yi-Shin Lin
options
        escape
        centerzero
        bitmapdir bitmapssearch
        resolution 1152 864
        parallelport out 2 3 4 5 6 7 8 9 # allow set parallel port to send out high voltage
        vsync_off
        cedrus

bitmaps
        instruction
        prcFullsetCompleted
        takeABreak
        lastOne
        finish3Sections
        finish5Sections
        finish7Sections
        finish9Sections
        target
        distractor1
        distractor2
        empty
        fixpoint
        frame
        tooearly
```

```
        toolate


table feature # This table is created in search.R and searchTabFun.R
# Table has been omitted. To compile this programme, you need to insert a
# new stimulus table


task search
    table feature
    c-begin
        set_pin(LP_PIN02);
        psy_show_centered_bitmap(fixpoint, 0, 0, 1);  // image1
        psy_delay(500);
        set_pin(LP_PIN03);
        psy_show_centered_bitmap(empty, 0, -30, 1);   // image2
        /* the cue bitmap stays on the screen for 200 ms.*/
        psy_delay(200);
        psy_delay(feature[tablerow].c4);                    // The duration of cue-
offset to target-onset is @4, defined at isiSet in R file.
    c-end
    cedrus clear       # clean any cedrus guess press before presenting the target set
    draw off
                show bitmap @4 @5 @6              # 3; target; [bitmapname] (xpos
ypos)(width height)
                show bitmap @7 @8 @9              # 4
                show bitmap @10 @11 @12 # 5
                show bitmap @13 @14 @15 # 6
                show bitmap @16 @17 @18 # 7
                show bitmap @19 @20 @21 # 8
```

```
                show bitmap @22 @23 @24 # 9
                show bitmap @25 @26 @27 # 10
                show bitmap @28 @29 @30 # 11
                show bitmap @31 @32 @33 # 12
                show bitmap @34 @35 @36 # 13
                show bitmap @37 @38 @39 # 14
                show bitmap @40 @41 @42 # 15
                show bitmap @43 @44 @45 # 16
                show bitmap @46 @47 @48 # 17
                show bitmap @49 @50 @51 # 18
                show bitmap @52 @53 @54 # 19
                show bitmap @55 @56 @57 # 20
                show bitmap @58 @59 @60 # 21
                show bitmap @61 @62 @63 # 22
        c-begin
          set_pin(LP_PIN04);
        c-end
 draw on
 c-begin
 /*  Record RT, using Cedrus RB-830. If longer than 4 seconds, terminate the trial. */
 keystatus = psy_cedrus_readkey(feature[tablerow].c2, 4000);
 set_pin(LP_PIN05);
 /*-------------------------------------------------------------------------*/
 psy_clear_screen();
 /* Reset all parallel pin back to 0 voltage. This defines the trial finish event */
 clear_pin(LP_PIN02 | LP_PIN03 | LP_PIN04 | LP_PIN05 | LP_PIN06 | LP_PIN07 |
LP_PIN08 | LP_PIN09);
 psy_delay(800);
```

```
 /*   c1-counter(); c2-target present(present-1/absent-2); c3-setsize;
      keystatus.key records which cedrus key (7 or 1) was pressed;
      keystatus.status records correct, incorrect, too early or timeout;
      keystatus.externaltime1 (EXTRT) record RT from the Cedrus key pad
 */
 fprintf(datafile, "%s %d %d %d %d %d \n", blockname,
           feature[tablerow].c1, feature[tablerow].c2,
           feature[tablerow].c3, keystatus.key + 1, keystatus.status,
           keystatus.externaltime1);
 c-end

#---------------------------------------------------------------------
# blocks
# ---------------------------------------------------------------------

block prc
 pager instruction
 delay 1500
 tasklist
  search 8 fixed
 end
 system R CMD BATCH dataDump.r

message prcFullsetCompleted

block feature1
 delay 1500
```

```
 tasklist
  search 80 fixed
 end
 system R CMD BATCH dataDump.r

message takeABreak

block feature2
 delay 1500
 tasklist
  search 80 fixed
 end
 system R CMD BATCH dataDump.r

message takeABreak

block feature3
 delay 1500
 tasklist
  search 80 fixed
 end
 system R CMD BATCH dataDump.r

message finish3Sections

block feature4
 delay 1500
```

```
 tasklist
  search 80 fixed
 end
 system R CMD BATCH dataDump.r

message takeABreak

block feature5
 delay 1500
 tasklist
  search 80 fixed
 end

message finish5Sections

block feature6
 delay 1500
 tasklist
  search 80 fixed
 end
 system R CMD BATCH dataDump.r

message takeABreak

block feature7
 delay 1500
 tasklist
```

```
   search 80 fixed
  end
  system R CMD BATCH dataDump.r

message finish7Sections

block feature8
  delay 1500
  tasklist
    search 80 fixed
  end
  system R CMD BATCH dataDump.r

message takeABreak

block feature9
  delay 1500
  tasklist
    search 80 fixed
  end
  system R CMD BATCH dataDump.r

message lastOne

block feature10
  delay 1500
  tasklist
```

```
  search 80 fixed
end
system R CMD BATCH dataDump.r
bitmap_from_file ./bitmapssearch/completed.png
wait_for_key
```

Appendix C Fitting Other Three-parameter Functions

This appdneix illustrates one way to fit a three-parameter probability function other than the Weibull function, using the BUGS model code in the thesis. This illustration tests whether the Weibull function fit better than the gamma function. I built a 3-parameter gamma function in the HBM framework. Because a typical gamma function contains two parameters and there is no a pre-built 3-parameter gamma function in Bayesian inference Using Gibbs Sampling (BUGS) programme, I used Johnson, Kotz, and Balakrishnan's (1994, pp 337, eq. 17.1) equation to implement the gamma function directly. The BUGS code is to change the Weibull density function to (see Appendix B.2 for the complete BUGS source code):

```
#--------------------------#
# Gamma density            #
#--------------------------#
term1[i,j] <- beta[i]*log(theta[i]) + (y[i,j] - psi[i])/theta[i] + loggam(beta[i])
term2[i,j] <- (beta[i]-1)*log(y[i,j] - psi[i])
```

Similar to the way in implementing other functions, I assessed the parameters via a minus log-likelihood and used the pseudo-Poisson (zero) trick (Spiegelhalter, Thomas, Best, & Lunn, 2007). This implementation resulted in unstable, non-converged estimations. Take the shape parameter as an example. When estimating the parameters in a participant's present trial responses in the spatial configuration search (display size 6), the estimation yielded three different posterior distributions and the trace plots from the three chains unstably oscillated

around different ranges (Figure C-1 to Figure C-3) [29]. In addition, the autocorrelation plots indicated a problem and this did not abate with increasing iterations. The diagnostic plots showed that the gamma function does not converge when fitted in the HBM framework.

[29] See Chapter 4 for the details of this study.

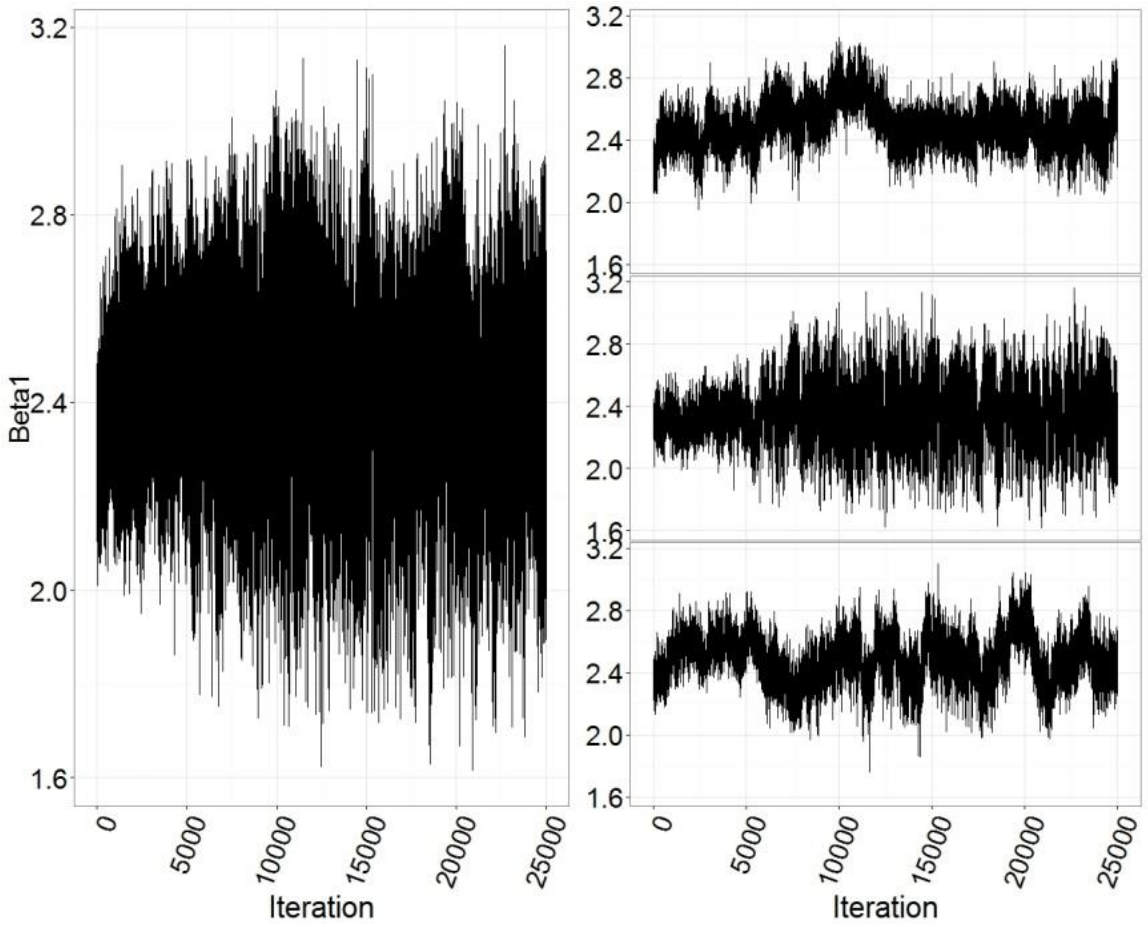


Figure C-1. The trace plots show non-converged parameter estimation using gamma function. Right panel plotted three simulation chains separately.

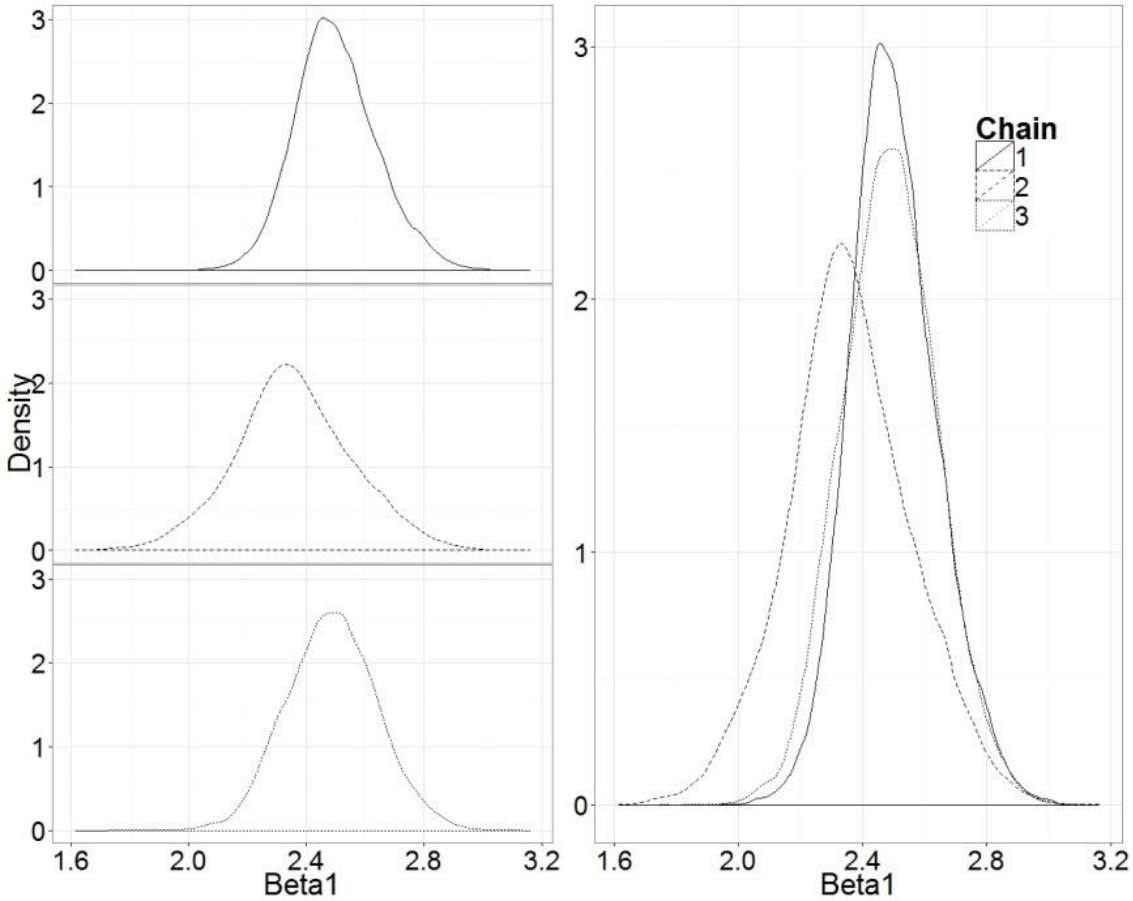


Figure C-2. The HBM gamma fit suggests three possible posterior density distributions. The result suggests that there are three underlying distributions that are able to generate the modelled dataset. Left panel plotted three simulation chains separately.

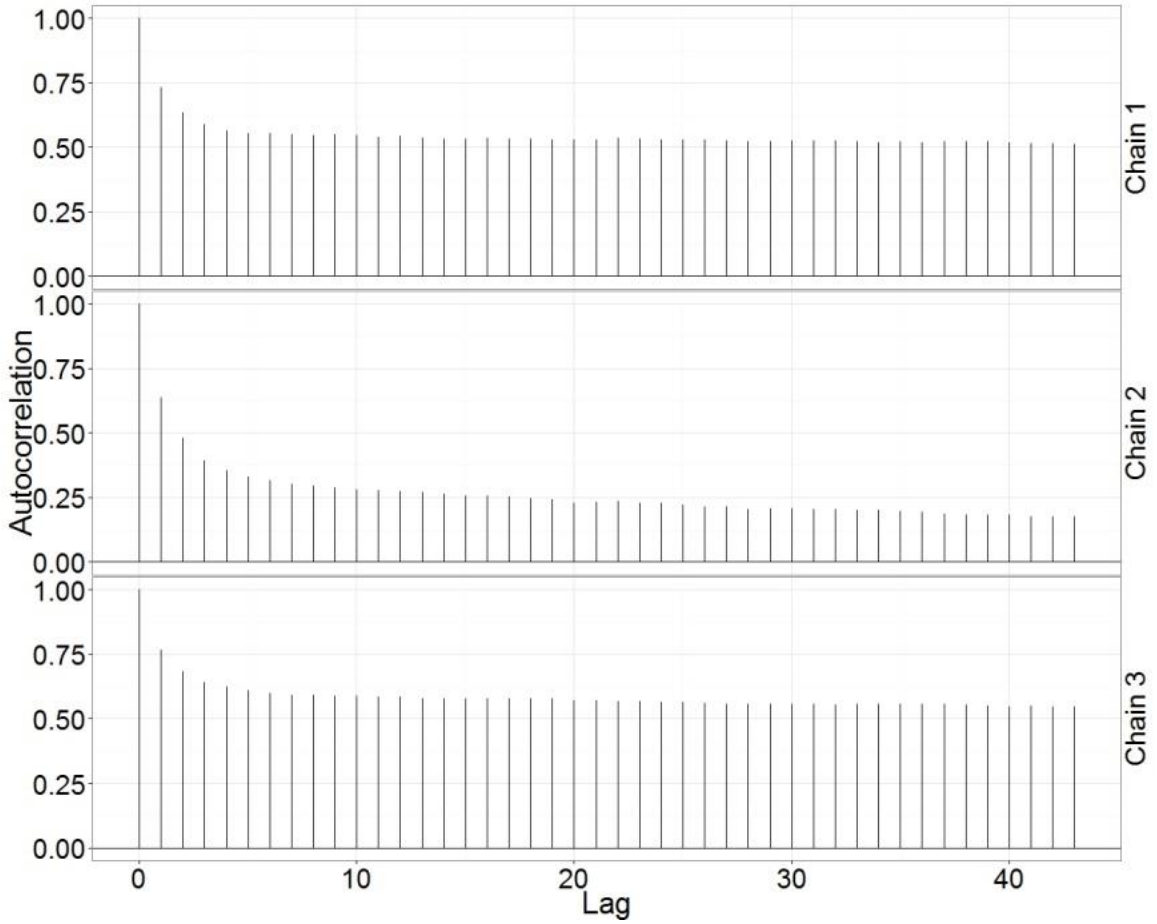


Figure C-3. The autocorrelation plots show correlated estimations even after long iterations, using the gamma function.

In summary, the non-converged gamma fit could be due to that (1) it is not suitable for HBM in this context, and/or (2) the gamma function indeed fits worse than Weibull function (as the DIC suggests). When fitting the gamma function in HBM, I set a high number of iteration (i.e., 105000) and reasonable thinning length, but this setting still cannot resolve the non-converged problem.

Reference

Aad, G., Abajyan, T., Abbott, B., Abdallah, J., Abdel Khalek, S., Abdelalim, A. A., … Zwalinski, L. (2012). Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Physics Letters B*, *716*(1), 1–29. http://doi.org/10.1016/j.physletb.2012.08.020

Anderson, G. M., Heinke, D., & Humphreys, G. W. (2010). Featural guidance in conjunction search: The contrast between orientation and color. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(5), 1108–1127. http://doi.org/10.1037/a0017179

Ansorge, U., & Becker, S. I. (2014). Contingent capture in cueing: the role of color search templates and cue-target color relations. *Psychological Research-Psychologische Forschung*, *78*(2), 209–221. http://doi.org/10.1007/s00426-013-0497-5

Arita, J. T., Carlisle, N. B., & Woodman, G. F. (2012). Templates for rejection: Configuring attention to ignore task-irrelevant features. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(3), 580–584. http://doi.org/10.1037/a0027885

Ashby, F. G., Tein, J.-Y., & Balakrishnan, J. D. (1993). Response Time Distributions in Memory Scanning. *Journal of Mathematical Psychology*, *37*(4), 526–555. http://doi.org/10.1006/jmps.1993.1033

Balota, D. A., & Spieler, D. H. (1999). Word frequency, repetition, and lexicality effects in word recognition tasks: Beyond measures of central tendency. *Journal of Experimental Psychology: General*, *128*(1), 32–55. http://doi.org/10.1037/0096-3445.128.1.32

Balota, D. A., & Yap, M. J. (2011). Moving Beyond the Mean in Studies of Mental Chronometry The Power of Response Time Distributional Analyses. *Current Directions in Psychological Science*, *20*(3), 160–166. http://doi.org/10.1177/0963721411408885

Balota, D. A., Yap, M. J., Cortese, M. J., & Watson, J. M. (2008). Beyond mean response latency: Response time distributional analyses of semantic priming. *Journal of Memory and Language*, *59*(4), 495–523. http://doi.org/10.1016/j.jml.2007.10.004

Barate, R., De Bonis, I., Decamp, D., Ghez, P., Goy, C., Jezequel, S., … Zobernig, G. (2000). Observation of an excess in the search for the Standard Model Higgs boson at ALEPH. *Physics Letters B*, *495*(1–2), 1–17. http://doi.org/10.1016/S0370-2693(00)01269-7

Bayes, T. (1970). An essay towards solving a problem in the doctrine of chances. *Studies in the History of Statistics and Probability*, *1*, 134–153.

Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., & Sayers, E. W. (2012). GenBank. *Nucleic Acids Research*, gks1195. http://doi.org/10.1093/nar/gks1195

Bergen, J., & Julesz, B. (1983). Parallel Versus Serial Processing in Rapid Pattern-Discrimination. *Nature*, *303*(5919), 696–698. http://doi.org/10.1038/303696a0

Berkes, P., Orbán, G., Lengyel, M., & Fiser, J. (2011). Spontaneous Cortical Activity Reveals Hallmarks of an Optimal Internal Model of the Environment. *Science*, *331*(6013), 83–87. http://doi.org/10.1126/science.1195870

Beutter, B. R., Eckstein, M. P., & Stone, L. S. (2003). Saccadic and perceptual performance in visual search tasks. I. Contrast detection and discrimination.

JOSA A, *20*(7), 1341–1355.

Bogacz, R., Brown, E., Moehlis, J., Holmes, P., & Cohen, J. D. (2006). The physics of optimal decision making: A formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological Review*, *113*(4), 700–765. http://doi.org/10.1037/0033-295X.113.4.700

Bravo, M. J., & Farid, H. (2012). Task demands determine the specificity of the search template. *Attention, Perception, & Psychophysics*, *74*(1), 124–131. http://doi.org/10.3758/s13414-011-0224-5

Bravo, M. J., & Farid, H. (2014). Informative cues can slow search: The cost of matching a specific template. *Attention Perception & Psychophysics*, *76*(1), 32–39. http://doi.org/10.3758/s13414-013-0532-z

Bricolo, E., Gianesini, T., Fanini, A., Bundesen, C., & Chelazzi, L. (2002). Serial Attention Mechanisms in Visual Search: A Direct Behavioral Demonstration. *Journal of Cognitive Neuroscience*, *14*(7), 980–993. http://doi.org/10.1162/089892902320474454

Brown, S., & Heathcote, A. (2005). A ballistic model of choice response time. *Psychological Review*, *112*(1), 117–128. http://doi.org/10.1037/0033-295X.112.1.117

Brown, S., & Heathcote, A. (2008). The simplest complete model of choice response time: linear ballistic accumulation. *Cognitive Psychology*, *57*(3), 153–178. http://doi.org/10.1016/j.cogpsych.2007.12.002

Bruhn, P., & Bundesen, C. (2012). Anticipation of visual form independent of knowing where the form will occur. *Attention Perception & Psychophysics*, *74*(5), 930–941. http://doi.org/10.3758/s13414-012-0296-x

Bundesen, C. (1990). A Theory of Visual-Attention. *Psychological Review*, *97*(4), 523–547. http://doi.org/10.1037//0033-295X.97.4.523

Bundesen, C., Habekost, T., & Kyllingsbæk, S. (2005). A Neural Theory of Visual Attention: Bridging Cognition and Neurophysiology. *Psychological Review*, *112*(2), 291–328. http://doi.org/10.1037/0033-295X.112.2.291

Burnham, K. P., & Anderson, D. R. (2004). Multimodel Inference Understanding AIC and BIC in Model Selection. *Sociological Methods & Research*, *33*(2), 261–304. http://doi.org/10.1177/0049124104268644

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

Cassey, P., Heathcote, A., & Brown, S. D. (2014). Brain and Behavior in Decision-Making. *PLoS Comput Biol*, *10*(7), e1003700. http://doi.org/10.1371/journal.pcbi.1003700

Castelhano, M. S., Pollatsek, A., & Cave, K. R. (2008). Typicality aids search for an unspecified target, but only in identification and not in attentional guidance. *Psychonomic Bulletin & Review*, *15*(4), 795–801. http://doi.org/10.3758/PBR.15.4.795

Cavanagh, J. F., Wiecki, T. V., Cohen, M. X., Figueroa, C. M., Samanta, J., Sherman, S. J., & Frank, M. J. (2011). Subthalamic nucleus stimulation reverses mediofrontal influence over decision threshold. *Nature Neuroscience*, *14*(11), 1462–1467. http://doi.org/10.1038/nn.2925

Cave, K. R., & Wolfe, J. M. (1990). Modeling the role of parallel processing in visual search. *Cognitive Psychology*, *22*(2), 225–271. http://doi.org/10.1016/0010-

0285(90)90017-X

Chelazzi, L. (1999). Serial attention mechanisms in visual search: A critical look at the evidence. *Psychological Research*, *62*(2-3), 195–219. http://doi.org/10.1007/s004260050051

Christie, L. S., & Luce, R. D. (1956). Decision structure and time relations in simple choice behavior. *The Bulletin of Mathematical Biophysics*, *18*(2), 89–112. http://doi.org/10.1007/BF02477834

Chun, M. M., & Wolfe, J. M. (1996). Just Say No: How Are Visual Searches Terminated When There Is No Target Present? *Cognitive Psychology*, *30*(1), 39–78. http://doi.org/10.1006/cogp.1996.0002

Chun, M. M., & Wolfe, J. M. (2001). Visual attention. In E. B. Goldstein (Ed.), *Blackwell handbook of perception* (pp. 272–310). Oxford, UK; Malden, Mass., USA: Blackwell.

Cleveland, W. S., Grosse, E., & Shyu, W. M. (1992). Local regression models. In T. J. Hastie & J. M. Chambers (Eds.), *Statistical models in S* (pp. 309–376). Pacific Grove: Wadsworth & Brooks-Cole advanced books & software.

Cousineau, D., Brown, S., & Heathcote, A. (2004). Fitting distributions using maximum likelihood: Methods and packages. *Behavior Research Methods, Instruments, & Computers*, *36*(4), 742–756. http://doi.org/10.3758/BF03206555

Cousineau, D., & Shiffrin, R. M. (2004). Termination of a visual search with large display size effects. *Spatial Vision*, *17*(4), 327–352. http://doi.org/10.1163/1568568041920104

Crawley, M. J. (2002). *Statistical computing: an introduction to data analysis using S-Plus*. Chichester, West Sussex, England; New York: Wiley.

Curran, P. J., & Hussong, A. M. (2009). Integrative data analysis: The simultaneous analysis of multiple data sets. *Psychological Methods*, *14*(2), 81–100. http://doi.org/10.1037/a0015914

Czerwinski, M., Lightfoot, N., & Shiffrin, R. M. (1992). Automatization and Training in Visual Search. *The American Journal of Psychology*, *105*(2), 271–315. http://doi.org/10.2307/1423030

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

Dawson, M. R. W. (1988). Fitting the ex-Gaussian equation to reaction time distributions. *Behavior Research Methods, Instruments, & Computers*, *20*(1), 54–57. http://doi.org/10.3758/BF03202603

Desimone, R., & Duncan, J. (1995). Neural Mechanisms of Selective Visual-Attention. *Annual Review of Neuroscience*, *18*, 193–222. http://doi.org/10.1146/annurev.neuro.18.1.193

Diederich, A. (1995). Intersensory Facilitation of Reaction Time: Evaluation of Counter and Diffusion Coactivation Models. *Journal of Mathematical Psychology*, *39*(2), 197–215. http://doi.org/10.1006/jmps.1995.1020

Di Lollo, V., Enns, J. T., & Rensink, R. A. (2000). Competition for consciousness among visual events: The psychophysics of reentrant visual processes. *Journal of Experimental Psychology: General*, *129*(4), 481–507. http://doi.org/10.1037/0096-3445.129.4.481

Dolan, C. V., van der Maas, H. L. J., & Molenaar, P. C. M. (2002). A framework for ML estimation of parameters of (mixtures of) common reaction time distributions given optional truncation or censoring. *Behavior Research Methods,*

Instruments, & Computers, *34*(3), 304–323. http://doi.org/10.3758/BF03195458

Donkin, C., Brown, S., & Heathcote, A. (2009). The overconstraint of response time models: Rethinking the scaling problem. *Psychonomic Bulletin & Review*, *16*(6), 1129–1135. http://doi.org/10.3758/PBR.16.6.1129

Donkin, C., Brown, S., & Heathcote, A. (2011). Drawing conclusions from choice response time models: A tutorial using the linear ballistic accumulator. *Journal of Mathematical Psychology*, *55*(2), 140–151. http://doi.org/10.1016/j.jmp.2010.10.001

Dosher, B. A., Han, S., & Lu, Z.-L. (2010). Information-limited parallel processing in difficult heterogeneous covert visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(5), 1128–1144. http://doi.org/10.1037/a0020366

Dowd, E. W., Kiyonaga, A., Egner, T., & Mitroff, S. R. (2015). Attentional guidance by working memory differs by paradigm: An individual-differences approach. *Attention, Perception, & Psychophysics*, *77*(3), 704–712. http://doi.org/10.3758/s13414-015-0847-z

Downing, P. E. (2000). Interactions Between Visual Working Memory and Selective Attention. *Psychological Science*, *11*(6), 467–473. http://doi.org/10.1111/1467-9280.00290

Draper, N. R., & Smith, H. (1998). *Applied regression analysis*. New York: Wiley.

Duncan, J. (1985). Visual search and visual attention. In M. I. Posner, O. S. M. Marin, & International Symposium on Attention and Performance (Eds.), *Attention and performance XI* (pp. 85–106). Hillsdale, N.J.: L. Erlbaum Associates.

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity.

Psychological Review, *96*(3), 433–458. http://doi.org/10.1037/0033-295X.96.3.433

Duncan, J., Humphreys, G., & Ward, R. (1997). Competitive brain activity in visual attention. *Current Opinion in Neurobiology*, *7*(2), 255–261. http://doi.org/10.1016/S0959-4388(97)80014-1

Dzhafarov, E. N. (1992). The structure of simple reaction time to step-function signals. *Journal of Mathematical Psychology*, *36*(2), 235–268. http://doi.org/10.1016/0022-2496(92)90038-9

Dzhafarov, E. N. (2003). Selective influence through conditional independence. *Psychometrika*, *68*(1), 7–25. http://doi.org/10.1007/BF02296650

Dzhafarov, E. N., & Gluhovsky, I. (2006). Notes on selective influence, probabilistic causality, and probabilistic dimensionality. *Journal of Mathematical Psychology*, *50*(4), 390–401. http://doi.org/10.1016/j.jmp.2006.03.003

Egeth, H. E., Virzi, R. A., & Garbart, H. (1984). Searching for conjunctively defined targets. *Journal of Experimental Psychology: Human Perception and Performance*, *10*(1), 32–39. http://doi.org/10.1037/0096-1523.10.1.32

Enoch, J. M. (1959). Effect of the Size of a Complex Display upon Visual Search. *Journal of the Optical Society of America*, *49*(3), 280–285. http://doi.org/10.1364/JOSA.49.000280

Enoch, J. M. (1960). Natural tendencies in visual search of a complex display. In *Symposium on Visual Search Techniques. Publication* (Vol. 712).

Farrell, S., & Ludwig, C. J. H. (2008). Bayesian and maximum likelihood estimation of hierarchical response time models. *Psychonomic Bulletin & Review*, *15*(6), 1209–1217. http://doi.org/10.3758/PBR.15.6.1209

Fechner, G. (1860). *Elemente der Psychophysik*. (H. Adler, Trans.). Leipzig; Amsterdam; New York: Breitkopf & Härtel; Bonset; Holt, Rinehart & Winston.

Fecteau, J. H., & Enns, J. T. (2005). Visual letter matching: Hemispheric functioning or scanning biases? *Neuropsychologia*, *43*(10), 1412–1428. http://doi.org/10.1016/j.neuropsychologia.2005.01.006

Feige, B., Biscaldi, M., Saville, C. W. N., Kluckert, C., Bender, S., Ebner-Priemer, U., … Klein, C. (2013). On the Temporal Characteristics of Performance Variability in Attention Deficit Hyperactivity Disorder (ADHD). *PLoS ONE*, *8*(10), e69674. http://doi.org/10.1371/journal.pone.0069674

Feller, W. (1971). *An Introduction to Probability Theory and Its Applications* (2 edition, Vol. 2). New York: Wiley.

Fific, M., Townsend, J. T., & Eidels, A. (2008). Studying visual search using systems factorial methodology with target-distractor similarity as the factor. *Perception & Psychophysics*, *70*(4), 583–603.

Found, A., & Müller, H. J. (1996). Searching for unknown feature targets on more than one dimension: Investigating a "dimension-weighting" account. *Perception & Psychophysics*, *58*(1), 88–101. http://doi.org/10.3758/BF03205479

Geisler, W. S. (2011). Contributions of ideal observer theory to vision research. *Vision Research*, *51*(7), 771–781. http://doi.org/10.1016/j.visres.2010.09.027

Gelman, A., & Hill, J. (2006). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press.

Gholson, B., & Hohle, R. H. (1968a). Choice Reaction Times to Hues Printed in Conflicting Hue Names and Nonsense Words. *Journal of Experimental Psychology*, *76*(3, Pt.1), 413–418. http://doi.org/10.1037/h0021284

Gholson, B., & Hohle, R. H. (1968b). Verbal reaction times to hues vs hue names and forms vs form names. *Perception & Psychophysics*, *3*(3), 191–196.

Grasman, R. P. P. P., Wagenmakers, E.-J., & van der Maas, H. L. J. (2009). On the mean and variance of response times under the diffusion model with an application to parameter estimation. *Journal of Mathematical Psychology*, *53*(2), 55–68. http://doi.org/10.1016/j.jmp.2009.01.006

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

Grethe, J. S., Van Horn, J. D., Woodward, J. B., Inati, S., Kostelec, P. J., Aslam, J. A., … Gazzaniga, M. S. (2001). The fMRI Data Center: An introduction. *Neuroimage*, *13*(6), S135–S135.

Gu, S.-L. H., Gau, S. S.-F., Tzang, S.-W., & Hsu, W.-Y. (2013). The ex-Gaussian distribution of reaction times in adolescents with attention-deficit/hyperactivity disorder. *Research in Developmental Disabilities*, *34*(11), 3709–3719. http://doi.org/10.1016/j.ridd.2013.07.025

Heathcote, A., Brown, S., & Cousineau, D. (2004). QMPE: Estimating Lognormal, Wald, and Weibull RT distributions with a parameter-dependent lower bound. *Behavior Research Methods, Instruments, & Computers*, *36*(2), 277–290. http://doi.org/10.3758/BF03195574

Heathcote, A., & Love, J. (2012). Linear Deterministic Accumulator Models of Simple Choice. *Frontiers in Psychology*, *3*. http://doi.org/10.3389/fpsyg.2012.00292

Heathcote, A., Popiel, S. J., & Mewhort, D. J., J. D. (1991). Analysis of response time distributions: An example using the Stroop task. *Psychological Bulletin*, *109*(2), 340–347. http://doi.org/10.1037/0033-2909.109.2.340

Heath, R. A., & Willcox, C. H. (1990). A stochastic model for inter-keypress times in a typing task. *Acta Psychologica*, *75*(1), 13–39. http://doi.org/10.1016/0001-6918(90)90064-M

Heinke, D., & Backhaus, A. (2011). Modelling Visual Search with the Selective Attention for Identification Model (VS-SAIM): A Novel Explanation for Visual Search Asymmetries. *Cognitive Computation*, *3*(1), 185–205. http://doi.org/10.1007/s12559-010-9076-x

Heinke, D., & Humphreys, G. W. (2003). Attention, spatial representation, and visual neglect: Simulating emergent attention and spatial memory in the selective attention for identification model (SAIM). *Psychological Review*, *110*(1), 29–87. http://doi.org/10.1037/0033-295X.110.1.29

Heinke, D., & Humphreys, G. W. (2004). Computational models of visual selective attention: A review. In G. Houghton (Ed.), *Connectionist Models in Cognitive Psychology* (pp. 273–312). Psychology Press.

Heitz, R. P., & Schall, J. D. (2012). Neural Mechanisms of Speed-Accuracy Tradeoff. *Neuron*, *76*(3), 616–628. http://doi.org/10.1016/j.neuron.2012.08.030

Heitz, R. P., & Schall, J. D. (2013). Neural chronometry and coherency across speed–accuracy demands reveal lack of homomorphism between computational and neural mechanisms of evidence accumulation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *368*(1628), 20130071. http://doi.org/10.1098/rstb.2013.0071

Hockley, W. E., & Corballis, M. C. (1982). Tests of serial scanning in item recognition. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, *36*(2), 189–212. http://doi.org/10.1037/h0080637

Hodsoll, J., & Humphreys, G. W. (2001). Driving attention with the top down: The relative contribution of target templates to the linear separability effect in the size dimension. *Perception & Psychophysics*, *63*(5), 918–926. http://doi.org/10.3758/BF03194447

Hodsoll, J. P., & Humphreys, G. W. (2005). The effect of target foreknowledge on visual search for categorically separable orientation targets. *Vision Research*, *45*(18), 2346–2351. http://doi.org/10.1016/j.visres.2005.03.017

Hoffman, J. E. (1979). A two-stage model of visual search. *Perception & Psychophysics*, *25*(4), 319–327. http://doi.org/10.3758/BF03198811

Hoffman, M. D., & Gelman, A. (2011). The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. *arXiv:1111.4246 [cs, Stat]*. Retrieved from http://arxiv.org/abs/1111.4246

Hohle, R. H. (1965). Inferred components of reaction times as functions of foreperiod duration. *Journal of Experimental Psychology*, *69*(4), 382–386. http://doi.org/10.1037/h0021740

Hohle, R. H. (1967). Component process latencies in reaction times of children and adults. *Advances in Child Development and Behavior*, *3*, 225–261.

Hsiang, S. M. (2013). *Visually-weighted regression. SSRN working paper*.

Hsu, Y.-F., & Chen, Y.-H. (2009). Applications of nonparametric adaptive methods for simple reaction time experiments. *Attention, Perception, & Psychophysics*, *71*(7), 1664–1675. http://doi.org/10.3758/APP.71.7.1664

Humphreys, G. W., & Müller, H. J. (1993). SEarch via Recursive Rejection (SERR): A Connectionist Model of Visual Search. *Cognitive Psychology*, *25*(1), 43–110. http://doi.org/10.1006/cogp.1993.1002

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, *40*(10–12), 1489–1506. http://doi.org/10.1016/S0042-6989(99)00163-7

Itti, L., Koch, C., & Niebur, E. (1998). A Model of Saliency-Based Visual Attention for Rapid Scene Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *20*(11), 1254–1259. http://doi.org/10.1109/34.730558

Johnson, N. L., Kotz, Samuel, Balakrishnan, N, Kotz, S., & Balakrishnan, N. (1994). *Continuous univariate distributions Vol. 1* (2nd ed.). New York: Wiley.

Julesz, B. (1984). A brief outline of the texton theory of human vision. *Trends in Neurosciences*, *7*(2), 41–45. http://doi.org/10.1016/S0166-2236(84)80275-1

Julesz, B., & Bergen, J. R. (1983). Human factors and behavioral science: Textons, the fundamental elements in preattentive vision and perception of textures. *Bell System Technical Journal, The*, *62*(6), 1619–1645. http://doi.org/10.1002/j.1538-7305.1983.tb03502.x

Kahneman, D., & Treisman, A. (1984). Changing views of attention and automaticity. In P. R. & D. D. R. (Eds.), *Varieties of attention* (pp. 29–61). Orlando, FL: New York: Academic Press.

Kandel, E. R., Schwartz, J. H., & Jessell, T. M. (2000). *Principles of neural science*. New York: McGraw-Hill, Health Professions Division.

Kastner, S., Pinsk, M. A., De Weerd, P., Desimone, R., & Ungerleider, L. G. (1999). Increased Activity in Human Visual Cortex during Directed Attention in the Absence of Visual Stimulation. *Neuron*, *22*(4), 751–761. http://doi.org/10.1016/S0896-6273(00)80734-5

Kirk, R. E. (1995). *Experimental design: Procedures for the behavioral sciences*.

Pacific Grove, Calif.: Brooks/Cole Publishing Company.

Knapp, W. H., & Abrams, R. A. (2012). Fundamental differences in visual search with verbal and pictorial cues. *Vision Research*, *71*, 28–36. http://doi.org/10.1016/j.visres.2012.08.015

Koch, C., & Tsuchiya, N. (2007). Attention and consciousness: two distinct brain processes. *Trends in Cognitive Sciences*, *11*(1), 16–22. http://doi.org/10.1016/j.tics.2006.10.012

Koch, C., & Tsuchiya, N. (2008). Response to Mole: Subjects can attend to completely invisible objects. *Trends in Cognitive Sciences*, *12*(2), 44–45. http://doi.org/10.1016/j.tics.2007.11.002

Krueger, L. E. (1978). A theory of perceptual matching. *Psychological Review*, *85*(4), 278–304. http://doi.org/10.1037/0033-295X.85.4.278

Kruschke, J. K. (2010). What to believe: Bayesian methods for data analysis. *Trends in Cognitive Sciences*, *14*(7), 293–300. http://doi.org/10.1016/j.tics.2010.05.001

Kwak, H.-W., Dagenbach, D., & Egeth, H. (1991). Further evidence for a time-independent shift of the focus of attention. *Perception & Psychophysics*, *49*(5), 473–480. http://doi.org/10.3758/BF03212181

Lacouture, Y., & Cousineau, D. (2008). How to use MATLAB to fit the ex-Gaussian and other probability functions to a distribution of response times. *Tutorials in Quantitative Methods for Psychology*, *4*(1), 35–45.

Lamme, V. A. F. (2003). Why visual attention and awareness are different. *Trends in Cognitive Sciences*, *7*(1), 12–18. http://doi.org/10.1016/S1364-6613(02)00013-X

Lamy, D. F., & Kristjánsson, Á. (2013). Is goal-directed attentional guidance just

intertrial priming? A review. *Journal of Vision*, *13*(3), 14. http://doi.org/10.1167/13.3.14

Lee, M. D., & Wagenmakers, E.-J. (2013). *Bayesian cognitive modeling: a practical course*.

Levi, D. M. (2008). Crowding - An essential bottleneck for object recognition: A mini-review. *Vision Research*, *48*(5), 635–654. http://doi.org/10.1016/j.visres.2007.12.009

Link, S. W. (1975). The relative judgment theory of two choice response time. *Journal of Mathematical Psychology*, *12*(1), 114–135. http://doi.org/10.1016/0022-2496(75)90053-X

Lin, Y.-S., Heinke, D., & Humphreys, G. W. (2015). Modeling visual search using three-parameter probability functions in a hierarchical Bayesian framework. *Attention, Perception, & Psychophysics*, *77*(3), 985–1010. http://doi.org/10.3758/s13414-014-0825-x

Li, Z. (2002). A saliency map in primary visual cortex. *Trends in Cognitive Sciences*, *6*(1), 9–16. http://doi.org/10.1016/S1364-6613(00)01817-9

Loft, S., Bowden, V. K., Ball, B. H., & Brewer, G. A. (2014). Fitting an ex-Gaussian function to examine costs in event-based prospective memory: Evidence for a continuous monitoring profile. *Acta Psychologica*, *152*, 177–82. http://doi.org/10.1016/j.actpsy.2014.08.010

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, *95*(4), 492–527. http://doi.org/10.1037/0033-295X.95.4.492

Luce, R. D. (1986). *Response Times*. Oxford University Press.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features

and conjunctions. *Nature*, *390*(6657), 279–281. http://doi.org/10.1038/36846

Lunn, D., Jackson, C., Best, N., Thomas, A., & Spiegelhalter, D. J. (2013). *The BUGS book: a practical introduction to Bayesian analysis*.

Lunn, D., Spiegelhalter, D., Thomas, A., & Best, N. (2009). The BUGS project: Evolution, critique and future directions. *Statistics in Medicine*, *28*(25), 3049–3067. http://doi.org/10.1002/sim.3680

Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: a user's guide*. Cambridge [England]; New York: Cambridge University Press.

Malinowski, P., & Hübner, R. (2001). The effect of familiarity on visual-search performance: Evidence for learned basic features. *Perception & Psychophysics*, *63*(3), 458–463. http://doi.org/10.3758/BF03194412

Marr, D., & Nishihara, H. K. (1978). Representation and Recognition of the Spatial Organization of Three-Dimensional Shapes. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, *200*(1140), 269–294. http://doi.org/10.1098/rspb.1978.0020

Matzke, D., Dolan, C. V., Logan, G. D., Brown, S. D., & Wagenmakers, E.-J. (2013). Bayesian parametric estimation of stop-signal reaction time distributions. *Journal of Experimental Psychology: General*, *142*(4), 1047–1073. http://doi.org/10.1037/a0030543

Matzke, D., Love, J., Wiecki, T. V., Brown, S. D., Logan, G. D., & Wagenmakers, E.-J. (2013). Release the BEESTS: Bayesian Estimation of Ex-Gaussian STop-Signal reaction time distributions. *Frontiers in Psychology*, *4*. http://doi.org/10.3389/fpsyg.2013.00918

Matzke, D., & Wagenmakers, E.-J. (2009). Psychological interpretation of the ex-

Gaussian and shifted Wald parameters: A diffusion model analysis. *Psychonomic Bulletin & Review*, *16*(5), 798–817. http://doi.org/10.3758/PBR.16.5.798

Mavritsaki, E., Heinke, D., Allen, H., Deco, G., & Humphreys, G. W. (2011). Bridging the Gap Between Physiology and Behavior: Evidence From the sSoTS Model of Human Visual Attention. *Psychological Review*, *118*(1), 3–41. http://doi.org/10.1037/a0021868

Maxfield, J. T., & Zelinsky, G. J. (2012). Searching through the hierarchy: How level of target categorization affects visual search. *Visual Cognition*, *20*(10), 1153–1163. http://doi.org/10.1080/13506285.2012.735718

McGill, W. J., & Gibbon, J. (1965). The general-gamma distribution and reaction times. *Journal of Mathematical Psychology*, *2*(1), 1–18. http://doi.org/10.1016/0022-2496(65)90014-3

Merkle, E., & Van Zandt, T. (2005). WinBUGS Tutorial. Retrieved from http://maigret.psy.ohiostate.edu/~trish/Downloads/.

Metropolis, N., & Ulam, S. (1949). The Monte Carlo Method. *Journal of the American Statistical Association*, *44*(247), 335–341. http://doi.org/10.1080/01621459.1949.10483310

Meyers, L., & Rhoades, R. (1978). Visual-Search of Common Scenes. *Quarterly Journal of Experimental Psychology*, *30*(MAY), 297–310. http://doi.org/10.1080/14640747808400677

Moran, J., & Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. *Science*, *229*(4715), 782–784. http://doi.org/10.1126/science.4023713

Moran, R., Zehetleitner, M., Müller, H. J., & Usher, M. (2013). Competitive guided

search: Meeting the challenge of benchmark RT distributions. *Journal of Vision*, *13*(8), 24. http://doi.org/10.1167/13.8.24

Morey, R. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorial in Quantitative Methods for Psychology*, *4*(2), 61–64.

Motter, B. C., & Belky, E. J. (1998). The guidance of eye movements during active visual search. *Vision Research*, *38*(12), 1805–1815. http://doi.org/10.1016/S0042-6989(97)00349-0

Myung, I. J. (2003). Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology*, *47*(1), 90–100. http://doi.org/10.1016/S0022-2496(02)00028-7

Näätänen, R. (1985). Selective attention and stimulus processing: Reflections in event-related potentials, magnetoencephalogram, and regional cerebral blood flow. In P. M. I. & M. O. S. M. (Eds.), *Attention and performance XI* (pp. 355–373). Hillsdale, N.J.: Erlbaum.

Najemnik, J., & Geisler, W. S. (2005). Optimal eye movement strategies in visual search. *Nature*, *434*(7031), 387–391. http://doi.org/10.1038/nature03390

Nakayama, K., & Silverman, G. (1986a). Serial and Parallel Encoding of Feature Conjunctions. *Investigative Ophthalmology and Visual Science*, *27*(3 SUPPL), 182.

Nakayama, K., & Silverman, G. (1986b). Serial and Parallel Processing of Visual Feature Conjunctions. *Nature*, *320*(6059), 264–265. http://doi.org/10.1038/320264a0

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

Nelder, J. A., & Mead, R. (1965). A Simplex Method for Function Minimization. *The Computer Journal*, *7*(4), 308–313. http://doi.org/10.1093/comjnl/7.4.308

O'Connor, D. H., Fukui, M. M., Pinsk, M. A., & Kastner, S. (2002). Attention modulates responses in the human lateral geniculate nucleus. *Nature Neuroscience*, *5*(11), 1203–1209. http://doi.org/10.1038/nn957

Olivers, C. N. L. (2009). What drives memory-driven attentional capture? The effects of memory type, display type, and search type. *Journal of Experimental Psychology. Human Perception and Performance*, *35*(5), 1275–1291. http://doi.org/10.1037/a0013896

Olivers, C. N. L., Peters, J., Houtkamp, R., & Roelfsema, P. R. (2011). Different states in visual working memory: when it guides attention and when it does not. *Trends in Cognitive Sciences*, *15*(7), 327–334. http://doi.org/10.1016/j.tics.2011.05.004

Palmer, E. M., Horowitz, T. S., Torralba, A., & Wolfe, J. M. (2011). What are the shapes of response time distributions in visual search? *Journal of Experimental Psychology: Human Perception and Performance*, *37*(1), 58–71. http://doi.org/10.1037/a0020747

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

Palmer, J., Verghese, P., & Pavel, M. (2000). The psychophysics of visual search. *Vision Research*, *40*(10–12), 1227–1268. http://doi.org/10.1016/S0042-6989(99)00244-8

Pashler, H. (1987). Target-distractor discriminability in visual search. *Perception & Psychophysics*, *41*(4), 285–292. http://doi.org/10.3758/BF03208228

Payne, B. R., & Stine-Morrow, E. A. L. (2014). Adult Age Differences in Wrap-Up During Sentence Comprehension: Evidence From Ex-Gaussian Distributional Analyses of Reading Time. *Psychology and Aging*, *29*(2), 213–228. http://doi.org/10.1037/a0036282

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in S and S-PLUS*. New York: Springer. Retrieved from http://site.ebrary.com/id/10002187

Pleskac, T. J., & Busemeyer, J. R. (2010). Two-stage dynamic signal detection: A theory of choice, decision time, and confidence. *Psychological Review*, *117*(3), 864–901. http://doi.org/10.1037/a0019737

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In *Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003). March* (pp. 20–22).

Purcell, B. A., Heitz, R. P., Cohen, J. Y., Schall, J. D., Logan, G. D., & Palmeri, T. J. (2010). Neurally constrained modeling of perceptual decision making. *Psychological Review*, *117*(4), 1113–1143. http://doi.org/10.1037/a0020311

Purcell, B. A., Schall, J. D., Logan, G. D., & Palmeri, T. J. (2012). From Salience to Saccades: Multiple-Alternative Gated Stochastic Accumulator Model of Visual Search. *The Journal of Neuroscience*, *32*(10), 3433–3446. http://doi.org/10.1523/JNEUROSCI.4622-11.2012

Quinlan, P. T., & Humphreys, G. W. (1987). Visual search for targets defined by combinations of color, shape, and size: An examination of the task constraints on feature and conjunction searches. *Perception & Psychophysics*, *41*(5), 455–472.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*(2), 59–108.

http://doi.org/10.1037/0033-295X.85.2.59

Ratcliff, R. (1979). Group reaction time distributions and an analysis of distribution statistics. *Psychological Bulletin*, *86*(3), 446–461. http://doi.org/10.1037/0033-2909.86.3.446

Ratcliff, R. (1981). A theory of order relations in perceptual matching. *Psychological Review*, *88*(6), 552–572. http://doi.org/10.1037/0033-295X.88.6.552

Ratcliff, R. (2008). The EZ diffusion method: Too EZ? *Psychonomic Bulletin & Review*, *15*(6), 1218–1228. http://doi.org/10.3758/PBR.15.6.1218

Ratcliff, R. (2014). Measuring psychometric functions with the diffusion model. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(2), 870–888. http://doi.org/10.1037/a0034954

Ratcliff, R., Gomez, P., & McKoon, G. (2004). A diffusion model account of the lexical decision task. *Psychological Review*, *111*(1), 159–182. http://doi.org/10.1037/0033-295X.111.1.159

Ratcliff, R., & McKoon, G. (2007). The Diffusion Decision Model: Theory and Data for Two-Choice Decision Tasks. *Neural Computation*, *20*(4), 873–922. http://doi.org/10.1162/neco.2008.12-06-420

Ratcliff, R., & McKoon, G. (2008). The Diffusion Decision Model: Theory and Data for Two-Choice Decision Tasks. *Neural Computation*, *20*(4), 873–922. http://doi.org/10.1162/neco.2008.12-06-420

Ratcliff, R., & Murdock, B. (1976a). Retrieval Processes in Recognition Memory. *Psychological Review*, *83*(3), 190–214. http://doi.org/10.1037//0033-295X.83.3.190

Ratcliff, R., & Murdock, B. B. (1976b). Retrieval processes in recognition memory.

Psychological Review, *83*(3), 190–214. http://doi.org/10.1037/0033-295X.83.3.190

Ratcliff, R., Philiastides, M. G., & Sajda, P. (2009). Quality of evidence for perceptual decision making is indexed by trial-to-trial variability of the EEG. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(16), 6539–6544. http://doi.org/10.1073/pnas.0812589106

Ratcliff, R., & Rouder, J. N. (1998). Modeling Response Times for Two-Choice Decisions. *Psychological Science*, *9*(5), 347–356. http://doi.org/10.1111/1467-9280.00067

Ratcliff, R., & Smith, P. L. (2004). A Comparison of Sequential Sampling Models for Two-Choice Reaction Time. *Psychological Review*, *111*(2), 333–367. http://doi.org/10.1037/0033-295X.111.2.333

Ratcliff, R., & Starns, J. J. (2009). Modeling confidence and response time in recognition memory. *Psychological Review*, *116*(1), 59–83. http://doi.org/10.1037/a0014086

Ratcliff, R., & Tuerlinckx, F. (2002). Estimating parameters of the diffusion model: Approaches to dealing with contaminant reaction times and parameter variability. *Psychonomic Bulletin & Review*, *9*(3), 438–481. http://doi.org/10.3758/BF03196302

Ratcliff, R., Van Zandt, T., & McKoon, G. (1999). Connectionist and diffusion models of reaction time. *Psychological Review*, *106*(2), 261–300. http://doi.org/10.1037/0033-295X.106.2.261

R Core Team, & others. (2012). R: A language and environment for statistical computing.

Reeves, A., Santhi, N., & Decaro, S. (2005). A random-ray model for speed and accuracy in perceptual experiments. *Spatial Vision*, *18*(1), 73–83.

Riddoch, J. M., & Humphreys, G. W. (1987). Perceptual and Action Systems in Unilateral Visual Neglect. *Advances in Psychology*, *45*, 151–181.

Roe, R. M., Busemeyer, J. R., & Townsend, J. T. (2001). Multialternative decision field theory: A dynamic connectionst model of decision making. *Psychological Review*, *108*(2), 370–392. http://doi.org/10.1037/0033-295X.108.2.370

Rohrer, D., & Wixted, J. T. (1994). An analysis of latency and interresponse time in free recall. *Memory & Cognition*, *22*(5), 511–524. http://doi.org/10.3758/BF03198390

Rouder, J. N., & Lu, J. (2005). An introduction to Bayesian hierarchical models with an application in the theory of signal detection. *Psychonomic Bulletin & Review*, *12*(4), 573–604. http://doi.org/10.3758/BF03196750

Rouder, J. N., Lu, J., Morey, R. D., Sun, D., & Speckman, P. L. (2008). A hierarchical process-dissociation model. *Journal of Experimental Psychology: General*, *137*(2), 370–389. http://doi.org/10.1037/0096-3445.137.2.370

Rouder, J. N., Lu, J., Speckman, P., Sun, D., & Jiang, Y. (2005). A hierarchical model for estimating response time distributions. *Psychonomic Bulletin & Review*, *12*(2), 195–223. http://doi.org/10.3758/BF03257252

Rouder, J. N., & Speckman, P. L. (2004). An evaluation of the Vincentizing method of forming group-level response time distributions. *Psychonomic Bulletin & Review*, *11*(3), 419–427. http://doi.org/10.3758/BF03196589

Rouder, J. N., Sun, D., Speckman, P. L., Lu, J., & Zhou, D. (2003). A hierarchical bayesian statistical framework for response time distributions. *Psychometrika*,

68(4), 589–606. http://doi.org/10.1007/BF02295614

Rouder, J. N., Yue, Y., Speckman, P. L., Pratte, M. S., & Province, J. M. (2010). Gradual growth versus shape invariance in perceptual decision making. *Psychological Review*, *117*(4), 1267–1274. http://doi.org/10.1037/a0020793

Schafer, J. L. (2010). *Analysis of incomplete multivariate data*. CRC press.

Schmidt, J., & Zelinsky, G. J. (2009). Search guidance is proportional to the categorical specificity of a target cue. *The Quarterly Journal of Experimental Psychology*, *62*(10), 1904–1914. http://doi.org/10.1080/17470210902853530

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, *84*(1), 1–66. http://doi.org/10.1037/0033-295X.84.1.1

Schönbrodt, F. (2012). Visually weighted/ Watercolor Plots, new variants: Please vote! Retrieved from http://www.nicebread.de/visually-weighted-watercolor-plots-new-variants-please-vote/

Schwarz, W. (2001). The ex-Wald distribution as a descriptive model of response times. *Behavior Research Methods, Instruments, & Computers*, *33*(4), 457–469. http://doi.org/10.3758/BF03195403

Shapiro, K. L. (1994). The Attentional Blink: The Brain's "Eyeblink." *Current Directions in Psychological Science*, *3*(3), 86–89.

Shiffrin, R. M., Lee, M. D., Kim, W., & Wagenmakers, E.-J. (2008). A Survey of Model Evaluation Approaches With a Tutorial on Hierarchical Bayesian Methods. *Cognitive Science*, *32*(8), 1248–1284. http://doi.org/10.1080/03640210802414826

Shi, H., & Yang, Y. (2007). A computational model of visual attention based on saliency

maps. *Applied Mathematics and Computation*, *188*(2), 1671–1677. http://doi.org/10.1016/j.amc.2006.11.036

Smith, A. F., & Roberts, G. O. (1993). Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society. Series B (Methodological)*, 3–23.

Smith, P. L. (1995). Psychophysically principled models of visual simple reaction time. *Psychological Review*, *102*(3), 567–593. http://doi.org/10.1037/0033-295X.102.3.567

Smith, P. L., & Ratcliff, R. (2004). Psychology and neurobiology of simple decisions. *Trends in Neurosciences*, *27*(3), 161–168. http://doi.org/10.1016/j.tins.2004.01.006

Smith, P. L., & Sewell, D. K. (2013). A competitive interaction theory of attentional selection and decision making in brief, multielement displays. *Psychological Review*, *120*(3), 589–627. http://doi.org/10.1037/a0033140

Soto, D., Heinke, D., Humphreys, G. W., & Blanco, M. J. (2005). Early, involuntary top-down guidance of attention from working memory. *Journal of Experimental Psychology. Human Perception and Performance*, *31*(2), 248–261. http://doi.org/10.1037/0096-1523.31.2.248

Soto, D., Hodsoll, J., Rotshtein, P., & Humphreys, G. W. (2008). Automatic guidance of attention from working memory. *Trends in Cognitive Sciences*, *12*(9), 342–348. http://doi.org/10.1016/j.tics.2008.05.007

Soto, D., & Humphreys, G. W. (2007). Automatic guidance of visual attention from verbal working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(3), 730–737. http://doi.org/10.1037/0096-

1523.33.3.730

Sperling, G. (1967). Successive approximations to a model for short term memory. *Acta Psychologica*, *27*, 285–292. http://doi.org/10.1016/0001-6918(67)90070-4

Spiegelhalter, D., Thomas, A., Best, N., & Lunn, D. (2007). OpenBUGS user manual, version 3.0. 2. *MRC Biostatistics Unit, Cambridge*.

Sternberg, S. (1966). High-speed scanning in human memory. *Science (New York, N.Y.)*, *153*(3736), 652–654.

Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. *Acta Psychologica*, *30*, 276–315. http://doi.org/10.1016/0001-6918(69)90055-9

Sternberg, S. (1975). Memory Scanning - New Findings and Current Controversies. *Quarterly Journal of Experimental Psychology*, *27*(FEB), 1–32. http://doi.org/10.1080/14640747508400459

Stoet, G. (2010). PsyToolkit: A software package for programming psychological experiments using Linux. *Behavior Research Methods*, *42*(4), 1096–1104. http://doi.org/10.3758/BRM.42.4.1096

Sturtz, S., Ligges, U., & Gelman, A. E. (2005). R2WinBUGS: A Package for Running WinBUGS from R. *Journal of Statistical Software*, *12*(3), 1–16.

Sui, J., & Humphreys, G. W. (2013). The boundaries of self face perception: Response time distributions, perceptual categories, and decision weighting. *Visual Cognition*, *21*(4), 415–445. http://doi.org/10.1080/13506285.2013.800621

Theeuwes, J. (1989). Effects of Location and Form Cueing on the Allocation of Attention in the Visual-Field. *Acta Psychologica*, *72*(2), 177–192. http://doi.org/10.1016/0001-6918(89)90043-7

Theeuwes, J. (1992). Perceptual selectivity for color and form. *Perception & Psychophysics*, *51*(6), 599–606. http://doi.org/10.3758/BF03211656

Theeuwes, J. (2010). Top–down and bottom–up control of visual selection. *Acta Psychologica*, *135*(2), 77–99. http://doi.org/10.1016/j.actpsy.2010.02.006

Thornton, T. L., & Gilden, D. L. (2007). Parallel and serial processes in visual search. *Psychological Review*, *114*(1), 71–103. http://doi.org/10.1037/0033-295X.114.1.71

Toeroek, A., Kolozsvari, O., Viragh, T., Honbolygo, F., & Csepe, V. (2014). Effect of stimulus intensity on response time distribution in multisensory integration. *Journal on Multimodal User Interfaces*, *8*(2), 209–216. http://doi.org/10.1007/s12193-013-0135-y

Towal, R. B., Mormann, M., & Koch, C. (2013). Simultaneous modeling of visual saliency and value computation improves predictions of economic choice. *Proceedings of the National Academy of Sciences*, *110*(40), E3858–E3867. http://doi.org/10.1073/pnas.1304429110

Townsend, J. T. (1971). A note on the identifiability of parallel and serial processes. *Perception & Psychophysics*, *10*(3), 161–163. http://doi.org/10.3758/BF03205778

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*(1), 46–54. http://doi.org/10.1111/j.1467-9280.1990.tb00067.x

Townsend, J. T., & Ashby, F. G. (1983). *The stochastic modeling of elementary psychological processes*. Cambridge [Cambridgeshire]; 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*(4), 321–359. http://doi.org/10.1006/jmps.1995.1033

Townsend, J. T., & Wenger, M. J. (2004). The serial-parallel dilemma: A case study in a linkage of theory and method. *Psychonomic Bulletin & Review*, *11*(3), 391–418. http://doi.org/10.3758/BF03196588

Treisman, A. M. (1986). Features and Objects in Visual Processing. *Scientific American*, *255*(5), B114–&.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*(1), 97–136. http://doi.org/10.1016/0010-0285(80)90005-5

Treisman, A. M., & Sato, S. (1990). Conjunction search revisited. *Journal of Experimental Psychology: Human Perception and Performance*, *16*(3), 459–478. http://doi.org/10.1037/0096-1523.16.3.459

Tse, C.-S., & Altarriba, J. (2012). The effects of first- and second-language proficiency on conflict resolution and goal maintenance in bilinguals: Evidence from reaction time distributional analyses in a Stroop task. *Bilingualism: Language and Cognition*, *15*(03), 663–676. http://doi.org/10.1017/S1366728912000077

Turner, B. M., Forstmann, B. U., Wagenmakers, E.-J., Brown, S. D., Sederberg, P. B., & Steyvers, M. (2013). A Bayesian framework for simultaneously modeling neural and behavioral data. *NeuroImage*, *72*, 193–206. http://doi.org/10.1016/j.neuroimage.2013.01.048

Turner, B. M., & Sederberg, P. B. (2014). A generalized, likelihood-free method for posterior estimation. *Psychonomic Bulletin & Review*, *21*(2), 227–250. http://doi.org/10.3758/s13423-013-0530-0

Turner, B. M., Sederberg, P. B., Brown, S. D., & Steyvers, M. (2013). A method for efficiently sampling from distributions with correlated dimensions. *Psychological Methods*, *18*(3), 368–384. http://doi.org/10.1037/a0032222

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: the leaky, competing accumulator model. *Psychological Review*, *108*(3), 550–592.

Vandekerckhove, J., Tuerlinckx, F., & Lee, M. D. (2011). Hierarchical diffusion models for two-choice response times. *Psychological Methods*, *16*(1), 44–62. http://doi.org/10.1037/a0021765

van der Heijden, A. H. C., Malhas, M. S. M., & van den Roovaart, B. P. (1984). An empirical interletter confusion matrix for continuous-line capitals. *Perception & Psychophysics*, *35*(1), 85–88. http://doi.org/10.3758/BF03205927

Van Horn, J. D., & Gazzaniga, M. S. (2013). Why share data? Lessons learned from the fMRIDC. *Neuroimage*, *82*, 677–682. http://doi.org/10.1016/j.neuroimage.2012.11.010

Van Ravenzwaaij, D., & Oberauer, K. (2009). How to use the diffusion model: Parameter recovery of three methods: EZ, fast-dm, and DMAT. *Journal of Mathematical Psychology*, *53*(6), 463–473. http://doi.org/10.1016/j.jmp.2009.09.004

Van Zandt, T. (2000). How to fit a response time distribution. *Psychonomic Bulletin & Review*, *7*(3), 424–465. http://doi.org/10.3758/BF03214357

Van Zandt, T., Colonius, H., & Proctor, R. W. (2000). A comparison of two response

time models applied to perceptual matching. *Psychonomic Bulletin & Review*, *7*(2), 208–256. http://doi.org/10.3758/BF03212980

Vickery, T. J., King, L.-W., & Jiang, Y. (2005). Setting up the target template in visual search. *Journal of Vision*, *5*(1), 8. http://doi.org/10.1167/5.1.8

Vincent, S. B. (1912). *The functions of the vibrissae in the behavior of the white rat ...* University of Chicago.

Voss, A., Rothermund, K., & Brandtstädter, J. (2008). Interpreting ambiguous stimuli: Separating perceptual and judgmental biases. *Journal of Experimental Social Psychology*, *44*(4), 1048–1056. http://doi.org/10.1016/j.jesp.2007.10.009

Voss, A., Rothermund, K., Gast, A., & Wentura, D. (2013). Cognitive processes in associative and categorical priming: A diffusion model analysis. *Journal of Experimental Psychology: General*, *142*(2), 536–559. http://doi.org/10.1037/a0029459

Voss, A., & Voss, J. (2007). Fast-dm: A free program for efficient diffusion model analysis. *Behavior Research Methods*, *39*(4), 767–775. http://doi.org/10.3758/BF03192967

Voss, A., & Voss, J. (2008). A fast numerical algorithm for the estimation of diffusion model parameters. *Journal of Mathematical Psychology*, *52*(1), 1–9. http://doi.org/10.1016/j.jmp.2007.09.005

Wagenmakers, E.-J. (2009). Methodological and empirical developments for the Ratcliff diffusion model of response times and accuracy. *European Journal of Cognitive Psychology*, *21*(5), 641–671. http://doi.org/10.1080/09541440802205067

Wagenmakers, E.-J., & Brown, S. (2007). On the linear relation between the mean and

the standard deviation of a response time distribution. *Psychological Review*, *114*(3), 830–841. http://doi.org/10.1037/0033-295X.114.3.830

Wagenmakers, E.-J., Maas, H. L. J. V. D., & Grasman, R. P. P. P. (2007). An EZ-diffusion model for response time and accuracy. *Psychonomic Bulletin & Review*, *14*(1), 3–22. http://doi.org/10.3758/BF03194023

Wagenmakers, E.-J., Ratcliff, R., Gomez, P., & McKoon, G. (2008). A diffusion model account of criterion shifts in the lexical decision task. *Journal of Memory and Language*, *58*(1), 140–159. http://doi.org/10.1016/j.jml.2007.04.006

Wagenmakers, E.-J., van der Maas, H. L. J., Dolan, C. V., & Grasman, R. P. P. P. (2008). EZ does it! Extensions of the EZ-diffusion model. *Psychonomic Bulletin & Review*, *15*(6), 1229–1235. http://doi.org/10.3758/PBR.15.6.1229

Ward, R., & McClelland, J. L. (1989). Conjunctive search for one and two identical targets. *Journal of Experimental Psychology: Human Perception and Performance*, *15*(4), 664–672. http://doi.org/10.1037/0096-1523.15.4.664

Watson, D. G., Humphreys, G. W., & Olivers, C. N. L. (2003). Visual marking: using time in visual selection. *Trends in Cognitive Sciences*, *7*(4), 180–186.

Weidner, R., Pollmann, S., Müller, H. J., & Cramon, D. Y. von. (2002). Top-down Controlled Visual Dimension Weighting: An Event-related fMRI Study. *Cerebral Cortex*, *12*(3), 318–328. http://doi.org/10.1093/cercor/12.3.318

Wiecki, T. V., Sofer, I., & Frank, M. J. (2013). HDDM: Hierarchical Bayesian estimation of the Drift-Diffusion Model in Python. *Frontiers in Neuroinformatics*, *7*. http://doi.org/10.3389/fninf.2013.00014

Wilcox, R. R., Erceg-Hurn, D. M., Clark, F., & Carlson, M. (2014). Comparing two independent groups via the lower and upper quantiles. *Journal of Statistical*

Computation and Simulation, *84*(7), 1543–1551. http://doi.org/10.1080/00949655.2012.754026

Wilschut, A., Theeuwes, J., & Olivers, C. N. L. (2013). The time it takes to turn a memory into a template. *Journal of Vision*, *13*(3), 8. http://doi.org/10.1167/13.3.8

Wilschut, A., Theeuwes, J., & Olivers, C. N. L. (2014). Priming and the guidance by visual and categorical templates in visual search. *Frontiers in Psychology*, *5*. http://doi.org/10.3389/fpsyg.2014.00148

Wolfe, J. M. (1994). Guided Search 2.0 A revised model of visual search. *Psychonomic Bulletin & Review*, *1*(2), 202–238. http://doi.org/10.3758/BF03200774

Wolfe, J. M. (1998a). Visual search. In H. E. Pashler (Ed.), *Attention* (pp. 13–73). Psychology Press.

Wolfe, J. M. (1998b). What Can 1 Million Trials Tell Us About Visual Search? *Psychological Science*, *9*(1), 33–39. http://doi.org/10.1111/1467-9280.00006

Wolfe, J. M. (2007). Guided search 4.0: Current progress with a model of visual search. In W. D. Gray (Ed.), *Integrated models of cognitive systems* (pp. 99–119). Oxford; New York: Oxford University Press ; University Press.

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*(3), 419–433. http://doi.org/10.1037/0096-1523.15.3.419

Wolfe, J. M., & Gancarz, G. (1996). Guided search 3.0: a model of visual search catches up with Jay Enoch 40 years later. In V. Lakshminarayanan (Ed.), *Basic and clinical applications of vision science* (pp. 189–192). Kluwer Academic,

Dordrecht, Netherlands.

Wolfe, J. M., & Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience*, *5*(6), 495–501. http://doi.org/10.1038/nrn1411

Wolfe, J. M., & Horowitz, T. S. (2008). Visual search. *Scholarpedia*, *3*(7), 3325.

Wolfe, J. M., Horowitz, T. S., Kenner, N., Hyle, M., & Vasan, N. (2004). How fast can you change your mind? The speed of top-down guidance in visual search. *Vision Research*, *44*(12), 1411–1426. http://doi.org/10.1016/j.visres.2003.11.024

Wolfe, J. M., Palmer, E. M., & Horowitz, T. S. (2010). Reaction time distributions constrain models of visual search. *Vision Research*, *50*(14), 1304–1311. http://doi.org/10.1016/j.visres.2009.11.002

Wolfe, J. M., Võ, M. L.-H., Evans, K. K., & Greene, M. R. (2011). Visual search in scenes involves selective and nonselective pathways. *Trends in Cognitive Sciences*, *15*(2), 77–84. http://doi.org/10.1016/j.tics.2010.12.001

Wolfe, J. M., Yee, A., & Friedman-Hill, S. R. (1992). Curvature is a basic feature for visual search tasks. *Perception*, *21*(4), 465 – 480. http://doi.org/10.1068/p210465

Woodman, G. F., & Luck, S. J. (2003). Serial deployment of attention during visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *29*(1), 121–138. http://doi.org/10.1037/0096-1523.29.1.121

Woodman, G. F., & Luck, S. J. (2007). Do the contents of visual working memory automatically influence attentional selection during visual search? *Journal of Experimental Psychology. Human Perception and Performance*, *33*(2), 363–377. http://doi.org/10.1037/0096-1523.33.2.363

Woodman, G. F., Luck, S. J., & Schall, J. D. (2007). The Role of Working Memory Representations in the Control of Attention. *Cerebral Cortex*, *17*(suppl 1), i118–i124. http://doi.org/10.1093/cercor/bhm065

Yang, H., & Zelinsky, G. J. (2009). Visual search is guided to categorically-defined targets. *Vision Research*, *49*(16), 2095–2103. http://doi.org/10.1016/j.visres.2009.05.017