Modeling visual search using three-parameter probability functions in a hierarchical Bayesian framework

Lin, Yi-Shin; Heinke, Dietmar; Humphreys, Glyn

DOI:
10.3758/s13414-014-0825-x

Attention, Perception, & Psychophysics

Modeling visual search using three-parameter probability functions in a hierarchical Bayesian framework

Journal:	*Attention, Perception, & Psychophysics*
Manuscript ID:	PP-ORIG-14-088.R2
Manuscript Type:	Original Manuscript
Date Submitted by the Author:	n/a
Complete List of Authors:	Lin, Yi-Shin; University of Birmingham, Psychology Heinke, Dietmar; University of Birmingham, Psychology Humphreys, Glyn; University of Oxford, Experimental Psychology;
Keywords:	Bayesian modeling, visual search, response time models

Running head: HIERARCHICAL BAYESIAN MODELING OF VISUAL SEARCH

Modeling visual search using three-parameter probability functions in a hierarchical Bayesian framework

Yi-Shin Lin, and Dietmar Heinke

University of Birmingham, UK and

Glyn W. Humphreys

University of Oxford, UK

ABSTRACT

This study applied Bayesian-based distributional analyses to examine the shape of response time (RT) distributions in three visual search paradigms, varying in task difficulty. The paradigms investigated two common observations in visual search: the effect of display size and variations in search efficiency across different task conditions, following a design used in previous studies (Palmer, Horowitz, Torralba, & Wolfe, 2011; Wolfe, Palmer, & Horowitz, 2010), where the parameters of the response distributions were measured. The study showed that the distributional parameters in an experimental condition can be reliably estimated by moderate sample sizes when Monte Carlo simulation techniques are applied. More importantly, analysing trial RTs, the study was able to extract paradigm-dependent shape changes in the RT distributions which could be accounted for together using the EZ2 diffusion model. The study shows that Bayesian-based RT distribution analyses can provide an important means to investigate underlying cognitive processes in search including stimulus grouping and the bottom-up guidance of attention.

Modeling visual search paradigm using three-parameter probability functions in a hierarchical Bayesian framework

Distributional analyses are becoming an increasingly popular method of analyzing performance in cognitive tasks (e.g., Balota & Yap, 2011; Heathcote, Popiel, & Mewhort, 1991; Hockley & Corballis, 1982; Ratcliff & Murdock, 1976; Sui & Humphreys, 2013; Tse & Altarriba, 2012). 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 response time (RT) data 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 hierarchical Bayesian modeling (HBM), a method that blends Bayesian statistics and hierarchical modeling. 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 modeling 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 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 modeling 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 modeling 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, 2007). 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 modeling in vain. Bayesian 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, Lu, Speckman, Sun, & Jiang, 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; modeled the RTs in N^{th}-1 trial as prior distributions) influence the current search performance (modeled 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 analog 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 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), rather than the shape, of RT distributions. A scale effect, interpreted together with other symbolic distance findings using a diffusion process or a random walk, implies a general enhancement of response speed, including perceptual and motor times, as opposed to a change merely in a late-acting cognitive process such as mental rechecking (Rouder, Lu, Speckman, Sun, & Jiang, 2005).

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, we 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 & Humphreys, 2003; Heinke & Backhaus, 2011).

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, 1998).

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 we adopt HBM (see Rouder et al., 2005 as well as Rouder & Lu, 2005) in the present study are because: (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, and (3) it takes into account the differences between individual performances. Notably, the response variability across different trials is no longer assumed to constitute random noise but rather it is treated as crucial information that must be modeled.

Our 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 (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 color 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).

The 3-parameter probability functions

Our study adopted four three-parameter probability – lognormal, Wald, Weibull and gamma[1] – functions (Johnson, Kotz, & Balakrishnan, 1994) to estimate RT distributions using the HBM. Differing from the frequently used ex-Gaussian function, the 3-parameter probability

[1] 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, we constructed the comparable 3-parameter HBM to test if other functions gain a substantial better fit using hierarchical Bayesian approach than the Weibull function. We thank Evan Palmer for this suggestion.

functions describe an RT distribution with the parameters, shift, scale and shape that characterize the pattern of a distribution. An increase in the scale parameter shortens the central location a distribution and thickens its tail. This implies that the responses originally accumulated around the central part become slower and thus been moved to the tail side. An increase in the shape parameter makes the tail thinner, because those originally slow responses are moved from the tail to the central location. 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 of shift parameter preserves the general pattern of a distribution. That is, an identical curve is moved rightwards (see Figure 1 for an illustration).

Figure 1 should be inserted around here

The study assumed that changes in RT distributions reflect unobservable cognitive processes (a similar argument also made by Heathcote et al., 1991). As illustrated in Figure 1, the factors that affect quick, moderate and slow responses evenly will show a selective effect on the shift parameter. The effect on the scale parameter will be from the factors that alter only the proportion of responses that are moved from the central location to the tail part of a distribution (or vice versa). Lastly, the effect on the shape parameter may result from the factors that affect both the central and tail parts of a distribution and effectively increase the response density between them.

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 et al., 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 we investigated here). We 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, Logan, Brown, & Wagenmakers, 2013 and suggested also in Rouder et al., 2005).

To understand how our distribution-based HBM correlates with underlying cognitive processes, we compared the HBM parameters with those estimated from the EZ2 diffusion model (Wagenmakers, van der Maas, & Grasman, 2007; Wagenmakers, van der Maas, Dolan, & Grasman, 2008) which is a closed-form and simplified variant of Ratcliff's diffusion model (1978). The diffusion model conceptualizes decision-making in a two-alternative forced choice (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 (motor times) (Ratcliff & McKoon, 2007).

The diffusion model has been applied to various 2AFC paradigms and so far both psychophysical 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; see Ratcliff & McKoon, 2007 for a review). 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 parts that are influenced by decision-related processes or by non-decision-related processes, 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 the left-most panel in Figure 1 shows, a scale increase shortens a distribution and thickens its tail. 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 diffusion processes – the evidence accumulator, the boundary separation, and the non-decision process – are operated at the stage of stimulus comparison in a search trial. We used the EZ2 model to estimate the means across trials of the diffusion parameters in each condition. The Weibull HBM on the other hand summarizes the shapes of RT distributions in each condition. The RT distributions thus are the aggregated outputs from the diffusion processes. The dual-modeling approach, on the one hand, assumes one search response is driven by the diffusion process, and on 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 trials into account, the EZ2 estimations were still be able to account for the descriptive model, because the benchmark paradigms produced 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, we 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.

Method

Participants

Forty volunteers took part, aged 18 to 22 years old (Mean ± SE = 18.9 ± 1.01; 33 females; 5 left-handers). All volunteers reported normal or corrected-to-normal vision and signed a consent form before taking part in the study. One participant was excluded from the analysis because of chance-level responses. The procedure was reviewed and granted permission to proceed by the Ethics Review Committee at the University of Birmingham.

Design

The study used a similar design to Wolfe et al. (2010) with a slight modification. Specifically, we used a circular display layout with a viewing area of 7.59 × 7.59 degrees of visual angle, which allocates 25 locations to hold search items. Wolfe and colleagues (2010) used a viewing area of 22.5 × 22.5 degrees of visual angle (also with 25 search locations) and each search item subtended around 3.5 to 4.1 visual angle. Relative to Wolfe et al.'s study, our 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 2 should be inserted around here

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). To minimise one of the possible experimenter biases related to the analysis of null hypothesis significance testing (NHST; Kruschke, 2010), the study set a target sample size (20 in each group) before collecting data. The target sample size was determined based on commonly used sample sizes (approximately 5 to 20 participants) in visual

search literature. We did not analyze the data from participants who withdrew and completed only part of the tasks; these participants were replaced with other individuals.

In the feature search task, each observer looked for a dark square amongst varying numbers of gray squares (both were 0.69 × 0.69 visual angle). In the conjunction search task, observers looked for a vertical, dark bar (0.33 × 0.96 visual angle) amongst two types of distractors, vertical gray bars (0.33 × 0.96 visual angle) and horizontal dark bars (0.96 × 0.33 visual angle). In the spatial configuration search task, each observer looked for the digit 2 amongst digit 5s (both are 0.33 × 0.58 visual angle) (see Figure 1 for one of the example trials in each task).

Before the search display was presented, a 500-ms fixation cross appeared at the center of the screen, followed by a 200-ms blank duration. A trial was terminated when the observer pressed the response key. The search tasks were programmed by using PsyToolkit (Stoet, 2010), complied by GNU C compiler on a PC equipped with Linux hard real-time kernel 2.6.31-11-rt and NVidia GeForce 8500 GT graphic card, which rendered the visual stimuli on an invisible circle in black or gray color onto a gray background (RGB, 190, 190, 190). All stimuli were presented on a Sony CPD-G420 CRT monitor at the resolution of 1152 × 864 pixels with a refresh rate set at 100 Hz. The visible area contained the entire screen (i.e., 1152 × 864 pixels), but the relevant stimuli were all drawn within the viewing area of 7.59 × 7.59 visual angle. Volunteers were asked to give speeded responses without compromising their accuracy and responses were made using a Cedrus RB-830 response pad. Each volunteer completed 800 trials, where one experimental condition contained 100 trials. The volunteers carrying out the feature and conjunction search tasks completed the tasks in a counter-balanced sequence.

Hierarchical Bayesian Model (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. We 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 JAGS (Plummer, 2003). Readers who are interested in the programming details could visit the authors' GitHub at https://github.com/yxlin/HBM-Approach-Visual-Search.

The Weibull function was used to model the individual RT observations, assuming that each of them was a random variable generated by the Weibull function. The function comprises three parameters, shape (i.e., β, describing the shape of a RT distribution), scale (i.e., θ, describing the general enhancement of the magnitude and variability in a RT distribution), and 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, $\eta 1$ and $\eta 2$, 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 for 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, we replaced the Weibull function with the 3-parameter gamma, lognormal, and Wald functions (Johnson, Kotz, & Balakrishnan, 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 our data and to Wolfe et al.'s data (2010) to help a direct comparison.

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 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[2] following Wagenmakers et al. (2008; see also other possible solutions in Macmillan & Creelman, 2005) for the conditions where an observer committed no error. Present or absent responses were modeled 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 the 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, 1998). 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 ($3 \times 2 \times 4$) data points for each parameter. The analyses assessed the variability across individuals in visually-weighted regression lines, using a non-parametric bootstrapping procedure, implemented by Schönbrodt (2012) for Hsiang's visually-weighted regression method (2013)[3].

[2] 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, $P_c = 1 - (1/2n)$.

[3] The technique was discussed and implemented in the blogsphere before it was formally published in the 2013 technical report.

Results

We report the data in four sections. Firstly, we report standard search analyses, using mean measures of performance for individuals across trials. Next, we 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)[4]. In the last section, we 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 we go on to discuss the factors contributing to the RT shape, shift and scale parameters, based on how these parameters change across the different search conditions and contrast them with the decision parameters from the diffusion model. The appendix 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 Markov chain Monte Carlo procedure.

We 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 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

[4] We thank Jeremy Wolfe and Evan Palmer for their permission.

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. We nevertheless append also standard analyses for absent trials in all the figures.

Mean RTs and error rates

As is typically done for the aggregation RT analyses, we 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.

A two-way ANOVA[5] showed reliable main effects of display size, $F(3, 165) = 176.107$, $\eta^2_p = .762$, $p = 1 \times 10^{-13}$, and search task, $F(2, 55) = 108.385$, $\eta^2_p = .798$, $p = 1 \times 10^{-13}$, as well as an interaction between these factors, $F(6, 165) = 68.633$, $\eta^2_p = .714$, $p = 1 \times 10^{-13}$. 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 tests for the feature search task showed a significant display size effect, $F(3, 54) = 7.494$, $\eta^2_p = .294$, $p = 2.78 \times 10^{-4}$. RTs were slower for display sizes 18 and 12 when compared

[5] 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.

with display size 3 ($t = 6.37$, $p = 3.22 \times 10^{-5}$, 95% CI, 11.61~31.82 ms; $t = 4.03$, $p = 4.67 \times 10^{-3}$, 95% CI, 4.43~28.95 ms). There was a reliable main effect of display size also for conjunction search, $F(3, 54) = 103.15$, $\eta^2_p = .851$, $p = 1 \times 10^{-13}$, and spatial configuration search tasks, $F(3, 57) = 113.8$, $\eta^2_p = .857$, $p = 1 \times 10^{-13}$. Post-hoc t tests for the conjunction task showed reliable differences across all display sizes (510, 552, 615 and 667 ms), $p = 2.63 \times 10^{-7}$, 9.70×10^{-9}, 2.67×10^{-9}, 4.98×10^{-6}, 6.08×10^{-8}, 4.19×10^{-5} (3 vs. 6, 3 vs. 12, 3 vs. 18, 6 vs. 12, 6 vs. 18, & 12 vs. 18; Bonferroni correction for multiple comparisons). Similar effects were present for the spatial configuration search too (679 ms, 809 ms, 1011 ms vs. 1154 ms), $p = 5.14 \times 10^{-7}$, 5.15×10^{-9}, 4.10×10^{-9}, 1.42×10^{-7}, 1.09×10^{-8}, 2.33×10^{-7} (3 vs. 6, 3 vs. 12, 3 vs. 18, 6 vs. 12, 6 vs. 18, & 12 vs. 18; Bonferroni correction for multiple comparisons; Figure 3).

The error rates showed a similar pattern as the average RT, consistent with there being no trade-off between the speed and accuracy of responses. A two-way ANOVA revealed reliable main effects of display size, $F(3, 165) = 38.09$, $\eta^2_p = .409$, $p = 1 \times 10^{-13}$ and search task, $F(2, 55) = 5.75$, $\eta^2_p = .173$, $p = .005$ as well as their interaction, $F(6, 165) = 10.867$, $\eta^2_p = .283$, $p = 3.52 \times 10^{-10}$. The spatial configuration search (error $rate_{mean} = 11.80$ %) was more difficult than the conjunction search task (8.62 %), but the difference did not exceed significant level after Bonferroni correction (the difference of mean error rate = 3.18 %, 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 = 3.621 %, 95% CI = -1.396~8.628 %, $p = .241$; again the difference was not significant). The only reliable difference of error rates was between the spatial configuration search and the feature search tasks (the difference of mean error rate = 6.801 %, 95% CI = 1.847~11.755 %, $p = .004$)

For the feature search, the effect of display size was not reliable, $F(3, 54) = 1.517$, $\eta^2_p = .078$, $p = .221$, while there was a reliable effect of display size for both the conjunction task, $F(3, 54) = 6.075$, $\eta^2_p = .252$, $p = .001$, and the spatial configuration task, $F(3, 57) = 41.426$, $\eta^2_p = .686$, $p = 1.24 \times 10^{-13}$ (lower panel in Figure 3). Post-hoc t tests indicated that in the conjunction search

task participants committed more errors at display size 18 (13.05 %) than at display sizes 12 (8.84 %; p = .028) and at 6 (6.79 %; 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 = .161).

Figure 3 should be inserted around here

Error analysis

To test if the shape change in an RT distribution is due to an increase of miss errors (Wolfe et al., 2010), we also analyzed 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).

A two-way ANOVA at the miss error rate showed reliable main effects of display size, F(3, 165) = 38.08, η^2_p = .409, p = 1×10^{-13} and search task, F(2, 55) = 5.75, η^2_p = .173, p = .005 as well as an interaction between these factors, F(6, 165) = 10.85, η^2_p = .283, p = 3.62×10^{-10}. Both the spatial configuration, F(3, 57) = 41.37, η^2_p = .685, p = 1.25×10^{-13}, and the conjunction search task, F(3, 54) = 6.08, η^2_p = .253, p = .001, showed increasing miss errors as the display size increased, but not the feature search task, F(3, 54) = 1.52, η^2_p = .078, p = .221. False alarms showed only a display size effect, F(3, 165) = 3.94, η^2_p = .067, p = .010. The reliable effect of false alarm errors was observed in both feature, F(3, 54) = 2.81, η^2_p = .135, p = .048 and conjunction search, F(3, 54) = 2.96, η^2_p = .141, p = .040, but not in spatial configuration search, F(3, 57) = 1.14, η^2_p = .057, p = .340 (Figure 4).

Figure 4 should be inserted around here

Distributional analysis

Figure 3 shows also the distributions of the means of RTs and error rates across the display sizes and tasks. Three noticeable characteristics are evident. Firstly, performance in the feature search task changed little across the display sizes. Secondly, in the two inefficient search tasks (conjunction & spatial configuration), increases in the display size not only delayed central RTs within the distribution (i.e., the estimates that median and mean results aim to capture), but there was also a shift in the entire response distribution. Thirdly, the increases in task difficulty affected not only central RTs but also the variability of the distribution. There were also some differences between the conjunction and spatial configuration tasks. The widely distributed RTs for the spatial configuration task elongated the central measures of performance as well as the long latency responses. Notably, the difference between the effects of the different display sizes at the long end of the response distribution was exacerbated for the spatial configuration search task.

The box-and-whisker plot for the error rates showed a similar pattern across the display sizes to that for the mean of the RT data, although the effects were relatively modest in magnitude.

Figure 5 should be inserted around here

Figure 5 shows the RT distributions at the different display sizes and search tasks. The distributions were constructed based on the mean RTs ($N_{feature}$ and $N_{conjunction}$ = 19 and $N_{spatial}$ = 20; 464 data points). The feature search showed a leptokurtic (i.e., peaky) distribution and the quantile-quantile plots indicated clear deviations at both ends of the distributions. The conjunction and spatial configuration search tasks at the small display sizes, however, showed only moderate signs of violation of the normality assumption, though at the large display sizes, the distributions were platykurtic (flat) and the long RT latencies showed signs of deviation from a normal distribution.

Figure 6 & Figure 7 should be inserted around here

Figure 6 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 7 shows the empirical cumulative distributions, drawn based on trial RTs (43485 and 109036 data points in our 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.

Contrasts with prior data

We compared our data with those of Wolfe et al.'s (2010). A comparison of the mean RT and error rates indicated similar patterns across the studies (Figure 3), as is suggested also by the probability and cumulative density plots, shown in Figure 7 -8.

With only a small number of participants, it is difficult to rule out the normality assumption when examining the mean RTs (see sub-plots in Figure 8), but the data for the trial RTs reveal a skewed distribution (Figure 9).

Figure 8 & Figure 9 should be inserted around here

The HBM estimates

In this section, we firstly presented each parameter separately for the respective ANOVA results, and we compared the data for the three search tasks at the different display sizes, modeled by the HBM. Next, we 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. We used deviance information criterion (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 and Wald functions showed the smallest DIC, the DICs across the four fitted functions were close. Moreover, the diagnostic of the gamma HBM suggests its posterior distributions did not converge. Excluding the non-converged gamma function, we 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).

Table 1 should be inserted around here

Shift. A two-way (task × display size) ANOVA[6] revealed a significant effect of task, $F(2, 55) = 129.748$, $p = 1.0 \times 10^{-13}$, $\eta^2_p = .825$, and display size, $F(3, 165) = 9.031$, $p = 1.43 \times 10^{-5}$, $\eta^2_p = .141$, but there was no reliable interaction, $F(6, 165) = 1.14$, $p = .34$, $\eta^2_p = .040$. Post-hoc t tests showed that the feature search had a smaller shift value than conjunction search, which also had a

[6] For the same reason as footnote 3, we analysed the three level of task as a between-participant factor.

smaller value than the spatial configuration search (246 ms, 342 ms, vs. 436 ms; $p = 2.37 \times 10^{-10}$, & 2.83×10^{-10}).

The plot in the upper panel of Figure 10 shows two important characteristics for target present trials. First, the non-parametric regression lines show that the shift parameter varies little across the participants in the four display sizes within a task. Second, each task demonstrates a different magnitude of the shift parameter, suggesting that varying the search process gives more weight to this parameter than varying display sizes.

Scale. The two-way (task × display size) ANOVA was significant for the task, $F(2, 55) = 161.70$, $p = 1.0 \times 10^{-13}$, $\eta^2_p = .855$, the display size, $F(3, 165) = 39.75$, $p = 1.0 \times 10^{-13}$, $\eta^2_p = .420$, and for the task × display size interaction, $F(6, 165) = 19.31$, $p = 1.0 \times 10^{-13}$, $\eta^2_p = .413$.

Separate ANOVAs showed that there were reliable display size effects for both the conjunction task, $F(3, 54) = 10.000$, $p = 2.42 \times 10^{-5}$, $\eta^2_p = .357$ (206, 257, 301 and 334 ms) and the spatial configuration task, $F(3, 57) = 33.47$, $p = 1.42 \times 10^{-12}$, $\eta^2_p = .638$ (302, 444, 607 and 760 ms), but not for the feature search task, $F(3, 54) = .084$, $p = .968$, $\eta^2_p = .005$ (201, 207, 206 and 205 ms). Post-hoc t tests showed that there were significant differences between all display sizes in spatial configuration search, $p = 7.59 \times 10^{-3}$, 9.34×10^{-6}, 1.34×10^{-7}, .021, 1.56×10^{-4}, & .04 (3 vs. 6, 3 vs. 12, 3 vs. 18, 6 vs. 12, 6 vs. 18, & 12 vs. 18; Bonferroni correction for multiple comparisons). This held for conjunction search only for the 3 vs. 12, and 3 vs. 18 comparisons (ps = .001; Bonferroni correction for multiple comparisons). No significant differences were observed in feature search.

The lower panel of Figure 10 shows two important characteristics. First, the regression lines indicate increasing variability (i.e., decreasing ribbon density) as the display sizes increased for conjunction and spatial configuration search, but not for the feature search. Second, the display size effect only became noticeable for the inefficient search tasks, in line with the RT mean.

Shape. The two-way (task × display size) ANOVA revealed significant effects of task, $F(2, 55) = 23.50$, $p = 4.21 \times 10^{-8}$, $\eta^2_p = .461$, and marginally display size, $F(3, 165) = 2.44$, $p = .067$, $\eta^2_p = .042$ and their interaction, $F(6, 165) = 3.45$, $p = .003$, $\eta^2_p = .111$.

Separate ANOVAs showed reliable display size effects for both conjunction search, $F(3, 54) = 4.21$, $p = .009$, $\eta^2_p = .190$ (1496, 1731, 1695 and 1702 ms) and spatial configuration search, $F(3, 57) = 4.45$, $p = .007$, $\eta^2_p = .190$ (1573, 1541, 1397 vs. 1529 ms), but not for feature search, $F(3, 54) = 2.13$, $p = .106$, $\eta^2_p = .106$ (1702, 1819, 1976 vs. 1850 ms). Post-hoc t tests showed significant display size differences at 3 vs. 6, 3 vs. 12, and 3 vs. 18, $p = .022$, .018, and .009 in the conjunction search. In the spatial configuration search, the display size differences were observed at 3 vs. 12, 6 vs. 12, and 12 vs. 18, $p = .013$, .047, and .003 (Bonferroni correction for multiple comparisons).

The plots in the middle panel of Figure 10 show two important characteristics. First, the regression lines indicate differences between the search conditions only at large display sizes (i.e., 6, 12 and 18). Second, there is a U-shaped function for the spatial configuration task – both for the magnitude and variability of the shape parameter. Interestingly, these results are not evident in Wolfe et al.'s (2010) data. The emergent decreases in the mean shape parameter and the associated increase in the variability suggest that additional factors influenced search at the larger display sizes here – which we suggest reflect grouping between the elements. We elaborate on this proposal in the General Discussion.

Figure 10 should be inserted around here

Diffusion model

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

Drift rate. The two-way (task × display size) ANOVA revealed a significant effect of task, $F(2, 55) = 9.47$, $p = 2.92 \times 10^{-4}$, $\eta^2_p = .256$, but not display size, $F(3, 165) = 0.472$, $p = .703$, $\eta^2_p = .009$, and there was no interaction, $F(6, 165) = 1.27$, $p = 0.28$, $\eta^2_p = .044$. Post-hoc t tests showed that the feature search (0.323) drifted faster than the conjunction search (0.265; marginally significant, 95 % CI, -0.117 to .001, $p = .057$) and the spatial configuration search (0.220; 95 % CI, 0.044 to 0.161, $p = 1.81 \times 10^{-4}$). No difference was found between the conjunction and spatial configuration searches.

The drift rate, shown in the upper left panel in Figure 11, manifests two critical characteristics. First, for both the feature and the conjunction search tasks, the drift rate evolves at a constant rate across the display sizes. The second noticeable characteristic is a clear separation of the drift rate across the three tasks, suggesting differences in the rate at which sensory evidence accumulates in the different tasks. There was a tendency also for the drift rate to rise at the large display size in the spatial configuration task (Figure 11), suggesting that there was an emergent factor, such as the grouping of homogeneous distractor elements, which increased the drift rate – though the variability across observers suggests that this was not universally the case for all participants. This was not evident in absent trials[7]. This upward trend was also not present in the data of Wolfe et al. (2010).

Non-decision time. The two-way (task × display size) ANOVA was significant for the main effect of task, $F(2, 55) = 5.64$, $p = .006$, $\eta^2_p = .170$ and the interaction, $F(6, 165) = 4.16$, $p = .001$, $\eta^2_p = .131$. 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 to 38.1 ms, $p = .008$) and conjunction search (61 ms; 95 % CI, 0.707 to 34.2 ms, $p = .038$). There were reliable display size effects for the spatial configuration task, $F(3, 57) = 6.886$, $p = 4.89 \times 10^{-4}$, $\eta^2_p = .266$ (60.59, 80.54, 89.50 vs. 84.23 ms), but not at feature or conjunction search tasks.

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

Boundary separation. The two-way (task × display size) ANOVA revealed a significant effect of the task, $F(2, 55) = 31.75$, $p = 6.81 \times 10^{-10}$, $\eta^2_p = .536$, the display size, $F(3,165) = 7.6$, $p = 8.61 \times 10^{-5}$, $\eta^2_p = .121$ and a task × display size interaction, $F(6, 165) = 4.76$, $p = 1.69 \times 10^{-4}$, $\eta^2_p = .147$. The value of the boundary separation for feature search (0.111) was smaller than that for spatial configuration search (0.192, $p = 1.01 \times 10^{-9}$, 95 % CI, 0.055 to 0.107), and this was not different from that found for conjunction search (0.132). The conjunction search task also demonstrated a reliable difference from the spatial configuration condition ($p = 1.49 \times 10^{-6}$, 95 % CI, 0.034 to 0.086). Separate ANOVAs showed reliable display size effects for the spatial configuration search, $F(3, 57) = 6.73$, $p = .001$, $\eta^2_p = .262$ (0.148, 0.170, 0.201 vs. 0.249; 3, 6, 12, vs. 18), but not for the feature or conjunction searches.

Figure 11 and Table 2 should be inserted around here

General Discussion

The study applied an integrated approach to modeling visual search data. We 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 allows us to reveal 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. Our 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. We used conventional distributional analyses to examine empirical RT distributions and the associated parameters were complemented with Bayesian-based hierarchical modeling to optimise estimates. Moreover, we examined those distributional parameters against a plausible computational model – the EZ2 diffusion model – to link the distributional parameters to underlying psychological processes.

Replicating many previous findings in the search literature our 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 reviews). The display size effect present in the feature search (415, 426, 432 and 437 ms) suggests some limitations on selecting feature targets but the analyses based on mean RTs do not differentiate if the effect ($\eta^2_p = .294$) 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 diffusion 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 at 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, our findings indicated that spatial configuration search was particularly difficult (Bricolo, Gianesini, Fanini, Bundesen, & Chelazzi, 2002; Kwak, Dagenbach, & Egeth, 1991; Woodman & Luck, 2003). This could reflect either a reduction in the guidance of search from spatial configuration compared with simple orientation and color information, or the length of time taken to identify each item after it had been attended. Interestingly, although when compared with the standard deviation of the conjunction search (9.68 ms), configuration search generally showed a larger value across participants (24.54 ms), the standard deviations 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 does facilitate search, a point we return to below (Bergen & Julesz, 1983; Chelazzi, 1999; Duncan & Humphreys, 1989; Heinke & Humphreys, 2003; Heinke & Backhaus, 2011).

Methodological issues

The analyses for the mean RTs however do not always accord with the analyses of trial RTs. For example, the density plots at mean RTs (Figure 5) suggest that the data are distributed symmetrically, contrasting with the common notion that an RT density curve tends to be positively distributed towards long latencies (Luce, 1986). However, the analyses of the trial RTs (Figure 6) reveal clearly skewed RT distributions. This is because the procedure of determining a representative value using a central location parameter (the mean in the case of our data) from each observer's RT distribution of a condition (individual curves in Figure 6) is affected greatly by the weight of slow RTs. The conditions and observers that contribute the slow responses tend to move the central location towards long latencies within a distribution, hence we observe more symmetrical and sub-Gaussian (i.e., flat) density curves for the mean RTs. Additionally, because

the density curve for the mean RTs is usually constructed by a biased central location parameter (with respect to a skewed RT distribution), the nature of the RT distribution (e.g., if there is a majority of quick responses and a minority of slow responses) is hidden by an unrepresentative central location parameter. A solution has been proposed recently by using some variants of distributional analyses (Balota & Yap, 2011; Bricolo et al., 2002; Heathcote et al., 1991) and these have been applied to various cognitive tasks (Palmer et al., 2011; Sui & Humphreys, 2013; Tse & Altarriba, 2012; Wolfe et al., 2010). Essentially, the distributional approach constructs an empirical distribution using trial RTs from each individual in a condition and uses a plausible distributional function (such as Weibull or ex-Gaussian) to extract distributional parameters, with the parameters averaged across participants then being compared across the different conditions. This approach descriptively dissects an RT distribution into multiple components (e.g., mu, sigma, & tau), each potentially reflecting contrasting psychological process (Balota & Yap, 2011). However, the link between the component and the underlying process can be elusive (Matzke & Wagenmakers, 2009) without directly modeling of the underlying factors. We address this issue by contrasting the empirical data modeled, respectively, by both a distributional approach (HBM) and a computational model (the EZ2 diffusion model).

On top of the analyses of mean performance, the integration of hierarchical Bayesian and EZ2-diffusion modeling helped to throw 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, Võ, Evans, & Greene, 2011). The EZ2-diffusion 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 conceptualization 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.

Model-based analysis

The EZ2 diffusion model and the HBM suggest that distributional parameters reflect different aspects of search. First, the shift parameter varied across the search tasks and display sizes, a pattern that is in line with our illustration and the ideal analysis (see Figure 1 & Appendix B). This parameter reflects the psychological processes influencing evenly all ranges of RTs. One of the diffusion processes likely to influence the shift changes is the drift rate, which showed only the main effect of the task. As the drift rate aims to model the rate of information accumulation determined by the *goodness-of-match* between templates and search stimuli, the shift parameter appears to result from the change in the quality of the memory match. This is a plausible account, because the three search tasks demand contrasting matching processes, from (i) feature search requiring only pre-attentive parallel processing to extract just one simple salient feature, to (ii) conjunction search, where binding two simple features must be bound to facilitate a good match, and to (iii) spatial configuration search, demanding both features binding and coding of the configuration of the features. The lack of interaction with display size further supports our argument that the shift reflects the factors affects the RT distribution equally. The weak display size effect can be readily explained by the crowded layout we used; it was not observed [$F(3,75) = 0.016$, $p = 0.997$] in Wolfe et al.'s data (2010). This weak effect at the shift parameter is further accounted for by our visually weighted plot in the drift rate parameter, showing a clear split of trends and an increase of between-observer variation at the large display size. Specifically, a subset of participants adopted a strategy similar as those participants in Wolfe and colleagues' study. They did not assemble similar a search unit, so the predicted drift rate decreases at the large display sizes,

whereas the other subset of participants benefited from the crowded homogeneous distractors and thus increased drift rate at the large display sizes.

Another account for the strong task effect, but the weak effect of display size, is that it reflects a process such as the recursive rejection of distractors proposed by Humphreys and Müller (1993) in their SERR model of visual search (see also Heinke & Humphreys, 2005). Humphreys and Müller (1993) argued that search can reflect the grouping and then recursive rejection of distractors. The process here may reflect the strength of grouping rather than the number of distractors since multiple distractors may be rejected together in a group – indeed effects of the number of distractors may be non-linear as grouping can increase at larger display sizes. Grouping and group selection both reflect the similarity of targets and distractors and the similarity of the distractors themselves, and these two forms of similarity vary in opposite directions in conjunction and spatial configuration search (relative to a feature search condition as employed here, there is stronger target-distractor grouping and weaker distractor-distractor grouping; see Duncan & Humphreys, 1989). If the process of distractor rejection is more difficult in conjunction and configuration search, compared with feature search, then there will be effects on a parameter reflecting this process, and this may not vary directly with display size, as we observed.

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 display size of the spatial configuration task is in line with a sudden increase of the drift rate standard deviation (0.080, 0.050, 0.054, 0.344), suggesting

either (1) a change in the quality of a match between the stimuli and the template or (2) a variable grouping unit (amongst different observers) affecting the recursive rejection process.

In addition, we observed a general increase in the values of the shape parameter from 1.73 of the display size 3, 1.86 of the display size 6, 2.05 of the display size 12, to 1.96 of display size 18 on absent trials in the spatial configuration task, $F(3, 57) = 6.13$, $p = .001$, $\eta^2_p = .244$. 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 (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 increase shape parameters. We, 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, our 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), we 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 display (i.e., the drift rate changes). This implies the factors contributing a change in different parts of an RT distribution will result in its shape changes. As our simulation study shows (Appendix B), doubling the shape parameter results in a decrease at

the boundary separation (in line with the miss-error account) and an increase at the drift rate (in line with the *goodness-of-match* account). Both of the diffusion parameters likely are the processes driven the changes at the distributional shape.

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 our 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.

Distributional parameters reflect underlying processes

The RT distributional parameters were posited, under the framework of the stage model of information processing, to reflect different aspects of peripheral and central processing. The shift parameter was associated with the speed of peripheral processes (i.e., irreducible minimum response latency, Dzhafarov, 1992), the scale parameter with the speed of executing central processes, and the shape together with the scale parameters related to the insertion of additional stages into the central processing (Rouder et al., 2005).

Using the benchmark paradigms of visual search (Wolfe et al., 2000), our data indicate that the shift parameter, instead of reflecting the speed of peripheral processes, may be associated with the process of distractor rejection and the quality of the match between a template and a search display. This is supported by the analysis using the EZ2 diffusion model. As we argued previously, the shift parameter captures the factors that influence the entire RT distribution equally. A possible situation that the peripheral process may result in a clear shift change is when the other two parameters are kept constant. That is, when no factor influences the decision-making process and when the shape of an RT distribution is unchanged. We suggest that the data better reflect a process such as the recursive rejection of the grouped distractors and the quality of the match to a target template, which, when accurate, contributes to an entire RT distribution.

Our results for the scale parameter are consistent with those of Rouder and colleagues (2005) in suggesting that it reflects the speed of execution in a central decision-making process. As the execution speed closely links with the decision boundaries and the initial state of sensory information an observer sets for a response trial, we observed a similar pattern in the scale parameter, the boundary separation and the non-decision time. The pattern in the non-decision time is readily accounted for by the fact that EZ2 diffusion model absorbs the parameter reflecting the initial state of sensory evidence into the non-decision time. The distance between the decision boundary and the initial state of sensory evidence can then be taken as reflecting changes in the response criteria and hence altering the scale of an RT distribution.

For the shape parameter we observed an emergent effect of perceptual grouping at the large display size in the spatial configuration search. This is in line with the drift rate data in that the drift rate was slower for the spatial configuration search task relative to the two simple search tasks both in our data (0.323, 0.265 vs. 0.220) and those of Wolfe et al. (2010) (0.341, 0.299 vs. 0.203). In Palmer et al.'s analysis (2011) no task effect was found in the shape parameter. Using the HBM we observed a significant task effect, $F(2, 55) = 23.50$, $p = 4.21 \times 10^{-8}$, $\eta^2_p = .461$, suggesting that the

previous result might reflect a lack of power. The observations of shape invariance in Palmer et al.'s analysis could also be interpreted in term of a memory match account (Ratcliff & Rouder, 2000). This account presumes that, when the integrity of a memory match between the template and search items is still intact, the evidence strength is strong enough to permit a correct decision (Smith, Ratcliff, & Wolfgang, 2004; Smith & Sewell, 2013). Since the previous study recruited fewer participants and some might find strategies to conduct the difficult searches still using the same processing stages as the feature search task, the shape parameter reflects only a marginal effect.

Another possible factor that may explain the different finding at the shape parameter is illustrated by the drift rate visually weighted plot. The visually weighted regression lines predict two groups of participants accumulating sensory evidence at different rates, but indicate only one homogeneous group in Wolfe et al.'s data. As our simulated study shows (Appendix B), the drift rate can also weigh-in to change the shape parameter. This suggests that some of our participants took advantage of the crowded layout to facilitate search using the spatial configuration task, and some did not. This did not happen in Wolfe et al.'s sparse layout, and likely contribute also to our significant finding at the shape parameter on present trials.

Limitation

The analytic approach we 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, we selected the Weibull function, because it permits a reliable converged posterior distribution, has broad application to memory as well as visual search (Logan, 1992; see also Hsu & Chen, 2009) and application to the hierarchical Bayesian framework (Rouder et al., 2005).

Conclusion

In conclusion, our study shows how the HBM-based distributional analysis, complemented with the EZ2 diffusion model, can help to clarify processes mediating visual search. 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 (e.g., effects on non-decision factors and the separation of decision boundaries). We 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.

References

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. doi:10.1177/0963721411408885

Bergen, J. R., & Julesz, B. (1983). Parallel versus serial processing in rapid pattern discrimination. *Nature*, *303*(5919), 696–698.

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. doi:10.1162/089892902320474454

Brooks, S. P., & Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, *7*(4), 434–455.

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. doi:10.1038/nn.2925

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

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. doi: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 J. M. Chambers & T. J. Hastie (Eds.), *Statistical Models in S* (pp. 309–376). New York, USA: Chapman and Hall.

Cousineau, D., Brown, S., & Heathcote, A. (2004). Fitting distributions using maximum likelihood: Methods

and packages. *Behavior Research Methods, Instruments, & Computers*, *36*(4), 742–756.

Crawley, M. J. (2002). *Statistical Computing: An Introduction to Data Analysis using S-Plus*. Chichester, West Sussex, England; 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. Erlbaum Associates.

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, *96*(3), 433–458. doi:10.1037/0033-295X.96.3.433

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

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

Gelman, A. (2004). *Bayesian data analysis*. Boca Raton, Fla.: Chapman & Hall/CRC.

Gelman, A., & Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge; New York: Cambridge University Press.

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. doi:10.1016/j.jmp.2009.01.006

Gu, S. L., Gau, S. S., 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. doi:10.1016/j.ridd.2013.07.025

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

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. doi: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. doi:10.1037/0033-295X.110.1.29

Heinke, D., & Humphreys, G. W. (2005). Computational models of visual selective attention: A review. In G. Houghton (Ed.), *Connectionist Models in Cognitive Psychology* (Vol. 1, pp. 273–312). England; New York: Hove; Psychology press.

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. doi:10.1037/h0080637

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. Doi:10.3758/APP.71.7.1664.

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

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

Kruschke, J. K. (2010). What to believe: Bayesian methods for data analysis. *Trends in Cognitive Sciences*, *14*(7), 293–300. doi: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. doi:10.3758/BF03212181

Lamy, D. F., & Kristjánsson, A. (2013). Is goal-directed attentional guidance just intertrial priming? A review. *Journal of Vision*, *13*(3), 14. doi:10.1167/13.3.14

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

Logan, G. D. (1992). Shapes of reaction-time distributions and shapes of learning curves: A test of the instance theory of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(5), 883–914. doi:10.1037/0278-7393.18.5.883

Luce, R. D. (1986). *Response Times: Their Role in Inferring Elementary Mental Organization*. New York; Oxford: Oxford University Press ; Clarendon Press.

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. doi:10.1002/sim.3680

Macmillan, N. A., & Creelman, C. D. (2004). *Detection Theory: A User's Guide*. Psychology press.

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. doi:10.3758/PBR.16.5.798

Merkle, E., & van Zandt, T. (2005, August). WinBUGS Tutorial. Retrieved from http://maigret.psy.ohio-state.edu/~trish/Downloads/.

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. doi:10.1167/13.8.24

Nelder, J. A., & Mead, R. (1965). A simplex method for function minimization. *Computer Journal*, *7*(4), 308–313.

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. doi: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. doi:10.1111/1467-8721.ep10772534

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).

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*(2), 59–108. doi:10.1037/0033-295X.85.2.59

Ratcliff, R., & McKoon, G. (2007). The Diffusion Decision Model: Theory and Data for Two-Choice Decision Tasks. *Neural Computation*, *20*(4), 873–922. doi:10.1162/neco.2008.12-06-420

Ratcliff, R., & Murdock, B. B. (1976). Retrieval processes in recognition memory. *Psychological Review*, *83*(3), 190–214. doi:10.1037/0033-295X.83.3.190

Ratcliff, R., & Rouder, J. N. (2000). A diffusion model account of masking in two-choice letter identification. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(1), 127–140. doi:10.1037/0096-1523.26.1.127

Ratcliff, R., & Smith, P. L. (2004). A Comparison of Sequential Sampling Models for Two-Choice Reaction Time. *Psychological Review*, *111*(2), 333–367. doi:10.1037/0033-295X.111.2.333

Riddoch, J. M., & Humphreys, G. W. (1987). Perceptual and Action Systems in Unilateral Visual Neglect. *Advances in Psychology*, *45*, 151–181.

Rohrer, D., & Wixted, J. T. (1994). An analysis of latency and interresponse time in free recall. *Memory & Cognition*, *22*(5), 511–524. doi: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. doi: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. doi: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. doi:10.3758/BF03257252

Rouder, J. N., & Speckman, P. L. (2004). An evaluatin of the Vincentizing method of forming group-level response time distributions. *Psychonomic Bulletin & Review*, *11*(3), 419–427. doi:10.3758/BF03196589

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. doi:10.1037/a0020793

Schwarz, W. (2001). The ex-Wald distribution as a descriptive model of response times. *Behavior Research Methods, Instruments, & Computers*, *33*(4), 457–469. doi:10.3758/BF03195403

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/

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. doi:10.1080/03640210802414826

Smith, P. L., Ratcliff, R., & Wolfgang, B. J. (2004). Attention orienting and the time course of perceptual decisions: response time distributions with masked and unmasked displays. *Vision Research*, *44*(12),

1297–1320. doi:10.1016/j.visres.2004.01.002

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. doi:10.1037/a0033140

Stoet, G. (2010). PsyToolkit: A software package for programming psychological experiments using Linux. *Behavior Research Methods*, *42*(4), 1096–1104. doi: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. doi:10.1080/13506285.2013.800621

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. doi:10.1073/pnas.1304429110

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*(3), 663–676. doi:10.1017/S1366728912000077

Van Zandt, T. (2000). How to fit a response time distribution. *Psychonomic Bulletin & Review*, *7*(3), 424–465. doi:10.3758/BF03214357

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. doi: 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. doi:10.3758/BF03194023

Wagenmakers, E.-J., Maas, H. L. J. van der, 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. doi: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. doi:10.1037/0096-1523.15.4.664

Wolfe, J. M. (1998). What can 1 million trials tell us about visual search? *Psychological Science*, 9(1), 33-39. doi: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., & Horowitz, T. S. (2008). Visual search. *Scholarpedia*, *3*(7), 3325. doi:10.4249/scholarpedia.3325

Wolfe, J. M., Palmer, E. M., & Horowitz, T. S. (2010). Reaction time distributions constrain models of visual search. *Vision Research*, *50*(14), 1304–1311. doi: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. doi: 10.1016/j.tics.2010.12.001

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. doi:10.1037/0096-1523.29.1.121

Author Note

Yi-Shin Lin and Dietmar Heinke, School of Psychology, University of Birmingham; Glyn W. Humphreys, Department of Experimental Psychology, University of Oxford.

Correspondence concerning this article should be addressed to Yi-Shin Lin, School of Psychology, University of Birmingham, Edgbaston, B15 2TT, UK; e-mail: yxl193@bham.ac.uk, or to Glyn W. Humphreys, Department of Experimental Psychology, University of Oxford, South Parks Road, Oxford, OX1 3UD, UK; email: Glyn.humphreys@psy.ox.ac.uk

Appendix A

HBM simulations

The stimulation study was to examine the estimation biases on three standard distributional parameters– mean, variance, and skew–when various probability functions were fitted under different sample sizes per experimental condition and when true distributions generating RTs are known. The study helped to clarify whether the per-condition trial number is sufficient to allow reliable parameter estimation using the HBM. The conclusion from this simulation suggested that (1) no difference between HBM and maximum likelihood estimation (MLE) when sample size was larger than 120; (2) HBM was better than MLE to estimate variance when the sample size was small; (3) the specification of equally plausible probability function is crucial only when it matches the true distribution that generates the RT data.

The simulations[8] examined four scenarios, alternatively assuming that the true distribution followed a normal, an ex-Gaussian, a Wald or a Weibull function (which adopted Cousineau, Brown & Heathcote's parameter values, 2004). Specifically, we used the parameter values listed in Table 3 of Courineau et al.'s report to construct four true distributions, which in turn repeatedly generated randomly simulated RTs. The simulations generated twenty homogeneous participants; each participant contributed RT observations in 10 different sample sizes ranging from 20 to 470 with a step size of 50. The simulated data were then submitted separately to the HBM and the maximum likelihood estimation (MLE) to estimate shift, scale, and shape parameters. Both methods assumed that RTs are random variables generated by the Weibull function. Those parameters were then analytically converted to the mean, the variance and the skewness to evaluate the performance of the two methods.

[8] We used R routines – rnorm, rexGAUS, rinvgauss, and rweibull3 – to generate those computer-simulated data.

Figure 12 should be inserted around here

Figure 12 shows the absolute value of the difference between the true value and the estimates at the mean, variance and skewness of a distribution. In general, no differences were observed between the two methods when estimating the mean. The only factor that improved the estimation was per-condition sample size, $F(9, 1520) = 64.46$, $p < .0000$. The more observations are in a condition, the higher the precision of the estimate is. The bias dropped rapidly when the trial number surpassed 100, from 17.59 ms at 20 observations to 6.74 ms at 120 observations and it decreased at a slower rate when the trial size was over 120 observations (an average, 5.4 ms). The specification of the true distribution did not alter the precision of mean estimation, $F(3, 1520) = 1.912$, $p = .126$, even though both estimation methods assume that a Weibull function accounts for the RT data.

In contrast, the HBM demonstrated a clear advantage over the maximum likelihood method when recovering the variance, $F(1, 1520) = 9.345$, $p = .0023$. Again, a large number of observations helped to improve the fit, $F(1, 1520) = 29.84$, $p < .0000$. Importantly, the HBM estimated the parameters better than the MLE at the small sample size ($N = 20$), $F(1, 152) = 6.907$, $p = .0095$, though the difference was gradually resolved when the observation numbers exceeded 120, $F(1, 152) = .936$, $p = .335$ (i.e. the dash line in *Figure 12*). The mis-specification of the underlying distribution resulted in different estimations of variance, $F(3, 1520) = 7.635$, $p < .0000$. Both methods needed a sample size larger than 170 to resolve this issue. The parameter recovery was better when the true distribution followed the Weibull function.

As for variance, the skewness was again better estimated by the HBM than the MLE, $F(1, 1520) = 22.596$, $p < .0000$. The correct specification of the distribution played an important role in

estimating the skewness, $F(3, 1520) = 1818.549$, $p < .0000$. Specifically, when the true distribution followed either ex-Gaussian or Wald functions, there was no advantage to using HBM after a trial size of 70. In this case increasing the sample size did not mitigate the problem, whereas when the true distribution followed a Weibull function, the HBM showed consistently higher precision than the MLE (see Figure 13). Interestingly, the HBM also gave better estimates than the MLE when the true distribution was normal (i.e., skewness = 0). Increasing the sample size helped to resolve the inferiority of MLE, but for this there needed to be at least 220 observations in each condition.

Figure 13 should be inserted around here

Model diagnostics

In this section, we assess the performance of the HBM and the MLE estimation methods for the three parameters expressed in the Weibull function. Firstly, a goodness-of-fit index, the Anderson-Darlings statistic, was used to compare the fits between the two methods. Next, we present graphical and statistical diagnostics on the convergence and *stationarity* of the MCMC chains. *Stationarity* refers to whether the multi-chain process of MCMC converges to a reliable and single distribution after a long iteration process. Each step of MCMC process uses the pre-defined model (i.e., the Weibull function in our case) to fit the empirical data and predicts a posterior distribution. The posterior distribution is then used as a prior distribution in next step to fit a new posterior distribution. This process iterates itself until the pre-defined iteration times (105000 times in our setting). The final posterior distribution is the predicted distribution presuming as the true distribution underlying an RT distribution in the visual search paradigm we examined. We ran three separate independent MCMC processes (i.e., 3 chains) to test if all three converged to an identical final posterior distribution. If this is so, it indicates that our hierarchical Bayesian setting

and the Weibull function provide an appropriate approximation towards an RT distribution. A non-convergent MCMC process does not provide a reliable prediction for the final posterior distribution. That is, it predicts, even after a long iteration process, different distributions accounting for the empirical data.

Goodness of fit. As RT distributions are formed by a continuous rather than discrete random variable, we used the Anderson-Darling statistic, rather than chi-square, as the index of goodness-of-fit. Figure 14 shows that there were few differences between the MLE fit and the HBM fit. Both methods performed similarly and improved their fits when as the display size increased.

Figure 14 should be inserted around here

Bayesian model diagnosis. In this section, we examined if the estimated parameters converged to a stationary posterior distribution and evaluated if the setting for conducting MCMC sampling were appropriate to permit reliable inferences.

Figures 14 and 15 shows one of the examples of the diagnostic plots, indicating the convergence of β (i.e., the shape parameter), which was estimated from one of the participants performing the spatial configuration search at display size 18 in the target-present condition. The figures are the posterior density curve and the autocorrelation plot. We ran three Markov chains in parallel to approach a stationary posterior distribution, so there are three sets of parallel data to represent three processes. Three MCMC chains mixed quickly to a limited range after the iteration process began (after 5000 burn-in iterations), suggesting that the posterior distribution reached a

reliable stationary point[9]. This impression is supported by the posterior density distributions (Figure 15), showing all three chains predict overlapping distributions with very similar shapes and dispersions, consistent with the three chains predicting an identical posterior distribution. The autocorrelation plots (Figure 16) show the MCMC sampling interval (in our case, the computer program took one sample every 4 iterations) is appropriate to reduce the inter-iteration correlation rapidly within first 50 iterations (this is after first 5000 samples were discarded).

Figure 15 and Figure 16 should be inserted around here

The information from the graphical diagnoses is compatible with the non-parametric statistical tests. Figure 17 presents the graphical summary of the non-parametric tests for the stationarity across all the conditions and participants. The upper panel shows the Brooks-Gelman-Rubin (*BGR,* Brooks & Gelman, 1998) shrink factor. The recommended *BGR* upper limit is 1.1 (Gelman & Hill, 2007; Gelman, 2004), and values below it are deemed acceptable. We drew the reference line at the grand average 95 % confident interval to allow a clear inspection of the distribution of the statistic. Very few BGR diagnostics exceeded the grand average of 95% confidence interval. Although a few values fall outside the upper limit of the box-and-whisker plots, they are nevertheless all within the recommended upper limit. The BGR shrink factor provides no evidence of an unstable mixing of the three chains, confirming the observations from the trace plots.

Figure 17 should be inserted around here

The middle panel shows the Geweke Z score. This test examines the stationarity of the

[9] A non-stationary mix will manifest as three clearly separated lines.

posterior distribution. A Z score exceeding ±1.96 is considered problematic. Only a few cases at display size 6 in the conjunction search and at display size 3 in the spatial configuration search exceeded the two reference lines, drawn at Z =±1.96. In general, the distribution of the Geweke Z is compatible with the stationarity assumption that we have observed in the posterior density plots. In other words, the posterior distribution estimated from the three separate chains converged to an identical distribution.

The lower panel shows the distribution of the p values from Heidelberger and Welch's test. The reference line was drawn at .05 p value. The figure complements the observations found both in the graphical diagnostics and the Geweke Z test. None of the p values exceeded .05 levels. In summary, we found, from both graphical diagnoses and non-parametric tests, no obvious evidence, in all estimated parameters across all participants and conditions, against either the hypothesis of the stationarity of the posterior distributions, the ill-mixed Markov chains, or an unreliable convergence.

Appendix B

EZ2 diffusion model simulations

The stimulation study was to undertaken to examine how the parameters estimated from EZ2 diffusion model correlate with the Weibull parameters in a simple well-controlled situation. We performed 3 case studies that changed only one of the Weibull parameters. Note that the three distributional parameters jointly determine the general shape of a distribution. Thus a change in the drift rate may alter one or more Weibull parameters. The 3 studies respectively doubled the shift, the scale and the shape parameters in a Weibull function and simulated 200 RTs from 20 homogeneous observers.

The data were than submitted to the EZ2 model to estimate the drift rate, boundary separation and non-decision time. The result indicates that firstly doubling the shift parameter resulted in a near two-fold increase of non-decision time (416 vs. 827 ms), small increases in the drift rate (0.012 vs. 0.013), and small decreases of the boundary separation (4.89 vs. 4.57). Second, doubling the scale parameter resulted in a decrease of the drift rate (0.013 vs. 0.009), an increase of boundary separation (4.70 vs. 6.57), and a 10-ms increase at the non-decision time (407 vs. 417 ms). Finally, doubling the shape parameter resulted in an increase of the drift rate (0.013 vs. 0.018), a decrease of the boundary separation (4.57 vs. 3.39) and again a small increase at the non-decision time (410 vs. 507; although this increase was relatively larger than that of doubling the scale parameter). Figure 18 shows a comparison across the three case studies.

Figure 18 should be inserted around here

Appendix C

Why use Weibull function?

The Weibull probability function is one of the many plausible functions that can accommodate positively skewed RT distributions. We chose it to describe RT distributions, because of its parametric characteristics, permitting us to describe the shape of an RT distribution in an intuitive way. Nevertheless, there are other alternatives, such as gamma, log-normal, and Wald functions. All of them are capable of accommodating skewed RT distributions and provide similar descriptive parameters. Here, we delineate our reasons to fit the HBM Weibull function to RT data.

First reason is that the Weibull function is able to provide a concise way to summarize the shape of an RT distribution. As described in the main text and illustrated in Figure 1, changes in t6he three parameters, shift, scale and shape, are associated differently with increases/decreases of RT densities, allowing us to understand how an experimental factor may alter different areas of an RT distribution. Secondly, the 3-parameter gamma function does not converge when fitted with hierarchical Bayesian approach. The compatible 3-parameter gamma function shows signs of non-convergence and perhaps because of this, it fits the data slightly worse, indicated by the DICs. Third, we fitted also the other two 3-parameter functions: Wald and log normal. They provide the same set of descriptive parameters as the Weibull function. The DICs suggest these function fit both ours and Wolfe, Palmer and Horowitz's (2010) data slightly better than Weibull function. However, we have maintained the Weibull function because all four functions fit are very similar for each task, display size, target types, and data set (Figure 19). To test whether any function would work we fitted a Gaussian function. The Gaussian fit DICs (~ -3150) are far worse than the four plausible functions that can

accommodate positive-skewed RT. Below we reported the detailed procedure to fit gamma functions.

Figure 19 should be inserted around here

To test whether the Weibull function fit better than the gamma function, we built a 3-parameter gamma function in the HBM framework. Because currently BUGS does not implement a pre-built 3-parameter gamma function, we used Johnson, Kotz, and Balakrishnan's (1994, pp 337, eq. 17.1) equation to implement the gamma function. Practically, the BUGS code is to change the density function to:

```
#--------------------------#
# 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 we implemented the Weibull function, we assessed the parameters via minus log-likelihood and used pseudo-Poisson (zero) trick. This implementation resulted in unstable, non-converged estimations. Take the shape parameter as an example. In the spatial configuration search participant 3, the shape (beta) estimation yielded three different of posterior distributions and the trace plots from the three chains unstably oscillated around different ranges. In addition, the autocorrelation plots indicated a problem and this does not abate with increasing iterations. In summary, the diagnostics show the gamma function does not converge when fitted with a HBM framework.

This failure of gamma fit could be because (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). Note that we have run a high number of iteration (i.e., 105000) and reasonable thinning length and this still cannot resolve the non-converged gamma fit.

Figures

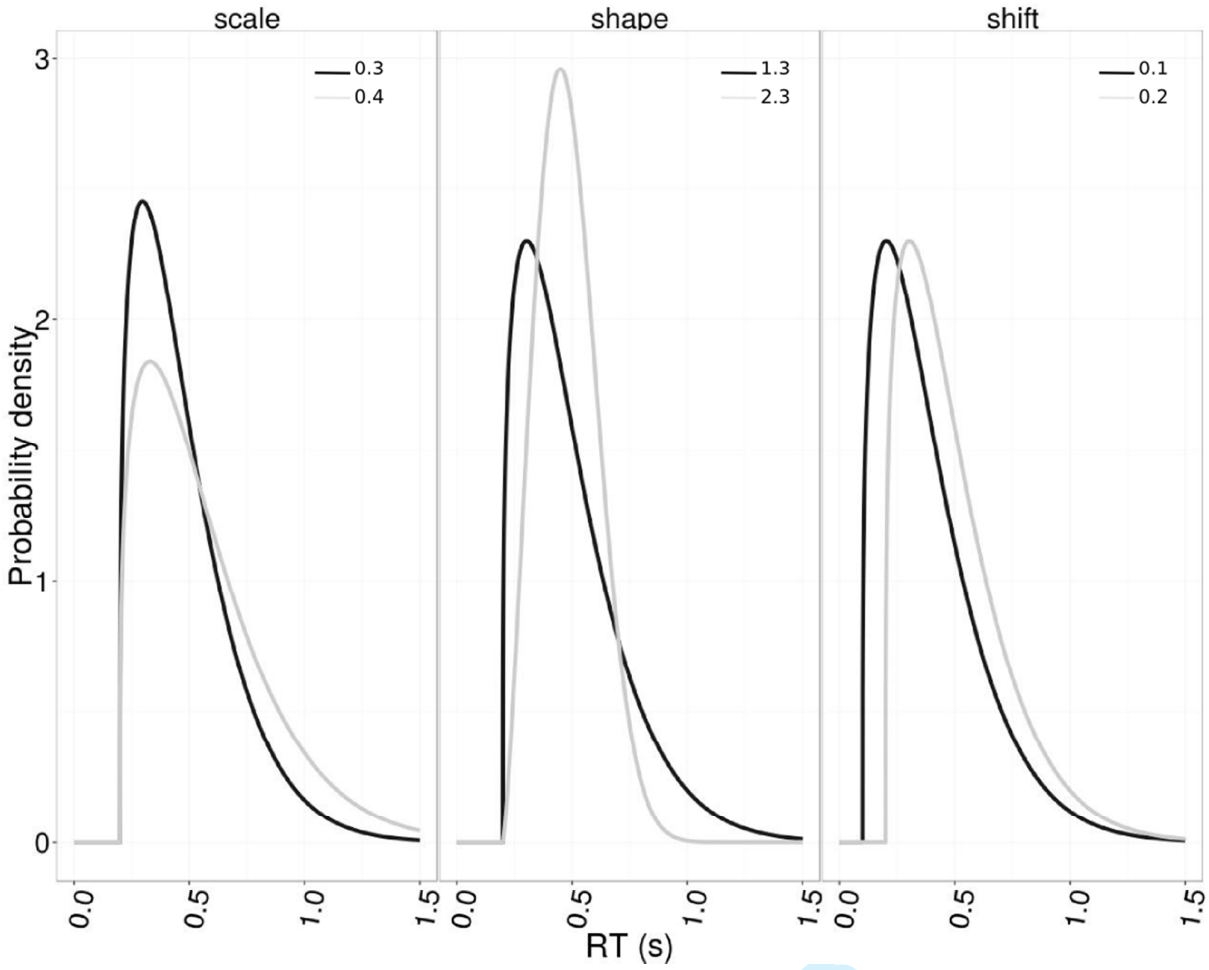


Figure 1. An illustration of the changes of the scale, shape, and shift parameters. The figure was simulated by a 3-parameter Weibull function. The legends in each panel show the extent to which the parameter is adjust while the others are kept constant

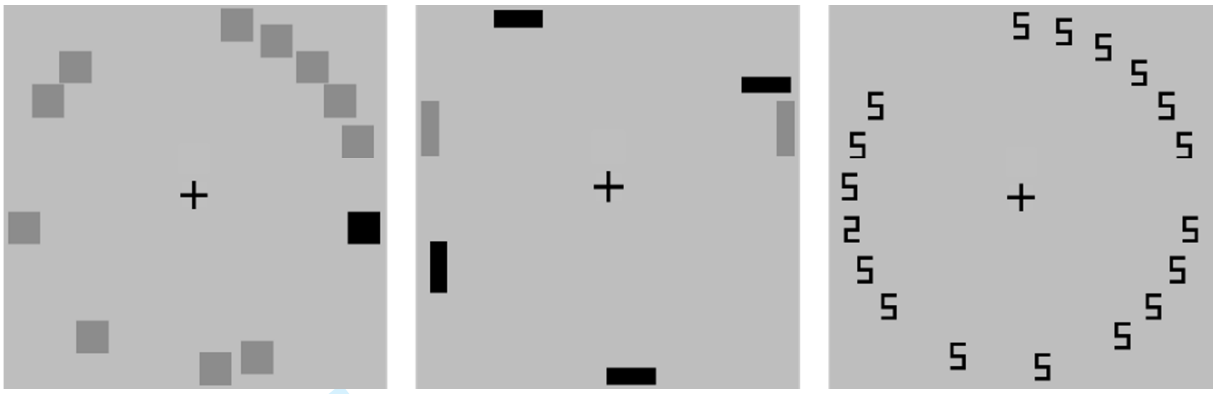


Figure 2. A schematic representation of the tasks; in each there was a target present (black item [feature]; black vertical [color-form conjunction] and the number 2 [spatial configuration]).

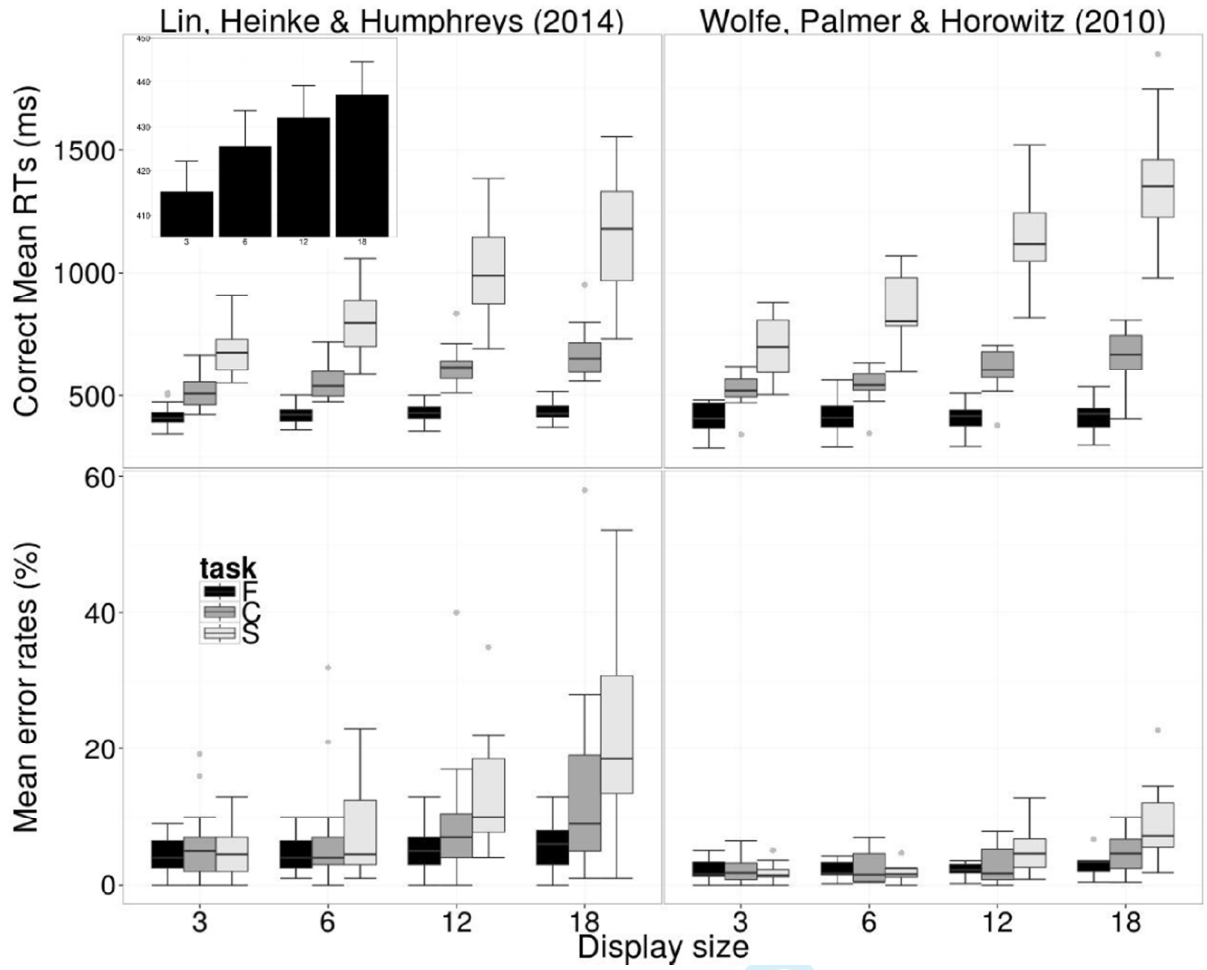


Figure 3. The box-and-whisker plots. The upper and lower panels show the means of RTs and error rates, respectively. The subplot in the upper-left panel shows a zoom-in view of the bar-plot of the feature search task (y axis ranging between 405 to 450 ms; x axis labelling four display sizes). The left and right panels present the analyses from the current and Wolfe et al.'s (2010) data sets, respectively.

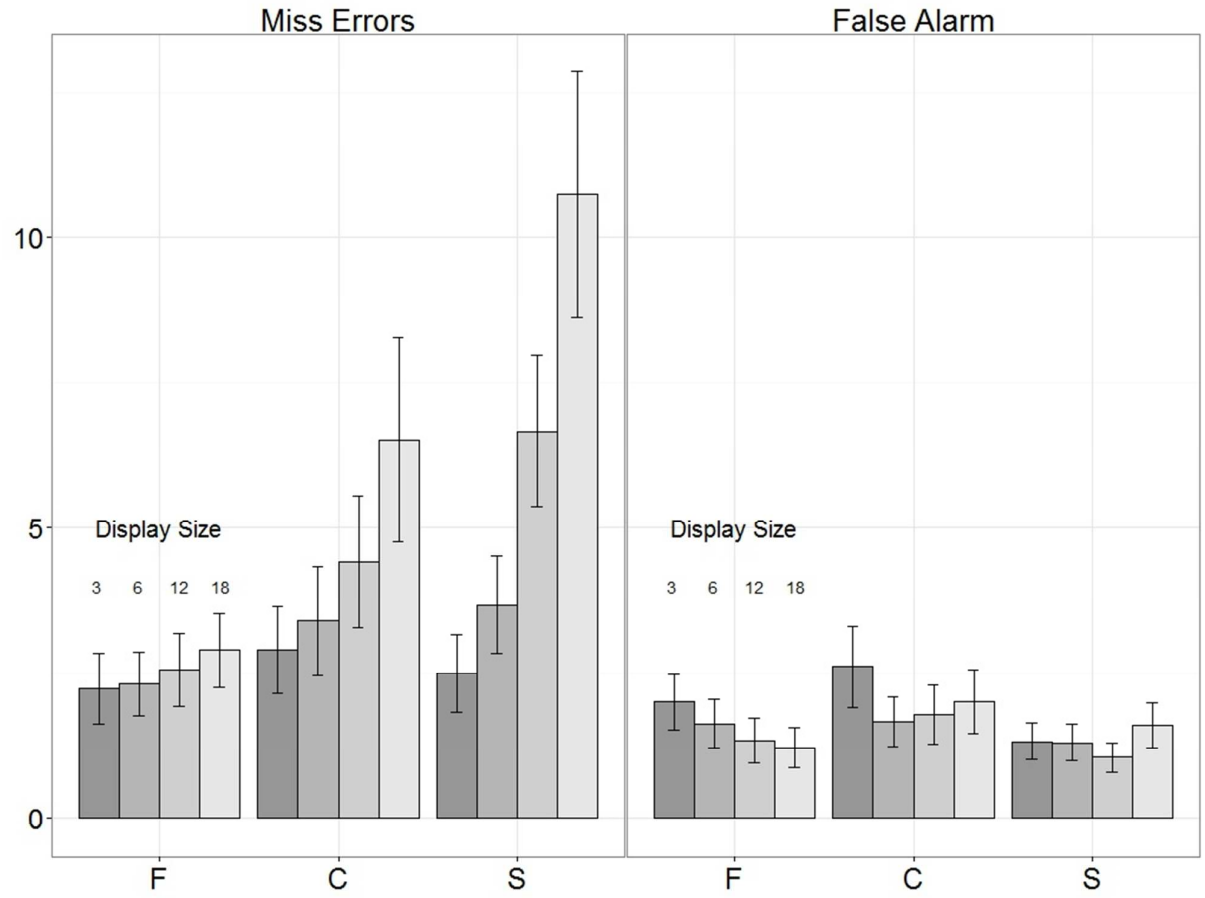


Figure 4. Mean rates of miss and false alarm errors. The error bars show one standard error of the mean. The y axis shows percentage of errors. F, C and S stand for feature, conjunction and spatial configuration searches.

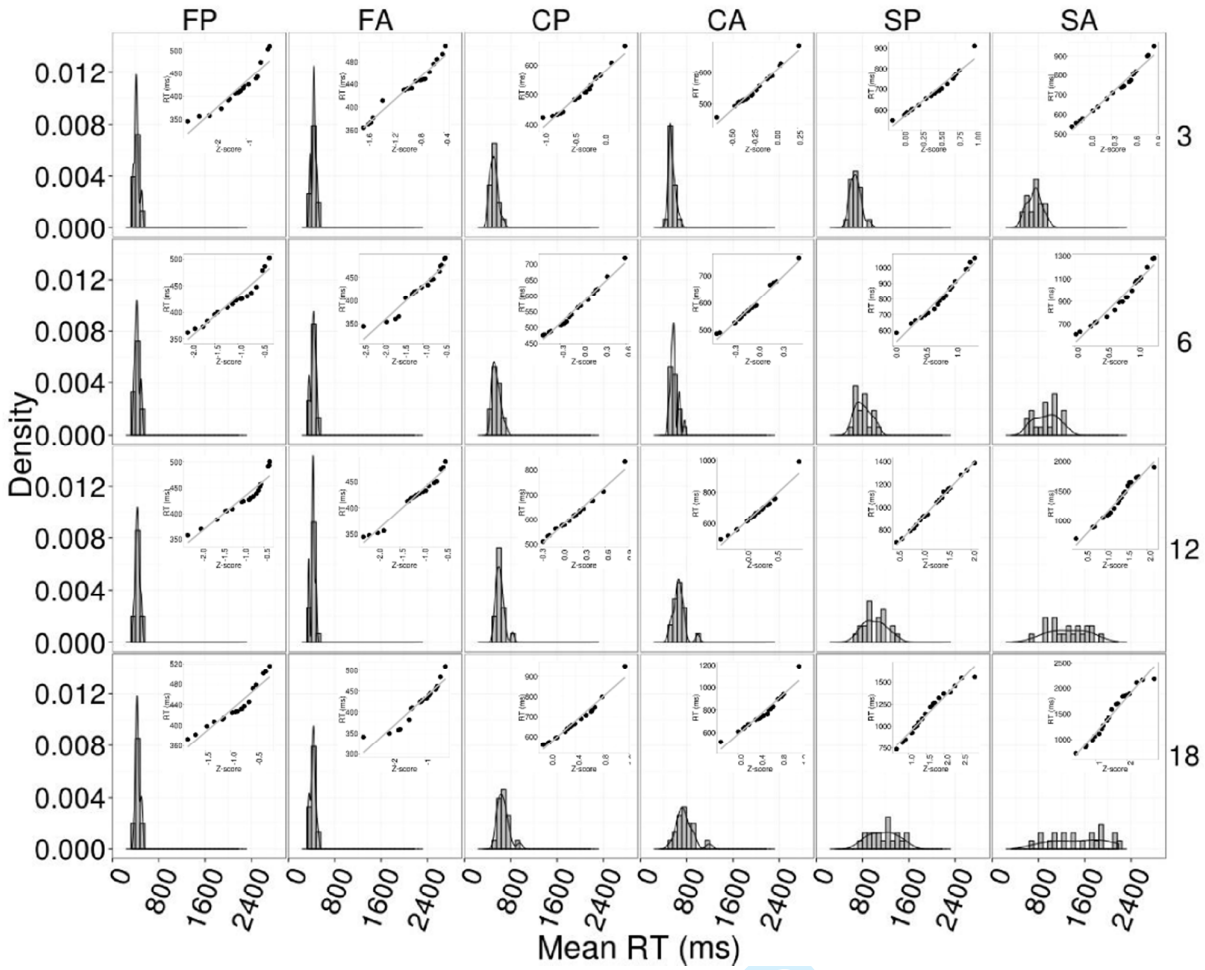


Figure 5. The mean-RT distributions. The sub-plots are quantile-quantile normalised plots, showing the deviations of data from the theoretical normal distribution. The y axis of the Q-Q normalised plots compares RT means [y axis label, RT (ms)] with normalised z score [x axis label, (Z-score)]. F, C, and S stand for feature, conjunction, spatial configuration tasks. P and A are present and absent trials, respectively.

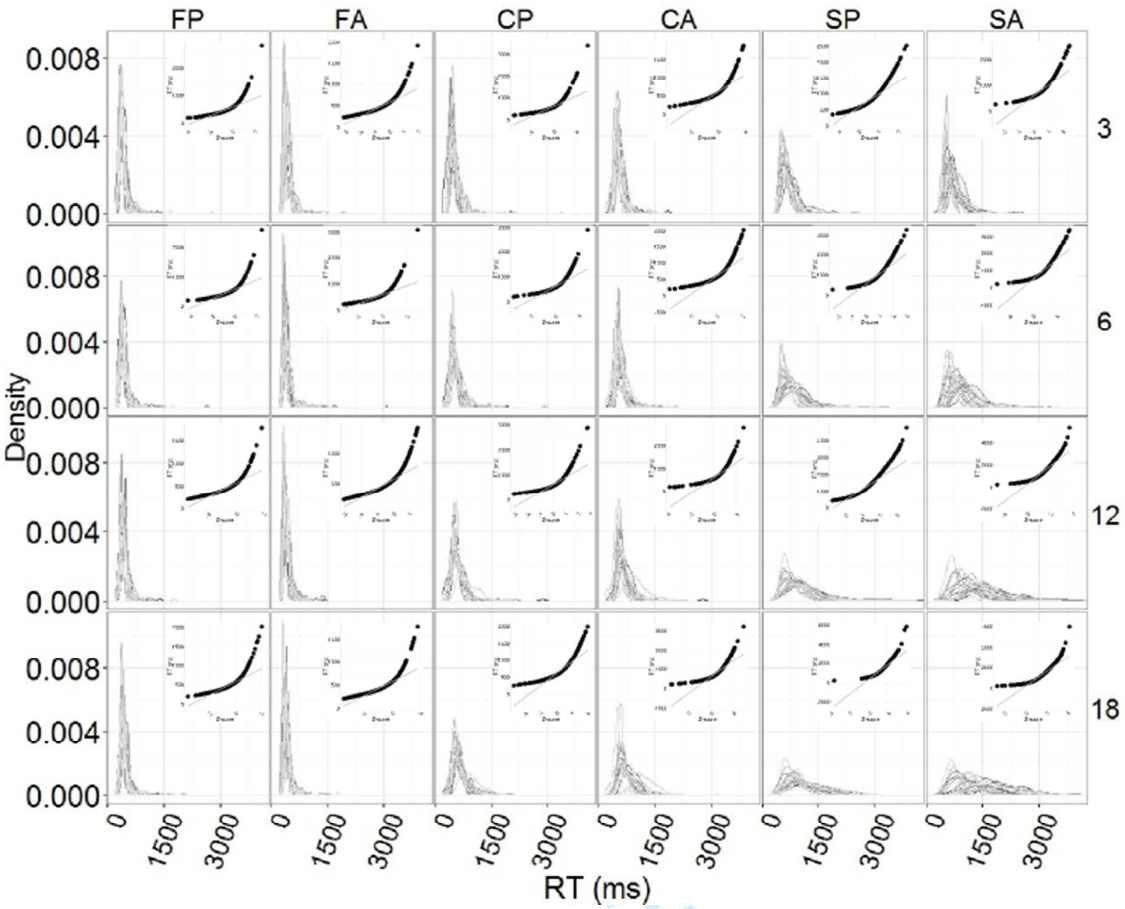


Figure 6. The trial-RT distributions. The y axis of the Q-Q normalised plots compares trial RTs [y axis label, RT (ms)] with normalised z scores [x axis label, (Z-score)]

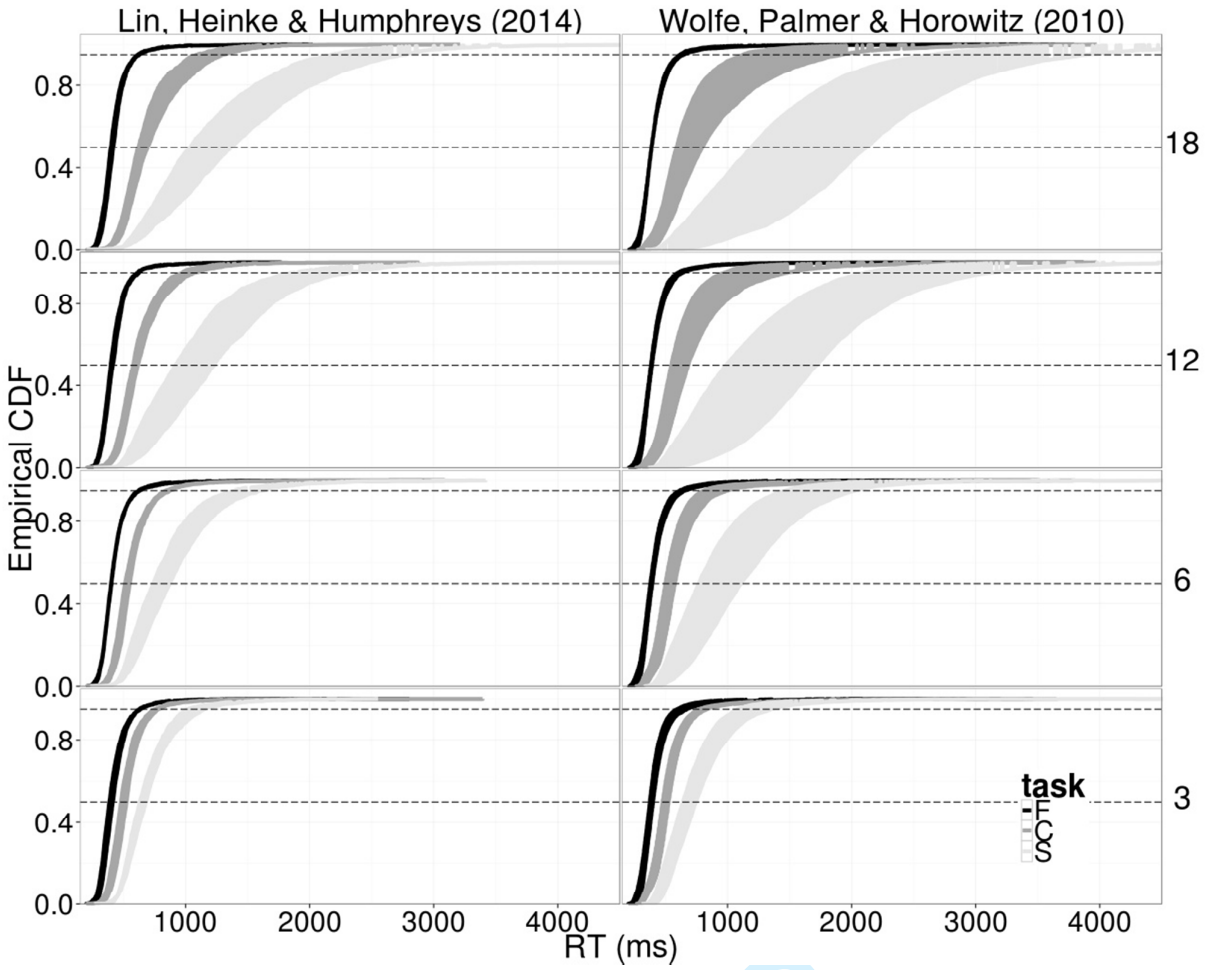


Figure 7. The empirical cumulative RT density curves drawn based on the trial RTs. The areas within each envelope represent the differences between target present and target absent trials for each task. The two dashed lines show the positions of the 50% and 95% cumulative densities. Long latencies (right border of envelopes) were consistently observed on target absent trials.

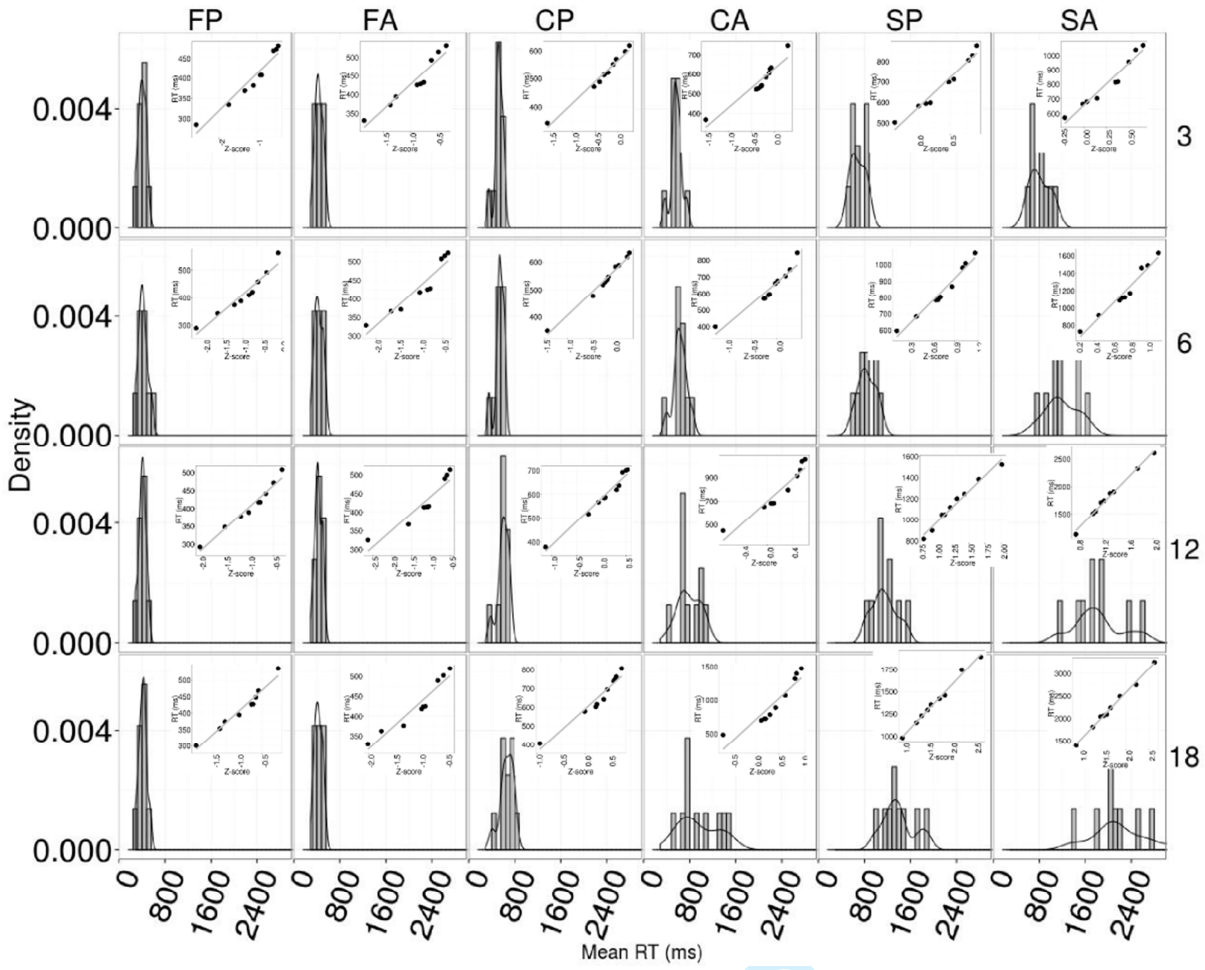


Figure 8. The mean-RT distributions. The y axis of the Q-Q normalised plots compares RT means [y axis label, RT (ms)] with normalised z score [x axis label, (Z-score)]. Data are from Wolfe et al (2010).

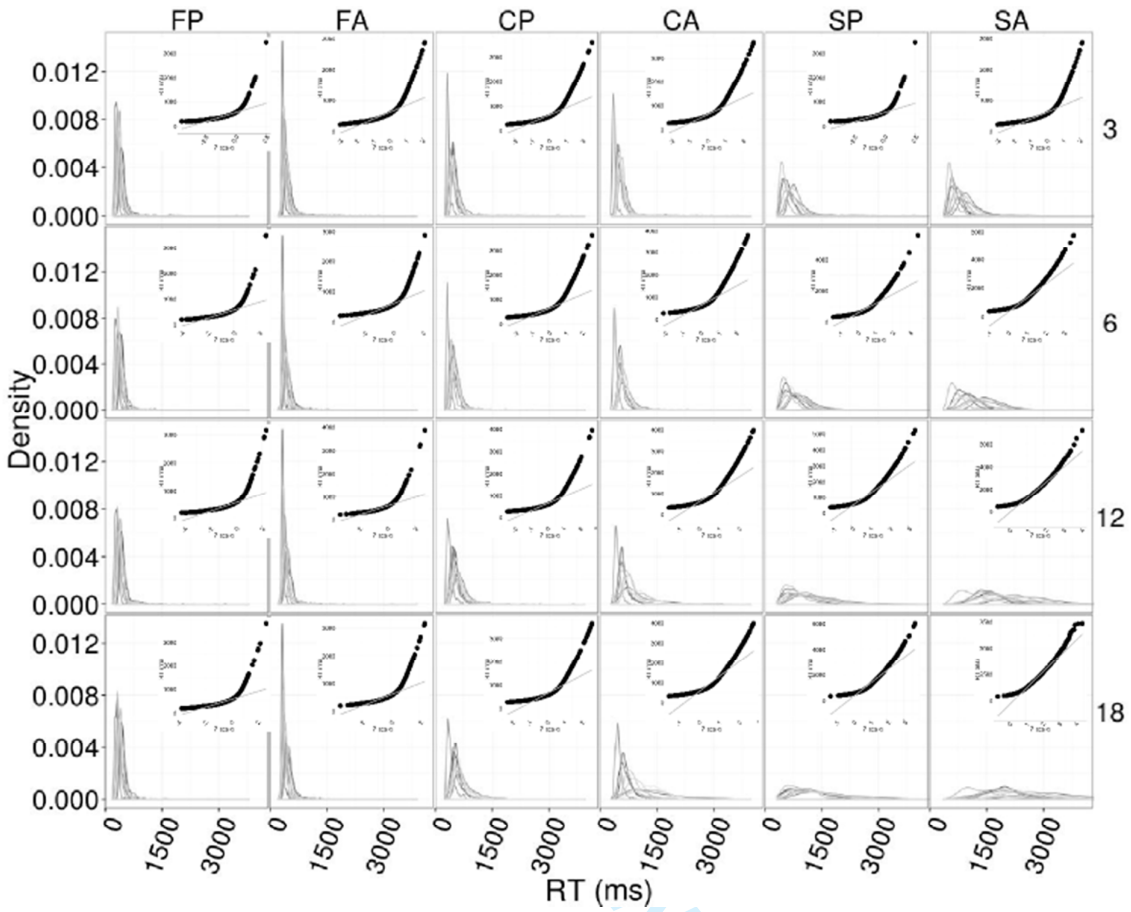


Figure 9. The trial-RT distributions. The y axis of the Q-Q normalised plots compares trial RTs [y axis label, RT (ms)] with normalised z scores [x axis label, (Z-score)]. Data are from Wolfe et al (2010).

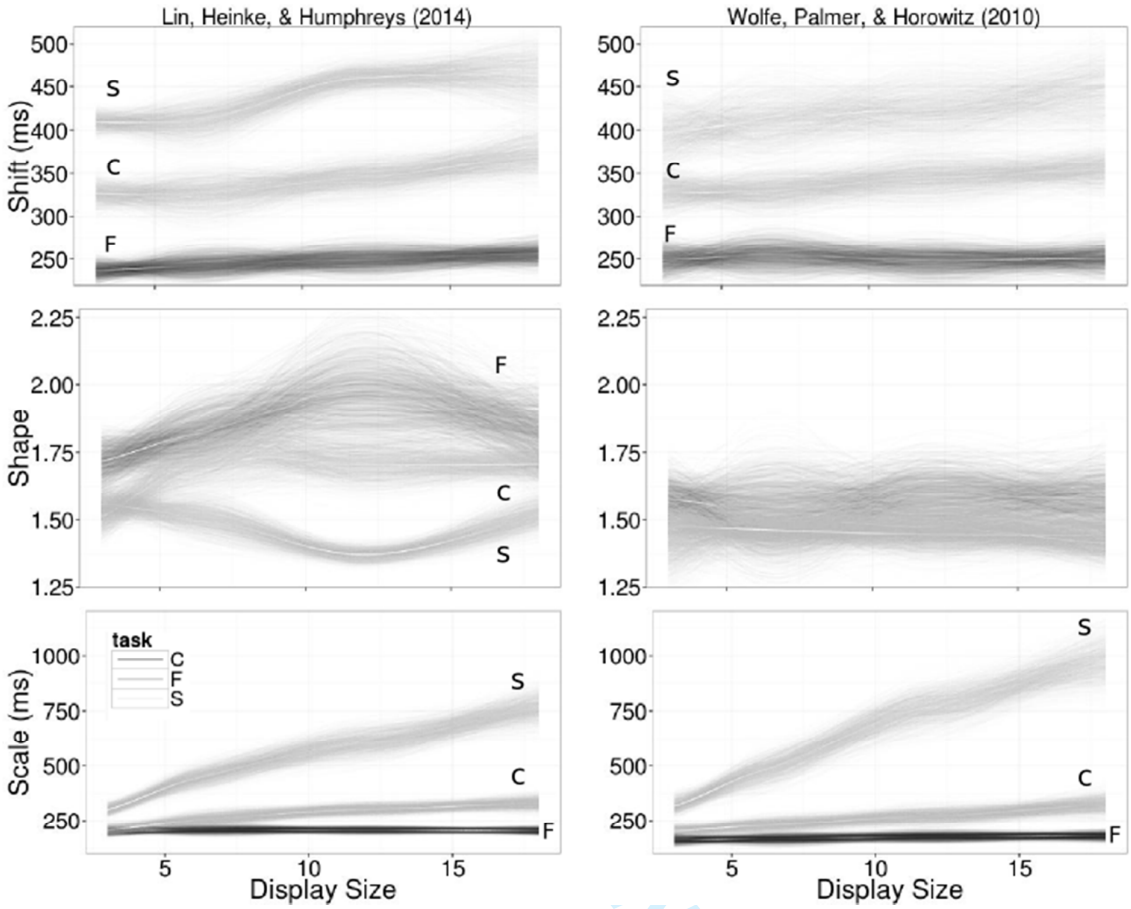


Figure 10. Visually-weighted regression (VWR) plots (Hsiang, 2012) for the three Weibull parameters. VWR performs regressions using display size as the continuous independent variable and Weibull function estimates (shift, shape or scale) as the predicted variables separately for the three search tasks. 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 (i.e., the three grey-scaled lines). Here we used locally weighted smoothing (Cleveland, Grosse, & Shyu, 1992). The density of lines and saturation of gray-scale lines 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.

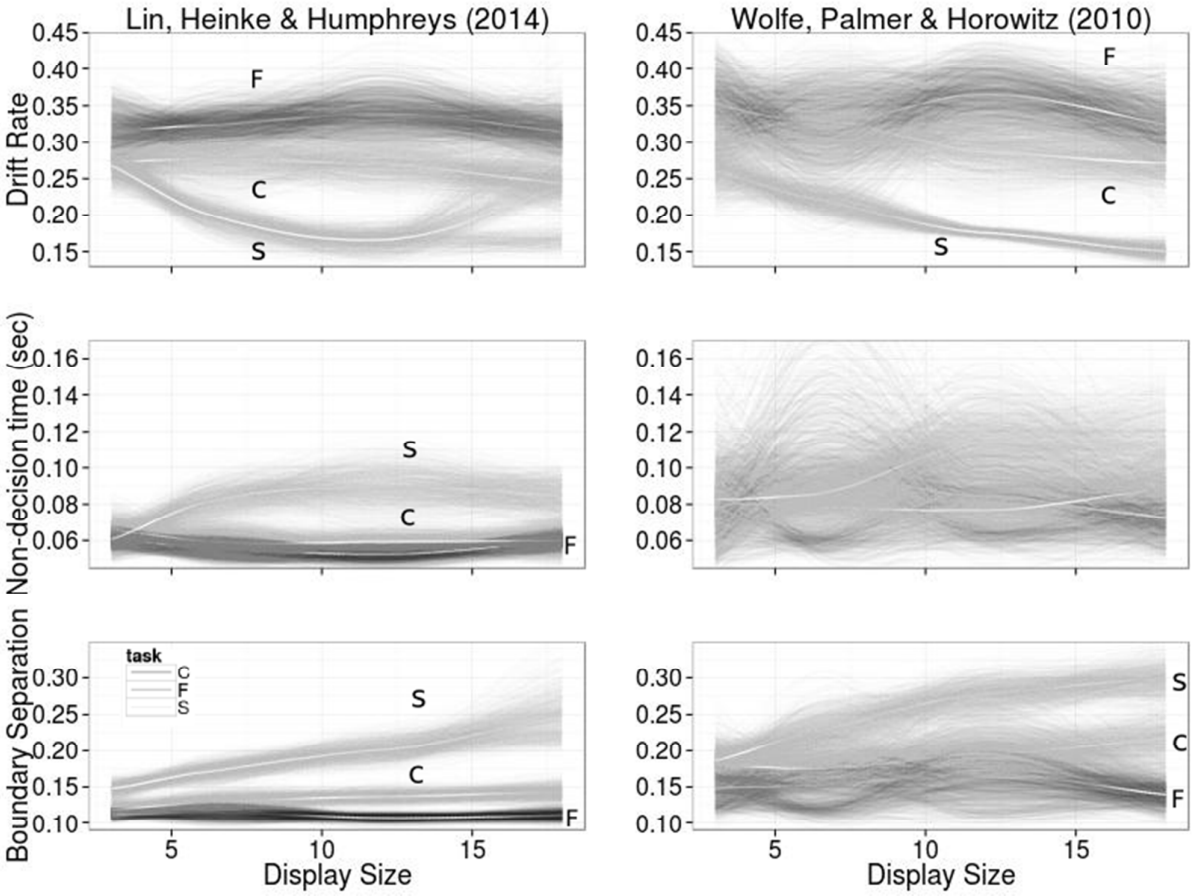


Figure 11. The visually-weighted regression plot for the EZ2 diffusion model parameters.

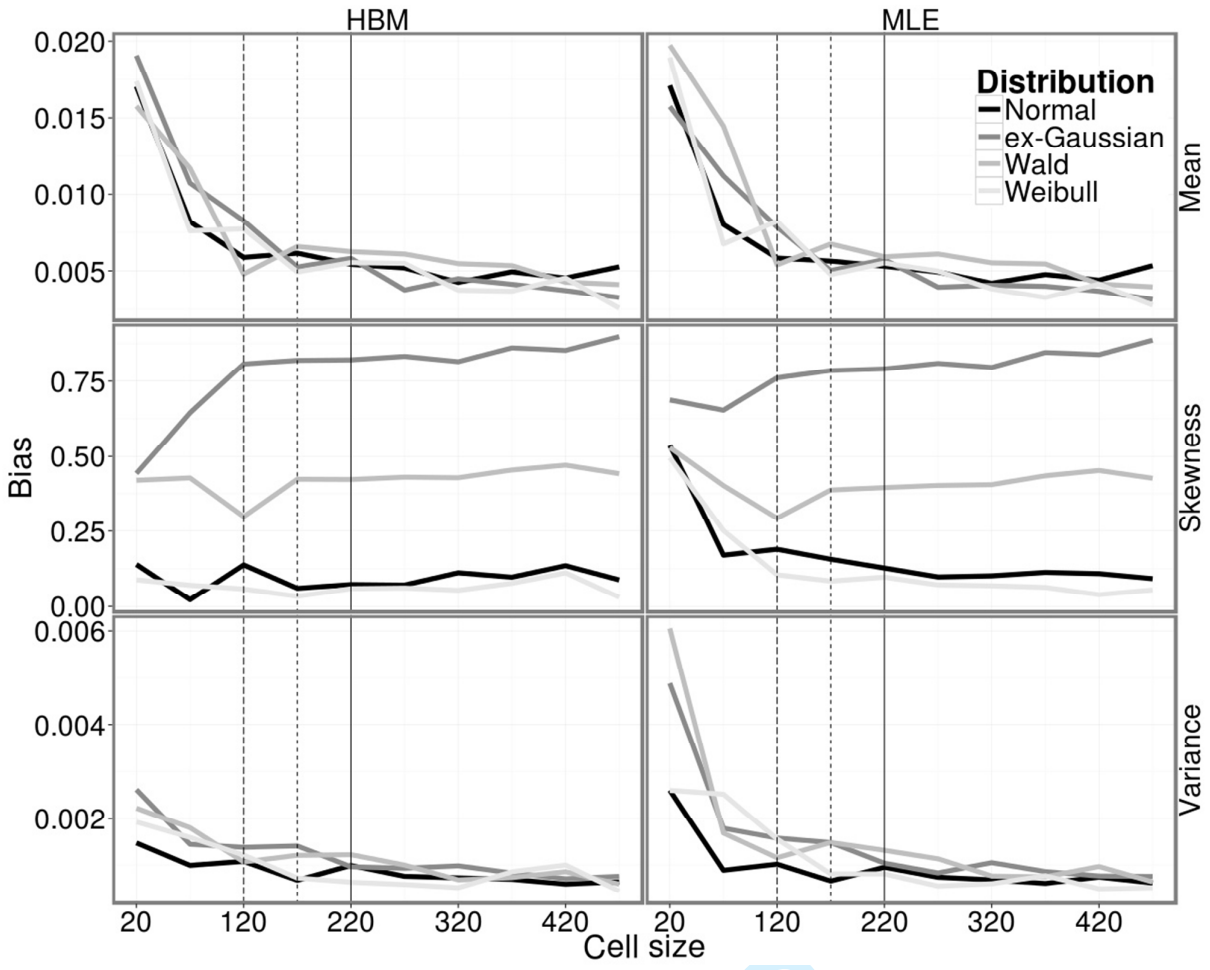


Figure 12. This figure compares the sample mean, variance and skewness to the true values that generated the simulated data. HBM and MLE stand for hierarchical Bayesian model and maximum likelihood estimate, respectively. The mean and variance are on the scale of seconds and square of seconds, respectively. The skewness was calculated by dividing the third moment about the RT mean [$m_3=(RT-RT_{mean})^3/N$] by the cube of the RT standard deviation, which renders it dimensionless (Crawley, 2002). The three dashed lines are drawn at the sample sizes, 120, 170, and 220 to show critical changes of the three parameters with regard to the sample sizes.

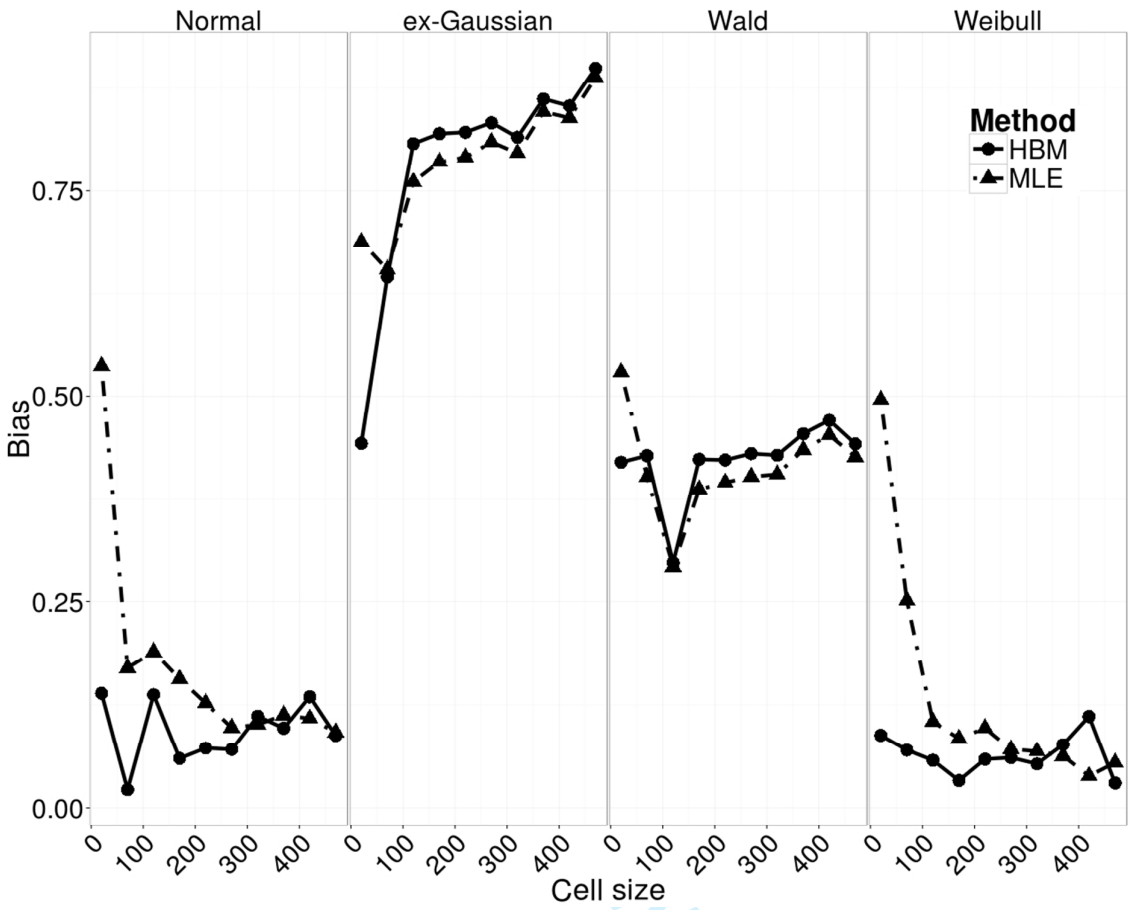


Figure 13. The estimation of skewness. The figure traces the difference between the HBM and MLE along different sample sizes in an experimental condition.

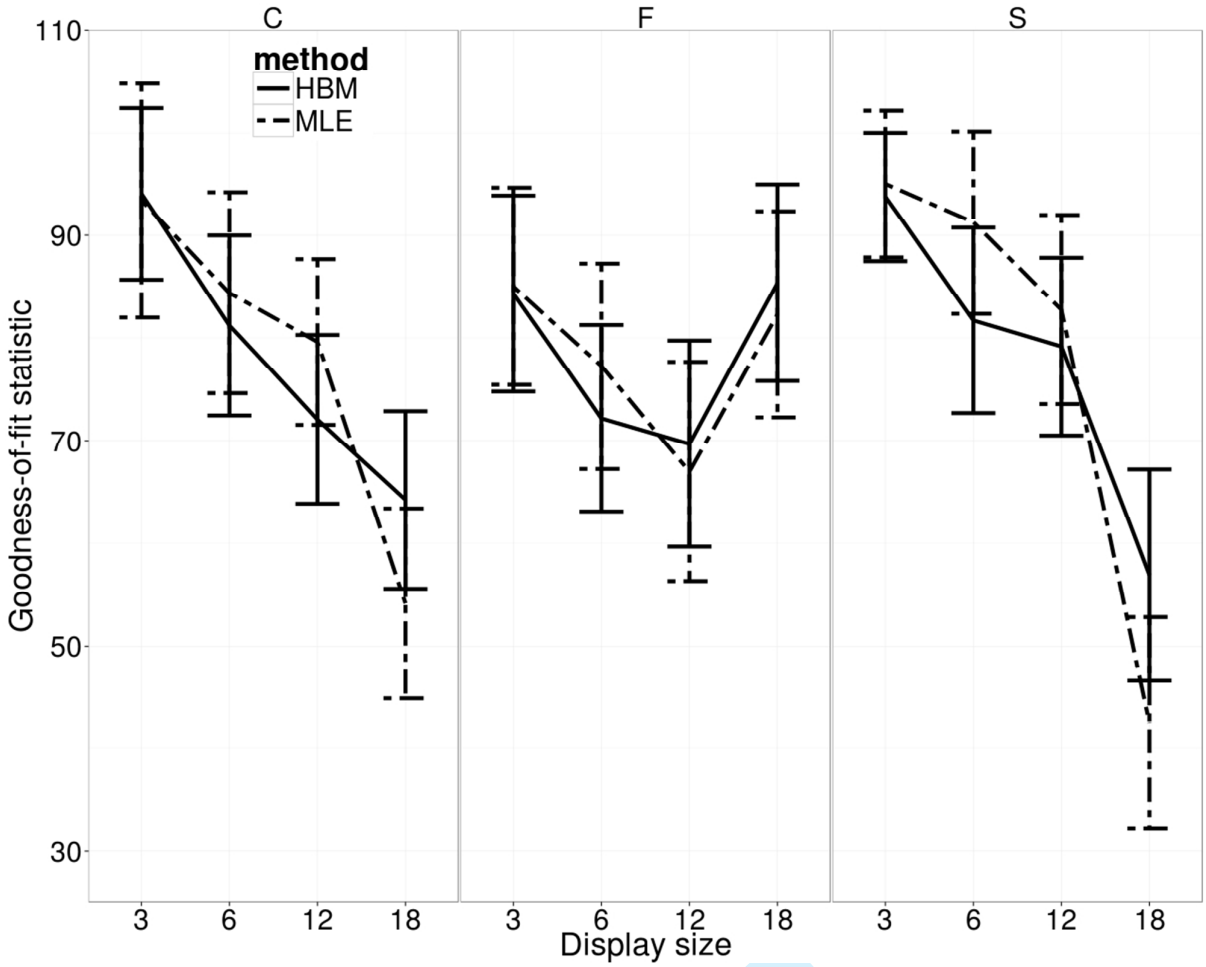


Figure 14. The figure shows the goodness-of-fit of the empirical data against the Weibull distribution, using maximum likelihood (MLE) and hierarchical Bayesian model (HBM) methods likelihood (MLE) and hierarchical Bayesian model (HBM) methods.

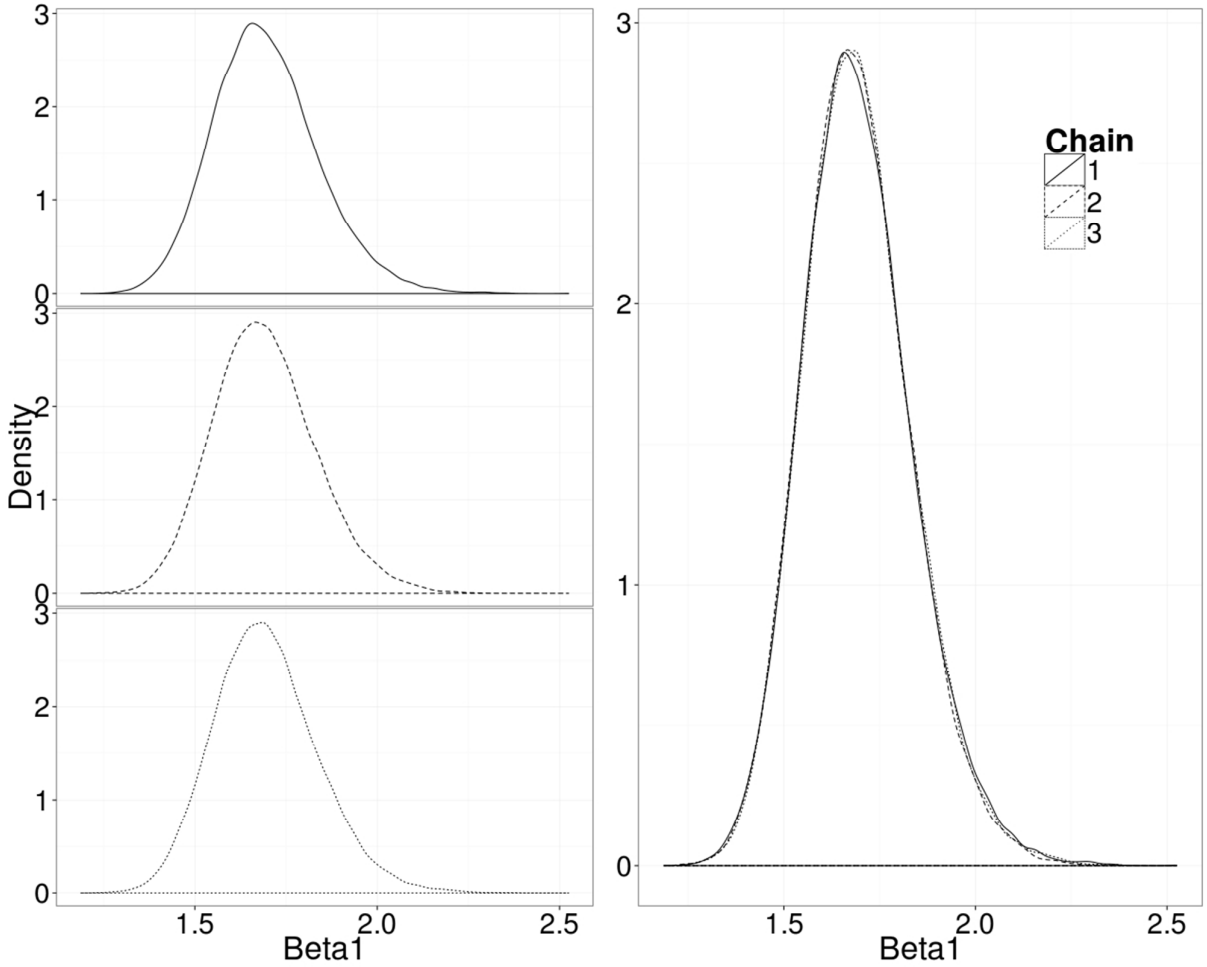


Figure 15. The diagnostic posterior density plots for the Markov Chain Monte Carlo (MCMC) process. The left panel shows the density curves, separately, for the three MCMC processes (i.e., Chain in the legend). The right panel draws them together to demonstrate that three processes reach a consistent estimation of the posterior density distribution, suggesting the MCMC process is reliable. The figure shows only β (shape) parameter from one of the participants in the target present condition of display size 18 in the spatial configuration search.

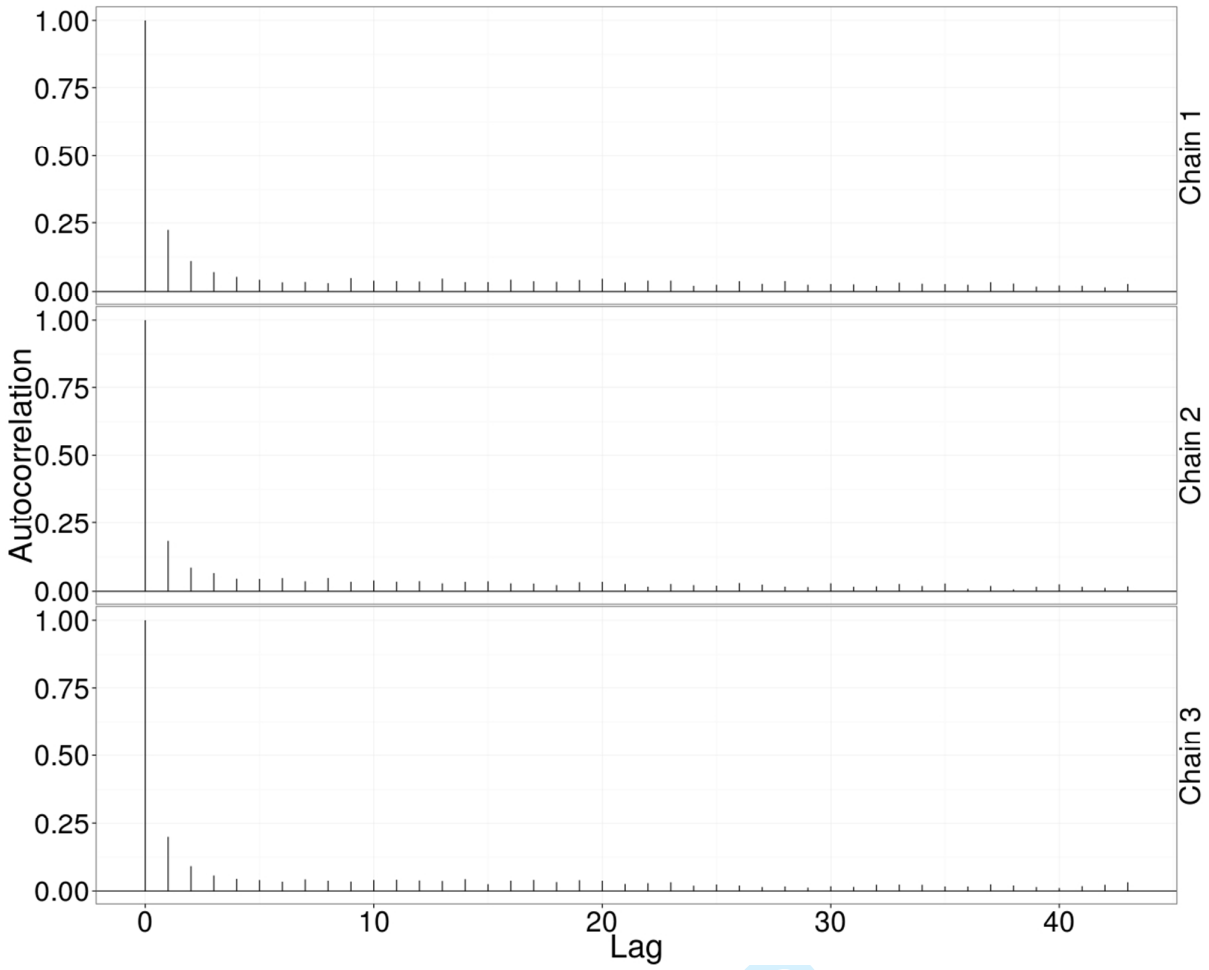


Figure 16. The diagnostic autocorrelation plots for the Markov Chain Monte Carlo (MCMC) process. The figure shows only β (shape) parameter from one of the participants in the target present condition of display size 18 in the spatial configuration search.

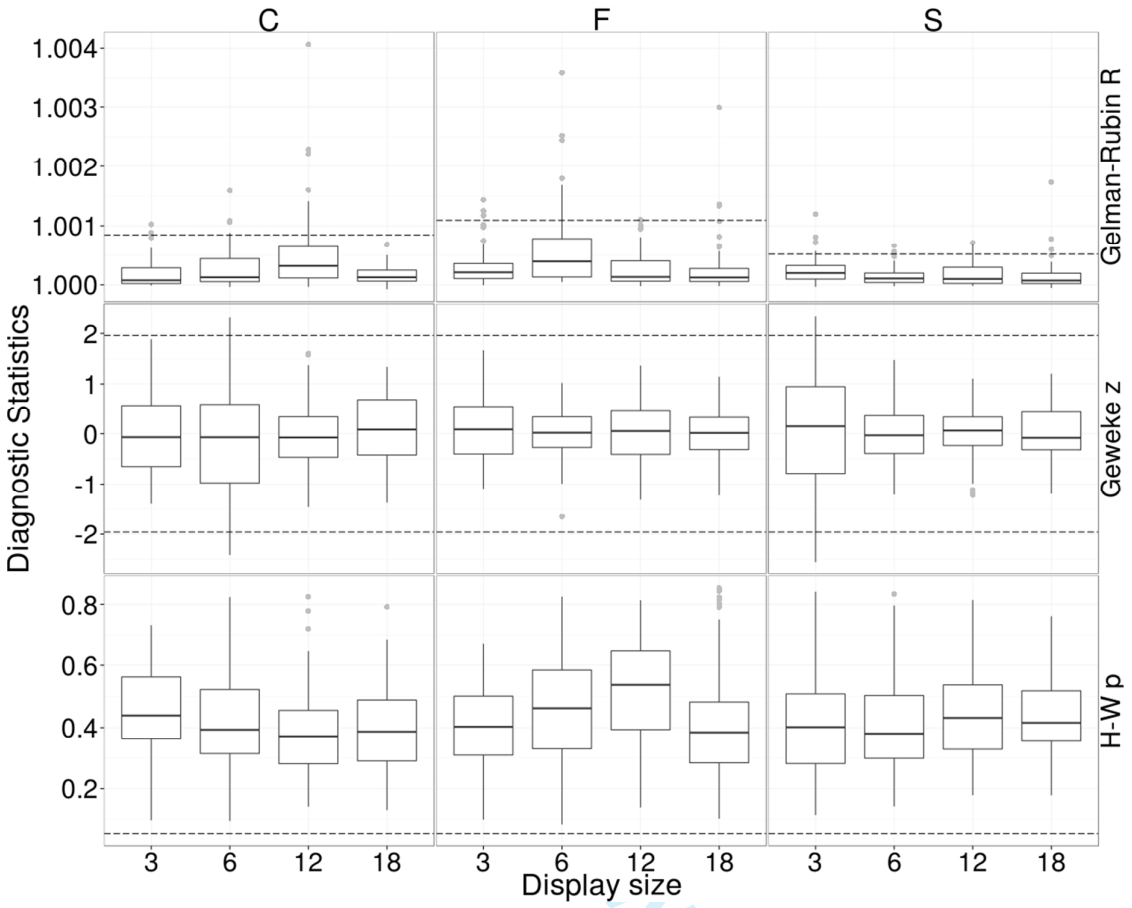


Figure 17. The figure shows three diagnostic tests for assessing MCMC convergence, using box-and-whisker plots to summarise the distribution of the three tested statistics. In each condition (display size × search task), each participant contribute 3 data points from the three Weibull parameters. The figure was drawn as a function of the diagnostic statistics against the display sizes. The three panel rows, from top to bottom, show the Brooks-Gelman-Rubin convergence test, the Geweke Z score, and the Heidelberger-Welch test, respectively. In the upper panel, the reference dashed line was drawn on the upper bound of the 95 % confident interval, separately for each search tasks. The middle panel used two reference lines at -1.96 and 1.96 (the upper and lower borders of 95% confident interval) to show that most of Geweke Z scores fell in the acceptable range. The lower panel shows the p values from the Heidelberger-Welch test. The reference line was drawn on the .05 alpha level.

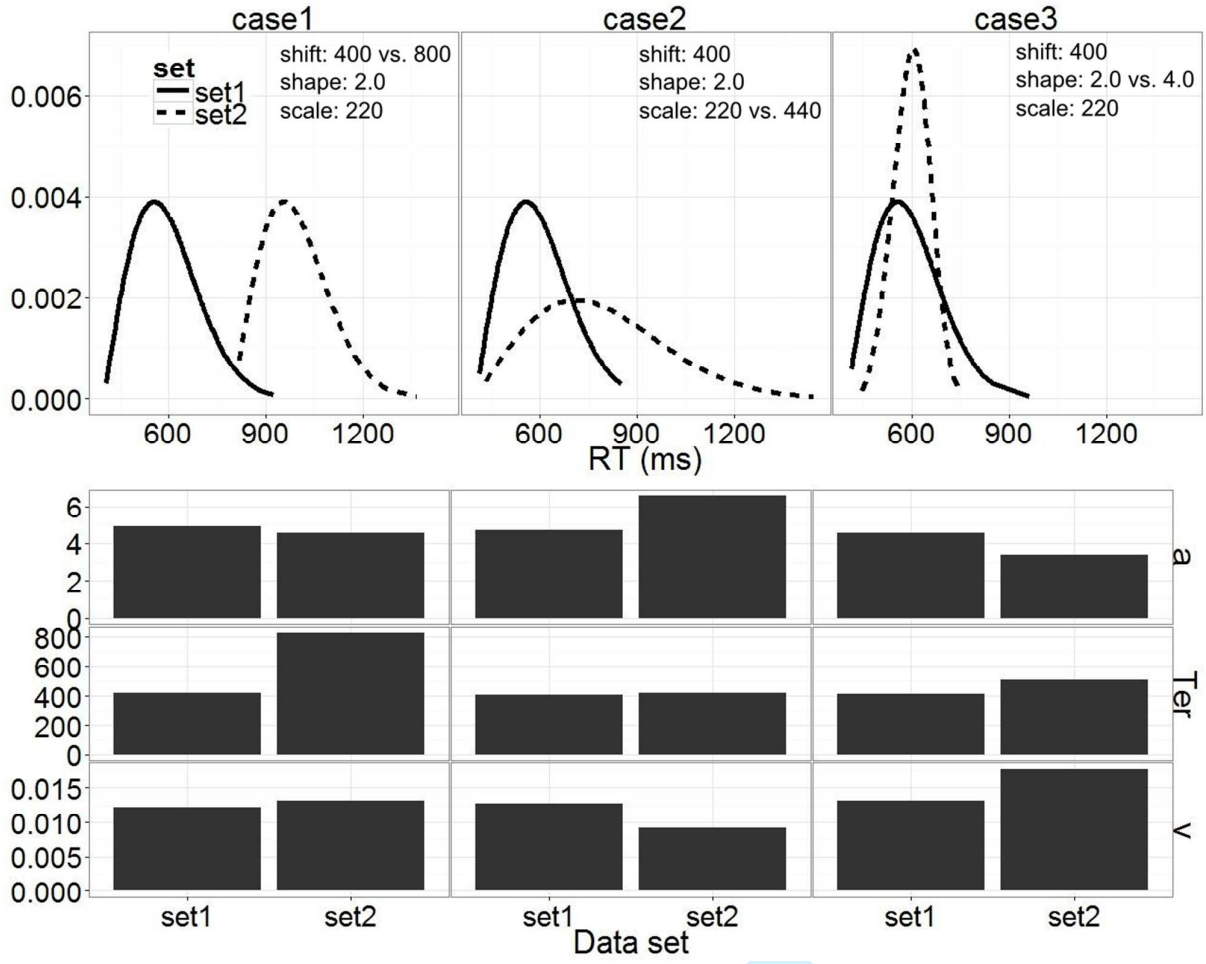


Figure 18. The figure shows how the change in a distributional parameter may alter the parameters estimated by the EZ2 diffusion model. Ter, a and v stand for the non-decision time, the boundary separation and the drift rate, respectively.

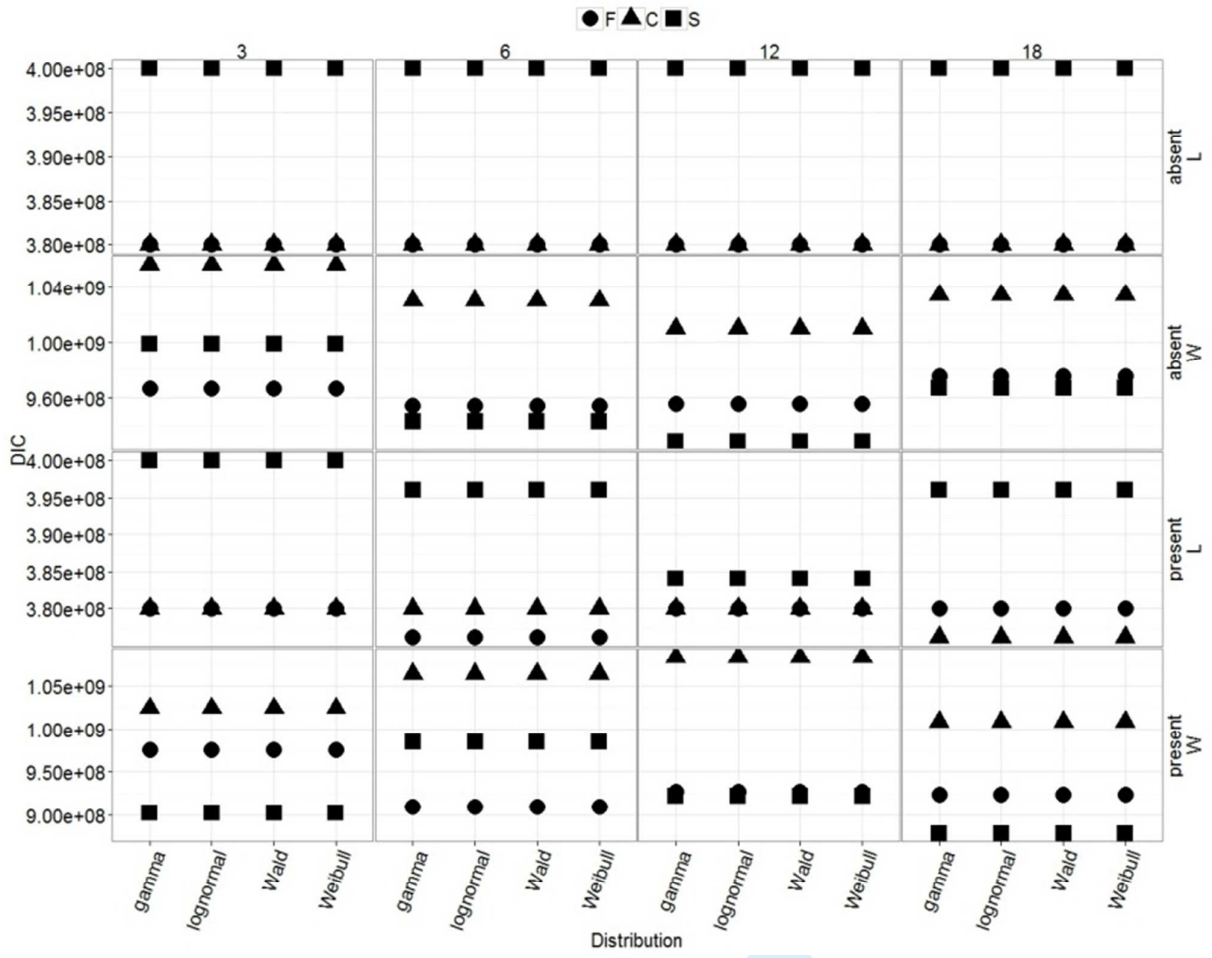


Figure 19. The figure compares the Bayesian DICs for the fitted 3-parameter probability functions across the data sets, search tasks, target types, and display sizes. L and W stands for ours and Wolfe, Palmer, and Horowitz's (2010) data sets.

Table

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

	Lin, Heinke, & Humphreys (2014)	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

Table 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

Response to Reviewers

Reviewer: 1

Comments to the Author

Overall the manuscript is much improved and the authors have responded well. The details of the Bayesian modelling are much clearer. Just a few minor points:

Comment 1

- Figure 14 may not be required. The reporting of diagnostic information of the MCMC process is very thorough, and a visual depiction of the chains is not necessary. Especially given the presence of Figure 15.

Thank you for this suggestion. Figure 14 (original) has been removed from the manuscript. The description in the Appendix A has also been modified to reflect this change.

Comment 2 and Comment 3

- My hunch is that JAGS and BUGS will be used for a long time to come, but that STAN and PyMC are the future. But the software landscape here is changing rapidly. The limitations you mention in terms of certain distributions (and others) are likely to be solved with these packages as they are under much more active development. But as you say, this is for future work.

- For future work, I suggest the authors consider including figures of the graphical model diagrams, as these can greatly increase the accessibly of the modelling.

Thank you for the suggestions.

Reviewer: 2

(There are no comments.)

Reviewer: 3 and Editor's summary

##########

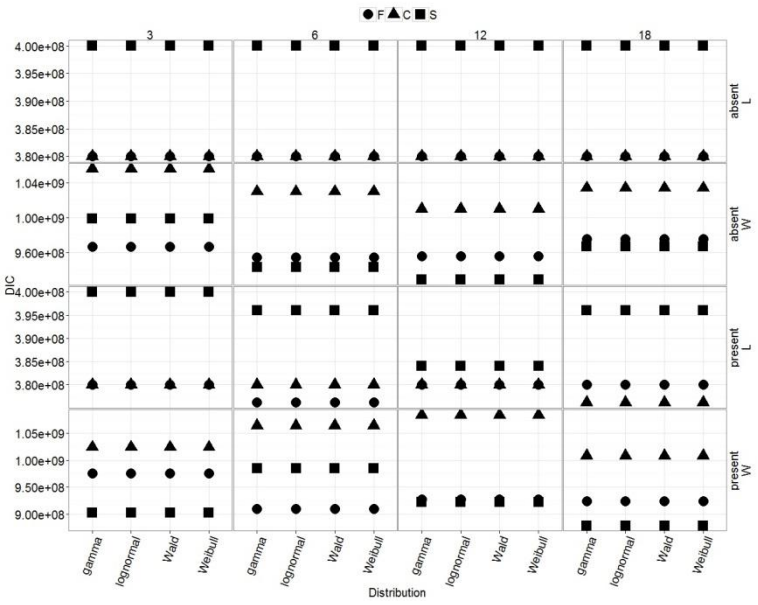
1. Reply to the issue of the justification of the Weibull function as the vehicle of major insights

##########

First reason is that the Weibull function is able to provide a concise way of summarising the shape of an RT distribution. We agree with Palmer's suggestion that mapping the Weibull parameters directly to psychological processes is difficult. Thus, we took this suggestion and followed Schwarz's (2001) work similarly using the Weibull function as a descriptive model of response time.

Secondly, we found that 3-parameter gamma function does not converge in HBM. We followed Palmer's suggestion in his previous comments to fit a gamma function. The compatible 3-parameter gamma function shows signs of non-convergence and perhaps because of this, it provides a slightly worse fit to the data (DIC).

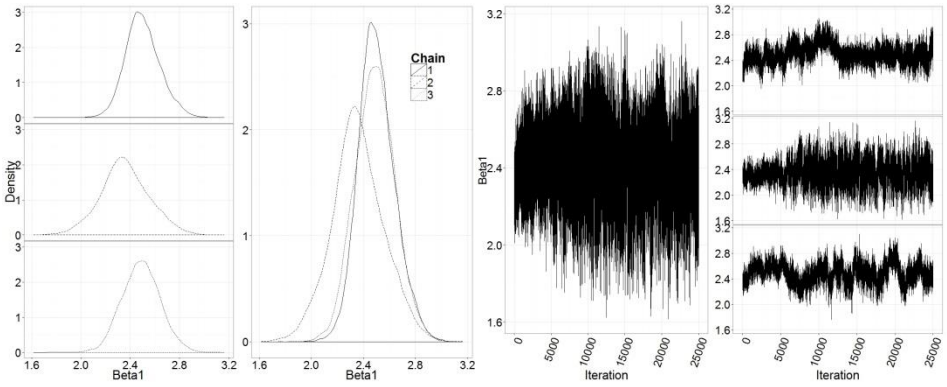
Thirdly, we fitted also the other two 3-parameter functions, Wald and log normal. They also provide descriptive parameters similar to the Weibull function. The DICs suggest that they fit both ours and Wolfe, Palmer and Horowitz's (2010) data better than the Weibull function. However, we still prefer the Weibull function, partly because we have found it to be highly robust and because all four function fits are very similar when examined separately for each task, display size, and target types (see below figure). We have fitted a Gaussian function to test whether our impression that all four 3-parameter functions provide better matches than other approaches. The Gaussian fit DICs are ~ -3150, which are far different from the four plausible functions that can accommodate positive-skewed RT. Below we reported the detailed procedure to fit gamma functions.



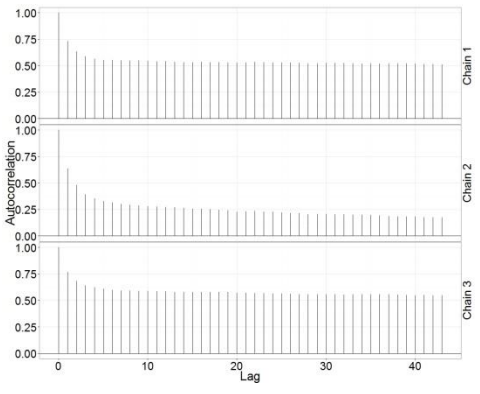
To test whether the Weibull function fit better than the gamma function, we built a 3-parameter gamma function in the HBM framework. Because currently BUGS does not implement a pre-built 3-parameter gamma function, we used Johnson, Kotz, and Balakrishnan's (1994, pp 337, eq. 17.1) equation to implement the gamma function. Practically, the BUGS code is to change the density function to:

```
#--------------------------#
# 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 we implemented the Weibull function, we assessed the parameters via minus log-likelihood and used the pseudo-Poisson (zero) trick. This implementation resulted in unstable, non-converged estimations. Take the shape parameter as an example. In the spatial configuration search participant 3, the shape (beta) estimation reported three different posterior distributions and the trace plots from the three chains oscillated unstably around different ranges.



In addition, the autocorrelation plots indicated a problem and this does not abate with increasing iterations.



In summary, the diagnostics indicate that the gamma function does not converge when fitted with a HBM framework. This failure could be because (1) the gamma function does not suit for HBM in this context, and/or (2) the gamma function indeed fits worse than Weibull function (as the DIC suggests). Note that we have run a high number of iterations (i.e., 105000) and reasonable thinning length (every 4 iterations) and this still cannot resolve the non-converged gamma fit.

As in the paper we report all four HBM fits, the Weibull model is no longer emphasised in the title (and elsewhere) – though, for the reasons we discussed above, it is still a major part of the paper (e.g., providing a useful and theoretical interesting way to describe RT distributions in visual search).

- Add text to this modification (In introduction section, after the sub-section of "The 3-parameter probability functions")

The 3-parameter probability functions

Our study adopted four three-parameter probability functions – lognormal, Wald, Weibull and gamma[1] (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 characterize the pattern of a distribution. An increase of scale parameters shortens the central location within a distribution and thickens its tail. This implies that the responses originally accumulated around the central part become slower and thus been moved to the tail side. An increase in the shape parameter makes the tail thinner, because those originally slow responses are moved from the tail to the central location. 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 1 for an illustration).

Figure 1 should be inserted around here

The study assumed that changes in RT distributions reflect unobservable cognitive processes (a similar argument also made by Heathcote et al., 1991). As illustrated in Figure 1, the factors that affect quick, moderate and slow responses evenly will show a selective effect on the shift parameter. The effect on the scale parameter will be from the factors that alter only the proportion of responses that are moved from the central location to the tail part of a distribution (or vice versa). Lastly, the effect on the shape may result from the factors that affect both the central and tail parts of a distribution and effectively increase the response density between them.

[1] 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, we constructed the comparable 3-parameter HBM to test if other functions gain a substantial better fit using hierarchical Bayesian approach than the Weibull function. We thank Evan Palmer for this suggestion.

- Modify the sentence in page 9 from "…in line with…" to "…is similar to…"

This is similar to some reports applying distributional analyses on RT data, attempting to link distributional parameters with psychological processes directly (e.g., Gu et al., 2013; Rohrer & Wixted, 1994).

- Add texts in sub-section, "Hierarchical Bayesian Model (HBM)" of the Method section (page 15)

… Likewise, we replaced the Weibull function with the 3-parameter gamma, lognormal, and Wald functions (Johnson, Kotz, & Balakrishnan, 1994), keeping similar prior parameter setting.

- Add texts in the sub-section of "The HBM estimates" in the Result section (page 23)

… We used deviance information criterion (DIC) to evaluate the function fit to the data. In general, the smaller the DIC, the better the fit (Lunn, Jackson, Best, Thomas, & Spiegelhalter, 2013). Although the lognormal and Wald functions showed the smallest DIC, the DICs across the four fitted functions were close. Moreover, the diagnostic of the gamma HBM suggests its posterior distributions did not converge. Excluding the non-converged gamma function, we 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).

- Add texts in the sub-section of "Limitation" in General discussion section

… other hand, we selected the Weibull function, because it permits a reliable converged posterior distribution, has broad application to memory … (Rouder et al., 2005).

- Add Appendix C

Appendix C

Why use Weibull function?

The Weibull probability function is one of the many plausible functions that can accommodate positively skewed RT distributions. We chose it to describe the RT distributions, because of its parametric characteristics, permitting us to describe the shape of an RT distribution in an intuitive way. Nevertheless, there are other alternatives, such as gamma, log-normal, and Wald functions. All of them are capable of accommodating skewed RT distributions and provide similar descriptive parameters. Here, we delineate our reasons to fit the HBM Weibull function to RT data.

First reason is that the Weibull function is able to provide a concise way to summarize the shape of an RT distribution. As described in the main text and illustrated in **Error! Reference source not found.**, changes in the three parameters, shift, scale and shape, are associated differently with increases/decreases of RT densities, allowing us to understand how an experimental factor may alter different areas of an RT distribution. Secondly, the 3-parameter gamma function does not converge when fitted with hierarchical Bayesian approach. The compatible 3-parameter gamma function shows signs of non-convergence and perhaps because of

this, it fits the data slightly worse, indicated by the DICs. Third, we fitted also two other two 3-parameter functions: Wald and log normal. They provide the same set of descriptive parameters as the Weibull function. The DICs suggest that these functions fit both ours and Wolfe, Palmer and Horowitz's (2010) data slightly better than the Weibull function. However, we have maintained the Weibull function because all four function fits are very similar for each task, display size, target types, and data set (Figure). To test whether any function would work we fitted a Gaussian function. The Gaussian fit DICs (~ -3150) are far worse than the four plausible functions that can accommodate positive-skewed RT. Below we reported the detailed procedure used to fit gamma functions.

Figure 19 should be inserted around here

To test whether the Weibull function fit better than the gamma function, we built a 3-parameter gamma function in the HBM framework. Because currently BUGS does not implement a pre-built 3-parameter gamma function, we used Johnson, Kotz, and Balakrishnan's (1994, pp 337, eq. 17.1) equation to implement the gamma function. Practically, the BUGS code is to change the density function to:

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

# 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 we implemented the Weibull function, we assessed the parameters via a minus log-likelihood and used the pseudo-Poisson (zero) trick. This implementation resulted in unstable, non-converged estimations. Take the shape parameter as an example. In the spatial configuration search for participant 3, the shape (beta) estimation yielded three different posterior distributions and the trace plots from the three chains unstably oscillated around different ranges. In addition, the autocorrelation plots indicated a problem and this did not abate with increasing iterations. In summary, the diagnostics show that the gamma function does not converge when fitted with a HBM framework.

This failure of gamma fit could be because (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). Note that we have run a high number of iteration (i.e., 105000) and reasonable thinning length and this still cannot resolve the non-converged gamma fit.

- Add Table 2

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

	Lin, Heinke, & Humphreys (2014)	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

- Add Figure 19

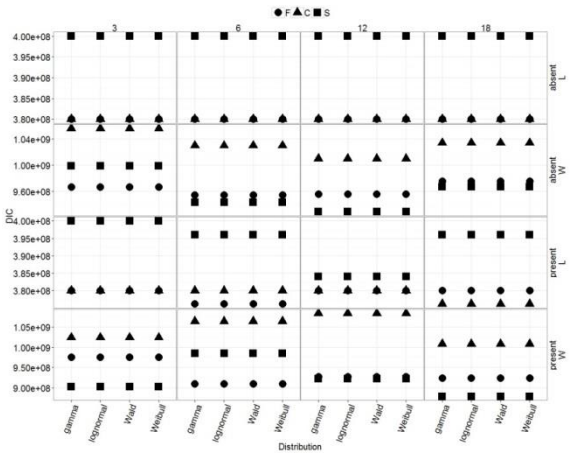


Figure 19. The figure compares the Bayesian DICs for the fitted 3-parameter probability functions across the data sets, search tasks, target types, and display sizes. L and W stands for ours and Wolfe, Palmer, and Horowitz's data sets.

##########

2. **Answer to the issues related to the splits of the drift parameter in the visually weighted regression plot of the diffusion parameter**

##########

We add interpretation, specifically mentioning the split prediction in the visually weighted regression. The following highlights the new interpretation extracted from the General Discussion section

---

This weak effect with the shift parameter is further accounted for by our visually weighted plot at the drift rate parameter, showing a clear split of trends and an increase of between-observer variation at the large display size. Specifically, a subset of participants adopted a strategy similar as those participants in Wolfe and colleagues' study. They did not assemble similar distractors as a search unit, so the predicted drift rate decreases at the large display sizes, whereas the other subset of participants benefited from the crowded homogeneous distractors and thus increased drift rate at the large display sizes.

⋮

Another possible factor that may explain the different findings for the shape parameter is illustrated by the drift rate visually weighted plot. The visually weighted regression lines predict two groups of participants accumulating sensory evidence at different rates, but indicate only one homogeneous group in Wolfe et al.'s data. As our simulated study shows (Appendix B), the drift rate can also weigh-in to change the shape parameter. This suggests that some of our participants took advantage of the crowded layout to facilitate search using the spatial configuration search, and some did not. This did not happen in Wolfe et al.'s

sparse layout, and likely contributed also to our significant finding for the shape parameter on present trials.

##

3. **Answer to the issue related to how the EZ2 model complements the result of Weibull function fit**

##

(1) The diffusion model's assumed three psychological processes, the evidence accumulator, the decision boundary, and the non-decision process, are operated at the stage of stimulus comparison in a search trial. We used the EZ2 model to estimate the means across trials of the three diffusion parameters in each condition. The Weibull HBM on the other hand summarizes the shapes of RT distributions in each condition. The RT distributions summarized by the Weibull HBM represent the outputs from the diffusion processes. Therefore, the complementary of the two modeling approach is because, on the one hand, we assume one search response is driven by the diffusion process, and on the other, all the responses in one experimental condition aggregate to form an RT distribution, described by the Weibull parameter. Even though the Weibull model takes only correct trials into account, the EZ2 estimation will still make sense, because the benchmark paradigms produce very high accuracy data.

(2) To make this interpretation more concrete, we ran 3 Monte Carlo simulations, using the 3-parameter Weibull function to simulate 6 20-participant datasets. Each participant contributed 200 trials. We then used EZ2 diffusion model to fit the simulated data. We show in Appendix B that, when doubling one Weibull parameter and keeping the others constant, how the EZ2 model fits the data and how the three diffusion parameters change. The General Discussion used the simulation result to interpret how the EZ2 model complements the Weibull HBM.

- Add texts in the Introduction section, page 11

The diffusion model was used to complement the distributional analysis. The three diffusion processes – the evidence accumulator, the boundary separation, and the non-decision process – are operated at the stage of stimulus comparison in a search trial. We used the EZ2 model to estimate the means across trials of the diffusion parameters in each condition. The Weibull HBM on the other hand summarizes the shapes of RT distributions in each condition. The RT distributions thus are the aggregated outputs from the diffusion processes. The dual-modeling approach, on the one hand, assumes one search response is driven by the diffusion process, and, on the other, all the responses in one experimental condition aggregate to form an RT distribution, describing by the Weibull parameters. Even though the Weibull model takes only correct trials into account, the EZ2 estimations were still able to account for the descriptive model, because the benchmark paradigms produced high accuracy responses.

- Revise texts and interpretation in the subsection, "Model-based analysis" in the General discussion section

Model-based analysis

	Previous version	Current version
	First, the shift parameter varied across the search tasks but was less affected by the display size, a pattern that was also observed in the drift rate.	First, the shift parameter varied across the search tasks and display sizes, a pattern that is in line with our illustration and the ideal analysis (see Figure 1 & Appendix B). This parameter reflects the psychological processes influencing evenly all ranges of RTs. One of the diffusion processes likely to influence the shift changes is the drift rate, which showed only the main effect of the task.
	As the drift rate is to model the rate of information accumulation determined by the *goodness-of-match* between templates and search stimuli, the shift parameter appears to reflect the quality of the memory match.	As the drift rate aims to model the rate of information accumulation determined by the *goodness-of-match* between templates and search stimuli, the shift parameter appears to result from the change in the quality of the memory match.
New	This is a plausible account, because the shift parameter captures the quick responses in a distribution. In a multi-item search array, it shows a correct decision made with a good template-stimulus match in the first few attempts to locate the target.	This is a plausible account, because the three search tasks demand contrasting matching processes, from (i) feature search requiring only pre-attentive parallel processing to extract just one simple salient feature, to (ii) conjunction search, where binding two simple features must be bound to facilitate a good match, and to (iii) spatial configuration search, demanding both features binding and coding of the configuration of the features. The lack of interaction with display size further supports our argument that the shift reflects the factors affects the RT distribution equally. The weak display size effect can be readily explained by the crowded layout we used; it was not observed [$F(3,75) = 0.016$, $p = 0.997$] in Wolfe et al.'s data (2010). This weak effect at the shift parameter is further accounted for by our visually weighted plot in the drift rate parameter, showing a clear split of trends and an increase of between-observer variation at the large display size. Specifically, a subset of

		participants adopted a strategy similar as those participants in Wolfe and colleagues' study. They did not assemble a single search unit, so the predicted drift rate decreases at the large display sizes, whereas the other subsets of participants benefited from the crowded homogeneous distractors and thus increased drift rate at the large display sizes.
same	Another account for the strong task effect, but the weak effect of display size, is that it reflects a process such as the recursive rejection of distractors proposed by Humphreys and Müller (1993) in their SERR model of visual search (see also Heinke & Humphreys, 2005).	Another account for the strong task effect, but the weak effect of display size, is that it reflects a process such as the recursive rejection of distractors proposed by Humphreys and Müller (1993) in their SERR model of visual search (see also Heinke & Humphreys, 2005).
same	Humphreys and Müller (1993) argued that search can reflect the grouping and then recursive rejection of distractors.	Humphreys and Müller (1993) argued that search can reflect the grouping and then recursive rejection of distractors.
same	The process here may reflect the strength of grouping rather than the number of distractors since multiple distractors may be rejected together in a group – indeed effects of the number of distractors may be non-linear as grouping can increase at larger display sizes.	The process here may reflect the strength of grouping rather than the number of distractors since multiple distractors may be rejected together in a group – indeed effects of the number of distractors may be non-linear as grouping can increase at larger display sizes.
same	Grouping and group selection both reflect the similarity of targets and distractors and the similarity of the distractors themselves, and these two forms of similarity vary in opposite directions in conjunction and spatial configuration search (relative to a feature search condition as employed here, there is stronger target-distractor grouping and weaker distractor-distractor grouping; see Duncan & Humphreys, 1989).	Grouping and group selection both reflect the similarity of targets and distractors and the similarity of the distractors themselves, and these two forms of similarity vary in opposite directions in conjunction and spatial configuration search (relative to a feature search condition as employed here, there is stronger target-distractor grouping and weaker distractor-distractor grouping; see Duncan & Humphreys, 1989).
same	If the process of distractor rejection is more difficult in conjunction and configuration search, compared with feature search, then there will be effects on a parameter reflecting this process, and this may not vary directly with display size, as we observed.	If the process of distractor rejection is more difficult in conjunction and configuration search, compared with feature search, then there will be effects on a parameter reflecting this process, and this may not vary directly with display size, as we observed.
same	In contrast to the shift parameter, the shape parameter showed marginal effect of the display size, a reliable effect at the	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.	task, and an interaction between these factors.
same	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.	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.
same	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 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), suggesting either a change in the quality of a match between the stimuli and the template or a variable grouping unit affecting the recursive rejection process.	This change in the shape parameter in the large display size of the spatial configuration task is in line with a sudden increase of the drift rate standard deviation (0.080, 0.050, 0.054, 0.344), suggesting either (1) a change in the quality of a match between the stimuli and the template or (2) a variable grouping unit (amongst different observers) affecting the recursive rejection process.
	In addition, we observed a general increase in the values of the shape parameter across the display sizes on 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).	In addition, we observed a general increase in the values of the shape parameter from 1.73 of the display size 3, 1.86 of the display size 6, 2.05 of the display size 12, to 1.96 of display size 18 on absent trials in the spatial configuration task, $F(3, 57) = 6.13$, $p = .001$, $\eta^2_p = .244$.
Same	The target absent-induced shape change in the spatial configuration task was observed also in Palmer and colleagues' analysis (2011).	The target absent-induced shape change in the spatial configuration task was observed also in Palmer and colleagues' analysis (2011).
Same	However, their data showed no reliable shape change across display sizes for present trials (Palmer et al., 2011).	However, their data showed no reliable shape change across display sizes for present trials (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 appears to be supported by their data showing high rate of miss errors in the	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

	spatial configuration task (Wolfe et al., 2010).	configuration task (Wolfe et al., 2010).
Same	The high rate of miss errors might reflect when an observer prematurely decides to give an absent response on a target present trial.	The high rate of miss errors might reflect when an observer prematurely decides to give an absent response on a target present trial.
same	This will in turn reduce the overall number of slow responses leading to an RT distribution with low skew.	This will in turn reduce the overall number of slow responses leading to an RT distribution with low skew.
same	This indicates that in the conditions with high miss errors, participants tended to set a low decision threshold for the target absent response.	This indicates that in the conditions with high miss errors, participants tended to set a low decision threshold for the target absent response.
same	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 increase shape parameters.	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 increase shape parameters.
	We, 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 searches.	We, 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.
same	Together with the miss error data, our data do indicate that a link between the miss errors and the shape of the RT distribution is plausible.	Together with the miss error data, our 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, we propose another explanation that, relative to the feature and the conjunction searches, the factor that changes the RT shape in the spatial configuration search is the *goodness-of-match* between the search template and the search display.	In addition to the explanation of participants abandoning search prematurely (i.e., a dynamic changes of boundary separation), we 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 display (i.e., the drift rate changes).
same	This implies the factors contributing a change in the structure of trial RTs will result in shape changes for the RT distributions.	This implies the factors contributing a change in different parts of an RT distribution will result in its shape.
New		As our simulation study shows (Appendix B), doubling the shape parameter results in a decrease at the boundary separation (in line with the miss-error account) and an increase at the drift rate (in line with the *goodness-of-match* account). Both of the diffusion parameters likely are the processes driven

		the changes at the distributional shape.
new		Both of the diffusion parameters likely are the processes driven the changes at the distributional shape.
same	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.	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 a general change in RT distributions.	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, as does the mean RTs (see a review in Wagenmakers & Brown, 2007).	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 our dataset was not sensitive to the display size, and a separate ANOVA failed to indicate a reliable effect in the feature search task.	Unlike the mean RTs, however, the scale parameter in our dataset was not sensitive to the display size in the feature search task.
same	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.	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.
same	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.	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.	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.

- Revise texts and interpretation in the subsection, "Distributional parameters reflect underlying processes" in the General discussion section

Distributional parameters reflect underlying processes

	Previous version	Current version
same	The RT distributional parameters were posited, under the framework of the stage model of information processing, to reflect different aspects of peripheral and central processing.	The RT distributional parameters were posited, under the framework of the stage model of information processing, to reflect different aspects of peripheral and central processing.
same	The shift parameter was associated with the speed of peripheral processes (i.e., irreducible minimum response latency, Dzhafarov, 1992), the scale parameter with the speed of executing central processes, and the shape together with the scale parameters related to the insertion of additional stages into the central processing (Rouder et al., 2005).	The shift parameter was associated with the speed of peripheral processes (i.e., irreducible minimum response latency, Dzhafarov, 1992), the scale parameter with the speed of executing central processes, and the shape together with the scale parameters related to the insertion of additional stages into the central processing (Rouder et al., 2005).
Same	Using the benchmark paradigms of visual search (Wolfe et al., 2000), our data indicate that the shift parameter, instead of reflecting the speed of peripheral processes, may be associated with the process of distractor rejection and the quality of the match between a template and a search display.	Using the benchmark paradigms of visual search (Wolfe et al., 2000), our data indicate that the shift parameter, instead of reflecting the speed of peripheral processes, may be associated with the process of distractor rejection and the quality of the match between a template and a search display.
same	This is supported by the analysis using the EZ2 diffusion model.	This is supported by the analysis using the EZ2 diffusion model.
	As we argued previously, the shift parameter descriptively captures the fast responses in an RT distribution and should not be taken as a peripheral process, because, even in these quick responses, observers have completed both peripheral and central operations in order to produce measurable responses.	As we argued previously, the shift parameter captures the factors that influence the entire RT distribution equally.
New		A possible situation that the peripheral process may result in a clear shift change is when the other two parameters are kept constant.
New		That is, when no factor influences the decision-making process and when the shape of an RT distribution is unchanged.
same	We suggest that the data better reflect a process such as the recursive rejection of the grouped distractors and the quality of the match to a target template, which, when accurate, contributes to the fast end of the RT distribution.	We suggest that the data better reflect a process such as the recursive rejection of the grouped distractors and the quality of the match to a target template, which, when accurate, contributes to an entire RT distribution.

same	Our results for the scale parameter are consistent with those of Rouder and colleagues (2005) in suggesting that it reflects the speed of execution in a central decision-making process.	Our results for the scale parameter are consistent with those of Rouder and colleagues (2005) in suggesting that it reflects the speed of execution in a central decision-making process.
same	As the execution speed closely links with the decision boundaries and the initial state of sensory information an observer sets for a response trial, we observed a similar pattern in the scale parameter, the boundary separation and the non-decision time.	As the execution speed closely links with the decision boundaries and the initial state of sensory information an observer sets for a response trial, we observed a similar pattern in the scale parameter, the boundary separation and the non-decision time.
same	The pattern in the non-decision time is readily accounted for by the fact that EZ2 diffusion model absorbs the parameter reflecting the initial state of sensory evidence into the non-decision time.	The pattern in the non-decision time is readily accounted for by the fact that EZ2 diffusion model absorbs the parameter reflecting the initial state of sensory evidence into the non-decision time.
same	The distance between the decision boundary and the initial state of sensory evidence can then be taken as reflecting changes in the response criteria and hence altering the scale of an RT distribution.	The distance between the decision boundary and the initial state of sensory evidence can then be taken as reflecting changes in the response criteria and hence altering the scale of an RT distribution.
same	For the shape parameter we observed an emergent effect of perceptual grouping at the large display size in the spatial configuration search.	For the shape parameter we observed an emergent effect of perceptual grouping at the large display size in the spatial configuration search.
same	This is in line with the drift rate data in that the drift rate was slower for the spatial configuration search task relative to the two simple search tasks both in our data (0.323, 0.265 vs. 0.220) and those of Wolfe et al. (2010) (0.341, 0.299 vs. 0.203).	This is in line with the drift rate data in that the drift rate was slower for the spatial configuration search task relative to the two simple search tasks both in our data (0.323, 0.265 vs. 0.220) and those of Wolfe et al. (2010) (0.341, 0.299 vs. 0.203).
same	In Palmer et al.'s analysis (2011) no task effect was found in the shape parameter.	In Palmer et al.'s analysis (2011) no task effect was found in the shape parameter.
	Using the HBM we observed a marginal task effect, $F(2, 25) = 3.22$, $p = .06$, $\eta^2_p = .205$, suggesting that the previous result might reflect a lack of power.	Using the HBM we observed a significant task effect, $F(2, 55) = 23.50$, $p = 4.21 \times 10^{-8}$, $\eta^2_p = .461$, suggesting that the previous result might reflect a lack of power.
same	The observations of shape invariance in Palmer et al.'s analysis could also be interpreted in term of a memory match account (Ratcliff & Rouder, 2000).	The observations of shape invariance in Palmer et al.'s analysis could also be interpreted in term of a memory match account (Ratcliff & Rouder, 2000).
	This account presumes that, when the integrity of a memory match between the template and search items is still intact, the evidence strength is strong enough to	This account presumes that, when the integrity of a memory match between the template and search items is still intact, the evidence strength is strong enough to

	permit a correct decision (Smith, Ratcliff, & Wolfgang, 2004; Smith & Sewell, 2013).	permit a correct decision (Smith, Ratcliff, & Wolfgang, 2004; Smith & Sewell, 2013).
same	Since the previous study recruited fewer participants and some might find strategies to conduct the difficult searches still using the same processing stages as the feature search task, the shape parameter reflects only a marginal effect.	Since the previous study recruited fewer participants and some might find strategies to conduct the difficult searches still using the same processing stages as the feature search task, the shape parameter reflects only a marginal effect.
New		Another possible factor that may explain the different finding at the shape parameter is illustrated by the drift rate visually weighted plot.
New		The visually weighted regression lines predict two groups of participants accumulating sensory evidence at different rates, but indicate only one homogeneous group in Wolfe et al.'s data.
New		As our simulated study shows (Appendix B), the drift rate can also weigh-in to change the shape parameter.
New		This suggests that some of our participants took advantage of the crowded layout to facilitate search using the spatial configuration task, and some did not.
New		This did not happen in Wolfe et al.'s sparse layout, and likely contribute also to our significant finding at the shape parameter on present trials.

- Added Appendix B

Appendix B

EZ2 diffusion model simulations

The stimulation study was undertaken to examine how the parameters estimated from EZ2 diffusion model correlate with the Weibull parameters in a simple well-controlled situation. We performed 3 case studies that changed only one of the Weibull parameters. Note that the three distributional parameters jointly determine the general shape of a distribution. Thus a change in the drift rate may alter one or more Weibull parameters. The 3 studies respectively doubled the shift, the scale and the shape parameters in a Weibull function and simulated 200 RTs from 20 homogeneous observers.

The data were than submitted to the EZ2 model to estimate the drift rate, boundary separation and non-decision time. The result indicated that firstly, doubling the shift parameter resulted in a near two-fold increase in the non-decision time (416 vs. 827 ms), small increases in the drift rate (0.012 vs. 0.013), and small decreases of the boundary separation (4.89 vs. 4.57). Second, doubling the scale parameter resulted in a decrease of the drift rate (0.013 vs. 0.009), an increase in boundary separation (4.70 vs. 6.57), and a 10-ms increase for the non-decision time (407 vs. 417 ms). Finally, doubling the shape parameter resulted in an increase of the drift rate (0.013 vs. 0.018), a decrease of the boundary separation (4.57 vs. 3.39) and again a small increase at the non-decision time (410 vs. 507; although this increase was relatively larger than that of doubling the scale parameter).

Figure 18 shows a comparison across the three case studies.

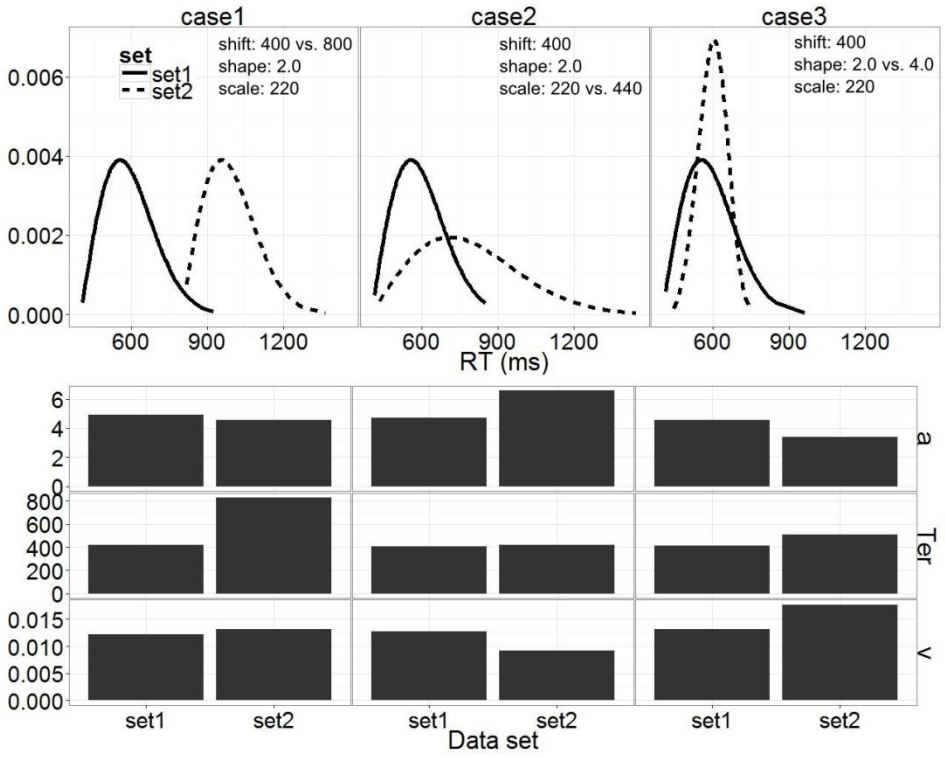


Figure 18. The figure shows how the change in a distributional parameter may alter the parameters estimated by the EZ2 diffusion model. Ter, a and v stand for the non-decision time, the boundary separation and the drift rate, respectively.