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

**Yi-Shin Lin · Dietmar Heinke · Glyn W. Humphreys**

**Abstract** In this study, we applied Bayesian-based distributional analyses to examine the shapes of response time (RT) distributions in three visual search paradigms, which varied in task difficulty. In further analyses we investigated two common observations in visual search—the effects of display size and of variations in search efficiency across different task conditions—following a design that had been used in previous studies (Palmer, Horowitz, Torralba, & Wolfe, *Journal of Experimental Psychology: Human Perception and Performance*, 37, 58–71, 2011; Wolfe, Palmer, & Horowitz, *Vision Research*, 50, 1304–1311, 2010) in which parameters of the response distributions were measured. Our 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, by analyzing trial RTs, we were able to extract paradigm-dependent shape changes in the RT distributions that could be accounted for by using the EZ2 diffusion model. The study showed that Bayesian-based RT distribution analyses can provide an important means to investigate the underlying cognitive processes in search, including stimulus grouping and the bottom-up guidance of attention.

**Keywords** Bayesian modeling · Visual search · Response time models

Y.-S. Lin (✉) · D. Heinke
School of Psychology, University of Birmingham, Edgbaston, B15 2TT Birmingham, UK
e-mail: yxl193@bham.ac.uk

G. W. Humphreys (✉)
Department of Experimental Psychology, University of Oxford, South Parks Road, Oxford OX1 3UD, UK
e-mail: glyn.humphreys@psy.ox.ac.uk

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 technique uses separate regressors to assess variations across trial RTs collected from a participant by estimating regression coefficients, contrary to conventional single-level analysis of variance (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 the 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, the 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 the 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. On the basis of Bayes's theorem, the outputs of prior distributions can then be used to calculate posterior distributions, which are conceptualized as the underlying functions for the parameters at 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 in aggregated RTs between observers. This differs from applying Bayesian statistics to account for how an observer identifies a search target by conceptualizing that his or her prior experiences (e.g., search history; modeling the RTs in the $[N-1]$th trial as the prior distribution) influence the current search performance (modeling the RTs in the $N$th trial as the posterior distribution).

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; for other examples, see Matzke & Wagenmakers, 2009; Rouder, Lu, Morey, Sun, & Speckman, 2008). In symbolic distance studies, observers may be asked to decide whether 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) than 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 results from HBM, however, suggest a further refinement for this interpretation, by showing that the locus of the effect resides in the scale (rate), rather than the shape, of the 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 et al., 2005).

## Application to visual search

In the present study, we 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 three 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) setup for the target can act to guide attention to stimuli whose features match those of the expected target. Depending on the relations between the target and the distractors, and also the relations between the distractors themselves (Duncan & Humphreys, 1989), performance is affected by several key factors, including the presence or absence of the target, and the similarity between the target and the distractor and the similarity between distractors (for a computational implementation of these effects based on stimulus grouping, see Heinke & Backhaus, 2011; Heinke & Humphreys, 2003).

The display size effect relates to how performance is affected by the number of distractors in the display. Effects of display size are frequently observed in tasks in which 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, e.g., 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 that we adopted HBM (see Rouder et al., 2005, as well as Rouder & Lu, 2005) in the present study are that (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.

In this study, we examined the effectiveness of distributional analyses and the HBM approach for understanding performance in three benchmark visual search tasks, which were modified from those of Wolfe, Palmer, and Horowitz (2010; a different set of analyses was also reported in Palmer, Horowitz, Torralba, & Wolfe, 2011; also see the computational model aiming at clarifying the mechanism of search termination in Moran, Zehetleitner, Müller, & Usher, 2013). In Wolfe et al.'s (2010) 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 in an invisible 5 × 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 three-parameter probability functions

For our study, we adopted four three-parameter probability—lognormal, Wald, Weibull, and gamma[1]—functions (Johnson, Kotz, & Balakrishnan, 1994) to estimate RT distributions using HBM. Unlike the frequently used ex-Gaussian function, the three-parameter probability functions describe an RT distribution with shift, scale, and shape parameters that characterize the pattern of a distribution. An increase in the scale parameter shortens the central location of a distribution and thickens its tail. This implies that the responses originally accumulated around the central part become slower, and thus move 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, an increase in the shape parameter not only changes the kurtosis, skewness, and variance, 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 rightward (see Fig. 1 for an illustration).

In this study, we assumed that changes in RT distributions reflect unobservable cognitive processes (a similar argument was made by Heathcote et al., 1991). As is illustrated in Fig. 1, factors that affect quick, moderate, and slow responses evenly will show a selective effect on the shift parameter. Factors that alter only the proportion of responses, moving it from the central location to the tail part of a distribution (or vice versa), will affect the scale parameter. Lastly, an effect on the shape parameter may result from 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 are not restricted to, 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 & Backhaus, 2011; Heinke & Humphreys, 2003; J. Palmer, 1995). Some previous work (e.g., Rouder et al., 2005) has suggested 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 that have applied distributional analyses to RT data, attempting to link distributional parameters with psychological processes directly (e.g., Gu, Gau, Tzang, & Hsu, 2013; Rohrer & Wixted, 1994). Although it is ambitious to posit links between distributional parameters and underlying

---

[1] These functions describe distributions with the same set of parameters: shape, scale, and shift. Because, relative to other functions, a previous analysis (Palmer et al., 2011) had reported a worse $\chi^2$ fit for the Weibull function, we constructed comparable three-parameter HBM analyses to test whether other functions would gain substantially better fit by using a hierarchical Bayesian approach rather than the Weibull function. We thank Evan Palmer for this suggestion.

psychological processes, a better strategy is to take advantage of the descriptive nature of distributional parameters (Schwarz, 2001), which permits 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, and shape). This enables researchers to examine RT data as an entirety, building on what can be provided by an analysis of mean RTs. However, one potential pitfall is uncertainty as to 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 by Matzke, Dolan, Logan, Brown, & Wagenmakers, 2013, and suggested also by 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, Dolan, & Grasman, 2008; Wagenmakers, van der Maas, & Grasman, 2007), which is a closed-form and simplified variant of Ratcliff's (1978) diffusion model. 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 evidence surpasses either the positive or the negative decision boundary on the *y*-axis, a decision is reached, and the time that 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 nondecision 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 process of template matching influences the 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 nondecision component is a residual time, calculated by subtracting the decision time (estimated by the diffusion model) from the total (recorded) RT; this may reflect

**Fig. 1** Illustration of changes in the scale, shape, and shift parameters, simulated by a three-parameter Weibull function. The legend in each panel shows the extent to which the parameter is adjusted while the others are kept constant
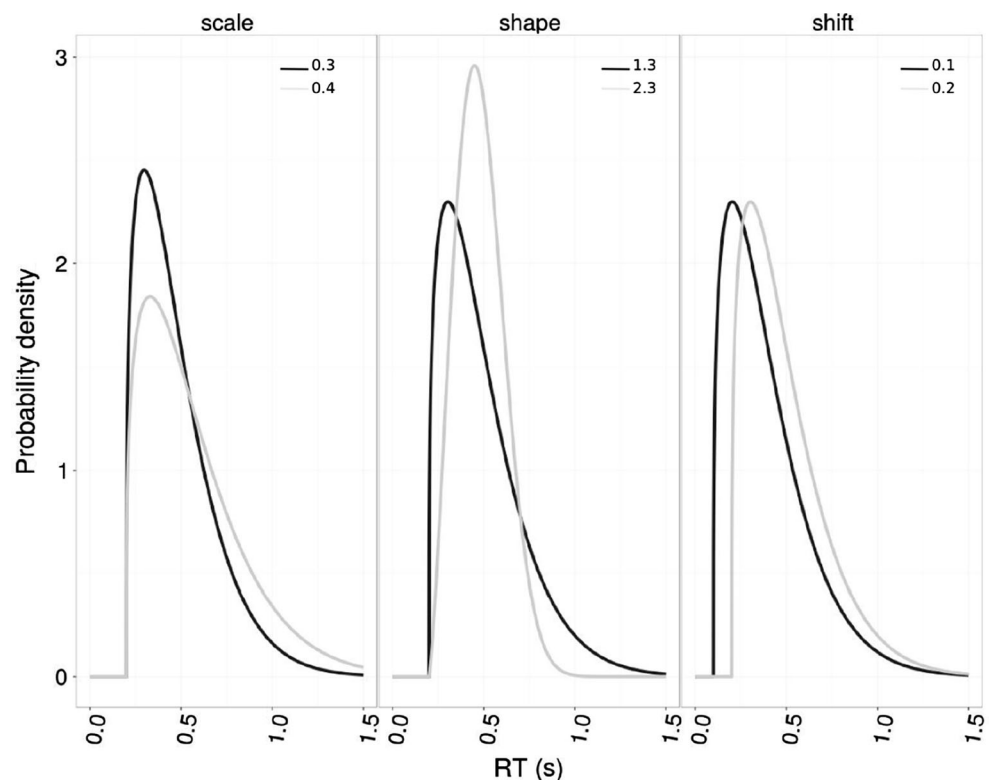
the time to encode stimuli (perceptual time) together with the time to produce a response output (motor time; Ratcliff & McKoon, 2007).

The diffusion model has been applied to various 2AFC paradigms, and so far both psychophysical and neurophysiological studies indicate its usefulness in probing the two latent decision-making processes and 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 type of simplification (Grasman, Wagenmakers, & van der Maas, 2009; though see a review for more complicated statistical decision models of visual search in Smith & Sewell, 2013) that 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 either decision-related or non-decision-related processes, the EZ2 model is able to account for the changes in RT distributions in a psychologically meaningful way. For instance, a factor that affects the nondecision process should reflect on the shift parameter, which 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, a factor that delays decision-related processes may consistently delay only the responses from the quick to the central band of an RT distribution, so it will result in an increase of the scale parameter. That is, as the leftmost panel in Fig. 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 an 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, boundary separation, and the nondecision 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. Weibull HBM, on the other hand, summarizes the shape of the RT distribution in each condition. The RT distributions thus are the aggregated outputs from the diffusion processes. The dual-modeling approach, on the one hand, assumes that one search response is driven by the diffusion process, and on the other, that all of 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 able to account for the descriptive model, because the benchmark paradigms produced high accuracy responses.

In summary, for this study we examined three questions related to the perceptual decision making during visual search. The first question was whether the demands of a 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 shown to be highly inefficient (Bricolo, Gianesini, Fanini, Bundesen, & Chelazzi, 2002; Kwak, Dagenbach, & Egeth, 1991; Woodman & Luck, 2003). It is reasonable to expect that this particular search task would change the shape of the RT distribution drastically. The second question examined was 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 items in a display merely adds to the burden on early perceptual process, we should expect to find no influences from the display size on any decision parameters, and thus on the RT shape. The third question examined was 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 nondecision parameters, and therefore manifest as an interaction effect on the shape of the RT distribution.

## Method

### Participants

Forty volunteers took part, from 18 to 22 years old ($M \pm SE = 18.9 \pm 1.01$; 33 females, seven males, 35 right- and five 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 design similar to that of Wolfe et al. (2010), with a slight modification. Specifically, we used a circular display layout with a viewing area of 7.59 × 7.59 deg of visual angle, in which 25 locations were allocated to hold search items. Wolfe et al. (2010) used a viewing area of 22.5 × 22.5 deg of visual angle (also with 25 search locations), and each search item subtended around 3.5 to 4.1 deg of visual angle. Relative to Wolfe et al.'s (2010) 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.

In the study, we investigated two factors, display size (3, 6, 12, and 18 items) and whether the target was present or absent, using a repeated measures within-subjects 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 minimize one of the possible experimenter biases related to the analysis of null hypothesis significance testing (Kruschke, 2010), we set a target sample size (20 in each group) before collecting data. The target sample size was determined on the basis of commonly used sample sizes (approximately 5–20 participants) in the 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 $0.69 \times 0.69$ deg of visual angle). In the conjunction search task, observers looked for a vertical, dark bar ($0.33 \times 0.96$ deg of visual angle) amongst two types of distractors, vertical gray bars ($0.33 \times 0.96$ deg of visual angle) and horizontal dark bars ($0.96 \times 0.33$ deg of visual angle). In the spatial configuration search task, each observer looked for the digit 2 amongst digit 5 s (both are $0.33 \times 0.58$ deg of visual angle) (see Fig. 2 for an example trial in each of the tasks).

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 a Linux hard real-time kernel 2.6.31-11-rt and an 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 $1,152 \times 864$ pixels with a refresh rate set at 100 Hz. The visible area included the entire screen (i.e., $1,152 \times 864$ pixels), but the relevant stimuli were all drawn within the viewing area of $7.59 \times 7.59$ deg of 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, in which each experimental condition comprised 100 trials. The volunteers carrying out the feature and conjunction search tasks completed the tasks in a counterbalanced sequence.

### Hierarchical Bayesian model (HBM)

The HBM framework is based on Rouder and Lu's (2005) R code, 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 may 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., $\beta$, describing the shape of an RT distribution), scale (i.e., $\theta$, describing the general enhancement of the magnitude and variability in an RT distribution), and shift (i.e., $\psi$, describing the possible minimal RT of a distribution). The $\beta$ parameter was then modeled by a $\gamma$ distribution with two hyperparameters, $\eta_1$ and $\eta_2$, and the $\theta$ and $\psi$ parameters were modeled by two uniform distributions. The former ($\theta$) was initialized as an uninformative distribution, whereas the latter ($\psi$) was set to the range from zero to minimal RTs for each respective condition and participant, because the $\psi$ parameter assumed a role as the nondecision component. The hyperparameters underlying the $\gamma$ distributions were then modeled by other $\gamma$ distributions with designated parameters, following Rouder and Lu (2005). Likewise, we replaced the Weibull function with the three-parameter gamma, log-normal, and Wald functions (Johnson et al., 1994), keeping similar prior parameter setting.

In HBM, correct RTs were modeled for each participant separately in each condition. The HBM consisted of three simultaneous iteration chains. Each of them iterated 105,000 times and sampled once every four iterations, in order to alleviate possible autocorrelation problems. The first 5,000 samples were considered to be arbitrary and were discarded (i.e., burn-in length). The same setting was applied both to our data and to Wolfe et al.'s (2010) data to allow for a direct comparison.
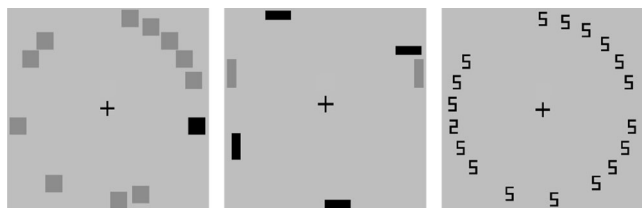


**Fig. 2** Schematic representation of the tasks. In each panel a target is present (from left to right, the black item [feature], the black vertical bar [color–form conjunction], and the number 2 [spatial configuration])

## Diffusion model

Analyses were also based on Grasman et al. (2009) EZ diffusion model, implemented in R's EZ2 package, to estimate the drift rate, boundary separation, and nondecision components separately for each participant in each condition. Following the assumptions of the EZ diffusion model (Wagenmakers et al., 2008), the across-trial variability associated with each of the drift rate, boundary separation, and nondecision 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 in which an observer committed no errors. "Present" and "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 were set according to the paradigm and the literature: (1) The paradigm permitted only two response options (the target was either present or absent) and (2) the search slope for the 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 nondecision component and the boundary separation were arbitrarily, but reasonably, set at 0.05 and 0.09. The initial values were simply educated guesses provided to allow the algorithm to approach reasonable estimations.

For both HBM and the diffusion model, the parameters were estimated on a per-condition, per-participant basis, so the 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 nonparametric bootstrapping procedure implemented by Schönbrodt (2012) for Hsiang's (2013) visually weighted regression method.[3]

## Results

We report the data in four sections. First, 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. Third, 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 HBM and the EZ2 diffusion model. These include the data for the Weibull and the diffusion model parameters, presented separately, with visually weighted nonparametric regression plots. From here we go on to discuss the factors contributing to the RT shape, shift, and scale parameters, on the basis of how these parameters change across the different search conditions, and contrast them with the decision parameters from the diffusion model. The appendix presents two simulation studies that we used to examine whether Weibull HBM estimates of the distributional parameters were reliable with a small sample size, and Bayesian diagnostics were used to verify the reliability of the Markov chain Monte Carlo procedure.

We focus on the data from target-present trials because target-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 by Moran et al., 2013). A decision in a target-absent trial is possibly reached on the basis, for example, of a termination rule that allows an observer to deem that the collected sensory evidence is strong enough to refute the presence of a target. Although it is likely that an observer, in a target-present trial, may also adopt an identical termination rule to infer the likelihood of target presence, he or she would rely on the stronger sensory evidence extracted from a target than from nontargets. This is likely when a target image is physically available in a trial and target foreknowledge is set up in an attentional template. Thus, the main aim of this report is to examine the role of factors such as the target–distractor grouping effect on the distribution of target-present responses in search. We nevertheless also append standard analyses for target-absent trials in all the figures.

## Mean RTs and error rates

As is typically done for analyses of aggregated RTs, we trimmed outliers by defining them as (1) incorrect responses or correct responses outside the range of 200 to 4,000 ms, for feature and conjunction searches, and 200 to 8,000 ms, for spatial configuration searches (though see Heathcote et al., 1991, for the downside of trimming RT data). The trimming scheme was the same that was used by Wolfe et al. (2010). This outlier trimming resulted in rejection rates of 9.2 %, 12 %, and 7.2 % of 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.

---

[2] When an observer made no error responses (i.e., 100 % accuracy, $P_c$), the accuracy was replaced with a value that corresponded 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.

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

**Fig. 3** Box-and-whisker plots. The upper and lower panels show means of the 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 labeling the four display sizes). The left and right panels present the analyses from the present and Wolfe et al.'s (2010) data sets, respectively
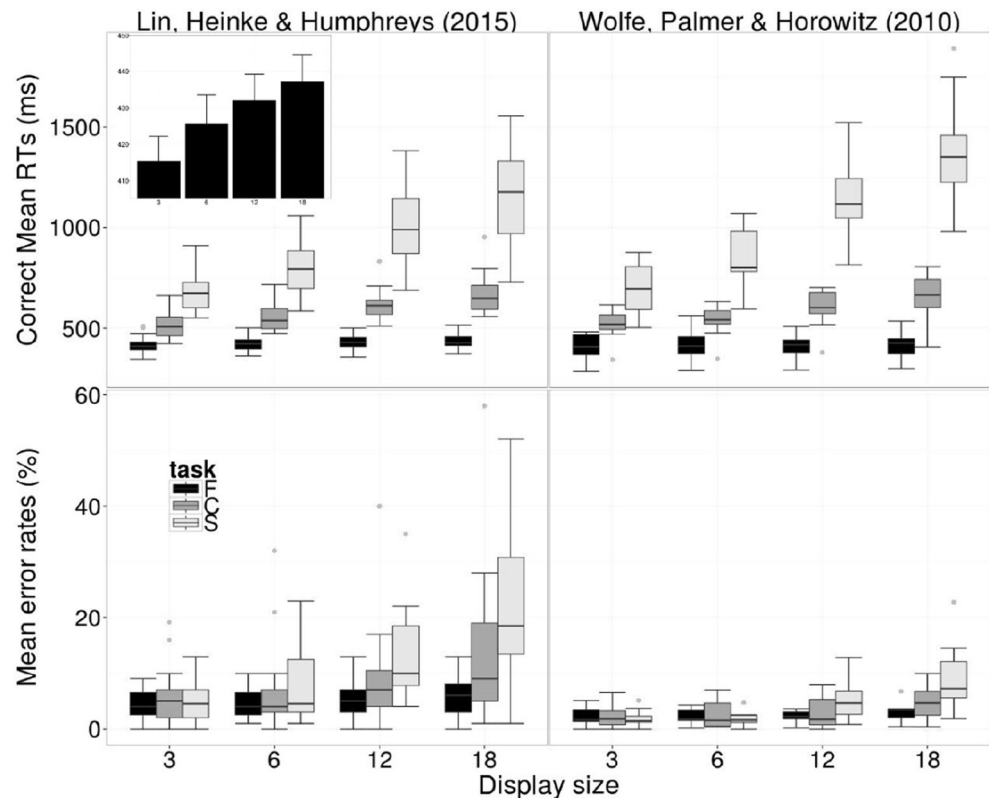


A two-way ANOVA[5] showed reliable main effects of display size, $F(3, 165) = 176.107$, $p = 1 \times 10^{-13}$, $\eta^2_p = .762$, and search task, $F(2, 55) = 108.385$, $p = 1 \times 10^{-13}$, $\eta^2_p = .798$, as well as an interaction between these factors, $F(6, 165) = 68.633$, $p = 1 \times 10^{-13}$, $\eta^2_p = .714$. The spatial configuration search ($RT_{mean} = 913$ ms) required reliably longer RTs than did 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 did 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, $F(3, 54) = 7.494$, $p = 2.78 \times 10^{-4}$, $\eta^2_p = .294$. RTs were slower for display sizes 18 and 12 when compared with display size 3 ($t = 6.37$, 95 % CI = ~11.61–31.82 ms, $p = 3.22 \times 10^{-5}$; $t = 4.03$, 95 % CI = ~4.43–28.95 ms, $p = 4.67 \times 10^{-3}$). We also found a reliable main effect of display size for conjunction search, $F(3, 54) = 103.15$, $p = 1 \times 10^{-13}$, $\eta^2_p = .851$, and spatial configuration search, $F(3, 57) = 113.8$, $p = 1 \times 10^{-13}$, $\eta^2_p = .857$, tasks. Post-hoc *t* tests for the conjunction task showed reliable differences across all display sizes (510, 552, 615, and 667 ms for 3, 6, 12,

and 18 items, respectively; $ps = 2.63 \times 10^{-7}$, $9.70 \times 10^{-9}$, $2.67 \times 10^{-9}$, $4.98 \times 10^{-6}$, $6.08 \times 10^{-8}$, $4.19 \times 10^{-5}$ for comparisons of 3 vs. 6, 3 vs. 12, 3 vs. 18, 6 vs. 12, 6 vs. 18, and 12 vs. 18, Bonferroni corrected for multiple comparisons). Similar effects were present for the spatial configuration search, too (679, 809, 1,011, and 1,154 ms for increasing display sizes; $ps = 5.14 \times 10^{-7}$, $5.15 \times 10^{-9}$, $4.10 \times 10^{-9}$, $1.42 \times 10^{-7}$, $1.09 \times 10^{-8}$, and $2.33 \times 10^{-7}$ for comparisons of 3 vs. 6, 3 vs. 12, 3 vs. 18, 6 vs. 12, 6 vs. 18, and 12 vs. 18, Bonferroni corrected for multiple comparisons; see Fig. 3).

The error rates showed a pattern similar to the average RTs, 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, $F(3, 165) = 38.09$, $p = 1 \times 10^{-13}$, $\eta^2_p = .409$, and search task, $F(2, 55) = 5.75$, $p = .005$, $\eta^2_p = .173$, as well as their interaction, $F(6, 165) = 10.867$, $p = 3.52 \times 10^{-10}$, $\eta^2_p = .283$. 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 the significance level after Bonferroni correction (the difference of mean error rate = 3.18 %, $p = .356$, 95 % CI = −1.774 % to ~8.134 %). The conjunction search task, in turn, was more difficult than the feature search task (5 % errors; difference in mean error rates = 3.621 %, $p = .241$, 95 % CI = −1.396 % to ~8.628 %; again, the difference was not significant). The only reliable difference in error rates was between the spatial configuration search and the feature search tasks (difference in

---

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

mean error rates = 6.801 %, $p$ = .004, 95 % CI = 1.847 % to ~11.755 %).

For the feature search, the effect of display size was not reliable, $F(3, 54)$ = 1.517, $p$ = .221, $\eta^2_p$ = .078, whereas there was a reliable effect of display size for both the conjunction search task, $F(3, 54)$ = 6.075, $p$ = .001, $\eta^2_p$ = .252, and the spatial configuration task, $F(3, 57)$ = 41.426, $p$ = 1.24 × 10$^{-13}$, $\eta^2_p$ = .686 (lower panel in Fig. 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 6 (6.79 %; $p$ = .043, Bonferroni corrected for multiple comparisons). In the spatial configuration search, there were differences across all display size pairings except for 3 versus 6 ($p$ = .161; $ps$ = 5.90 × 10$^{-5}$, 9.85 × 10$^{-6}$, 3.58 × 10$^{-4}$, 6.80 × 10$^{-6}$, and 1.21 × 10$^{-5}$ for 3 vs. 12, 3 vs. 18, 6 vs. 12, 6 vs. 18, and 12 vs. 18, Bonferroni corrected for multiple comparisons).

### Error analysis

To test whether the shape change in an RT distribution was due to an increase of miss errors (Wolfe et al., 2010), we also analyzed two types of errors: misses (i.e., participants pressed the "absent" key in target-present trials) and false alarms (i.e., participants pressed the "present" key in target-absent trials).

A two-way ANOVA on miss error rates showed reliable main effects of display size, $F(3, 165)$ = 38.08, $p$ = 1 × 10$^{-13}$, $\eta^2_p$ = .409, and search task, $F(2, 55)$ = 5.75, $p$ = .005, $\eta^2_p$ =

.173, as well as an interaction between these factors, $F(6, 165)$ = 10.85, $p$ = 3.62 × 10$^{-10}$, $\eta^2_p$ = .283. Both the spatial configuration, $F(3, 57)$ = 41.37, $p$ = 1.25 × 10$^{-13}$, $\eta^2_p$ = .685, and conjunction search, $F(3, 54)$ = 6.08, $p$ = .001, $\eta^2_p$ = .253, tasks showed increasing miss errors as the display size increased, but the feature search task did not, $F(3, 54)$ = 1.52, $p$ = .221, $\eta^2_p$ = .078. False alarms showed only a display size effect, $F(3, 165)$ = 3.94, $p$ = .010, $\eta^2_p$ = .067. The reliable effect of false alarm errors was observed in both feature search, $F(3, 54)$ = 2.81, $p$ = .048, $\eta^2_p$ = .135, and conjunction search, $F(3, 54)$ = 2.96, $p$ = .040, $\eta^2_p$ = .141, but not in spatial configuration search, $F(3, 57)$ = 1.14, $p$ = .340, $\eta^2_p$ = .057 (Fig. 4).

### Distributional analysis

Figure 3 also shows the distributions of the means of RTs and error rates across the display sizes and tasks. Three noticeable characteristics are evident. First, performance in the feature search task changed little across the display sizes. Second, in the two inefficient search tasks (conjunction and 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 also shifted the entire response distribution. Third, 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



**Fig. 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
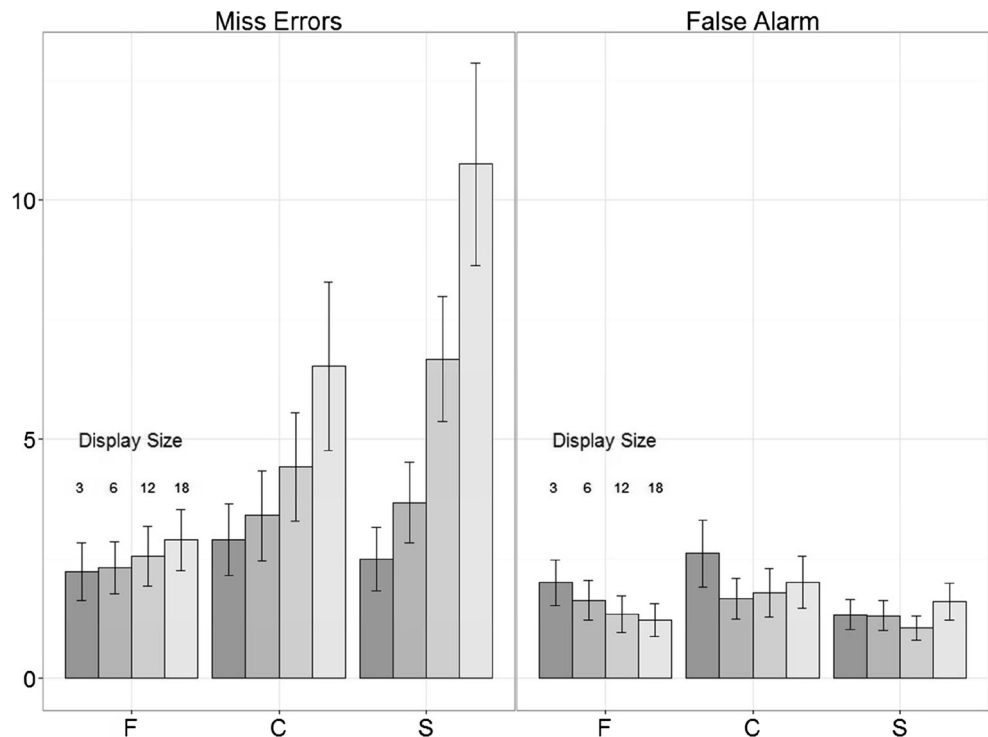
central measures of performance as well as the long-latency responses. Notably, the difference between the effects of the different display sizes at the long end of the response distribution was exacerbated for the spatial configuration search task.

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

Figure 5 shows the RT distributions at the different display sizes and search tasks. The distributions were constructed on the basis of the mean RTs ($N_{feat}$ and $N_{conj}$ = 19, $N_{spat}$ = 20; 464 data points). The feature search showed a leptokurtic 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-latency RTs showed signs of deviation from a normal distribution.

Figure 6 shows RT distributions and quantile–quantile plots. The distributions were constructed on the basis of the trial RTs (43,485 data points). Each density line represents the data from one participant. Evidently, the normality assumption was untenable

across all of the conditions. All subplots showed that the data clearly deviated from the theoretical normal lines. It is also apparent that individual differences played 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 on the basis of trial RTs (43,485 and 109,036 data points in our and Wolfe et al.'s, 2010, 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 and 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 is to produce a response that falls in the right tail of the RT distribution. This observation again cautions us against a reliance solely on using measurements of the central location when investigating visual search performance.

Contrasts with prior data

We compared our data with those of Wolfe et al. (2010). A comparison of the mean RT and error rates indicated similar



Fig. 5 Mean RT distributions. The subplots within each panel are quantile–quantile (Q–Q) normalized plots showing deviations of the data from the theoretical normal distribution. The Q–Q normalized plots compare RT means [y-axis label, "RT (ms)"] with normalized z scores [x-axis label, "Z-score"]. F, C, and S stand for feature, conjunction, and spatial configuration tasks, and P and A are target-present and -absent trials, respectively
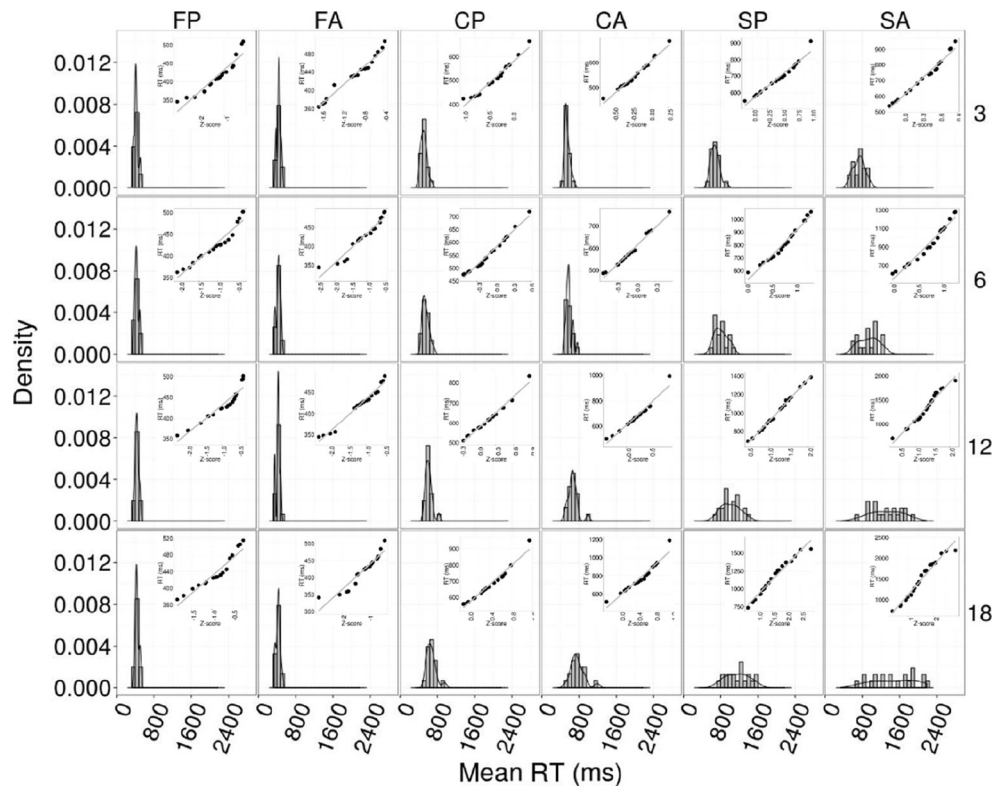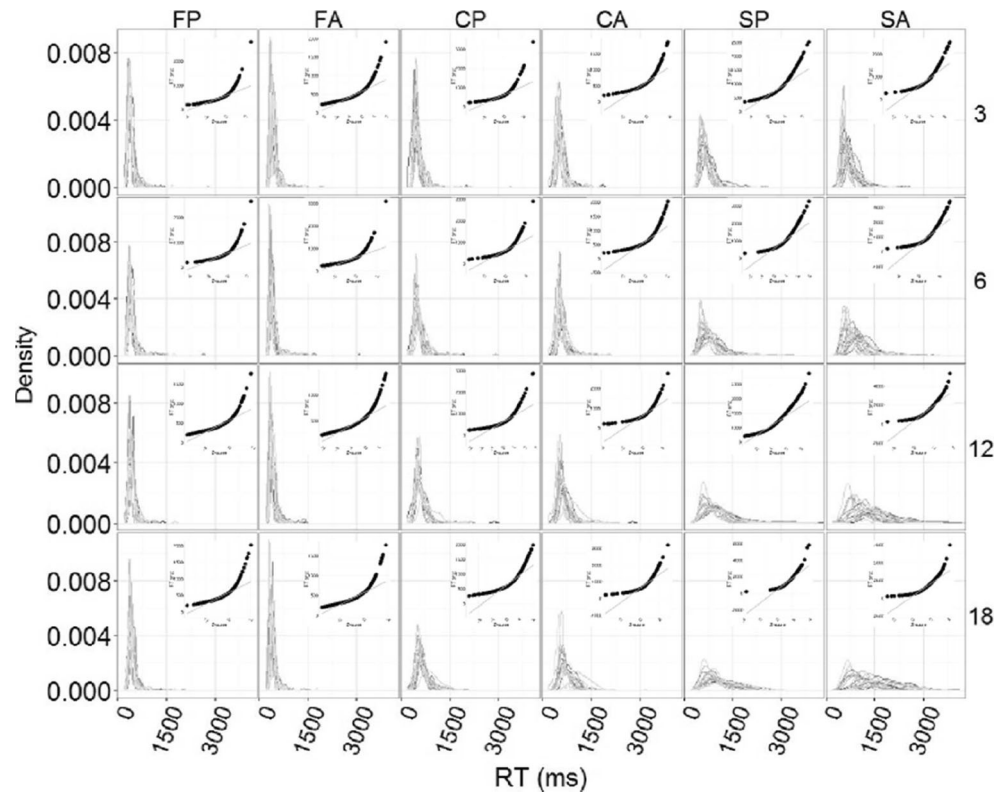
**Fig. 6** Trial RT distributions. The quantile–quantile normalized subplot within each panel compares trial RTs [*y*-axis label, "RT (ms)"] with normalized *z* scores [*x*-axis label, "Z-score"]



patterns across the studies (see Fig. 3), as is also suggested by the cumulative density and RT plots shown in Figs. 7 and 8.

With only a small number of participants, it is difficult to rule out the normality assumption when examining the mean
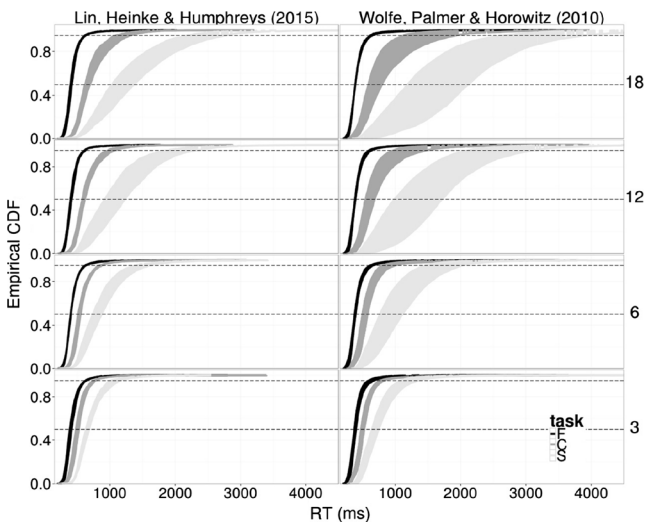


**Fig. 7** Empirical cumulative RT density curves, drawn on the basis of the trial RTs. The areas within each envelope represent the differences between target-present and target-absent trials for each task. The two dashed lines show the positions of the 50 % and 95 % cumulative densities. Long latencies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

RTs (see the subplots in Fig. 8), but the data for the trial RTs reveal a skewed distribution (Fig. 9).

### HBM estimates

In this section, we first present each parameter separately for the respective ANOVA results, and compare the data for the three search tasks at the different display sizes, modeled by HBM. Next, we describe a nonparametric bootstrap regression to assess the relationship between display size and the difficulty of the search task. The analysis focused on target-present trials. We used the deviance information criterion (DIC) to evaluate each function's fit to the data (see Table 1). 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 DICs, the DICs across the four fitted functions were close. Moreover, the diagnostics of the gamma HBM suggests that its posterior distributions did not converge. Excluding the nonconverged gamma function, we arbitrarily report estimates from the Weibull HBM, given that prior work has shown that this function provides a highly robust account, not strongly moderated by noise in the data (for a specific pathology of the Weibull function, see Rouder & Speckman, 2004, pp.
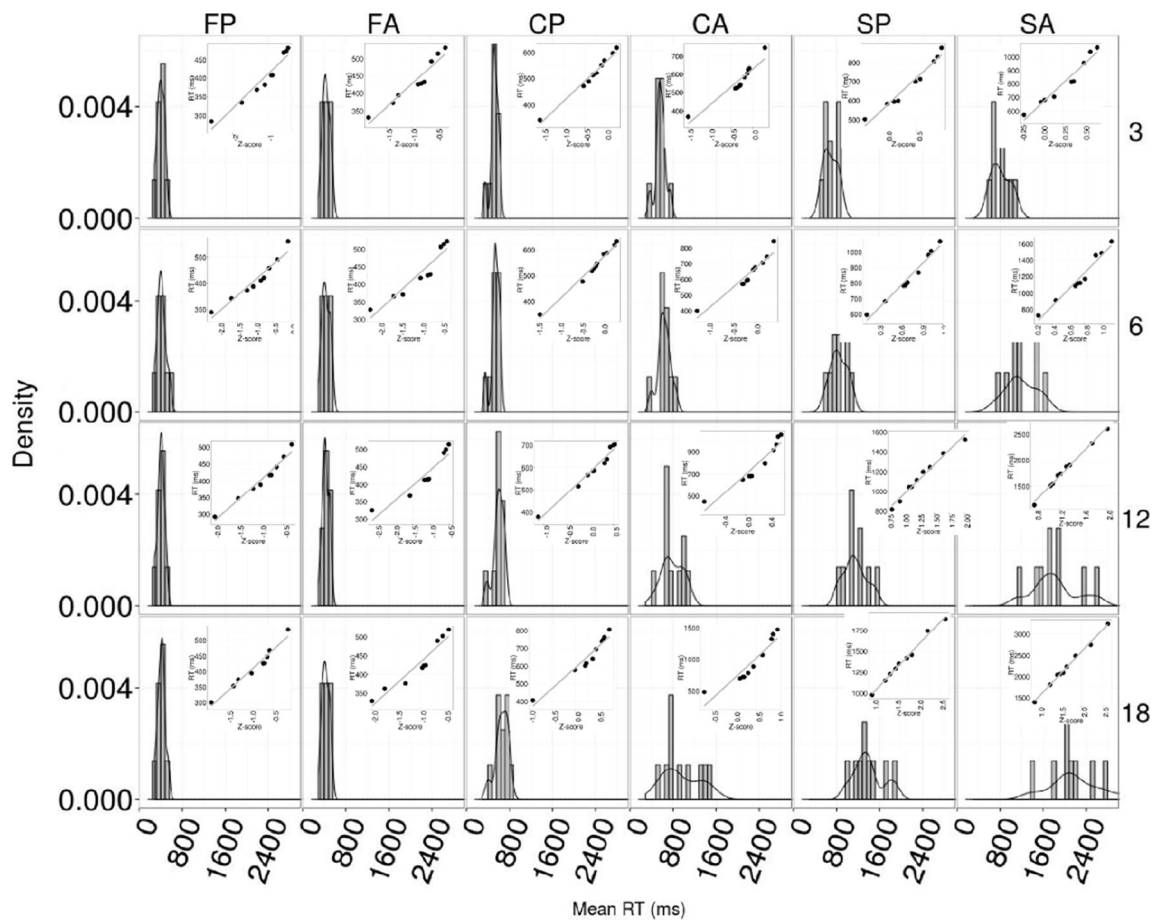
**Fig. 8** Mean RT distributions. The quantile–quantile normalized subplot within each panel compares RT means [*y*-axis label, "RT (ms)"] with normalized *z* scores [*x*-axis label, "Z-score"]. The data are taken from Wolfe et al. (2010)

424–425, and for how HBM resolves this problem, see Rouder et al., 2005, p. 203).

*Shift* A two-way (Task × Display Size) ANOVA[6] revealed significant effects 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 did conjunction search, which also had a smaller value than the spatial configuration search (246 and 342 ms vs. 436 ms, $ps = 2.37 \times 10^{-10}$ and $2.83 \times 10^{-10}$). For a summary of these and of all ANOVA results from this study, see Table 2.

The plot in the upper left panel of Fig. 10 shows two important characteristics for target-present trials. First, the non-parametric regression lines show that the shift parameter varied little across 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 does varying display sizes.

*Scale* The two-way (Task × Display Size) ANOVA was significant for the task, $F(2, 55) = 161.70$, $p = 1.0 \times 10^{-13}$, $\eta^2_p = .855$, 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 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 with increasing display size) 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, $F(3, 54) = 0.084$, $p = .968$, $\eta^2_p = .005$ (201, 207, 206, and 205 ms). Post-hoc *t* tests showed significant differences between all display sizes in spatial configuration search ($ps = 7.59 \times 10^{-3}$, $9.34 \times 10^{-6}$, $1.34 \times 10^{-7}$, .021, $1.56 \times 10^{-4}$, and .04 for the comparisons of 3 vs. 6, 3 vs. 12, 3 vs. 18, 6 vs. 12, 6 vs. 18, and 12 vs. 18; Bonferroni corrected for multiple comparisons). This held for conjunction

---

[6] For the same reason as in note 5, we analyzed the three levels of Task as a between-subjects factor.
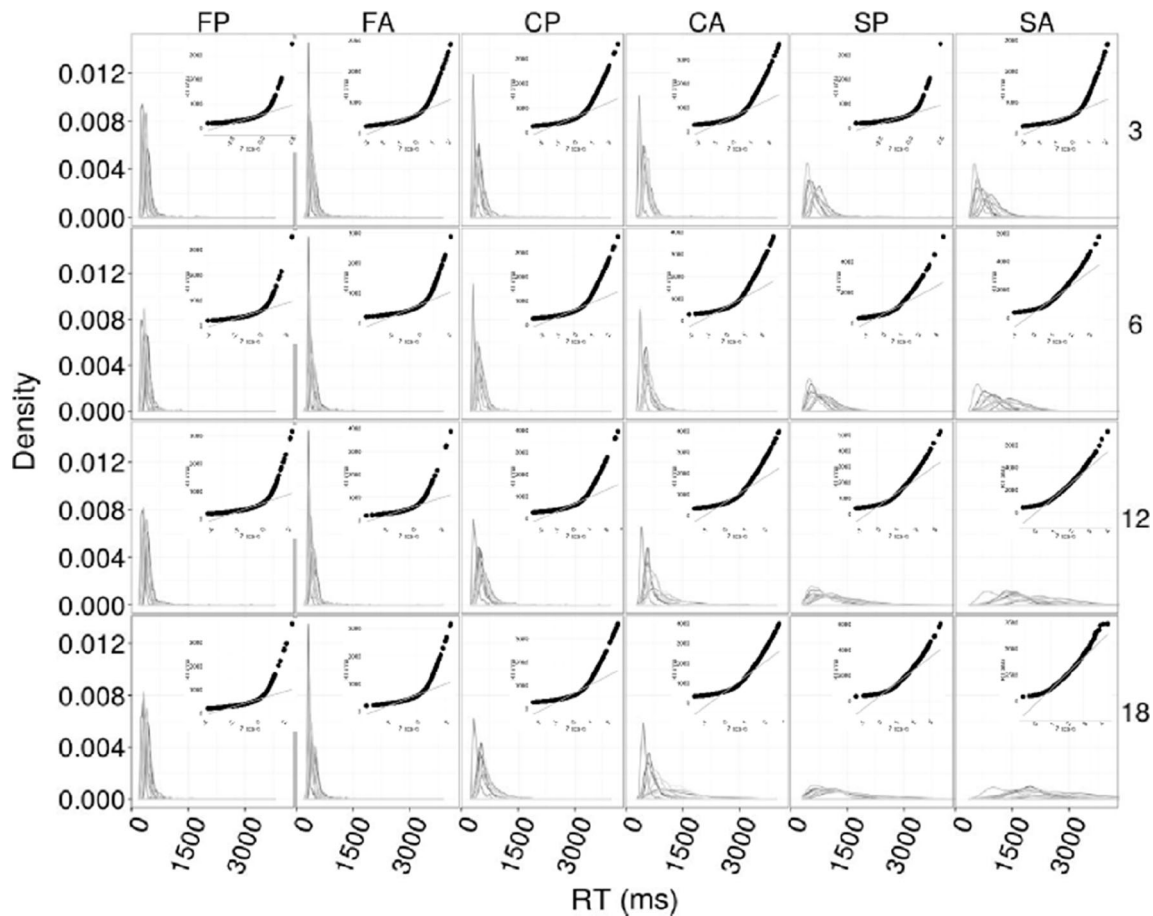
**Fig. 9** Trial RT distributions. The quantile–quantile normalized subplot within each panel compares trial RTs [y-axis label, "RT (ms)"] with normalized z scores [x-axis label, "Z-score"]. The data are taken from Wolfe et al. (2010)

search only for the 3 versus 12, and 3 versus 18 comparisons ($p$s = .001, Bonferroni corrected for multiple comparisons). No significant differences were observed in feature search.

The lower left panel of Fig. 10 shows two important characteristics. First, the regression lines indicate increasing variability (i.e., decreasing ribbon density) as the display sizes increase for conjunction and spatial configuration search, but not for feature search. Second, the display size effect only becomes noticeable

**Table 1** Deviance information criteria of the four fitted functions

|  | Present Study | Wolfe et al. (2010) |
| --- | --- | --- |
| Gamma | 385,348,342 | 975,871,147 |
| Lognormal | 385,348,002 | 975,870,279 |
| Wald | 385,348,026 | 975,870,358 |
| Weibull | 385,348,139 | 975,871,078 |

The criteria are averaged across target-absent and -present trials, tasks, and display sizes.

for the inefficient search tasks, in line with the RT mean results.

*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 significant results for 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 (1,496, 1,731, 1,695, and 1,702 with increasing display size), $F(3, 54) = 4.21$, $p = .009$, $\eta^2_p = .190$, and spatial configuration search (1,573, 1,541, 1,397, and 1,529), $F(3, 57) = 4.45$, $p = .007$, $\eta^2_p = .190$, but not for feature search (1,702, 1,819, 1,976, and 1,850), $F(3, 54) = 2.13$, $p = .106$, $\eta^2_p = .106$. Post-hoc $t$ tests showed significant display size differences for 3 versus 6, 3 versus 12, and 3 versus 18 items, $p$s = .022, .018, and .009, in the conjunction search. In the spatial configuration search, the display size differences were observed at 3 versus 12, 6 versus 12, and 12 versus 18

**Table 2**   Summary table for the significance of two-way analyses of variance for all tested parameters

|              | Mean RT | Error Rate | RT Shift | RT Shape | RT Scale | Drift Rate | Nondecision 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                   |

V indicates a significance level, $p < .01$.

items, $p$s = .013, .047, and .003 (Bonferroni corrected for multiple comparisons).

The plots in the middle left panel of Fig. 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—for both 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 variability suggest that additional factors influenced search at the large display sizes here—which we suggest reflects grouping between the elements. We elaborate on this proposal in the General Discussion.

Diffusion model

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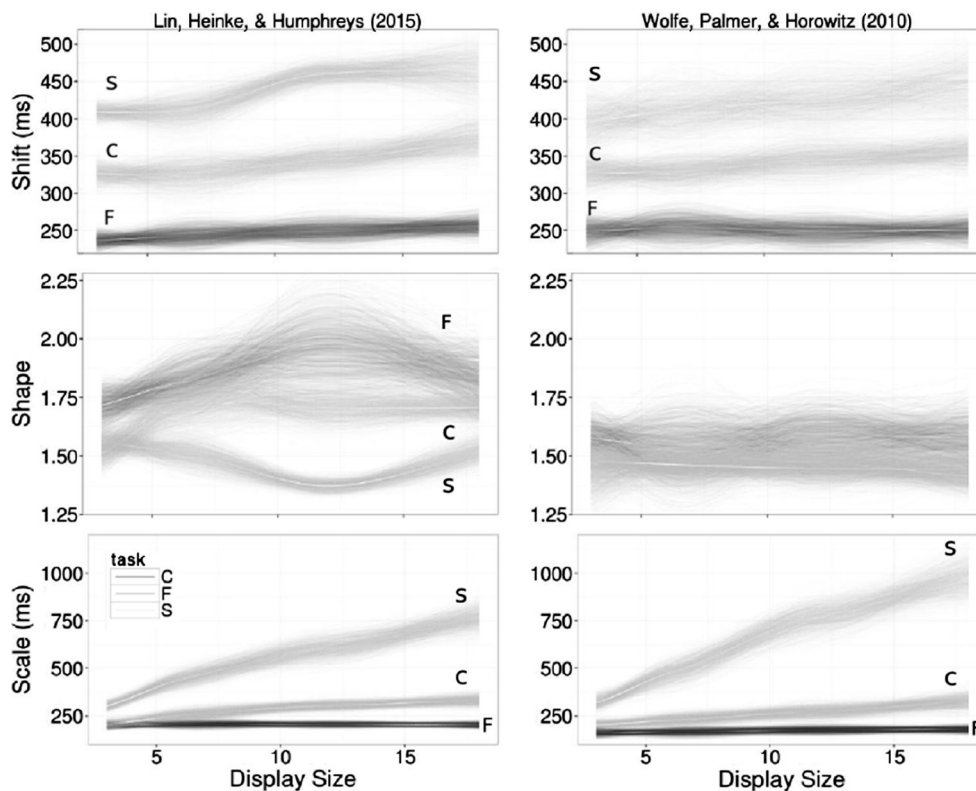


**Fig. 10** Visually weighted regression (VWR) plots (Hsiang, 2013) for the three Weibull parameters. VWR performs regressions using display size as the continuous independent variable and Weibull function estimates (shift, shape or scale) as the predicted variables separately for the three search tasks. 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 grayscaled lines). Here we used locally weighted smoothing (Cleveland, Grosse, & Shyu, 1992). The density of lines and saturation of grayscale 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

**Fig. 11** Visually weighted regression plot for the EZ2 diffusion model parameters
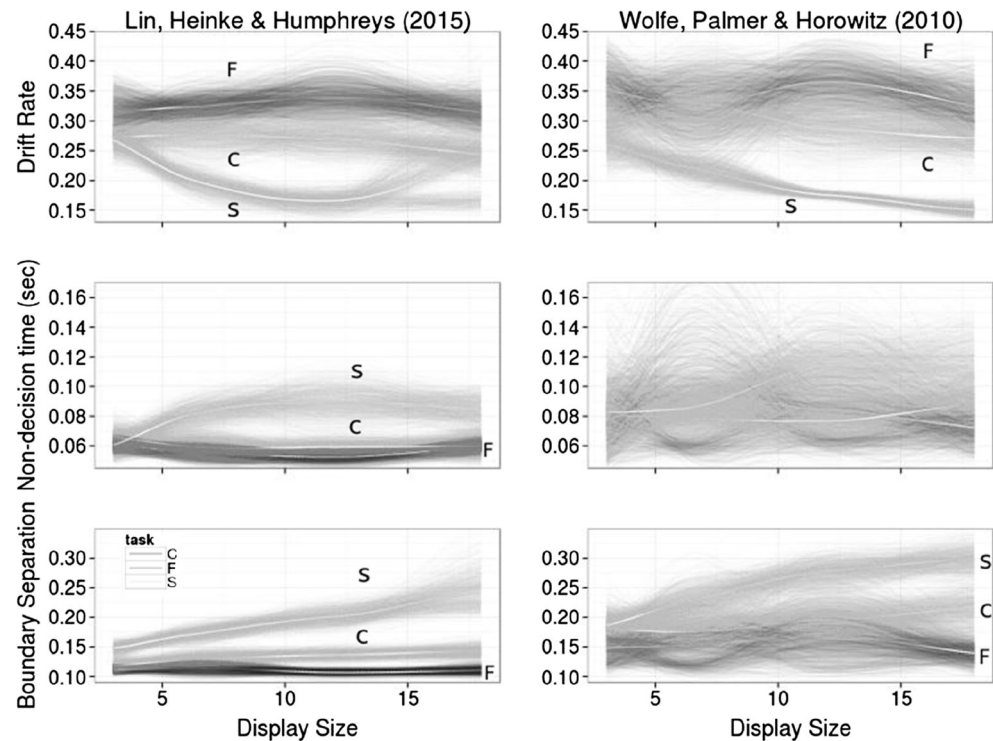


*Drift rate* The two-way (Task × Display Size) ANOVA revealed a significant effect of task, $F(2, 55) = 9.47$, $p = 2.92 \times 10^{-4}$, $\eta^2_p = .256$, but not of display size, $F(3, 165) = 0.472$, $p = .703$, $\eta^2_p = .009$, and no interaction, $F(6, 165) = 1.27$, $p = .28$, $\eta^2_p = .044$. Post-hoc $t$ tests showed that the feature search (0.323) drifted faster than did the conjunction search (0.265; marginally significant, $p = .057$, 95 % CI = −0.117 to .001) and the spatial configuration search (0.220; $p = 1.81 \times 10^{-4}$, 95 % CI = 0.044 to 0.161). No difference was found between the conjunction and spatial configuration searches.

The drift rate, shown in the upper left panel in Fig. 11, manifests two critical characteristics. First, for both the feature and 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 rates across the three tasks, suggesting differences in the rates at which sensory evidence accumulates in the different tasks. There is also a tendency for the drift rate to rise at the large display size in the spatial configuration task (Fig. 11), suggesting that an emergent factor, such as the grouping of homogeneous distractor elements, 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 target-absent trials.[7] This upward trend was also not present in the data of Wolfe et al. (2010).

---

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

*Nondecision 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 nondecision time than were feature search (57 ms, $p = .008$, 95 % CI = 4.53 to 38.1 ms) and conjunction search (61 ms, $p = .038$, 95 % CI = 0.707 to 34.2 ms). We observed 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, and 84.23 ms with increasing display size), but not for feature or conjunction search tasks.

*Boundary separation* The two-way (Task × Display Size) ANOVA revealed significant effects of the task, $F(2, 55) = 31.75$, $p = 6.81 \times 10^{-10}$, $\eta^2_p = .536$, and the display size, $F(3, 165) = 7.6$, $p = 8.61 \times 10^{-5}$, $\eta^2_p = .121$, as well as 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 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 (0.148, 0.170, 0.201, and 0.249 for 3, 6, 12, and 18 items, respectively), $F(3,$

57) = 6.73, $p = .001$, $\eta^2_p = .262$, but not for the feature or conjunction searches.

## General discussion

In this study, we applied an integrated approach to the modeling of 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 allowed us to reveal the possible accounts of the three distributional parameters—shift, shape, and scale—by associating them with the nondecision time, drift rate, and boundary separation estimated from the diffusion model. Our study goes farther than most previous ones (Balota & Yap, 2011; Heathcote et al., 1991; Sui & Humphreys, 2013; Tse & Altarriba, 2012) that have applied distributional analyses 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 optimize the 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 showed efficient search for feature targets and inefficient search when targets could only be distinguished from nontargets by conjoining multiple features (shape and color, or shape only; see Chelazzi, 1999, and Chun & Wolfe, 2001, for reviews). The display size effect present in feature search (415, 426, 432, and 437 ms) suggests some limitations on selecting feature targets, but the analyses based on mean RTs did not differentiate whether the effect ($\eta^2_p = .294$) was due to postselection reporting (Duncan, 1985; Riddoch & Humphreys, 1987) or to an involvement of focal attention in feature search. This question was addressed by examining the estimates from HBM together with those from the EZ2 diffusion model. The lack of display size effects for nondecision time suggests that the increasing trend in the mean RTs was unlikely to be due to a delay in peripheral processes, such as motor or early perceptual times. Neither drift rate showed a reliable effect of the display size for feature search. The only possible difference was an unreliable display size effect ($p = .106$), together with an increase of variation in the shape parameter at display size 18. This result appears to favor 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 et al., 2002; Kwak et al., 1991; Woodman & Luck, 2003). This could reflect either a reduction in the guidance of search from spatial configuration, as compared with simple orientation and color information, or in 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, and 20.49 ms). This last result suggests that high-density homogeneous configurations of distractors do facilitate search, a point that we return to below (Bergen & Julesz, 1983; Chelazzi, 1999; Duncan & Humphreys, 1989; Heinke & Backhaus, 2011; Heinke & Humphreys, 2003).

### Methodological issues

The analyses of the mean RTs, however, did not always accord with the analyses of trial RTs. For example, the density plots of mean RTs (Fig. 5) suggest that the data were distributed symmetrically, contrasting with the common notion that an RT density curve tends to be positively distributed toward long latencies (Luce, 1986). However, the analyses of the trial RTs (Fig. 6) revealed clearly skewed RT distributions. This was 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 for a condition (individual curves in Fig. 6) is affected greatly by the weight of the slow RTs. The conditions and observers that contribute the slow responses tend to move the central location toward longer latencies within a distribution; hence, we observed 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 means of a biased central-location parameter (with respect to a skewed RT distribution), the nature of the RT distribution (e.g., if there are 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 of 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 by using the 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 being averaged across participants and then compared

across the different conditions. This approach descriptively dissects an RT distribution into multiple components (e.g., mu, sigma, and tau), each potentially reflecting a 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 addressed this issue by contrasting the empirical data modeled 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 et al. (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 with the response rate, and the shape parameter with postattentive response selection (Wolfe, Võ, Evans, & Greene, 2011). The EZ2 diffusion model directly estimates 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, 2004, terms); (2) the boundary separation, reflecting the response criterion (Wagenmakers et al., 2007); and (3) the nondecision time, reflecting the time that an observer requires to encode stimuli and execute 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 directly set a minimal threshold for responses and rule out the possibility of negative responses. This suggests an association between the RT shift and nondecision time parameters.

Model-based analysis

The EZ2 diffusion model and HBM results suggest that distributional parameters reflect different aspects of search. First, the shift parameter varied across the search tasks and display sizes, a pattern that was in line with our illustration and the ideal analysis (see Fig. 1 and Appendix 2). This parameter reflects the psychological processes that evenly influence all ranges of RTs. One of the diffusion processes likely to influence the shift changes is the drift rate, which showed only a main effect of task. Since the drift rate aims to model the rate of information accumulation determined by the goodness of match between the templates and search stimuli, the shift parameter appears to result from a 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, which requires only preattentive parallel processing to extract just one simple salient feature, to (ii) conjunction search, in which two simple features must be bound

to facilitate a good match, to (iii) spatial configuration search, which demands both feature binding and coding the configuration of the features. The lack of an interaction with display size further supports our argument that the shift reflects factors that affect the entire RT distribution equally. The weak display size effect can be readily explained by the crowded layout that we used; it had not been observed [$F(3, 75) = 0.016$, $p = .997$] in Wolfe et al.'s (2010) data. This weak effect of 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 to those of the participants in Wolfe and colleagues' (2010) study. These participants did not assemble a similarity search unit, so the predicted drift rate decreased at 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 this pattern 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 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 nonlinear, since 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 such as the one employed here, the two types of search respectively reflect weaker distractor–distractor grouping and stronger target–distractor grouping; see Duncan & Humphreys, 1989). If the process of distractor rejection is more difficult in conjunction and configuration search, as compared with feature search, then the effects on a parameter will reflect this process, and this may not vary directly with display size, as we observed.

In contrast to the shift parameter, the shape parameter showed a marginal effect of display size, a reliable effect of 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 demonstrated a U-shaped function for the spatial configuration search. This last result is consistent with 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, and 0.344 with increasing display size), 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 at display size 3, to 1.86 at display size 6, 2.05 at display size 12, and 1.96 at display size 18, on target-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 also observed in Palmer et al.'s (2011) analysis. However, their data showed no reliable shape change across display sizes for target-present trials (Palmer et al., 2011). Following Wolfe et al.'s (2010) suggestion, Palmer and colleagues speculated that the display size effect for the shape parameter might result from the premature abandoning of search, a view that was supported by Wolfe et al.'s (2010) data showing a high rate of miss errors in the spatial configuration task. The high rate of miss errors might reflect an observer prematurely deciding to give an "absent" response on a target-present trial. This would, in turn, reduce the overall number of slow responses, leading to an RT distribution with low skewness. This indicates that in the conditions with high miss errors, participants tended to set a low decision threshold for the "absent" response. The tendency might also appear in the target-absent trials, resulting in correct rejections by luck, a result leading to RT distributions in these trials with an increase in the shape parameter. We, applying a more sensitive method under the constraint of limited trial numbers, showed reliable display size effects on the RT shape in the target-present trials of the spatial configuration and conjunction searches. Together with the miss error data, our data do indicate that a link between 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: Relative to feature search, the factor that changes the RT shape in spatial configuration search is the *goodness of match* between the search template and the search display (i.e., the drift rate changes). This implies that factors contributing to change in different parts of an RT distribution will result in its shape changing. As our simulation study shows (see Appendix 2), doubling the shape parameter results in a decrease in boundary separation (in line with the miss error account) and an increase in the drift rate (in line with the goodness-of-match account). The two diffusion parameters are likely the processes driving changes in the shape of the RT distribution.

Among the three Weibull parameters, the scale parameter showed the highest correlation with mean RTs (Pearson's $r = .78$, $p = 2.20 \times 10^{-16}$), 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 change in the central location of the RT distributions. The scale parameter estimates an overall enhancement (or reduction) of response latency as well as response variance, as do the mean and variance of RTs (see the review in Wagenmakers & Brown, 2007). Unlike the mean RTs, however, the scale parameter in our data set was not sensitive to the display size in the feature search task. A cross-examination with the boundary separation in the diffusion model appeared to indicate that the scale parameter might reflect the influence of response criteria, with only the inefficient tasks showing a 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 in response latency and variance (the more reluctant one is 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 have been posited, under the framework of the stage model of information processing, to reflect different aspects of peripheral and central processing. The shift parameter has been associated with the speed of peripheral processes (i.e., the irreducible minimum response latency; Dzhafarov, 1992), the scale parameter with the speed of executing central processes, and the shape and scale parameters with the insertion of additional stages into central processing (Rouder et al., 2005).

Using benchmark paradigms of visual search (Wolfe et al., 2010), 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 in which a 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 grouped distractors and the quality of the match to a target template, which, when accurate, contributes to the entire RT distribution.

Our results for the scale parameter are consistent with those of Rouder and colleagues (2005) in suggesting that this parameter reflects the speed of execution in a central decision-making process. Since the execution speed closely links with the decision boundaries and the initial state of sensory

information that an observer sets for a response trial, we observed similar patterns in the scale parameter, the boundary separation, and the nondecision time. The pattern in the nondecision times is readily accounted for by the fact that the EZ2 diffusion model absorbs the parameter reflecting the initial state of sensory evidence into the nondecision 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 for 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 than for the two simple search tasks, in both our data (0.323 and 0.265 vs. 0.220) and those of Wolfe et al. (2010; 0.341 and 0.299 vs. 0.203). In Palmer et al.'s (2011) analysis, no task effect was found in the shape parameter. Using 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 terms 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 in the previous study fewer participants were recruited, and some might have found strategies through which to conduct the difficult searches while still using the same processing stages as in the feature search task, the shape parameter reflected only a marginal effect.

Another possible factor that may explain the different findings for the shape parameter is illustrated by the visually weighted plot of drift rate. The visually weighted regression lines indicate two groups of participants accumulating sensory evidence at different rates, where only one homogeneous group appeared in Wolfe et al.'s (2010) data. As our simulated study showed (Appendix 2), the drift rate can also change the shape parameter. This suggests that some of our participants took advantage of the crowded layout to facilitate search in the spatial configuration task, and some did not. This could not happen in Wolfe et al.'s (2010) sparse layout, and likely also contributed to our significant finding for the shape parameter on target-present trials.

Limitation

The analytic approach that we adopted assumes that individual RTs are generated by three-parameter probability functions. Our selection of the Weibull function was determined, on the one hand, by its probing three important aspects—the shift, scale, and shape—of an RT distribution, which differs from what the ex-Gaussian function describes (mu, tau, and sigma). On the other hand, we selected the Weibull function because it permits a reliable posterior distribution to converge, has broad application to memory as well as to visual search (Logan, 1992; see also Hsu & Chen, 2009), and also can apply to the hierarchical Bayesian framework (Rouder et al., 2005).

## Conclusion

In conclusion, our study shows that an HBM-based distributional analysis, complemented with the EZ2 diffusion model, can help to clarify the 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 those of display size in some cases (on the drift rate and shift parameters), but not in others (e.g., effects on nondecision factors and on the separation of decision boundaries). We have linked this dissociation to the involvement of distractor grouping and rejection (on the one hand) and to serial selection of the target and the setting of a response criterion (on the other). This approach goes beyond what can be done using standard analyses based on mean RTs.

## Appendix 1

HBM simulations

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

Through simulations[8] we 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, 2004, parameter values).

---

[8] We used R routines—rnorm, rexGAUS, rinvgauss, and rweibull3—to generate those computer-simulated data.
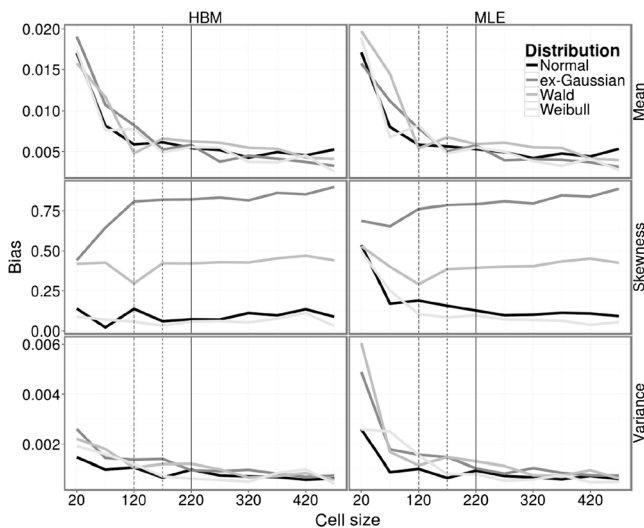
**Fig. 12** Sample mean, variance, and skewness values, compared to the true values that generated the simulated data. "HBM" and "MLE" stand for the hierarchical Bayesian model and maximum likelihood estimates, respectively. The mean and variance are in scales of seconds and squared seconds, respectively. 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 sample sizes of 120, 170, and 220 to show critical changes in the three parameters with regard to the sample sizes

Specifically, we used the parameter values listed in Table 3 of Cousineau et al.'s report to construct four true distributions, which in turn repeatedly generated randomly simulated RTs. The simulations generated 20 homogeneous participants; each participant contributed RT observations for ten different sample sizes, ranging from 20 to 470 with a step size of 50. The simulated data were then submitted separately to HBM and to
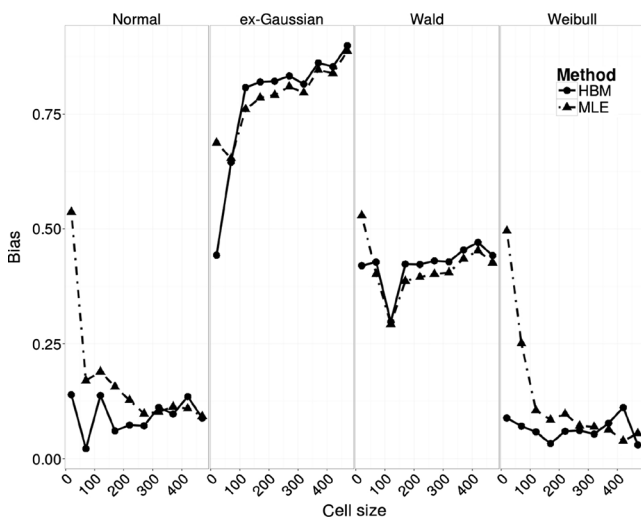
MLE to estimate the shift, scale, and shape parameters. Both methods assumed that RTs were random variables generated by the Weibull function. Those parameters were then analytically converted to the mean, variance, and skewness to evaluate the performance of the two methods.

Figure 12 shows the absolute values of the differences between the true values 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 the per-condition sample size, $F(9, 1520) = 64.46$, $p < .0001$: The more observations were in a condition, the higher was the precision of the estimate. 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 (on 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 accounted for the RT data.

In contrast, HBM demonstrated a clear advantage over the MLE method when recovering the variance, $F(1, 1520) = 9.345$, $p = .0023$. Again, a large number of observations helped to improve the fit, $F(1, 1520) = 29.84$, $p < .0001$. Importantly, HBM estimated the parameters better than MLE at the smallest sample size ($N = 20$), $F(1, 152) = 6.907$, $p = .0095$, though the difference was gradually resolved when the observation numbers exceeded 120, $F(1, 152) = 0.936$, $p = .335$ (i.e., the dashed line in Fig. 1). Misspecification of the underlying distribution resulted in different estimations of variance, $F(3, 1520) = 7.635$, $p < .0001$. Both methods needed a sample size larger than 170 to resolve this issue (the dotted line in Fig. 12). Parameter



**Fig. 13** The estimation of skewness. The figure traces the differences between hierarchical Bayesian modeling (HBM) and maximum likelihood estimation (MLE) at different sample sizes in an experimental condition
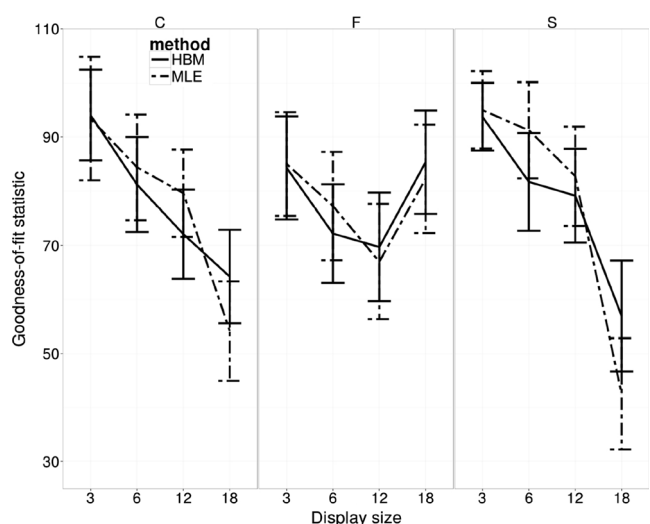


**Fig. 14** Goodness of fit of the empirical data against the Weibull distribution, using the maximum likelihood estimation(MLE) and hierarchical Bayesian modeling (HBM) methods
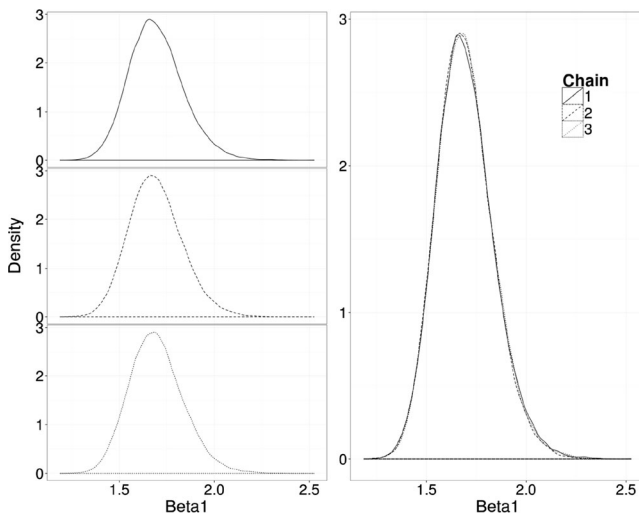
**Fig. 15** Diagnostic posterior density plots for the Markov chain Monte Carlo (MCMC) process. The left panel shows the density curves, separately, for the three MCMC processes (i.e., "Chain" in the legend). The right panel draws them together to demonstrate that the three processes reach a consistent estimation of the posterior distribution, suggesting that the MCMC process is reliable. The figure shows only the $\beta$ (shape) parameter from one of the participants in the target-present condition of display size 18 in the spatial configuration search

recovery was better when the true distribution followed the Weibull function.

As for variance, the skewness was again estimated better by HBM than by MLE, $F(1, 1520) = 22.596$, $p < .0001$. The correct specification of the distribution played an important role in estimating the skewness, $F(3, 1520) = 1,818.549$, $p < .0001$. Specifically, when the true distribution followed either an ex-Gaussian or a Wald function, there was no advantage to using HBM after a trial size of 70. In this case, increasing the
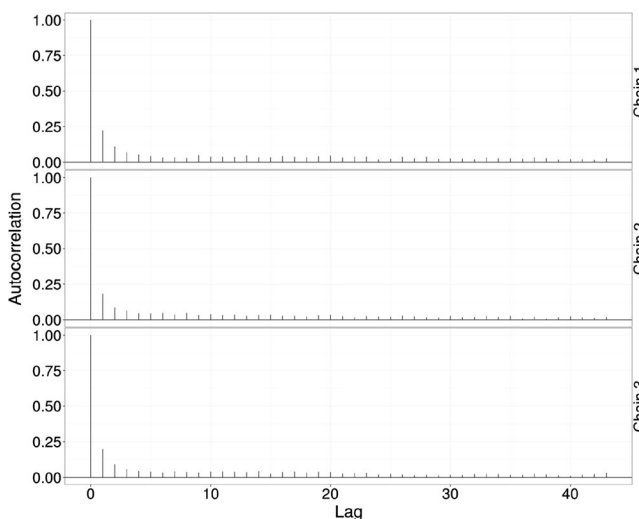


**Fig. 16** Diagnostic autocorrelation plots for the Markov chain Monte Carlo (MCMC) process. The figure shows only the $\beta$ (shape) parameter from one of the participants in the target-present condition of display size 18 in the spatial configuration search

sample size did not mitigate the problem, whereas when the true distribution followed a Weibull function, HBM showed consistently higher precision than MLE (see Fig. 13). Interestingly, HBM also gave better estimates than 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 at least 220 observations were needed in each condition.

Model diagnostics

In this section, we assess the performance of the HBM and MLE estimation methods for the three parameters expressed in the Weibull function. First, a goodness-of-fit index, the Anderson–Darling statistic, is 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 multichain process of MCMC converges to a reliable and single distribution after a long iteration process. Each step of the MCMC process uses the predefined model (i.e., the Weibull function, in our case) to fit the empirical data and
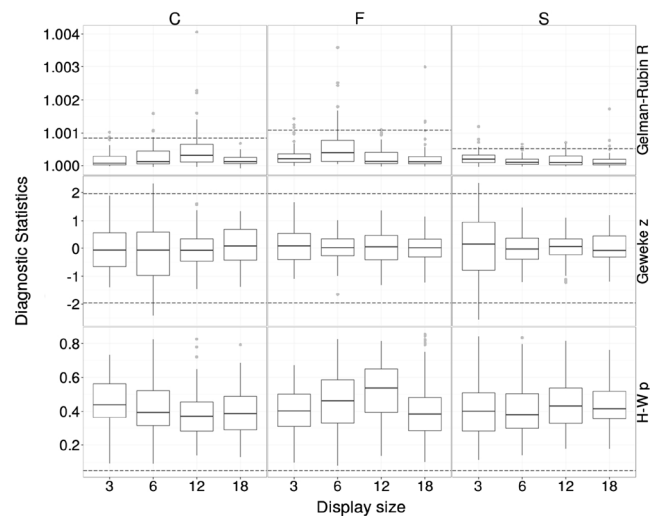


**Fig. 17** Three diagnostic tests for assessing Markov chain Monte Carlo convergence, using box-and-whisker plots to summarize the distributions of the three tested statistics. In each condition (Display Size × Search Task), each participant contributed three data points from the three Weibull parameters. 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 panels, the reference dashed line is drawn at the upper bound of the 95 % confidence interval, separately for each search task. In the middle panels, we used two reference lines at −1.96 and 1.96 (the upper and lower bounds of the 95 % confidence interval) to show that most Geweke z scores fell in the acceptable range. The lower panels show p values from the Heidelberger–Welch test. The reference line is drawn at the .05 alpha level

**Fig. 18** Figure showing how changes in the distributional parameters may alter the parameters estimated by the EZ2 diffusion model. Ter, *a*, and *v* stand for the nondecision time, boundary separation, and drift rate, respectively
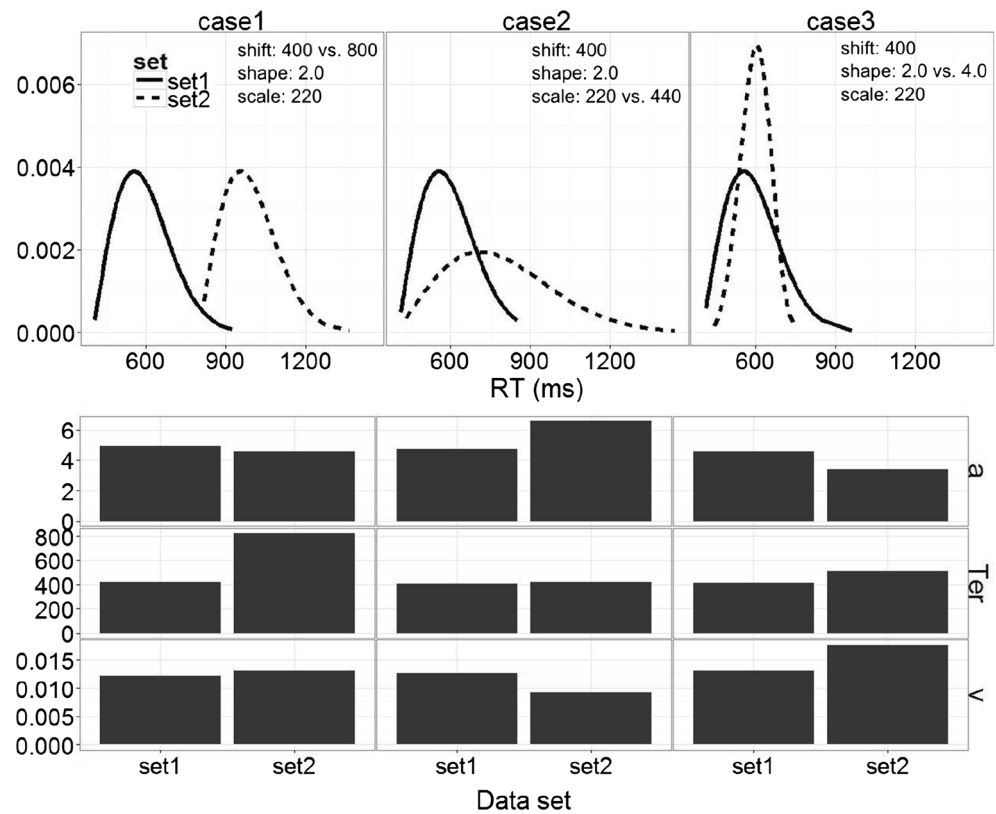


predict a posterior distribution. The posterior distribution is then used as a prior distribution in the next step to fit a new posterior distribution. This process iterates itself until the predefined number of iterations (105,000 in our setting). The final posterior distribution was the predicted distribution presumed to be the true distribution underlying an RT distribution in the visual search paradigm that we examined. We ran three separate independent MCMC processes (i.e., three chains) to test whether all three converged to an identical final posterior distribution. If this was so, it would indicate that our hierarchical Bayesian setting and the Weibull function provided an appropriate approximation of an RT distribution. A nonconvergent MCMC process would not provide a reliable prediction for the final posterior distribution. That is, it would predict, even after a long iteration process, different distributions accounting for the empirical data.

*Goodness of fit* Because RT distributions are formed by a continuous rather than a 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 and HBM fits: Both methods performed similarly and improved their fits as the display size increased.

*Bayesian model diagnosis* In this section, we examine whether the estimated parameters converged to a stationary posterior distribution and whether the setting for conducting MCMC sampling was appropriate for permitting reliable inferences.

Figures 15 and 16 shows one of the examples of the diagnostic plots, indicating the convergence of $\beta$ (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 three sets of parallel data represent the three processes. Three MCMC chains mixed quickly to a limited range after the iteration process began (after 5,000 burn-in iterations), suggesting that the posterior distribution reached a reliable stationary point.[9] This impression is supported by the posterior density distributions (Fig. 15), showing that all three chains predict overlapping distributions with very similar shapes and dispersions, consistent with the three chains predicting identical posterior distributions. The autocorrelation plots (Fig. 16) showed that the MCMC sampling interval (in our case, the computer program took one sample every

---

[9] A nonstationary mix will manifest as three clearly separated lines.

four iterations) is appropriate to reduce the interiteration correlation rapidly within first 50 iterations (after the first 5,000 samples have been discarded).

The information from the graphical diagnoses is compatible with the nonparametric statistical tests. Figure 17 presents the graphical summary of the nonparametric tests for the stationarity across all the conditions and participants. The upper panels show the Brooks–Gelman–Rubin (*BGR*; Brooks & Gelman, 1998) shrink factor. The recommended *BGR* upper limit is 1.1 (Gelman, 2004; Gelman & Hill, 2007), and values below it are deemed acceptable. We drew the reference line at the grand average 95 % confidence interval to allow a clear inspection of the distribution for the statistics. 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.

The middle panels show the Geweke *z* score. This test examines the stationarity of the 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 $Z = \pm1.96$. 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 panels show the distribution of the *p* values from Heidelberger and Welch's test. The reference line was drawn at a .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 nonparametric 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.
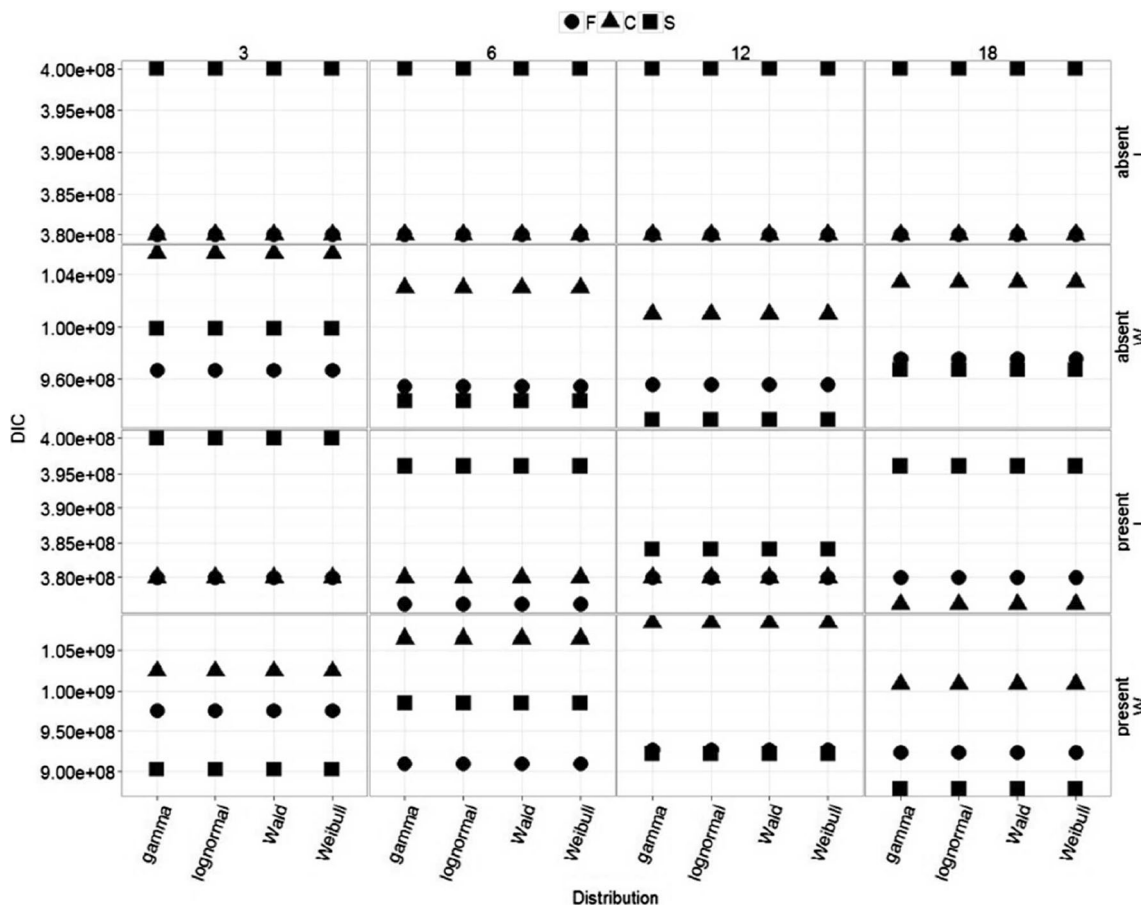


**Fig. 19** The figure compares the Bayesian deviance information criteria for the fitted three-parameter probability functions across the data sets, search tasks, target types, and display sizes. "L" and "W" stand for our and Wolfe et al. (2010) data sets

## Appendix 2: EZ2 diffusion model simulations

The stimulation study was undertaken to examine how the parameters estimated from the EZ2 diffusion model correlate with the Weibull parameters in a simple, well-controlled situation. We performed three case studies in which we 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 of the Weibull parameters. In the three studies, we doubled, respectively, 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 nondecision time. The results indicated that, first, doubling the shift parameter resulted in a near two-fold increase in nondecision time (416 vs. 827 ms), a small increase in drift rate (0.012 vs. 0.013), and a small decrease in boundary separation (4.89 vs. 4.57). Second, doubling the scale parameter resulted in a decrease in drift rate (0.013 vs. 0.009), an increase in boundary separation (4.70 vs. 6.57), and a 10-ms increase in the nondecision time (407 vs. 417 ms). Finally, doubling the shape parameter resulted in an increase in drift rate (0.013 vs. 0.018), a decrease in boundary separation (4.57 vs. 3.39), and again, a small increase in the nondecision time (410 vs. 507), although this increase was relatively larger than that from doubling the scale parameter. Figure 18 shows a comparison across the three case studies.

## Appendix 3: Why use the 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 the gamma, lognormal, 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 our RT data.

The first reason is that the Weibull function is able to provide a concise way to summarize the shape of an RT distribution. As is described in the main text and illustrated in Fig. 1, 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

three-parameter gamma function does not converge when fitted with hierarchical Bayesian approach. The compatible three-parameter gamma function shows signs of nonconvergence and perhaps because of this, it fits the data slightly worse, indicated by the DICs. Third, we fitted also the other two three-parameter functions: Wald and lognormal. They provide the same set of descriptive parameters as the Weibull function. The DICs suggest that these functions fit both our and Wolfe et al. (2010) data slightly better than the Weibull function. However, we have maintained the Weibull function because the fits of the four functionsare very similar for each task, display size, target type, and data set (Fig. 19). To test whether any function would work, we fitted a Gaussian function. The Gaussian fit DICs (approximately −3,150) are far worse than the four plausible functions that can accommodate positive-skewed RT. Below, we report the detailed procedure to fit gamma functions.

To test whether the Weibull function fits better than the gamma function, we built a three-parameter gamma function in the HBM framework. Because currently BUGS does not implement a prebuilt three-parameter gamma function, we used Johnson et al. (1994, p. 337, Eq. 17.1) equation to implement the gamma function. Practically, the density function in the BUGS code is changed 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, nonconverged 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 that did not abate with increasing iterations. In summary, the diagnostics show that the gamma function does not converge when fitted with an HBM framework.

This failure of gamma fit could have occurred because (1) the gamma function is not suitable for HBM in this context, and/or (2) the gamma function indeed fits worse than the Weibull function (as the DIC suggests). Note that we ran a high number of iterations (i.e., 105,000) with a reasonable thinning length, which still did not resolve the nonconverged gamma fit.

# 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,* 160–166. doi:10.1177/0963721411408885

Bergen, J. R., & Julesz, B. (1983). Parallel versus serial processing in rapid pattern discrimination. *Nature, 303,* 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,* 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,* 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,* 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,* 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,* 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: 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, NY: Chapman & Hall.

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

Crawley, M. J. (2002). *Statistical computing: An Introduction to data analysis using S-Plus*. Chichester, UK: Wiley.

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

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review, 96,* 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,* 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,* 1209–1217. doi:10.3758/PBR.15.6.1209

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

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge, UK: 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,* 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,* 3709–3719. doi:10.1016/j.ridd.2013.07.025

Heathcote, A., Popiel, S. J., & Mewhort, D. J. (1991). Analysis of response time distributions: An example using the Stroop task. *Psychological Bulletin, 109,* 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,* 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,* 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). New York, NY: Psychology Press.

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

Hsiang, S. M. (2013). *Visually-weighted regression* (SSRN Working Paper). Retrieved from papers.ssrn.com/sol3/papers.cfm?abstract_id=2265501

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

Humphreys, G. W., & Müller, H. J. (1993). SEarch via Recursive Rejection (SERR): A connectionist model of visual search. *Cognitive Psychology, 25,* 43–110. doi:10.1006/cogp.1993.1002

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

Kruschke, J. K. (2010). What to believe: Bayesian methods for data analysis. *Trends in Cognitive Sciences, 14,* 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,* 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,* 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,* 883–914. doi:10.1037/0278-7393.18.5.883

Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization*. Oxford, UK: 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*. Boca Raton, FL: Chapman & Hall/CRC.

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

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide*. Mahwah, NJ: Erlbaum.

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

Matzke, D., & Wagenmakers, E.-J. (2009). Psychological interpretation of the ex-Gaussian and shifted Wald parameters: A diffusion model analysis. *Psychonomic Bulletin & Review, 16,* 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,* 308–313. doi:10.1093/comjnl/7.4.308

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,* 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,* 118–123. doi:10.1111/1467-8721.ep10772534

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In K. Hornik, F. Leisch, & A. Zeileis (Eds.), *Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003)* (pp. 20–22). Vienna, Austria: Technische Universität Wien.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85,* 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,* 873–922. doi:10.1162/neco.2008.12-06-420

Ratcliff, R., & Murdock, B. B. (1976). Retrieval processes in recognition memory. *Psychological Review, 83,* 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,* 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,* 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,* 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,* 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,* 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,* 195–223. doi:10.3758/BF03257252

Rouder, J. N., & Speckman, P. L. (2004). An evaluation of the Vincentizing method of forming group-level response time distributions. *Psychonomic Bulletin & Review, 11,* 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,* 1267–1274. doi:10.1037/a0020793

Schönbrodt, F. (2012). Visually weighted/watercolor plots, new variants: Please vote! Retrieved from www.nicebread.de/visually-weighted-watercolor-plots-new-variants-please-vote/

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

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,* 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,* 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,* 589–627. doi:10.1037/a0033140

Stoet, G. (2010). PsyToolkit: A software package for programming psychological experiments using Linux. *Behavior Research Methods, 42,* 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,* 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,* 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,* 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,* 663–676. doi:10.1017/S1366728912000077

Van Zandt, T. (2000). How to fit a response time distribution. *Psychonomic Bulletin & Review, 7,* 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,* 830–841. doi:10.1037/0033-295X.114.3.830

Wagenmakers, E.-J., van der Maas, H. L. J., Dolan, C. V., & Grasman, R. P. P. P. (2008). EZ does it! Extensions of the EZ-diffusion model. *Psychonomic Bulletin & Review, 15,* 1229–1235. doi:10.3758/PBR.15.6.1229

Wagenmakers, E.-J., van der Maas, H. L. J., & Grasman, R. P. P. P. (2007). An EZ-diffusion model for response time and accuracy. *Psychonomic Bulletin & Review, 14,* 3–22. doi:10.3758/BF03194023

Ward, R., & McClelland, J. L. (1989). Conjunctive search for one and two identical targets. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 664–672. 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,* 33–39. doi:10.1111/1467-9280.00006

Wolfe, J. M., & Horowitz, T. S. (2008). Visual search. *Scholarpedia, 3,* 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,* 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,* 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,* 121–138. doi:10.1037/0096-1523.29.1.121